



# Kent Academic Repository

O'Connor, Rebecca (2016) *Reproductive Isolation, in Individuals and During Evolution, as Result of Gross Genomic Rearrangement in Pigs, Birds and Dinosaurs*. Doctor of Philosophy (PhD) thesis, University of Kent,.

## Downloaded from

<https://kar.kent.ac.uk/58390/> The University of Kent's Academic Repository KAR

## The version of record is available from

## This document version

UNSPECIFIED

## DOI for this version

## Licence for this version

UNSPECIFIED

## Additional information

## Versions of research works

### Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

### Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in *Title of Journal*, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

## Enquiries

If you have questions about this document contact [ResearchSupport@kent.ac.uk](mailto:ResearchSupport@kent.ac.uk). Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

# Kent Academic Repository

## Full text document (pdf)

### Citation for published version

O'Connor, Rebecca E. (2016) Reproductive Isolation, in Individuals and During Evolution, as Result of Gross Genomic Rearrangement in Pigs, Birds and Dinosaurs. Doctor of Philosophy (PhD) thesis, University of Kent,.

### DOI

### Link to record in KAR

<http://kar.kent.ac.uk/58390/>

### Document Version

UNSPECIFIED

#### Copyright & reuse

Content in the Kent Academic Repository is made available for research purposes. Unless otherwise stated all content is protected by copyright and in the absence of an open licence (eg Creative Commons), permissions for further reuse of content should be sought from the publisher, author or other copyright holder.

#### Versions of research

The version in the Kent Academic Repository may differ from the final published version.

Users are advised to check <http://kar.kent.ac.uk> for the status of the paper. **Users should always cite the published version of record.**

#### Enquiries

For any further enquiries regarding the licence status of this document, please contact:

[researchsupport@kent.ac.uk](mailto:researchsupport@kent.ac.uk)

If you believe this document infringes copyright then please contact the KAR admin team with the take-down information provided at <http://kar.kent.ac.uk/contact.html>

**Reproductive Isolation, in Individuals and During  
Evolution, as Result of Gross Genomic Rearrangement in  
Pigs, Birds and Dinosaurs**

**Rebecca E. O'Connor**

**16 August 2016**

**A thesis submitted to the University of Kent for the degree of**

**DOCTOR OF PHILOSOPHY**

**in Genetics in the Faculty of Science, Technology and Medical Studies**

**The School of Biosciences**

## **Declaration**

No part of this thesis has been submitted in support of an application for any degree or qualification of the University of Kent or any other University or Institute of learning

Rebecca E. O'Connor

16 August 2016



## Incorporation of Published Work

### Original Research Manuscripts:

- Lithgow PE\*, **O'Connor R.E.\***, et al. *'Novel tools for characterising inter and intra chromosomal rearrangements in avian microchromosomes'*. Chromosome research. 2014;22(1):85-97. \*Joint first authors
- Romanov MN, Farré M, Lithgow PE, Fowler KE, Skinner BM, **O'Connor R.E.**, et al. *'Reconstruction of gross avian genome structure, organization and evolution suggests that the chicken lineage most closely resembles the dinosaur avian ancestor'*. BMC genomics. 2014;15(1):1060.
- Schmid M, Smith J, Burt DW, Aken BL, Antin PB, Archibald AL,...**O'Connor R.E.**... et al. *'Third Report on Chicken Genes and Chromosomes 2015'*. Cytogenetic and Genome Research. 2015;145(2):78-179.
- Damas J.\* and **O'Connor, R.E.\*** et al. (Genome Research - submitted). *'Upgrading short read animal genome assemblies to chromosome level using comparative genomics and a universal probe set'*. \*Joint first authors
- **O'Connor R.E.** et al. (Animal Genetics - submitted). *'Isolation of subtelomeric sequences of porcine chromosomes reveals errors in the pig genome assembly'*.

### Published Abstracts:

- **O'Connor R.E.** et al. 2014. Identification of Chromosomal Translocations in Pigs using FISH with Subtelomeric Probes and the development of a novel screening tool for their application "21st International Colloquium on Animal Cytogenetics and Gene Mapping." Chromosome Res 22 (2014): 393-437.
- Griffin, D.K., **O'Connor R.E.** et al. 2014. Avian cytogenetics goes functional. "21st International Colloquium on Animal Cytogenetics and Gene Mapping." Chromosome Res 22 (2014): 393-437.
- Sadraie, M., Fowler K.E., **O'Connor R.E.**, et al., 2015. Evaluation of aneuploidy of autosome chromosomes in boar sperm samples. 13th Annual Preimplantation Genetic Diagnosis International Society (PGDIS) Meeting Abstracts." Chromosome Res 22 (2014): 573-643.
- Griffin, D.K., **O'Connor R.E.** et al., 2015 Avian Chromonomics goes functional. Chromosome Research. pp. 413–414
- **O'Connor, R.E.** et al., 2015. Reconstruction of the putative Saurian karyotype and the hypothetical chromosome rearrangements that occurred along the Dinosaur lineage. Chromosome Research, 23(2), p.343.
- Sadraie, M., Fowler K.E., **O'Connor R.E.**, et al., 2015. Evaluation of aneuploidy of autosome chromosomes in boar sperm. In Chromosome Research. pp. 384–385.
- Lithgow, P.E., **O'Connor R.E.** et al., 2015. Inter and intra chromosomal rearrangements in avian microchromosomes. Chromosome Research. pp. 413–414.
- Romanov, M.N.,... **O'Connor R.E.** et al., 2015. Avian ancestral karyotype reconstruction and differential rates of inter-and intrachromosomal change in different lineages. Chromosome Research, 23(2), p.414.

- **O'Connor R.E.** et al., 2016. Upgrading Molecular Cytogenetics to Study Reproduction and Reproductive Isolation in Mammals, Birds, and Dinosaurs. *Cytogenetic and Genome Research*, 148(2-3), pp.83–155.
- Damas, J., **O'Connor R.E.** et al. 2016. A Novel Hybrid Approach for Drafting Chromosome-Level Genome Assemblies Applied to Avian Genomes. In *Plant and Animal Genome XXIV Conference*. Plant and Animal Genome.

### Oral Research Presentations

- 22<sup>nd</sup> International Colloquium on Animal Cytogenetics and Genomics, Toulouse, France July 2016. *'Exceptional degree of avian microchromosome conservation revealed by cross species BAC mapping'*
- 22<sup>nd</sup> International Colloquium on Animal Cytogenetics and Genomics, Toulouse, France July 2016. *'Gross genome evolution in the Dinosauria'*
- Pig Breeders Round Table, University of Kent, April 2015. *'Screening for Hypoprolificacy in Pigs using Subtelomeric FISH probes'*
- 20<sup>th</sup> International Chromosome Conference, University of Kent, September 2014. *'The Saurian Karyotype and Chromosomal Rearrangements along the Dinosaur Lineage'*
- 21<sup>st</sup> International Colloquium on Animal Cytogenetics and Gene Mapping, (Ischia - Naples), June 2014. *'Screening for Hypoprolificacy in Pigs using FISH'*
- Pig Breeders Round Table, University of Kent, April 2013. *'Molecular Cytogenetic Approaches to Translocation Screening'*
- CISO R Reproduction in Animals Mini-symposium, University of Kent, June 2015. *'The Saurian Karyotype and Chromosomal Rearrangements along the Dinosaur Lineage'*
- Pig Breeders Round Table, University of Kent, April 2015. *'Screening for Hypoprolificacy in Pigs using FISH'*
- 21<sup>st</sup> Postgraduate Symposium, University of Kent, September 2014. *'Gross Genome Evolution in the Dinosauria?'*
- 5<sup>th</sup> Annual Research Festival, University of Kent, September 2014. *'Reconstructing the Dinosaur Genome'*

### Poster Presentations:

- **O'Connor R.E.** et al. 19th International Chromosome Conference, Bologna, September 2013. *'Identification of Chromosomal Translocations in Pigs using FISH with Subtelomeric Probes and the development of a novel screening tool for their application'*
- Lithgow P., **O'Connor R.E.** et al. 2014. International Chromosome Conference, Bologna, September 2013. *'Comparative genomics and genome evolution of avian microchromosomes through the use of novel chromosome painting tools'*
- Lithgow P., **O'Connor R.E.** et al. 2014. 20th International Chromosome Conference, Canterbury, September 2014. *'Inter and intra chromosomal rearrangements in avian microchromosomes'*
- Sadraie, M., Fowler K.E., **O'Connor R.E.**, et al. 20th International Chromosome Conference, Canterbury, September 2014. *'Evaluation of aneuploidy of autosome chromosomes in boar sperm samples'*
- Sadraie, M., Fowler K.E., **O'Connor R.E.**, et al. 13<sup>th</sup> Preimplantation Genetic Diagnosis International Society, University of Kent, April 2014. *'Evaluation of aneuploidy of chromosomes 1, 16, 12 and 18 in boar sperm samples'*
- Romanov M., **O'Connor R.E.** et al. 2014. 22<sup>nd</sup> Plant and Animal Genome Conference, San Diego, January 2014. *'In silico reconstruction of chromosomal rearrangements and an avian ancestral karyotype'*

- Romanov M., **O'Connor R.E.** et al. 2<sup>nd</sup> Annual Food, Nutrition and Agriculture Genomics Congress, London, April 2015. *'Comparative cytogenomics enhanced with bioinformatic tools provides further insights into genome evolution of birds and other amniotes'*
- Damas J., **O'Connor R.E.** et al. Royal Veterinary College Postgraduate Research Symposium, London, May 2015. *'Towards the Construction of Avian Chromosome Assemblies'*
- Martell H, **O'Connor R.E.** et al. 2<sup>nd</sup> Regional Student Group, International Society for Computational Biology Conference, Norwich, August 2015. *'Assembling and Comparing Avian Genomes by Molecular Cytogenetics'*
- Damas J., **O'Connor R.E.** et al. 23<sup>rd</sup> Plant and Animal Genome Conference, San Diego, January 2016. *'A Novel Hybrid Approach for Drafting Chromosome-Level Genome Assemblies Applied to Avian Genomes'*.
- **O'Connor R.E.** et al. 22<sup>nd</sup> International Colloquium on Animal Cytogenetics and Genomics, Toulouse, France July 2016. *'A novel approach for the physical mapping of avian genome assemblies to a chromosomal level'*
- **O'Connor R.E.** et al. 22<sup>nd</sup> International Colloquium on Animal Cytogenetics and Genomics, Toulouse, France July 2016. *'Development of a porcine chromosomal translocation screening device reveals errors in the pig genome assembly'*
- **O'Connor R.E.** et al. 21<sup>st</sup> International Chromosome Conference, Brazil July 2016. *'Upgrading Molecular Cytogenetics to Study Reproduction and Reproductive Isolation in Mammals, Birds, and Dinosaurs'*
- Martell H, **O'Connor R.E.** et al. 26<sup>th</sup> International Society for Computational Biology Conference, Florida, USA July 2016. *'Assembling and Comparing Avian Genomes by Molecular Cytogenetics'*

#### **Prizes Awarded:**

- Best Poster:  
22<sup>nd</sup> International Colloquium on Animal Cytogenetics and Genomics, Toulouse, France July 2016. *'Development of a porcine chromosomal translocation screening device reveals errors in the pig genome assembly'*

#### **Products Developed:**

- Chromoprobe Multiprobe® Porcine (Cytocell)
- Chromoprobe Multiprobe® Chicken (Cytocell)

## Acknowledgements

First and foremost, I would like to thank my good friend and supervisor Professor Darren Griffin, without whose support and guidance this thesis would not have been possible. Darren's enthusiasm for the subject and his confidence in my abilities has allowed me to make the most of the many wonderful opportunities that have come my way, every one of which he has been fully supportive of.

One of these many opportunities has been to work with highly respected collaborators. On the academic side, I have been privileged to work with Dr Denis Larkin and his team: Joana Damas and Dr Marta Farré, in what I hope will go on to be a long and fruitful collaboration. On the commercial side, I have been very fortunate to work with the team at Cytocell, in particular Dr Gothami Fonseka, Richard Frodsham and Dr Martin Lawrie who have been tremendously supportive throughout this project. On the pig breeding side, it has been very rewarding to work with Dr Grant Walling at JSR among many others. Dr Abullah Al-Mutery from Sharjah University in the United Arab Emirates has given me many opportunities to share the products developed during this thesis through teaching workshops I have performed in Sharjah and along with Sunitha Joseph, Dr Ulli Werner and Renate Werner has always been very hospitable and supportive of this work.

In the lab, there are so many people to thank (past and present) whose kindness, support and humour have made this PhD such a great experience. Kara, Claudia, Becki, Kate, Henry, Jacob, Giuseppe, Becca, Lucas, Sophie, Anjali, Mike, Peter and Ben Skinner - thank you so much to those of you who have helped generate the data presented here and to everyone for your words of wisdom (and words of madness!) and for making each day such a joy.

Last but by no means least, my endless thanks go to Joe, Daisy and Paddy whose love, patience and support has been extraordinary and more than I could ever have hoped for. Not only that, they are the funniest, kindest and most thoughtful people in the world and I feel very blessed that they are part of my life.

## Table of Contents

Declaration.....	i
Incorporation of Published Work.....	ii
Acknowledgements .....	v
Table of Contents .....	i
List of Figures .....	vii
List of Tables .....	xv
Abbreviations.....	xvii
Abstract .....	xviii
<b>1 Introduction.....</b>	<b>1</b>
<b>1.1 Methods to study chromosomes and their rearrangements .....</b>	<b>3</b>
1.1.1 Classical Cytogenetics.....	3
1.1.2 Molecular Cytogenetic Analysis using Fluorescence <i>in situ</i> Hybridisation .....	5
1.1.3 Microarrays and Comparative Genomic Hybridisation (Cytogenomics) .....	13
1.1.4 Genome Sequencing Technologies.....	15
1.1.5 Genome Sequencing Strategies.....	21
1.1.6 Avian Genome Sequencing in the New Era .....	29
1.1.7 Comparative Genomic Visualisation Tools at the Chromosome Level.....	36
1.1.8 Genome Reconstruction using Bioinformatics and FISH .....	38
<b>1.2 Chromosome rearrangements in medicine and evolution .....</b>	<b>41</b>
1.2.1 Numerical chromosome abnormalities.....	42
1.2.2 Structural Chromosome Abnormalities.....	44
<b>1.3 Karyotype Variation .....</b>	<b>53</b>
1.3.1 Mammalian Karyotype Diversity .....	53
1.3.2 Avian Karyotype Diversity.....	54
1.3.3 Reptilian Karyotype Diversity .....	56
<b>1.4 Genome Structure Variation .....</b>	<b>57</b>
1.4.1 Avian Genome Structure .....	57
1.4.2 Genomic Structure of Reptiles in Comparison to Birds .....	59
1.4.3 Sex Chromosome Evolution.....	60
1.4.4 Genome Structure in the Dinosaurs .....	62
<b>1.5 Mechanisms of Genome and Chromosome Evolution .....</b>	<b>63</b>
1.5.1 Non-allelic Homologous Recombination .....	63

1.5.2	Genetic Recombination .....	64
1.5.3	Homologous Synteny Blocks and Evolutionary Breakpoint Regions .....	65
1.5.4	Chromosomal Breakpoints - Random or Non-random? .....	67
1.5.5	Chromosomal Rearrangements and Speciation .....	68
1.5.6	Homoplasmy and Hemiplasy .....	70
<b>1.6</b>	<b>Evolution and Genomics: Amniotes, Birds and Dinosaurs .....</b>	<b>73</b>
1.6.1	Amniote Evolution .....	73
1.6.2	Evolution of the Dinosaurs .....	76
1.6.3	Speciation and Radiation of the Dinosaurs .....	76
1.6.4	Avian Evolution from the Dinosaurs .....	77
1.6.5	Chromosomal Studies in the Context of Evolution .....	80
<b>1.7</b>	<b>Specific Aims of this Thesis .....</b>	<b>83</b>
<b>2</b>	<b>Materials and methods .....</b>	<b>84</b>
<b>2.1</b>	<b>Chromosome Preparation .....</b>	<b>84</b>
2.1.1	Fibroblast Culturing .....	84
2.1.2	Chromosome Harvesting .....	85
2.1.3	Blood Culture and DNA Extraction .....	86
<b>2.2</b>	<b>Generation of Labelled FISH Probes .....</b>	<b>87</b>
2.2.1	Selection of BAC clones .....	87
2.2.2	Preparation of BAC Clones .....	88
2.2.3	Amplification and Labelling BACs .....	89
<b>2.3</b>	<b>Fluorescence <i>in situ</i> Hybridisation (FISH) .....</b>	<b>91</b>
2.3.1	Metaphase Slide Preparation and Hybridisation .....	91
2.3.2	Multiprobe Device Preparation .....	92
2.3.3	Microscopy .....	93
2.3.4	Image Analysis - Karyotyping .....	93
<b>2.4</b>	<b>Physical Genome Mapping .....</b>	<b>93</b>
2.4.1	Construction of PCFs using the RACA Algorithm (RVC) .....	94
2.4.2	Verification of PCFs (RVC) .....	94
2.4.3	Creation of a Refined Set of Pigeon and Peregrine PCFs (RVC) .....	95
<b>2.5</b>	<b>Ancestral Karyotype Reconstruction .....</b>	<b>95</b>
2.5.1	Avian Ancestor .....	95
2.5.2	Saurian Ancestor .....	95
2.5.3	Multiple Genomes Alignment and Identification of HSBs and EBRs .....	96

2.5.4	Ancestral Saurian and Avian Karyotype Reconstruction .....	96
2.5.5	Genome Rearrangement Analysis .....	97
2.5.6	Gene Ontology Analysis.....	97
<b>3</b>	<b>Specific aim 1: To isolate sub-telomeric sequences from pig and cattle genome assemblies and develop a multiprobe device for the screening of both overt and subtle chromosome rearrangements in these species.....</b>	<b>99</b>
<b>3.1</b>	<b>Background.....</b>	<b>99</b>
<b>3.2</b>	<b>Specific aims .....</b>	<b>101</b>
<b>3.3</b>	<b>Materials and Methods.....</b>	<b>101</b>
3.3.1	Blood culture .....	101
3.3.2	BAC Selection and FISH.....	101
<b>3.4</b>	<b>Results.....</b>	<b>102</b>
3.4.1	Specific aim 1a: To develop a practicable and commercially viable system for the screening of chromosome translocations in domestic male mammals by classical approaches .....	102
3.4.2	Specific Aim 1b: To isolate sub-telomeric chromosome identifier probes as tools for chromosome translocation detection in pigs and, in so doing test the hypothesis that genome assembly information in the porcine assembly accurately represents chromosomal position in the sub-telomeric regions.....	103
3.4.3	Specific Aim 1c: To develop a means of screening for porcine cryptic translocations for the whole karyotype in a single experiment .....	107
3.4.4	Specific Aim 1d: To apply the above system in the screening of sub-fertile boars and, in particular test the hypothesis that some sub-fertile boars have translocations that can be detected by this system but not by karyotyping alone .....	108
3.4.5	Specific Aim 1e: To test the hypothesis that translocation screening technology developed for boars can also be applied to other mammals of agricultural importance such as cattle	110
<b>3.5</b>	<b>Discussion.....</b>	<b>112</b>
3.5.1	Chromosome translocation detection using FISH .....	112
3.5.2	Genome Assembly Integrity .....	113
3.5.3	Extension of the FISH screening method to other agricultural species.....	114
<b>3.6</b>	<b>Conclusions.....</b>	<b>115</b>

<b>4</b>	<b>Specific aim 2: To apply the technology developed in specific aim 1 for the solving of previously intractable karyotypes and test the hypothesis that microchromosomal rearrangement is rare in avian evolution.....</b>	<b>116</b>
4.1	Background.....	116
4.2	Specific aims .....	117
4.3	Materials and Methods.....	117
4.4	Results .....	118
4.4.1	Specific Aim 2a: To develop a means of identifying all sequenced avian chromosomes in a single experiment by molecular cytogenetics.....	118
4.4.2	Specific Aim 2b: To test the above system on closely related avian species to identify patterns of inter chromosomal rearrangement during avian evolution .....	120
4.4.3	Specific Aim 2c: To develop a means of improving the success of cross species FISH using BAC clones.....	125
4.4.4	Specific Aim 2d: To adapt the system described in specific aim 2a using technology developed in specific aim 2c so that it can be applied to all bird karyotypes.....	128
4.4.5	Specific Aim 2e: To test the hypothesis that microchromosomal rearrangement is a rare event, compared to other vertebrate species, during avian genome evolution....	130
4.5	Discussion.....	136
4.5.1	Characterisation of all sequenced avian chromosomes in one experiment.....	136
4.5.2	Cross-species FISH Improvements.....	137
4.5.3	Interchromosomal rearrangement during avian evolution.....	138
4.5.4	Microchromosome Conservation .....	139
4.6	Conclusions.....	140
<b>5</b>	<b>Specific aim 3: To use the technology developed in specific aims 1 and 2 in conjunction with a bioinformatics approach developed by colleagues at the Royal Veterinary College, London to complete the cytogenetic mapping of scaffold based genome assemblies to full chromosomal level in two key (but karyotypically dissimilar) avian species (peregrine falcon and pigeon) .....</b>	<b>141</b>
5.1	Background.....	141
5.2	Specific Aims.....	143
5.3	Materials and methods .....	144
5.3.1	Bioinformatic Approach to Upgrading Genome Assemblies .....	144
5.3.2	BAC Selection and FISH.....	144
5.4	Results .....	145



5.4.1	Specific Aim 3a: To use the approach for BAC selection outlined in chapter 4 to select a larger panel (over 200) that map to PCFs of pigeon and peregrine genomes....	145
5.4.2	Specific aim 3b: To apply the above BAC set to pigeon ( <i>Columba livia</i> ) chromosomes to upgrade the scaffold-based assembly to chromosome level and map the intrachromosomal differences between pigeon and the reference genome .....	146
5.4.3	Specific aim 3c: To apply the above BAC set to peregrine falcon ( <i>Falco peregrinus</i> ) chromosomes to upgrade the scaffold-based assembly to chromosome level and map the evolutionary changes that led to the gross genomic rearrangements that occurred in this species	148
5.4.4	Specific Aim 3d: To supply raw data from mapping PCFs in both species to RVC for uploading to the Evolution Highway browser.....	150
<b>5.5</b>	<b>Discussion .....</b>	<b>152</b>
5.5.1	Generation of Chromosomal Level Genome Assemblies .....	152
5.5.2	Avian Interchromosomal Rearrangements .....	153
5.5.3	The Pigeon Genome .....	153
5.5.4	The Peregrine Falcon Genome .....	154
<b>5.6</b>	<b>Conclusions.....</b>	<b>155</b>
<b>6</b>	<b>Specific aim 4: To use bioinformatic tools to re-create the overall genome structure of both Saurian and Avian ancestors and to retrace the gross evolutionary changes that occurred along the dinosaur lineage. To perform gene ontology analysis of homologous synteny blocks and evolutionary breakpoint regions (EBRs) of chromosomes to test the hypothesis that there is an enrichment for genes that correspond to known phenotypic characteristics of the species in question. ....</b>	<b>156</b>
<b>6.1</b>	<b>Background.....</b>	<b>156</b>
<b>6.2</b>	<b>Specific Aims.....</b>	<b>160</b>
<b>6.3</b>	<b>Materials and Methods.....</b>	<b>160</b>
<b>6.4</b>	<b>Results .....</b>	<b>161</b>
6.4.1	Specific Aim 4a: To use bioinformatic tools to recreate the most likely ancestral avian karyotype through analysis of 6 avian genomes and to infer the evolution of chromosomal rearrangements from the divergence of the avian ancestor to extant avian karyotypes .....	161
6.4.2	Specific aim 4b: Gene Ontology Analysis of Avian microchromosomes.....	166

6.4.3	Specific Aim 4c: To use the cytogenetic tools (sequence conserved BACs) developed in chapter 4 to investigate genome conservation between avian and non-avian reptiles	167
6.4.4	Specific aim 4d: To use a similar approach to specific aim 4a to infer the arrangement of saurian ancestor chromosomes and trace the evolution of chromosomal rearrangements from the divergence of the saurian ancestor, to the appearance of the common avian ancestor and from there onwards to the extant chicken lineage.....	171
6.4.5	Specific aim 4e: To analyse gene ontology terms in the HSBs and EBRs of this dinosaur lineage to test the hypothesis that these relate to phenotypic characteristics associated with the dinosaurs .....	178
<b>6.5</b>	<b>Discussion .....</b>	<b>183</b>
6.5.1	Reconstruction of Ancestral Genomes .....	183
6.5.2	Ostrich Genome Inconsistencies .....	184
6.5.3	Chicken is Most Closely Related to the Dinosaur Ancestor .....	184
6.5.4	The Saurian Ancestor.....	185
6.5.5	Genome Evolution from 275 mya to 100 mya .....	187
6.5.6	Gene Ontology Analysis.....	189
6.5.7	Conclusions.....	191
<b>7</b>	<b>General Discussion .....</b>	<b>193</b>
7.1.1	Tools for Studying Chromosome Rearrangements in Mammals and Birds (Chapters 3-4).....	194
7.1.2	Formation of the Signature Avian Karyotype (Chapters 4-6) .....	196
7.1.3	Future Work: Mammals and Beyond .....	200
7.1.4	Personal Perspectives.....	203
<b>8</b>	<b>References: .....</b>	<b>204</b>
<b>9</b>	<b>Appendix.....</b>	<b>231</b>
<b>10</b>	<b>Publications arising from this work .....</b>	<b>242</b>

## List of Figures

Figure 1-1: Stages of mitosis, starting from interphase where DNA has replicated through to cytokinesis when two daughter cells are produced (© Clinical Tools Inc).....	1
Figure 1-2: Representative G-banded Karyotypes for (a) Pig, (b) Human, (c) American Rhea and (d) Mouse. ....	2
Figure 1-3: Chromosome banding revealed by a variety of staining techniques. (a) G-banded human chromosomes; (b) Q-banded human chromosomes; (c) R-banded human chromosomes; (d) C-banded human chromosomes (O'Connor C. 2008). ....	4
Figure 1-4: The principles of fluorescence <i>in situ</i> hybridisation (FISH). ....	6
Figure 1-5: Telomeric and sub-telomeric sequence organisation (Knight and Flint, 2000).....	7
Figure 1-6: Chromosome painting of gibbon chromosome paints onto human chromosomes (Ferguson-Smith & Trifonov, 2007). ....	8
Figure 1-7: Dual colour FISH of GGA macrochromosome paints GGA6 (red) and GGA7 (green) on <i>A. roseicollis</i> chromosomes (Nanda et al. 2007). ....	9
Figure 1-8: The principle behind the multiprobe device for use in FISH (Cytocell). ....	12
Figure 1-9: 24 colour FISH tested on an interphase nucleus derived from a human lymphocyte (Ioannou et al. 2011) .....	13
Figure 1-10: Schematic representation of array-CGH, where target and reference DNA are labelled and hybridised together with the addition of COT-1 DNA to block repetitive sequences, prior to analysis of the signal intensities produced by the sample of interest, allowing for quantification of DNA and copy number (Feuk et al. 2006). ....	14
Figure 1-11: Schematic representation of Sanger sequencing (Mardis 2013). ....	16
Figure 1-12: Schematic representation illustrating the differences between (a) BAC-by-BAC or hierarchical sequencing and (b) Shotgun sequencing (Commins et al. 2009). ....	22
Figure 1-13: Combined PacBio sequencing and third generation mapping to improve sequence contiguity (dnanexus.com). ....	25
Figure 1-14: Sequence Coverage of the Avian Phylogenetic Tree (Zhang G. et al. 2014a). ....	31
Figure 1-15: Overview of the RACA Algorithm. ....	35
Figure 1-16: Syntenic blocks between human and mouse superimposed on to the mouse genome, illustrating a significant level of genomic rearrangement (Waterston et al. 2002). ...	36
Figure 1-17: Screenshot of Evolution Highway, showing the alignments of multiple mammalian genomes against human chromosome 16 with the grey blocks representing HSBs and the chromosome number for each species listed against each block (Murphy et al. 2005). ....	37

Figure 1-18: Screenshot from Genomicus demonstrating how the gene PHOX2B compares to other closely related species with (A) showing PhyloView which illustrates the gene in a phylogenetic context and (B) showing AlignView which illustrates the PHOX2B and its orthologues in a greater number of species (Muffato et al. 2010). ..... 38

Figure 1-19: Vertebrate phylogenetic tree illustrating whole genome duplications and lineage specific chromosome fission and fusion events that occurred across the vertebrates (adapted from Nakatani et al. 2007). ..... 43

Figure 1-20: Chromosomal inversion ..... 44

Figure 1-21: Distribution of chromosomal inversions between humans (on the left) and the chimpanzee homolog (on the right) inferred from sequence comparisons (Feuk et al. 2005).. 46

Figure 1-22: Structural chromosome rearrangements. (a) Reciprocal translocation, illustrating exchange of genetic material between non-homologous chromosomes; (b) Robertsonian translocation. .... 47

Figure 1-23: Meiosis in an individual carrying a reciprocal translocation illustrating formation of a quadrivalent, subsequently leading to incorrect segregation during anaphase which results in unbalanced gametes (O'Connor C, 2008). ..... 49

Figure 1-24: Structural Chromosome Rearrangements: (a) Deletion; (b) Duplication..... 50

Figure 1-25: Different methods of diagnosing Cri du Chat syndrome. (a) The G-banded karyotype on the left has an arrow marking the small deletion seen on the p-arm of chromosome 5. (b) The FISH image illustrates the normal chromosome on the left with two signals and the abnormal chromosome on the right with an absence of a p-arm (green) signal illustrating the deletion (Sun et al. 2014). (c) Array-CGH output illustrating the deletion of DNA at the p-arm of chromosome 5 (the X axis represents the distance in Mb from the p-terminus and the Y axis represents the hybridisation ratio (Rickman et al. 2005). ..... 51

Figure 1-26: Chromosomal number variety in vertebrates. Diploid number is represented on the x-axis, with the y-axis indicating species in the different orders (Ruiz-Herrera et al. 2011). ..... 53

Figure 1-27: Variation in diploid chromosome number. a) Metaphase spread of the Indian Muntjac (*Muntiacus muntjak*) where  $2n=6/7$ . b) Metaphase spread of the Viscacha rat (*Tympanoctomys barrerae*) where  $2n=102$  (Graphodatsky et al. 2011). ..... 54

Figure 1-28: Typical avian karyotype organisation with few macrochromosomes and many, morphologically indistinguishable microchromosomes. From the top down, chicken, Japanese quail, turkey and duck are represented (Schmid et al. 2005). ..... 55

Figure 1-29: Giemsa stained karyotype of the Nile crocodile (*Crocodylus niloticus*; 2n=32) with chicken homology by chromosome number listed to the right of the chromosome (Kasai et al. 2012). ..... 56

Figure 1-30: Karyotype of the Burmese python (*Python morolus*; 2n=36) a) Giemsa stained karyotype b) C-banded sex chromosomes (Matsubara et al. 2006). ..... 56

Figure 1-31: Giemsa stained karyotype of the Chinese soft-shelled turtle (*Pelodiscus sinensis*; 2n=66) (Matsuda et al. 2005). ..... 57

Figure 1-32: A comparison of genomic elements, genome size, karyotype number and sex determination systems in reptiles and mammals. The width of clades is illustrative of species diversity (Janes et al. 2010). ..... 62

Figure 1-33: Schematic representation of HSBs and EBRs (adapted from Farré et al. 2011). .... 66

Figure 1-34: Speciation theory models illustrated using the example of an inversion. When (a) and (b) hybridise the F1 hybrid is heterokaryotypic. (c) illustrates the traditional speciation theory (hybrid dysfunction model) leading to underdominance. (d) illustrates the suppressed recombination model which states that suppressed recombination in the heterokaryotype will lead to (e) increased nucleotide rate change or f) the capture of alleles allowing ecotypic adaption in that region (Brown & O'Neill, 2010). ..... 70

Figure 1-35: Ancestral avian macrochromosome karyotype with chicken orthologues highlighted in red text (adapted from Griffin et al. 2007). ..... 72

Figure 1-36: A time tree of amniotes (adapted from Shedlock and Edwards, 2009; Brusatte et al. 2011 and Benton et al. 2014). ..... 75

Figure 1-37: *Archaeopteryx lithographica* fossil showing clear evidence of feathered wings. ... 77

Figure 1-38: The revised avian phylogenetic tree illustrating the phylogenetic placement of avian taxa and radiation time points before and after the K-Pg extinction event 66 mya (Jarvis et al. 2014). ..... 79

Figure 1-39: Schematic illustration of karyotype evolution among amniotes (Uno et al. 2012). 81

Figure 2-1: Schematic representation of probe on the porcine multiprobe device with each square containing a FITC labelled p-arm probe and a Texas Red labelled q-arm probe for each chromosome. .... 92

Figure 2-2: Schematic representation of probe layout for the avian multiprobe device with each square containing a FITC labelled p-arm probe and a Texas Red labelled q-arm probe for each of the microchromosomes with the shaded squares in the middle row containing macrochromosome paints labelled in multiple colours as represented by the font colour in the diagram. .... 93

Figure 2-3: Example of Evolution Highway alignment output that produced 96 msHSBs in four (zebra finch, duck, opossum and anole Lizard) genomes as aligned against GGA1 totalling 137.47 Mb out of 195.28 Mb (or ~70%)..... 96

Figure 3-1: Standard DAPI banded karyotype of boar carrying a 7:10 RCP. The affected chromosomes are highlighted with arrows..... 103

Figure 3-2: Clone ID PigE-108N22 labelled in Texas Red which should map to the distal end of SSC15 but appears halfway along this acrocentric chromosome. The FITC labeled probe mapped correctly. Scale bar 10µm..... 104

Figure 3-3: Screenshot from NCBI clone finder illustrating where BAC PigE-108N22 was expected to map on SSC15. .... 105

Figure 3-4: Labelled probes for SSC7 illustrating a reciprocal translocation between SSC7 and SSC10. Scale bar 10µm. .... 107

Figure 3-5: Labelled BAC clones for SSC5 (p-arm labelled in FITC and q-arm labelled in Texas Red) showing a translocation between chromosome 5 and 6 despite suboptimal chromosome preparation. Scale bar 10µm..... 109

Figure 3-6: Chromosome paints for SSC5 (Texas Red) and SSC6 (FITC) illustrating the cryptic translocation that had been previously undetectable from the karyotype. Scale bar 10µm. . 109

Figure 3-7: DAPI banded cattle karyotype (*Bos taurus*) where 2n=60 (Larkin et al. 2014). .... 110

Figure 3-8: Subtelomeric cattle probes for BTA24 p-arm labelled in FITC (CH240-382F1) and q-arm labelled in Texas Red (CH240-19L13). Scale bar 10µm. .... 112

Figure 4-1: Example of dual FISH results for Chicken (*Gallus gallus* – GGA) chromosome 12 to confirm correct mapping. The p-arm BAC (CH261-88K1) is labelled with FITC and the q-arm BAC (CH261-152H14) is labelled with Texas Red. Scale bar 10µm. .... 118

Figure 4-2: Chicken (*Gallus gallus* - GGA) macrochromosome paints for chromosomes 2 (FITC), 5 (Texas Red) and 8 (aqua) tested on the peafowl (*Pavo cristatus*) illustrating a fusion revealed by the aqua paint where the paint only covers around half of the chromosome. *Scale bar* 10µm. .... 120

Figure 4-3: Macrochromosome GGA paints 1 (FITC), 3 (Aqua) and 4 (Texas Red) hybridised to Chinese quail chromosomes illustrating the fused chromosome four. *Scale bar* 10µm..... 122

Figure 4-4: Chicken microchromosome probes (GGA23p - CH261-191G17 and GGA23q-arm - CH261-90K11) revealing a fusion to a macrochromosome in the peregrine falcon (*Falco peregrinus*- FPE). *Scale bar* 10µm..... 123

Figure 4-5: Microchromosomal conservation observed across a wide range of avian species as revealed by testing BACs from chicken chromosome 24 (CH261-103F4 FITC and CH261-65O4 Texas Red). Scale bar 10µm. .... 130

Figure 4-6: Ideogram for 12 of the 18 species tested here (duck, pigeon, ostrich, turkey, houbara, pharaoh eagle owl, collared dove, Eurasian woodcock, Japanese quail, Chinese quail, guinea fowl, blackbird and canary), illustrating microchromosomal conservation across all 12 species plus representation of the macrochromosomal rearrangements previously published and confirmed here (indicated with a red arrow for a fission and blue for a fusion). .... 132

Figure 4-7: Hybridisation of GGA11 BACs (CH261-121N21-Fitc and CH261-154H1-Texas red) to cockatiel (*Nymphus hollandicus*) metaphases illustrating fusion of ancestral microchromosome to a macrochromosome. Scale bar 10µm. .... 133

Figure 4-8: Ideogram representing the budgerigar (*Melopsittacus undulatus* -MUN) karyotype, illustrating a high degree of interchromosomal rearrangement in the macrochromosomes and a high degree of microchromosomal conservation with the exception of 3 microchromosome homologs (GGA10, 11, 13 and 14). .... 133

Figure 4-9: Ideogram representing karyotypic structure of the cockatiel (*Nymphus hollandicus* - NHO) illustrating an overall structure that is broadly in alignment with the budgerigar but with one less microchromosome fusion and fewer interchromosomal rearrangements in the macrochromosomes. .... 134

Figure 4-10: Hybridisation of GGA18 BACs (CH261-60N6-Fitc and CH261-72B18-Texas red) to peregrine falcon (*Falco peregrinus*) metaphases illustrating fusion of ancestral microchromosome to a macrochromosome. Scale bar 10µm. .... 135

Figure 4-11: Ideogram representing overall karyotypic structure of the peregrine falcon (*Falco peregrinus* - FPE) illustrating an extensive amount of interchromosomal rearrangement throughout the karyotype. .... 136

Figure 5-1: Ideogram representation of pigeon (CLI) chromosomes and their chicken (GGA) homologs. .... 146

Figure 5-2: Cytogenetic and PCF mapping of pigeon chromosome 2 (CLI2) using FISH; (a) Evolution highway alignment of zebra finch, chicken and pigeon genomes alongside the PCFs produced by RACA and the BACs that map in this region; (b) Cytogenetic map of BACs in the correct orientation on CLI2; (c) Physical mapping of BACs to CLI2 using FISH; (d) Ideogram illustration of GGA homolog to CLI2. .... 147

Figure 5-3: Ideogram representation of peregrine falcon (FPE) chromosomes and their chicken (GGA) homologs. .... 148

Figure 5-4: Cytogenetic and PCF mapping of peregrine falcon chromosome 5 (FPE5) using FISH; (a) Evolution highway alignment of zebra finch, chicken and peregrine genomes alongside the PCFs produced by RACA and the BACs that map in this region; (b) Cytogenetic map of BACs in the correct orientation on FPE5; (c) Physical mapping of BACs to FPE5 using FISH; (d) Ideogram illustration of GGA homologs to FPE5. .... 150

Figure 5-5: Screenshot from Evolution Highway illustrating the uploaded assembled peregrine falcon genome for chromosomes 1 to13 (<http://eh-demo.ncsa.uiuc.edu/birds-test/#/SynBlocks>). .... 151

Figure 6-1: Evolution Highway screenshot of 5 avian genomes and the outgroup (anole lizard) aligned to chromosomes 5 and 11 of the reference genome - the chicken. Pink blocks are indicative of an inversion compared to the reference (chicken) genome and numbers in the blocks reflect the chromosome number of that species or the scaffold number in the budgerigar and the ostrich. .... 161

Figure 6-2: Avian ancestor chromosome 5 and the likely rearrangements that have occurred along each lineage to the extant bird. Rainbow patterned arrows within the chromosomes represent the HSBs, red curved arrows indicate chromosome inversions, blue arrows indicate chromosome translocations, green outline indicates previous chromosome painting results (anole lizard arrangement also indicated). .... 162

Figure 6-3: Avian ancestor chromosome 11 and the likely rearrangements that have occurred along each lineage to the extant bird. As the arrangement for ostrich and Neognathae ancestors were the same, the avian ancestor could be derived (unlike for other chromosomes smaller than 5). \* in Budgerigar, FISH indicates fusion to a larger chromosome. Colour scheme and pattern consistent with Figure 6-2. .... 163

Figure 6-4: Number of chromosomal inversions that most parsimoniously explain the patterns seen in the 6 extant species as they diverged from the ancestor for chromosomes 1-5 (using the lizard outgroup). The greatest rates of change were seen in zebra finch and budgerigar. The phylogenetic is tree based on Jarvis et al. (2014). .... 165

Figure 6-5: Number of chromosomal inversions that most parsimoniously explain the patterns seen in the 5 extant species as they diverged from the ancestor for chromosomes 6-28 + Z (using the ostrich as an outgroup). .... 166

Figure 6-6: Dual colour FISH of GGA24 probes hybridised to 3 phylogenetically distant avian species (a) *Gallus gallus*, (b) *Falco peregrinus*, (c) *Melopsittacus undulatus* and two reptile species (d) *Anolis carolinensis* and (e) *Trachemys scripta*, illustrating an extraordinary degree of genome conservation. Scale bar 10µm. .... 169



Figure 6-7: Dual colour FISH of GGA26 illustrating a microchromosomal pattern in three representative avian species (a) *Columba livia*, (b) *Coturnix japonica*, (c) *Cyanoamphus novaezelandia* and the position of the same BACs hybridising to macrochromosomes of (d) *Anolis carolinensis* and (e) *Trachemys scripta* suggesting a macrochromosomal origin of the ancestral avian microchromosome 26. Scale bar 10µm. .... 170

Figure 6-8: Screenshot of pairwise alignments for zebra finch, duck, opossum and anole lizard aligned against chicken chromosome 5 using Evolution Highway. .... 171

Figure 6-9: HSB coverage relative to the chicken genome where blue bars represent GGA chromosome length and the red bars represent HSB coverage of each chromosome. .... 172

Figure 6-10: Ideogram of saurian ancestor CARs (SAA) derived from MGRA2 with their GGA homologs illustrating interchromosomal rearrangement of GGA homologs relative to hypothetical ancestor. .... 173

Figure 6-11: Intrachromosomal rearrangements between the saurian ancestor chromosome 1 (at the top of the image) and chicken chromosome 2 illustrating 5 inversions that have occurred between the two. The numbers in each block represent the msHSBs that have stayed together as a larger block of synteny without evidence of intra or interchromosomal rearrangement and a minus symbol indicates a change of direction. .... 174

Figure 6-12: Intrachromosomal rearrangements between the saurian ancestor chromosome 4 (at the top of the image) and chicken chromosome 4 illustrating 9 inversions that have occurred between the two. .... 175

Figure 6-13: Inter and intrachromosomal rearrangements between saurian ancestor chromosomes 6 and 7 resulting in a reciprocal translocation occurring between the two (highlighted in yellow) resulting in chromosomes 5 and 20 in the chicken. .... 177

Figure 6-14: Illustration of body size change among the theropod lineage from their emergence around 235mya to the divergence of modern birds (Lee et al. 2014). .... 191

Figure 7-1: Marketing material created by Cytocell for the Porcine multiprobe device. .... 195

Figure 7-2: Marketing material created by Cytocell for the Chicken multiprobe device. .... 196

Figure 7-3: Changes to diploid number along archosauromorpha lineage to birds. .... 197

Figure 7-4: Haploid genome sizes illustrating the significantly smaller genome sizes along the theropod lineage originating early in the evolution of archosaurs (Organ et al. 2007). .... 198

Figure 7-5: Dovetail dog scaffolds aligned to the reference assembly of dog (CFA – *Canis familiaris*) chromosomes 23-25. Numbers inside chromosomes represent scaffold IDs with black lines to the right of each chromosome indicating putative positions of BACs selected every 5 Mb

RE O'Connor

suggesting that the majority of Dovetail scaffolds could be ordered and oriented along the chromosomes. Image was provided by Harris Lewin (University of California, Davis). ..... 202

## List of Tables

Table 1-1: Summary of Zoo-FISH studies carried out to date on avian species. For the most part they involve hybridisation of chromosome paints 1-10+Z (Griffin et al. 1999) on to the chromosomes of other avian species.....	10
Table 1-2: Comparison of NGS sequencing technologies (adapted from Liu et al. 2012) .....	20
Table 1-3: Recently sequenced avian genomes with corresponding N50 sizes, genome sizes and total number of genes per species (Zhang G. et al. 2014b). .....	32
Table 1-4: Avian species that exhibit a fusion on chromosome 4 as revealed by chromosome painting with chicken macrochromosomal paints. ....	71
Table 3-1: Summary results from boar karyotyping for translocation screening.....	102
Table 3-2: Incorrectly mapped porcine BACs. ....	106
Table 3-3: Correctly mapped BACs for each porcine chromosome arm. ....	108
Table 3-4: Correctly mapped subtelomeric cattle BACs by chromosome from the CHORI-240 library. ....	111
Table 4-1: Chicken ( <i>Gallus gallus</i> - GGA) BACs selected from the CHORI-261 library for each chicken microchromosome with start and end positions in the chicken genome and the size of the BAC listed. ....	119
Table 4-2: Successful cross species hybridisations of chicken microchromosome BACs within the <i>Galliformes</i> order. Clear signals are denoted by an '✓' and 'No' where no signal was evident. ....	121
Table 4-3: Cross species hybridisation results for GGA subtelomeric microchromosome BACs on avian species tested beyond the <i>Galliformes</i> order.....	124
Table 4-4: Successful hybridisation results across 5 avian species using selected and non-selected BACs derived from the CHORI chicken BAC library. ....	125
Table 4-5: Successful hybridisation results across 5 avian species using selected and non-selected BACs derived from the TGMCBa Zebra finch BAC library.....	126
Table 4-6: Comparison of overall FISH success rates for selected versus non-selected BACs across 5 avian species. ....	126
Table 4-7: List of avian species tested using two selected BACs per reference (chicken) microchromosome. ....	128
Table 4-8: BACs chosen for multiple cross species analysis. Two BACs were chosen for each chromosome using the chicken genome as the reference. ....	129

Table 5-1: Statistics for the chromosome level assemblies of peregrine falcon ( <i>Falco peregrinus</i> ) and the pigeon ( <i>Columba livia</i> ). .....	145
Table 5-2: Lineage specific rearrangements detected in the pigeon ( <i>Columba livia</i> ) when compared to the chicken genome. ....	147
Table 5-3: Lineage specific rearrangements detected in the peregrine falcon ( <i>Falco peregrinus</i> ) when compared to the chicken genome.....	149
Table 6-1: Total numbers of inter- and intrachromosomal rearrangements in each species since their divergence from the avian ancestor 100 mya. ....	164
Table 6-2: Hybridisation success by GGA chromosome using GGA BACs on <i>Anolis carolinensis</i> (anole lizard) and <i>Trachemys scripta</i> (red-eared slider turtle). Clear signals are denoted by an '✓' and 'No' where no signal was evident; 'Proto' denotes GGA microchromosome BACs fused to a macrochromosome. ....	168
Table 6-3: Size in base pairs (bp) of each saurian ancestral CAR. ....	172
Table 6-4: Total number of inter and intrachromosomal rearrangements between the saurian ancestor and the avian ancestor and the saurian ancestor and the extant chicken genome..	176
Table 6-5: Number of the type of chromosomal rearrangement that occurred between the saurian ancestor, the avian ancestor and the extant chicken genome.....	177
Table 6-6: Sequence variants affecting diversity of height identified as having significant enrichments in the EBRs of the saurian karyotype (adapted from Gudbjartsson et al. 2008).	182

## Abbreviations

ACA	<i>Anolis carolinensis</i>
BAC	Bacterial Artificial Chromosome
BTA	<i>Bos taurus</i>
CAR	Contiguous ancestral region
CLI	<i>Columba livia</i>
DAPI	4',6 diamidino 2 phenylindole
ddH <sub>2</sub> O	Double Distilled Water
DTT	Dithiothreitol
EBR	Evolutionary Breakpoint Region
EH	Evolution Highway
FISH	Fluorescence <i>in situ</i> Hybridisation
FITC	Fluorescein isothiocyanate
FPE	<i>Falco peregrinus</i>
GGA	<i>Gallus gallus</i>
GO	Gene Ontology
GRIMM	Genome Rearrangements In Man and Mouse
HSB	Homologous Synteny Block
KCl	Potassium chloride
LB	Luria Bertani
MBG	Molecular Biology Grade
MGR	Multiple Genome Rearrangements
MGRA	Multiple Genome Rearrangements and Ancestors
msHSB	Multispecies Homologous Synteny Block
mya	Million years ago
PBS	Phosphate Buffered Saline
PCF	Predicted Chromosome Fragment
PCR	Polymerase Chain Reaction
RACA	Reference Assisted Chromosome Assembly
RCP	Reciprocal translocation
RVC	Royal Veterinary College, London
SF	Syntenic fragments
SSC	<i>Sus scrofa</i>
TSC	<i>Trachemys scripta</i>
UKC	University of Kent

## Abstract

Chromosomal (karyotypic) analysis in animals is performed for three primary reasons: to diagnose genetic disease; to map genes to their place in the genome and to retrace evolutionary events by cross species comparison. Technology for analysis has progressed from chromosome banding (cytogenetics), to fluorescence *in-situ* hybridisation (FISH - molecular cytogenetics) through to microarrays and ultimately whole genome sequence analysis (cytogenomics or chromonomics). Indeed, the past 10-15 years has seen a revolution in whole genome sequencing, first with the human genome project, followed by those of key model and agricultural species and, more recently, ~60 *de novo* avian genome assemblies. Whole genome analysis provides detailed insight into the biology of chromosome rearrangements that occur both in individuals (for diagnostic purposes) and at an evolutionary level. It permits the study of gene mapping, trait linkage, phylogenomics, and gross genomic organisation and change. An essential pre-requisite however is an unbroken length of contiguous DNA sequence along the length of each chromosome. Most recent *de novo* genome assemblies fall short of this level of resolution producing lengths of contiguous sequence that are sub-chromosomal in size (scaffolds).

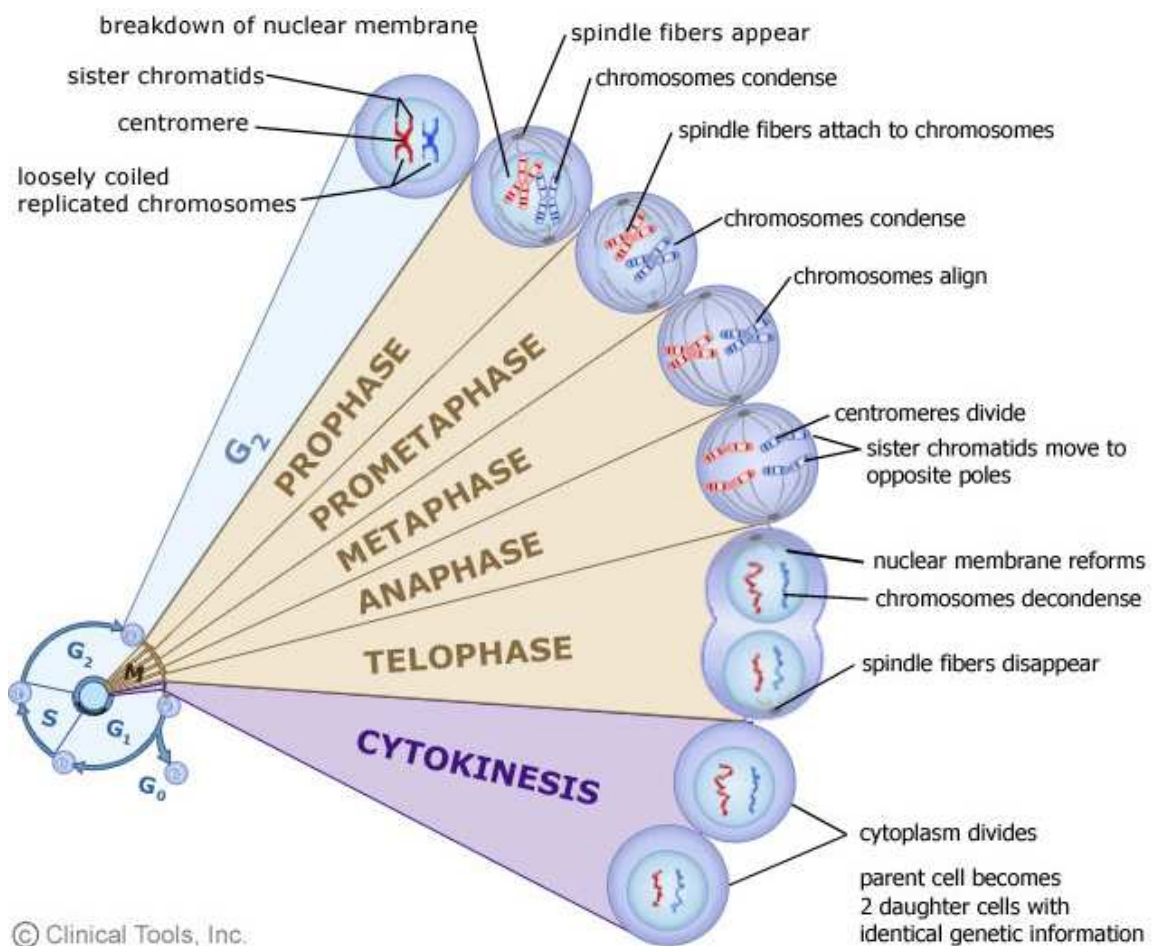
Chromosome rearrangements can affect reproductive capability at an individual level (causing reduced fertility) and at a population level leading to reproductive isolation and subsequent speciation. The purpose of this thesis was to implement a step change in the combination of FISH technology with genome sequence data to provide greater insight into the nature of chromosomal rearrangement at an individual and evolutionary level. It therefore had four specific aims:

The first was to isolate sub-telomeric sequences from the pig, cattle and chicken genome assemblies to develop a tool for the rapid screening of chromosome rearrangements. Now routinely used for porcine translocation screening (and in the future bovine screening), development work revealed serious integrity errors in the pig genome. The second aim was to isolate evolutionary conserved sequences from avian chromosomes to create a means of screening for macro- and microchromosomal rearrangements in birds. Results confirmed the hypothesis that microchromosomal rearrangements were rare in birds, except for previously known whole chromosomal fusions. The third was to use the above tools to complete scaffold based genome assemblies in two key avian species - the peregrine falcon and the pigeon. Finally, bioinformatic tools were used to infer the overall genome structure of hypothetical saurian and avian ancestors. Retracing of the evolutionary changes that occurred up until the emergence of birds allowed an assessment of chromosome evolution along the saurischia-maniraptora-avialae lineage. Analysis of evolutionary breakpoint regions (EBRs) allowed testing of the hypothesis that the ontology of genes within EBRs corresponded to measurable phenotypic change in the lineage under investigation. An enrichment of genes associated with body height corresponded to rapid size change in the dinosaur lineage that led to modern birds.

Taken together, these results paint a picture of a genome that, from about 260 million years ago formed a 'signature' highly successful avian-dinosaur karyotype that remained largely unchanged interchromosomally to the present day. These results represent significant insight into amniote genomic organization with the added benefit of developing tools that are widely applicable and transferrable for commercial animal breeding, for constructing *de novo* genome assemblies and for reconstructing, by inference, the overall genomic structure and evolution of extinct animals.

## 1 Introduction

Eukaryotic cells are characterised by the ability to package their DNA into coiled structures (chromosomes) to facilitate efficient packing, and subsequent segregation of identical copies of the genome to daughter cells during mitosis (as illustrated in Figure 1-1).

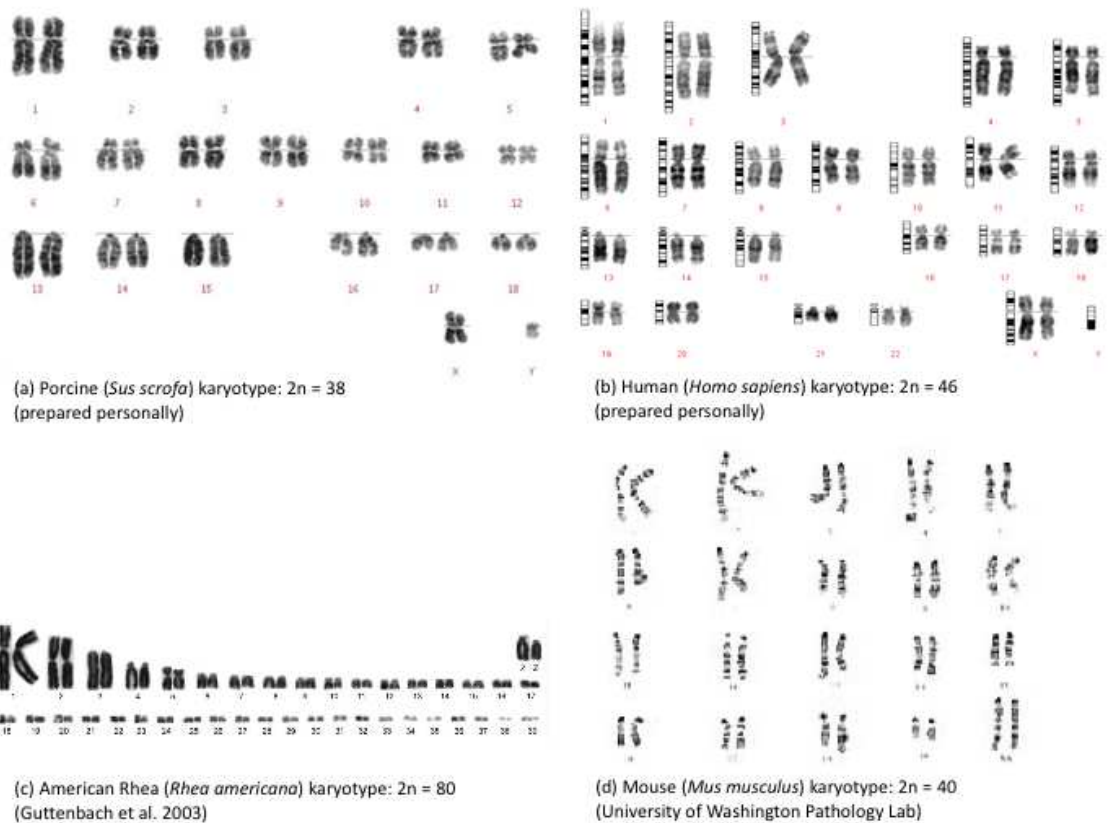


**Figure 1-1:** Stages of mitosis, starting from interphase where DNA has replicated through to cytokinesis when two daughter cells are produced (© Clinical Tools Inc).

Remarkably, this pattern of coiling is thought to be identical in virtually every cell division and remains consistent from individual to individual within the same species, with rare exceptions usually being indicative of a mutation. These species-specific patterns to chromosome coiling and segregation can be visualised as a 'karyotype'. Karyotypes can be thought of as a low resolution signature of an organism and can be prepared by inducing virtually any cell from the body to divide, arresting dividing cells in metaphase, swelling by osmosis, fixation to a glass slide then staining for visualisation purposes (for more details, see next section). In Figure 1-2 the



karyotypes of several species (pig, human, bird and mouse) are illustrated. In each case, an image of a stained metaphase preparation is taken, photographed and the chromosomes arranged according to a convention agreed by the scientific community.



**Figure 1-2: Representative G-banded Karyotypes for (a) Pig, (b) Human, (c) American Rhea and (d) Mouse.**

There are three primary reasons for studying a karyotype: the first is that deviations from the species norm are usually indicative of disease. For this reason, the human karyotype is more intensively studied than any other species, with catalogues of chromosome disorders being beyond the scope of even very large volumes of print-based media (Borgaonkar 1975; Schinzel 2001). The second is for the purposes of genomic mapping: in every eukaryotic species, each gene in the genome is located with reference to the chromosome on which it resides, the arm of the chromosome and the appropriate distance from the chromosome end; for example, the cystic fibrosis transmembrane regulator gene is defined as being 7q31, this means that it is found on the long (q) arm of human chromosome 7, major band 3, minor band 1 (Rommens et al. 1989). The third is from the perspective of genome evolution (Ferguson-Smith & Trifonov 2007): given that each species has a unique karyotype and that more closely related species usually have similar karyotypes to one another, any changes in the karyotype, and therefore

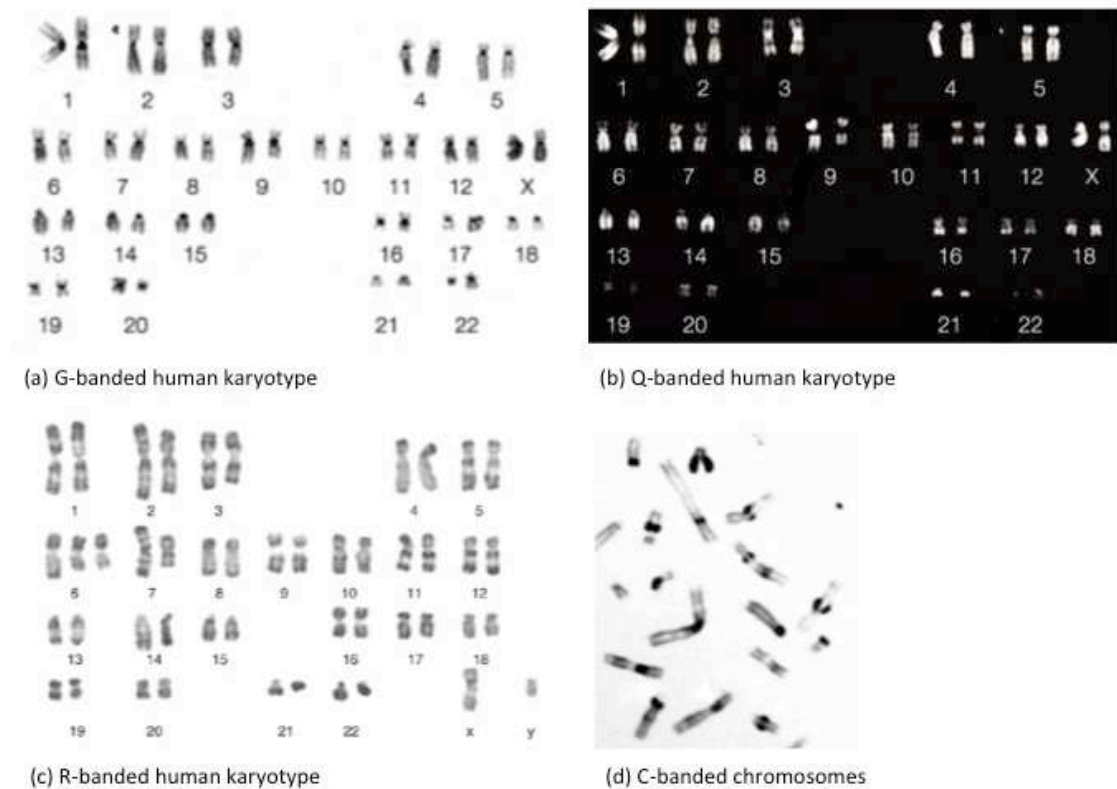
overall genome structure, represent the process of gross genomic evolution. Indeed, changes at a karyotypic level can impose species barriers, in which hybrids of species with two different karyotypes are compromised in their ability to reproduce, resulting in reproductive isolation (Brown & O'Neill 2010). An analogous effect can be seen in humans (and other animals) who are heterozygous for a chromosome rearrangement (e.g. a translocation or inversion) and, although phenotypically normal (as there is no gain or loss of DNA) display reduced fertility (Griffin & Finch 2005). Chromosomal changes that cause reproductive problems in individual animals therefore may also cause reproductive isolation at a species level. Furthermore, changes in gene order can have phenotypic consequences by bringing together new gene combinations and separating previously established ones (Larkin et al. 2009).

This thesis is concerned with aspects of karyotypic (cytogenetic) screening for compromised fertility, molecular cytogenetic mapping in a range of species and tracing of gross genomic (karyotype) evolution in extant and extinct species.

## **1.1 Methods to study chromosomes and their rearrangements**

### **1.1.1 Classical Cytogenetics**

As mentioned in the previous section, in order to create a karyotype, cultured cells are arrested at the metaphase stage of mitosis using a solution of the spindle disrupting agent colcemid prior to swelling the cells osmotically (usually with 75mM KCl). At this point the cells are treated with fixative (typically 3:1 methanol:glacial acetic acid) and are then dropped onto a glass slide, recently reviewed in O'Connor C. (2008). The cells are stained in order to differentiate certain areas of the chromosome producing the characteristic banding appearance seen in a karyotype. Staining techniques, as shown in Figure 1-3 include G-banding (staining with Giemsa), Q-banding (staining with the fluorescent dye quinacrine, Hoescht 33258 or DAPI – 4',6-diamidino-2-phenylindole), R-banding (reverse Giemsa staining), T-banding (identification of the R bands closest to the telomeres) or C-banding (similar to T-banding, but focused on the heterochromatin rich areas around the centromere) (Shaffer & Tommerup 2005).



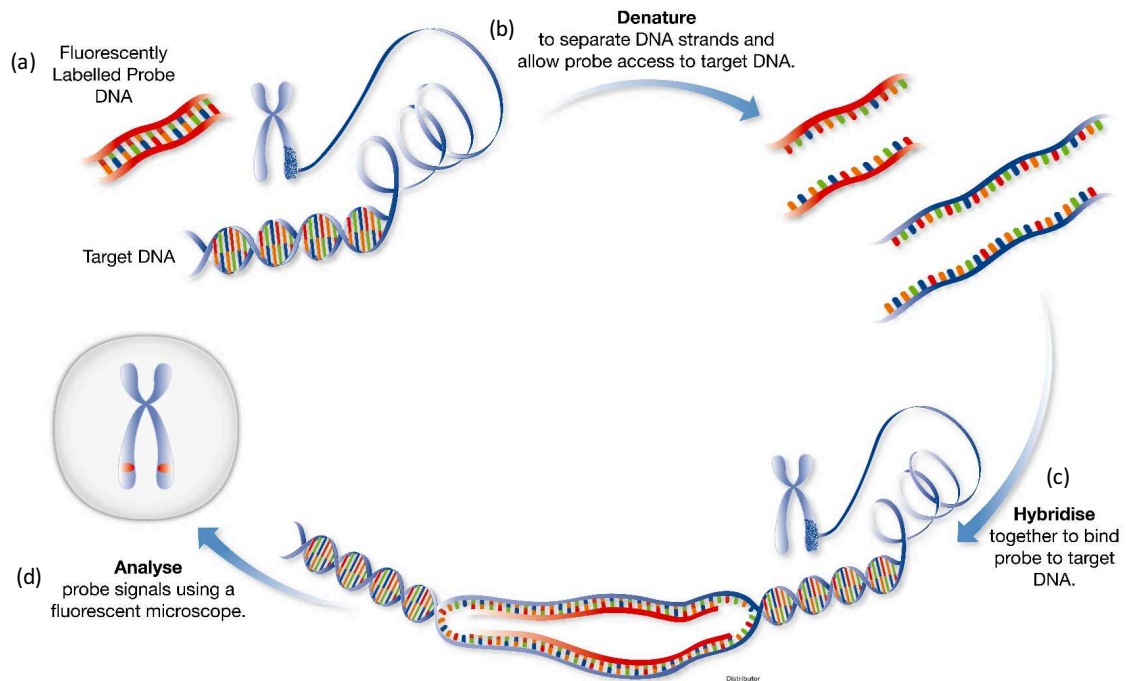
**Figure 1-3: Chromosome banding revealed by a variety of staining techniques. (a) G-banded human chromosomes; (b) Q-banded human chromosomes; (c) R-banded human chromosomes; (d) C-banded human chromosomes (O'Connor C. 2008).**

Traditionally, genome analysis and translocation screening is performed by G-banding followed by routine karyotyping. While this is simple and cost effective, it requires specialist knowledge of the karyotype of interest and is limited in its ability to detect rearrangements smaller than 5-10 Mb in size, particularly if bands of similar intensity are exchanged (Martin & Warburton 2015). Since the landmark discoveries in 1959 of the link between Down syndrome and the presence of an additional chromosome 21 by Lejeune and colleagues (Lejeune et al. 1959) along with the finding by Jacobs and Strong (1959) that an additional X chromosome leads to Klinefelter syndrome, karyotyping has been extensively used as a means of screening for chromosomal abnormalities. The identification of the Philadelphia chromosome (the genetic cause of chronic myeloid leukaemia-CML) soon after by Nowell and Hungerford (1960) followed by the introduction of Q-banding and G-banding for higher resolution analysis by Janet Rowley in 1973 (Rowley 1973) led to the identification of thousands of other chromosomal rearrangements associated with cancer (Mitelman 2005). The karyotype rapidly became a vital tool in the clinic extending to prenatal diagnosis of aneuploidies and many other inherited conditions (Trask 2002).

In the following years, while clinical diagnosis of genetic disorders by karyotyping became routine for human disease (Borgaonkar 1975; Schinzel 2001) it took until the latter part of the 20th century for it to become established in non-human animals, with pigs and cattle being the most studied species (Ducos et al. 2008). Since then, several continental European programmes of chromosomal screening have been established with a particular focus on screening for translocations in boars and bulls, with perhaps the best known translocation being the 1:29 Robertsonian translocation in bulls (Gustavsson, 1979). This has led to the identification of a significant number of further chromosomal rearrangements in otherwise phenotypically normal boars and bulls as well as leading to the birth of so called animal 'clinical' cytogenetics. The largest centre of animal chromosome screening is based in the National Veterinary School of Toulouse, France, however since the turn of the century there has been a gradual reduction in the number of laboratories that perform animal cytogenetics (with approximately 10-15 operating worldwide, mostly in Europe) (Ducos et al. 2008).

### **1.1.2 Molecular Cytogenetic Analysis using Fluorescence *in situ* Hybridisation**

At a higher resolution, molecular cytogenetic techniques such as FISH (fluorescence *in situ* hybridisation) enable specific regions of DNA to be identified either at metaphase or interphase, thereby allowing the identification of subtle translocations or mutations that may otherwise not be visible at the karyotype level. In order to visualise the required DNA region, a probe is created by selecting a clone, typically a BAC (bacterial artificial chromosome) that maps to the region of interest on the chromosome. The BAC (within a bacterial host) is grown and purified before labelling with a fluorophore either directly using nick translation or indirectly using PCR. The probe is then denatured alongside the target DNA, then hybridised to fixed metaphase or interphase chromosome preparations allowing for annealing of the probe DNA to the complementary DNA sequence on the chromosome of interest (Speicher & Carter 2005) as illustrated in Figure 1-4. The labelled sequence can then be visualised under a fluorescence microscope. Using probes in this manner allows for precise mapping of chromosomal regions and if used for clinical screening it allows for effective diagnosis and confirmation of any aberrations seen at the karyotype level.

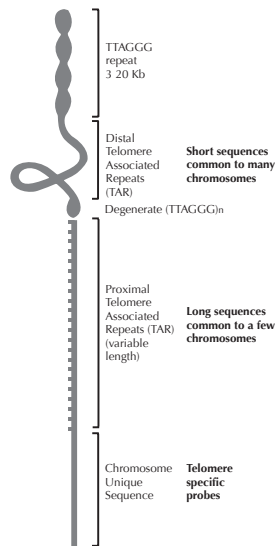


**Figure 1-4: The principles of fluorescence *in situ* hybridisation (FISH).**

**(a) Fluorescently labelled probe DNA and a target sequence are identified; (b) Probe and target DNA are co-denatured; (c) Probe and target DNA are hybridised together allow annealing of complementary DNA sequences. (d) Probe signals are visualisation using fluorescent microscopy (adapted from Cytocell promotional material).**

### **1.1.2.1 Sub-telomeric FISH Probes**

Sub-telomeric chromosome regions have been identified as highly gene rich (Saccone et al. 1992) and chromosomal abnormalities in these regions have been implicated in around 6% of cases of idiopathic mental retardation (Flint et al. 1995). Given their position in the very distal region of the chromosome they are also useful markers to delineate the end of the chromosome (as shown in Figure 1-5). Many translocations are undetectable on a standard karyotype, either because the exchanged regions are too small (cryptic translocations) or because the banding patterns are indistinguishable. Given that the nature of reciprocal translocations inevitably involves the ends of the chromosome, the ability to highlight these regions with a FISH probe enables visualisation of affected chromosomes. The coupling of sub-telomeric probes and FISH therefore simplifies both the identification of aneuploidies and balanced translocations by allowing detection of signal number and position (Knight & Flint 2000).



**Figure 1-5: Telomeric and sub-telomeric sequence organisation (Knight and Flint, 2000).**

To date, cryptic translocation detection using these probes has largely been applied to humans (e.g. Knight et al. 1996). It is, nonetheless feasible to adapt this technology for use in non-human species such as pigs and cattle. As with classical cytogenetics, this could have benefit in animal breeding however, to the best of my knowledge, has not yet been widely applied and as such is an issue addressed in this thesis.

#### **1.1.2.2 Chromosome Painting - Mammals**

Chromosome painting is an adaptation of the FISH technique that involves whole chromosome (or sub chromosomal region) libraries as probes. It is used widely in clinical cytogenetics (Telenius et al. 1992; Carter et al. 1992) and has been extensively used as a method for comparing genomes of distantly related species. Homologous DNA sequences between whole chromosomes can be detected using this method by hybridising DNA from single chromosomes of one species onto the chromosomes of another (also known as zoo-FISH) (Chowdary et al. 1998). The source DNA is typically derived from a fluid suspension of chromosomes that is sorted and separated using a dual laser flow cytometer (Ferguson-Smith 1997). The DNA is then amplified by a whole genome amplification protocol and labelled with fluorophores to enable detection under a fluorescence microscope (Rens et al. 2006). Regions of homology are therefore revealed, such as those shown in Figure 1-6, where paints derived from gibbon chromosomes have been hybridised onto human chromosomes (Ferguson-Smith & Trifonov 2007). Despite the wealth of results obtained using chromosome paints on mammalian species (Wienberg et al. 2000), the technique is ultimately limited by evolutionary distance with little

evidence of chromosome paints derived from eutherian species being detected on marsupial or monotreme chromosomes (Graphodatsky et al. 2012).

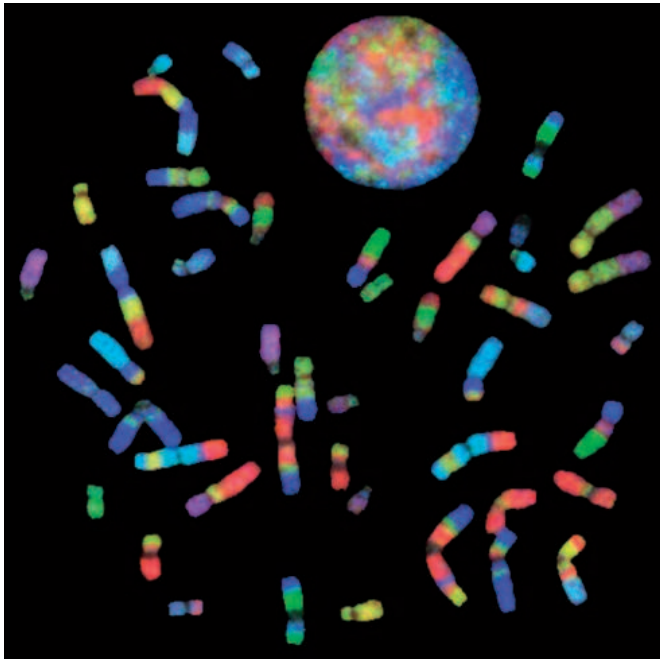
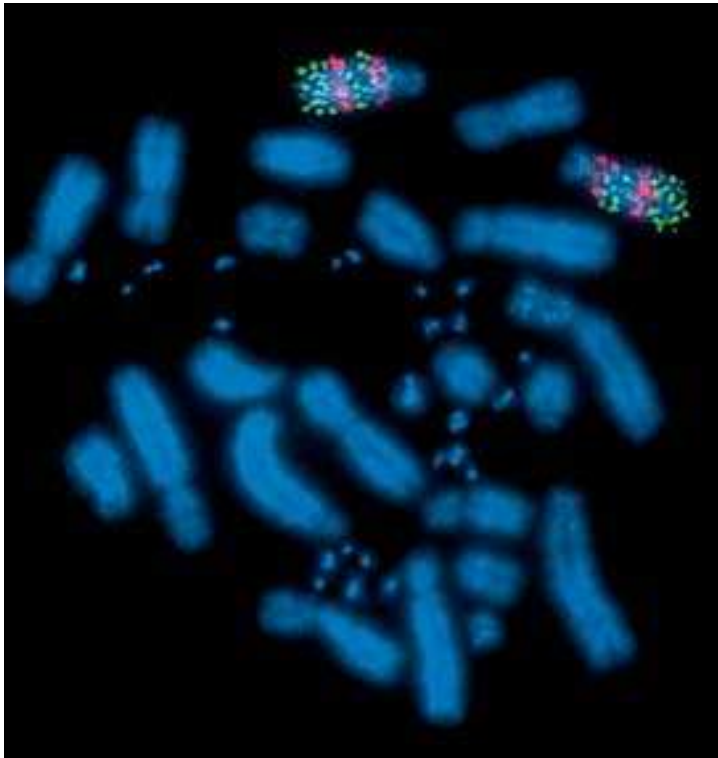


Figure 1-6: Chromosome painting of gibbon chromosome paints onto human chromosomes (Ferguson-Smith & Trifonov, 2007).

### 1.1.2.3 Chromosome Painting - Birds

As well as facilitating characterisation of the chicken karyotype (Griffin et al. 1999; Habermann et al. 2001; Masabanda et al. 2004), the generation of chromosome paints for chicken chromosomes 1-9 plus Z and W led to a surge in avian comparative genomics research (e.g. Shetty et al. 1999; Raudsepp et al. 2002; Shibusawa et al. 2002; Itoh & Arnold 2005; Griffin et al. 2007; Nanda et al. 1999; Nanda et al. 2011; Nishida et al. 2008). A high degree of success has been accomplished using these paints with results achievable in species as evolutionarily far removed from each other and from chicken as falcons, ostrich and emus (Nishida et al. 2008; Nishida et al. 2007; Shetty et al. 1999). A summary of the avian species investigated using chromosome paints is listed in Table 1-1 and an example of chromosome paints hybridised to peach-faced lovebird (*Agapornis roseicollis*) chromosomes is shown in Figure 1-7.



**Figure 1-7: Dual colour FISH of GGA macrochromosome paints GGA6 (red) and GGA7 (green) on *A. roseicollis* chromosomes (Nanda et al. 2007).**

Unlike the research performed on mammalian chromosomes, hybridisation across a greater evolutionary distance is possible with chicken chromosome paints. For example, homology has been detected between chicken, turtles and crocodiles, all of which last shared a common ancestor over 200 million years ago (mya) (Matsuda et al. 2005; Kasai et al. 2012). The use of microchromosomal paints, however, has been comparatively limited (Shetty et al. 1999; Griffin et al. 1999; Hansmann et al. 2009; Nie et al. 2009), largely due to the paints being divided into 'pools' of microchromosomes rather than being assigned to separate, entire chromosomes (Lithgow, O'Connor R.E. et al. 2014). Whilst able to define whole blocks of homology between species, orientation of the blocks cannot be defined using chromosome nor can intrachromosomal rearrangements be identified. To overcome both of these limitations, a BAC based approach is necessary, either in conjunction with paints or as a technique in its own right. This thesis is concerned with the development of such an approach.



Order	Common Name	Species Name	Author
<i>Accipitriformes</i>	Bearded vulture	<i>Gypaetus barbatus</i>	Nanda et al. 2006
<i>Accipitriformes</i>	Common buzzard	<i>Buteo buteo</i>	Nie et al. 2015
<i>Accipitriformes</i>	Grey hawk	<i>Asturina nitida</i>	de Oliveira et al. 2013
<i>Accipitriformes</i>	Griffon vulture	<i>Gyps fulvus</i>	Nanda et al. 2006
<i>Accipitriformes</i>	Griffon vulture	<i>Gyps fulvus</i>	Nie et al. 2015
<i>Accipitriformes</i>	Harpy eagle	<i>Harpia harpyja</i>	de Oliveira et al. 2005
<i>Accipitriformes</i>	Himalayan vulture	<i>Gyps himalayensis</i>	Nie et al. 2015
<i>Accipitriformes</i>	Mountain hawk-eagle	<i>Nisaetus nipalensis</i>	Nishida et al. 2013
<i>Accipitriformes</i>	Osprey	<i>Pandion haliaetus</i>	Nishida et al. 2014
<i>Accipitriformes</i>	Roadside hawk	<i>Rupornis magnirostris</i>	de Oliveira et al. 2013
<i>Accipitriformes</i>	Rüppell's vulture	<i>Gyps rueppellii</i>	Nanda et al. 2006
<i>Accipitriformes</i>	Savanna hawk	<i>Buteogallus meridionalis</i>	de Oliveira et al. 2013
<i>Accipitriformes</i>	White hawk	<i>Pseudastur albicollis</i>	de Oliveira et al. 2010
<i>Anseriformes</i>	Chinese goose	<i>Anser cygnoides</i>	Islam et al. 2014
<i>Anseriformes</i>	Common swan	<i>Coscoroba coscoroba</i>	Rodrigues et al. 2014
<i>Anseriformes</i>	Domestic duck	<i>Anas platyrhynchos</i>	Islam et al. 2014
<i>Anseriformes</i>	Greylag goose	<i>Anser anser</i>	Guttenbach et al. 2003
<i>Anseriformes</i>	Muscovy duck	<i>Cairina moschata</i>	Islam et al. 2014
<i>Casuariiformes</i>	Double-wattled cassowary	<i>Casuarius casuarius</i>	Nishida et al. 2007
<i>Casuariiformes</i>	Emu	<i>Dromaius novaehollandiae</i>	Shetty et al. 1999
<i>Cathartiformes</i>	California condor	<i>Gymnogyps californianus</i>	Raudsepp et al. 2002
<i>Cathartiformes</i>	Turkey vulture	<i>Cathartes aura</i>	Tagliarini et al. 2011
<i>Charadriiformes</i>	Herring gull	<i>Larus argentatus</i>	Hansmann et al. 2009
<i>Charadriiformes</i>	Stone curlew	<i>Burhinus oediconemus</i>	Hansmann et al. 2009
<i>Charadriiformes</i>	Stone curlew	<i>Burhinus oediconemus</i>	Nie et al. 2009
<i>Columbiformes</i>	Rock pigeon	<i>Columba livia</i>	Derjusheva et al. 2004
<i>Columbiformes</i>	Rock pigeon	<i>Columba livia</i>	Hansmann et al. 2009
<i>Falconiformes</i>	Kestrel	<i>Falco tinnunculus</i>	Nishida et al. 2008
<i>Falconiformes</i>	Merlin	<i>Falco columbarius</i>	Nishida et al. 2008
<i>Falconiformes</i>	Peregrine Falcon	<i>Falco peregrinus</i>	Nishida et al. 2008
<i>Galliformes</i>	California quail	<i>Callipepla californica</i>	Shibusawa et al. 2004
<i>Galliformes</i>	Chinese bamboo-partridge	<i>Bambusicola thoracica</i>	Shibusawa et al. 2004

Order	Common Name	Species Name	Author
<i>Galliformes</i>	Chinese quail	<i>Coturnix chinensis</i>	Shibusawa et al. 2004
<i>Galliformes</i>	Common peafowl	<i>Pavo cristatus</i>	Shibusawa et al. 2004
<i>Galliformes</i>	Golden pheasant	<i>Chrysolophus pictus</i>	Guttenbach et al. 2003
<i>Galliformes</i>	Guinea fowl	<i>Numidea meleagris</i>	Shibusawa et al. 2002
<i>Galliformes</i>	Guinea fowl	<i>Numidea meleagris</i>	Shibusawa et al. 2004
<i>Galliformes</i>	Japanese quail	<i>Coturnix japonica</i>	Guttenbach et al. 2003
<i>Galliformes</i>	Japanese quail	<i>Coturnix japonica</i>	Shibusawa et al. 2004
<i>Galliformes</i>	Lady Amherst's pheasant	<i>Chrysolophus amherstiae</i>	Shibusawa et al. 2004
<i>Galliformes</i>	Plain chachalaca	<i>Ortalis vetula</i>	Shibusawa et al. 2004
<i>Galliformes</i>	Ring necked pheasant	<i>Phasianus colchicus</i>	Guttenbach et al. 2003
<i>Galliformes</i>	Ring-necked pheasant	<i>Phasianus colchicus</i>	Shibusawa et al. 2004
<i>Galliformes</i>	Silver pheasant	<i>Lophura nycthemera</i>	Shibusawa et al. 2004
<i>Galliformes</i>	Silver pheasant	<i>Lophura nycthemera</i>	Guttenbach et al. 2003
<i>Galliformes</i>	Turkey	<i>Meleagris gallopavo</i>	Griffin et al. 2008
<i>Galliformes</i>	Turkey	<i>Meleagris gallopavo</i>	Shibusawa et al. 2004
<i>Galliformes</i>	Western capercaillie	<i>Tetrao urogallus</i>	Shibusawa et al. 2004
<i>Gruiformes</i>	Eurasian coot	<i>Fulica atra</i>	Hansmann et al. 2009
<i>Passeriformes</i>	Blackbird	<i>Turdus merula</i>	Guttenbach et al. 2003
<i>Passeriformes</i>	Chaffinch	<i>Fringilla coelebs</i>	Derjusheva et al. 2004
<i>Passeriformes</i>	Redwing	<i>Turdus iliacus</i>	Derjusheva et al. 2004
<i>Passeriformes</i>	Zebra Finch	<i>Taeniopygia guttata</i>	Itoh and Arnold. 2005
<i>Psittaciformes</i>	Budgerigar	<i>Melopsittacus undulatus</i>	Nanda et al. 2007
<i>Psittaciformes</i>	Cockatiel	<i>Nymphicus hollandicus</i>	Hansmann et al. 2009
<i>Psittaciformes</i>	Cockatiel	<i>Nymphicus hollandicus</i>	Nanda et al. 2007
<i>Psittaciformes</i>	Peach faced lovebird	<i>Agapornis roseicollis</i>	Nanda et al. 2007
<i>Rheiformes</i>	Greater rhea	<i>Rhea americana</i>	Nishida et al. 2007
<i>Rheiformes</i>	Lesser rhea	<i>Pterocnemia pennata</i>	Nishida et al. 2007
<i>Rheiformes</i>	Rhea	<i>Rhea americana</i>	Guttenbach et al. 2003
<i>Strigiformes</i>	Eagle owl	<i>Bubo bubo</i>	Guttenbach et al. 2003
<i>Strigiformes</i>	Great grey owl	<i>Strix nebulosa</i>	Hansmann et al. 2009
<i>Struthioniformes</i>	Ostrich	<i>Struthio camelus</i>	Nishida et al. 2007
<i>Tinamiformes</i>	Elegant crested tinamou	<i>Eudromia elegans</i>	Nishida et al. 2007

**Table 1-1: Summary of Zoo-FISH studies carried out to date on avian species. For the most part they involve hybridisation of chromosome paints 1-10+Z (Griffin et al. 1999) on to the chromosomes of other avian species.**

#### **1.1.2.4 Comparative mapping - BAC clones**

As mentioned above, FISH using BAC clones allows for the identification and mapping of precise regions in the genome of interest. Individual BAC clones of around 150kb in size are used to define targeted regions of the genome. Where BAC regions also span a gene (unless the gene has been deleted or become a pseudogene), the locus of that gene can be mapped directly to the chromosome for the purposes of cytogenetic assignment. In terms of cross-species BAC mapping, a similar approach to that used with chromosome paints can be applied, however in this case, given that the BAC regions are so small compared to that covered by a chromosome paint, the likelihood of sufficient sequence homology being present is reduced. Cross-species BAC mapping has therefore been relatively underrepresented in the literature, with success rates of around 70% being reported from chicken BACs used on the very closely related turkey (*Meleagris gallopavo*) (Griffin et al. 2008) reducing to less than 40% on Pekin duck (*Anas platyrhynchos*) chromosomes (Skinner et al. 2009) and 10-20% on zebra finch metaphases (Griffin, personal communication). Nonetheless, BAC selection on the basis of sequence homology has been successfully reported using cattle BACs on other species (Larkin et al. 2006). To the best of my knowledge however this has not yet been reported in birds.

#### **1.1.2.5 Multiple Hybridisation Tools for FISH**

Technology developed for the identification of genetic disease in humans has facilitated the detection of multiple hybridisations on a single microscope slide (Knight et al. 1996). Using a slide that has been divided into 8 or 24 even sized squares each of which has been applied with metaphase chromosome preparations, along with a coverslip device that has corresponding wells containing probes labelled in different colours for each chromosome of interest, 8 or 24 dual-colour hybridisations are possible on one slide (illustrated in Figure 1-8). At present, this technology has only been used for human chromosome analysis; theoretically however, there should be no species limitation to the use of such a device for this purpose.

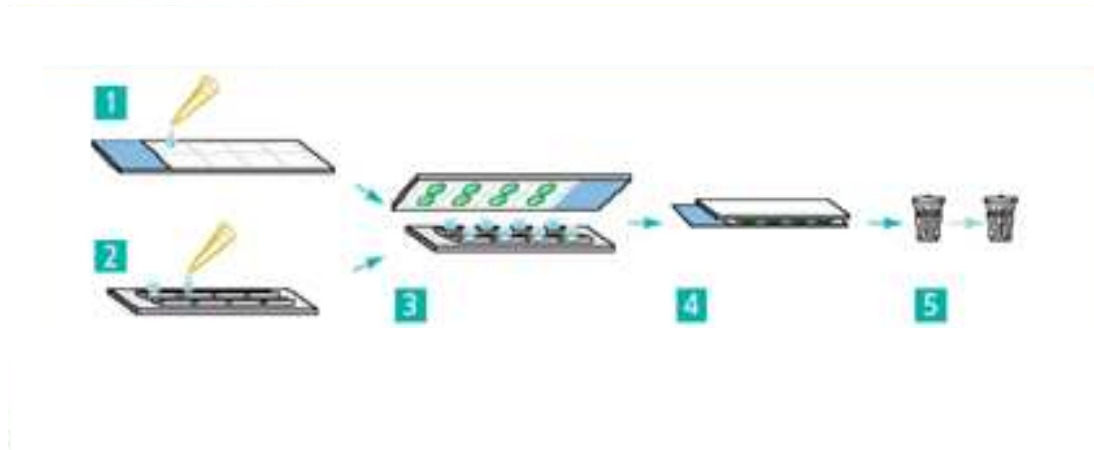


Figure 1-8: The principle behind the multiprobe device for use in FISH (Cytocell).

1. Template slide spotted with cell suspension; 2. Multiprobe device spotted with hybridisation solution; 3. Sample slide and device denatured together on hotplate; 4. Hybridise together overnight; 5. Rapid stringency washes then DAPI and fluorescence microscopy.

#### 1.1.2.6 Multiple Colour FISH Hybridisation

Initial work using FISH probes was limited to single colour experiments, largely due to the limitations of using indirect probe labelling which required that a single hapten (usually biotin) was incorporated into the probe and subsequently detected using a fluorophore in a separate step (Pinkel et al. 1986). In the early 1990s the addition of another hapten (usually digoxigenin) enabled detection of a second colour, although this was soon superseded with the development of direct labelling techniques that facilitated the visualisation of even more colours in a single experiment with 12 separate colours reported in one study (Dauwerse et al. 1992). Within a few years, several research groups extended this range of to include 24 colours (Speicher et al. 1996; Schrock et al. 1996). However, the number of colours available that are spectrally distinct from each other limited routine use of the technique. In order to overcome this limitation, Ioannou et al. (2011) developed a means of hybridising 6 colours in 4 layers enabling the detection of 24 probes in one assay (Ioannou et al. 2011) as illustrated in Figure 1-9. Originally tested on interphase nuclei of lymphocytes, sperm and blastomeres, the application of the technique was extended to include abnormal IVF embryos to determine whether variations in chromosome number were the underlying cause of post-zygotic errors (Ioannou et al. 2012). In principle, the use of multiple colour FISH is not limited to clinical diagnosis but instead can be utilised in all aspects of FISH including those of comparative genomics (as discussed in 1.1.2.2).

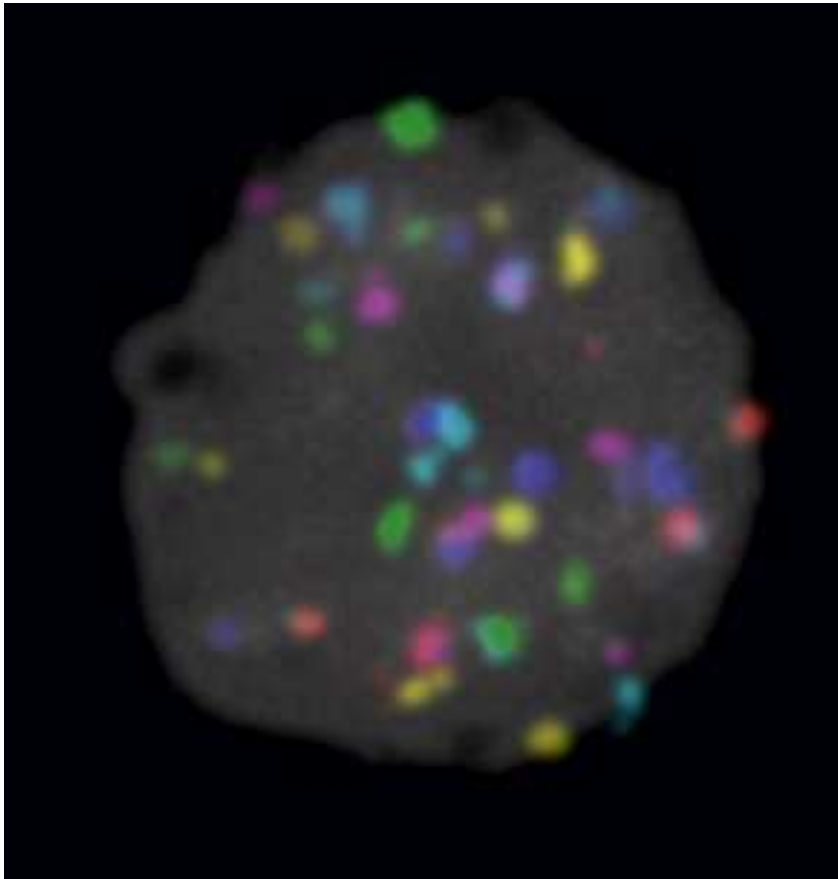


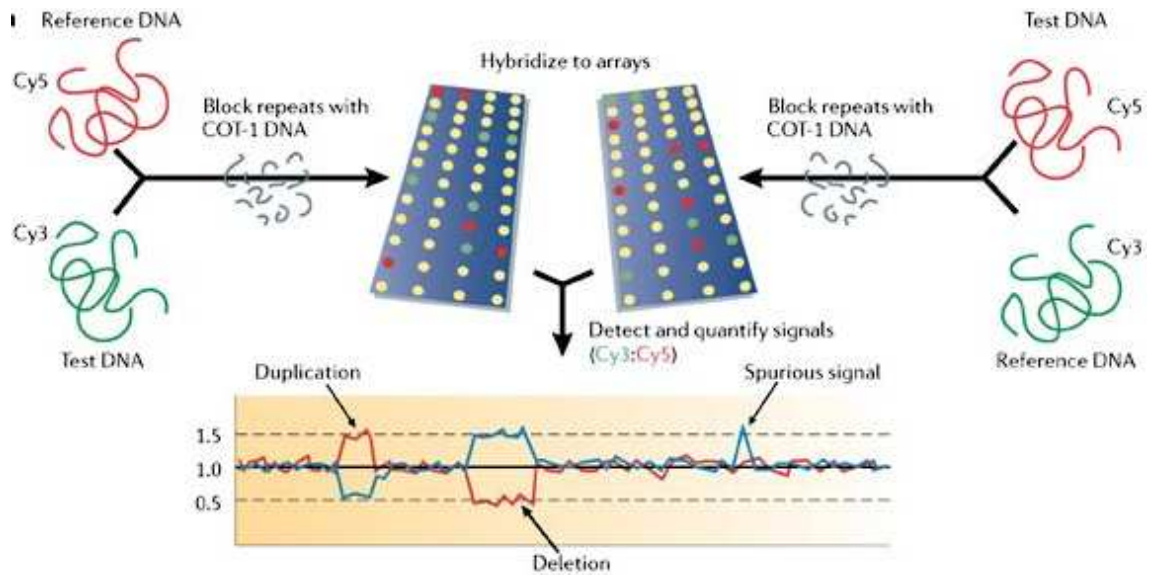
Figure 1-9: 24 colour FISH tested on an interphase nucleus derived from a human lymphocyte (Ioannou et al. 2011)

### 1.1.3 Microarrays and Comparative Genomic Hybridisation (Cytogenomics)

Comparative genomic hybridisation (CGH) has been extensively employed in medicine for the identification of disease and/or genetic gain or loss of chromosomal regions, with a particular early emphasis on their use for the investigation of tumour cells (Kallioniemi et al. 1992). In this technique, target genomic DNA and a reference DNA sample are labelled differentially with fluorophores, denatured and hybridised together and viewed on metaphase preparations using fluorescence microscopy. Differences in the signal intensity produced along the chromosome are indicative of DNA gains or losses, allowing the identification of unbalanced chromosomal abnormalities (Weiss et al. 1999).

Extension of this technique to include the use of a microarray (array-CGH) enables hybridisation of the test and reference DNA to a target such as BAC DNA, plasmid DNA or oligonucleotides that have been robotically printed onto a glass slide. This method allows for a much greater resolution, to the level of 100kb compared to the 5-10Mb possible in regular CGH. The direct quantification of signal intensities emitted from the hybridised fluorophores thereby enables

identification of copy number alterations in the test genome compared to the reference (Bejjani & Shaffer 2006) as illustrated in Figure 1-10.



**Figure 1-10: Schematic representation of array-CGH, where target and reference DNA are labelled and hybridised together with the addition of COT-1 DNA to block repetitive sequences, prior to analysis of the signal intensities produced by the sample of interest, allowing for quantification of DNA and copy number (Feuk et al. 2006).**

Array-CGH has proven to be a successful method of diagnosing sub-microscopic abnormalities (otherwise undetectable using FISH or CGH) that can lead to conditions such as Prader-Willi/Angelman syndrome. In addition, array-CGH has been a useful tool to detect aneuploidy in prenatal diagnosis. Despite the increased resolution afforded by array-CGH over CGH however, it is still limited by its inability to detect balanced chromosomal rearrangements (Evangelidou et al. 2013). Perhaps the most widespread use of array-CGH is in the detection of copy number variation (CNV). CNVs are classified as an intermediate sized structural variation larger than di- or trinucleotide repeats and smaller than is recognisable at a cytogenetic level (Zarrei et al. 2015). In a landmark paper, Redon and colleagues demonstrated that a significant proportion of the variation in the human genome is derived from CNVs (Redon et al. 2006). Since then, thousands of CNVs have been discovered using array-CGH (Conrad et al. 2010) contributing to the wealth of data being produced on structural variation of the human genome by projects such as the 1000 genomes project (McVean et al. 2012; Auton et al. 2015). Cross-species array-CGH has also been performed in animals including primates where 58 CNVs were identified between chimpanzee and humans (Perry et al. 2006). Hypothesised to play a role in speciation and adaptation, investigations into CNVs between birds have revealed the presence of 16

putative CNVs between turkey and chicken (Griffin et al. 2008), 32 between duck and chicken (Skinner et al. 2009), 20 between zebra finch and chicken (Völker et al. 2010) and ranging from 5.5 in Lady Amherst's pheasant to 39.75 in the red-legged partridge (Skinner et al. 2014). In particular, apparent copy number variation differences relating to fast-twitch muscle activity (e.g. MYOZ3) seen in falcons (renowned for speed of flight) (Frey and Olson. 2002) and respiratory function (e.g. MAPK8IP3) seen in quails (which exhibit unusual migratory patterns) (Borisov et al. 2003) have also been identified.

#### **1.1.4 Genome Sequencing Technologies**

DNA sequencing has created a revolution in biology, leading to the birth of many new disciplines including comparative genomics, the primary focus of this thesis. Sequencing methods are however currently only capable of producing relatively short sections of sequenced DNA. The development therefore of complex computational assembly methods along with the development of high throughput tools have allowed the process to be scaled up, resulting in DNA sequencing becoming a mainstream tool in the lab. The history and current status of sequencing technologies is discussed in the next section.

##### **1.1.4.1 Sanger Sequencing**

Genome sequencing began in the 1970s with the development of 'Chain termination sequencing', more commonly known as 'Sanger' sequencing. Briefly, reads are generated from sequencing random small cloned fragments from both directions of a genome. Template DNA, primers, DNA polymerase, deoxynucleosidetriphosphates (dNTPs) and di-deoxynucleosidetriphosphates (ddNTPs) are included in the elongation mix. ddNTPs are modified dNTPs to end DNA strand elongation, hence the name 'chain-termination sequencing'. The elongation reactions are run as 4 separate reactions one for each ddNTP and subsequently separated by size using polyacrylamide gel electrophoresis (Sanger et al. 1977). The resulting bands are visualised and their positions relative to each other read to identify the sequence (illustrated in Figure 1-11).

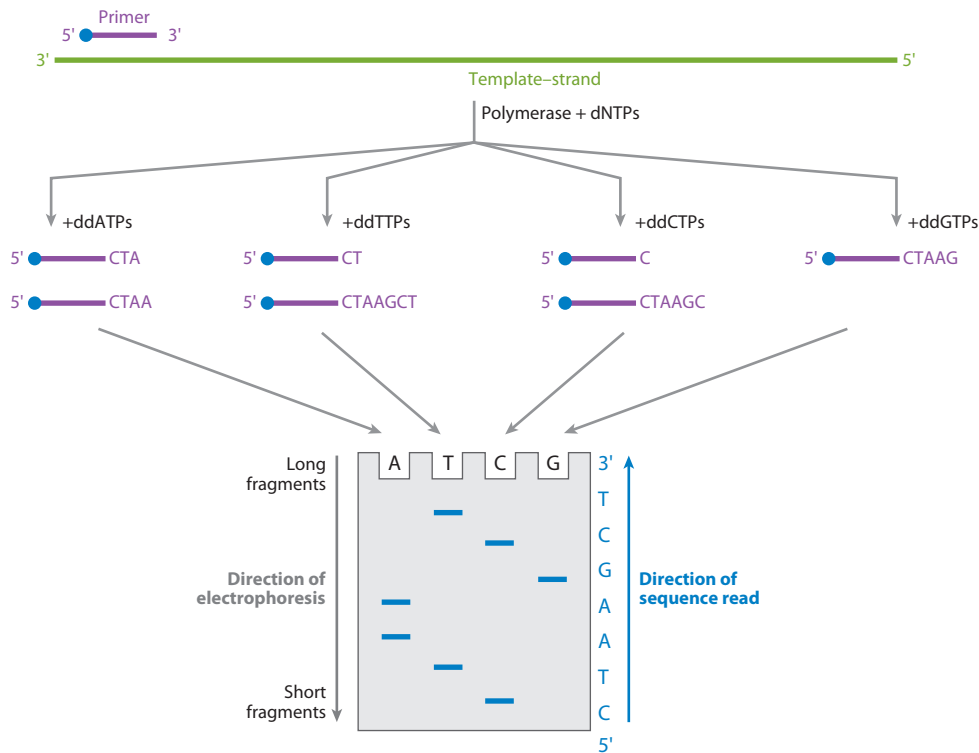


Figure 1-11: Schematic representation of Sanger sequencing (Mardis 2013).

Since modified by Smith and colleagues in 1986 to incorporate fluorescent dyes, this method remains the fundamental principle underlying modern sequencing technologies (Smith et al. 1986). The Sanger method has an advantage in that it can achieve read lengths of around 800-1000 base pairs compared to the much shorter read length of 100-500 bp produced by next generation sequencing (NGS) techniques, however it is low throughput and costly to perform (Shendure & Ji 2008).

#### 1.1.4.2 Next Generation Sequencing

In the late 1980s and early 1990s it became clear that the rate-limiting step in sequencing whole genomes was the chemistry of the Sanger sequencing technology itself. That is the process was too long, laborious and expensive to facilitate sequencing of multiple whole genomes. The demand for low-cost DNA sequencing led to the development of high-throughput sequencing methods (so called next-generation sequencing or NGS) that are not only simpler and cheaper but also can run many thousands of DNA sequence experiments in parallel, vastly improving the speed of production and volume of data produced (Miller et al. 2010). Towards the end of the last century, novel approaches to DNA sequencing were becoming commercially available. Tsien, Ross, Fahnestock and Johnston patented a stepwise ('base-by-base') sequencing protocol

making use of removable 3' blockers on arrays of DNA (1990; Patent No. WO9106678). Later Ronaghi and colleagues (1996) published an approach termed 'pyrosequencing' (Ronaghi et al. 1996) and Kawashima, Mayer and Farinelli patented so-called 'DNA colony sequencing' (1998; Patent No. WO9844152). Combined, these approaches combined now form the basis of the chemistry used in many Illumina systems (Mardis 2013). Now widely regarded as the first commercially available 'next-generation' sequencing method, Lynx Therapeutics published and marketed a technique in 2000 that they termed massively parallel signature sequencing (MPSS). Soon after in 2004, 454 Life Sciences (now Roche) launched a version of pyrosequencing that also involved massive parallel sequencing. From the outset, these newer methods were significantly less expensive than the Sanger approach. A synopsis of the key techniques is described below.

#### **1.1.4.2.1 Massively Parallel Signature Sequencing (MPSS)**

MPSS is a bead-based approach making use of adapter ligation and adapter decoding technology which reads the DNA sequence in blocks of four nucleotides. Prone to sequence-specific bias and specific sequence loss, the actual process of MPSS is now considered obsolete, although its basic properties are typical of later NGS data types, i.e. up to millions of short read DNA sequences generated at the same time (Brenner et al. 2000).

#### **1.1.4.2.2 454 Pyrosequencing**

As mentioned, a version of pyrosequencing that allowed parallel sequencing was developed by 454 Life Sciences. This approach takes DNA held in water droplets within an oil solution (emulsion PCR) and amplifies the DNA along with a DNA template attached to a single primer-coated bead. Multiple reactions can run at the same time thanks to numerous picolitre-volume wells, each of which contains a single bead along with the enzymes required for the sequencing reaction (Margulies et al. 2005).

#### **1.1.4.2.3 Polony Sequencing**

In 2005, polony sequencing was used to sequence the complete *E. coli* genome, combining an *in-vitro* paired-tag library with emulsion PCR, an automated epifluorescence microscope, and ligation-based sequencing chemistry. Despite providing an inexpensive yet accurate method of sequencing, only ~60 megabits of sequence was obtained from around 786 gigabits of image data collected (Shendure et al. 2005).



#### **1.1.4.2.4 Illumina (Solexa) Sequencing**

This Illumina sequencing method attaches DNA molecules and primers to a slide or flow cell which are then amplified using PCR to form 'DNA clusters'. To identify the DNA sequence, four types of reversible terminator base are added. Unincorporated nucleotides are then washed away before a camera takes images of the fluorescently labelled nucleotides. At this point, the dye, in conjunction with the ddNTP, is chemically removed from the DNA, allowing a subsequent cycle to begin anew. Unlike pyrosequencing, the DNA chains are extended one nucleotide at a time and image capturing can be staggered to allow for very large numbers of clustered DNA to be captured in a short time (Mardis 2013).

#### **1.1.4.2.5 SOLiD Sequencing**

SOLiD technology (Applied Biosystems - now Life Technologies) uses ligation based sequencing chemistry and emulsion PCR. The resulting beads, each consisting of many copies of the same DNA molecule, are put on a glass slide to which oligonucleotides are annealed and ligated. Preferential ligation by DNA ligase for matching sequences results in a signal indicative of the nucleotide at that position. The final result in terms of read length and quantity is similar to that of Illumina machines (Hedges et al. 2011).

#### **1.1.4.2.6 Ion Torrent Semiconductor Sequencing**

Based on a detection system in which hydrogen ions released during DNA polymerisation are detected, micro-wells containing template DNA are flooded with a single type of nucleotide (A, T, C or G) which, if complementary to the template is incorporated into a growing strand resulting in the release of a hydrogen ion which in turn triggers a sensor indicating when a reaction has occurred (Rothberg et al. 2011).

The NGS outputs from all of these platforms result in short read lengths than those achieved using the Sanger method, therefore reducing the length of contiguous sequence that can be assembled (Mardis, 2013). By generating multiple reads from each region a high level of coverage can be achieved helping to mitigate the effects of these short reads and allowing the reads to be assembled into contigs. These contigs can then be assembled into scaffolds using a computer aided approach (Miller et al. 2010). Sequence quality is assessed using the contig N50 value (defined as the shortest contig length of 50% of the genome) and the scaffold N50 value

(calculated in the same manner for scaffolds), with a higher N50 value and therefore longer contigs/scaffolds considered to indicate a high quality assembly.

#### **1.1.4.3 Third Generation Sequencing**

Increasingly sophisticated sequencing technologies known as third generation sequencing (TGS) aim to address some of the problems associated with short reads. The most established example of which is PacBio (Pacific Biosciences) which uses single molecule real-time (SMRT) technology to generate much longer reads, in the order of 10kb on average (Berlin et al. 2014). Combining nanotechnology with molecular biology and ultra-sensitive fluorescence detection, this method enables sequencing of single molecules (Eid et al. 2009). The longer read lengths generated here have the benefit of spanning many more repetitive sequence regions, therefore producing a more contiguous genome reconstruction compared to NGS approaches (Roberts et al. 2013). The main limitations of this method are that the cost is significantly higher than using NGS platforms (Lee et al. 2016).

Other third generation methods include Illumina TruSeq Synthetic Long Reads (Kuleshov et al. 2014) and Oxford Nanopore MinION device (Loman et al. 2015). The first of these produces highly accurate outputs but the reads are shorter than those of PacBio. The second of the two, the Oxford Nanopore kit, produces reads of a similar length to PacBio but of lower accuracy, although the handheld nature of the equipment has made it useful for studies in remote locations, in particular for studying Ebola outbreaks in West Africa (Quick et al. 2016).

A comparison of the most widely used first, next and third generation sequencing platforms is illustrated in Table 1-2.

Method	Read length	Accuracy	Reads per run	Time per run	Cost (USD) per 1 million bases	Advantages	Disadvantages
<b>Third Generation Sequencing:</b>							
Single-molecule real-time sequencing (Pacific Biosciences)	10,000 bp to 15,000 bp	87%	50,000 per SMRT cell	30 minutes to 4 hours	\$0.13–\$0.60	Longest read length. Fast.	Moderate throughput. Can be very expensive.
<b>Next Generation Sequencing:</b>							
Ion semiconductor (Ion Torrent sequencing)	up to 400 bp	98%	up to 80 million	2 hours	\$1	Less expensive equipment. Fast.	Homopolymer errors.
Pyrosequencing (454)	700 bp	99.90%	1 million	24 hours	\$10	Long read size. Fast.	Runs are expensive. Homopolymer errors.
Sequencing by synthesis (Illumina)	75 bp to 600 bp (depending on sequencer)	99.90%	1-3 billion (depending on sequencer)	1 to 11 days (depending on sequencer)	\$0.05 to \$0.15	Potential for high sequence yield	Equipment can be very expensive. Requires high concentrations of DNA.
Sequencing by ligation (SOLiD sequencing)	50+35 or 50+50 bp	99.90%	1.2 to 1.4 billion	1 to 2 weeks	\$0.13	Low cost per base.	Slower than other methods. Has issues sequencing palindromic sequences.
<b>First Generation Sequencing:</b>							
Chain termination (Sanger sequencing)	400 to 900 bp	99.90%	N/A	20 minutes to 3 hours	\$2400	Long individual reads	Expensive and impractical for larger sequencing projects. Also requires plasmid cloning or PCR.

Table 1-2: Comparison of NGS sequencing technologies (adapted from Liu et al. 2012)

### **1.1.5 Genome Sequencing Strategies**

There are two primary strategies for sequencing a genome. The first – ‘clone-by-clone’ sequencing (sometimes known as hierarchical shotgun sequencing or ‘BAC by BAC’ sequencing), uses a ‘map first, sequence second’ approach (Green 2001). Target DNA is first mapped using clone based physical mapping techniques to produce ‘contigs’ or series of overlapping series of clones each of which spans a large, contiguous region of the source genome from which a minimal tiling path is generated. The BACs within this tiling path are then sheared into fragments and sequenced using the techniques described above (and illustrated in Figure 1-12). These contigs then form the framework from which scaffolds are put together. The fragments produced often exceed the read length possible with current technology, and therefore sequence reads are generated from both ends of the fragment, resulting in a library of mate-pairs with a range of insert sizes – the space between the paired reads (Pop. 2009). This approach has the benefit of reducing the risk of misassemblies both across a long range and a short range by reducing the reliance on computational interpretation. It is however very time consuming and coverage can be incomplete (Kaiser et al. 2003). This method was used to produce the publicly funded human genome assembly in 2001 (described in section 1.1.5.2), which cost in excess of \$100 million (Drmanac et al. 2010).

In the second approach, known as ‘whole genome shotgun sequencing’ there is no initial mapping phase. Instead, the entire genome is fragmented into pieces of a specific size, that are sub-cloned into appropriate plasmid vectors which are then sequenced. This produces a collection of read pairs (mate-pairs) that are separated by a known distance (the size of the original fragment) (Pop. 2009). The resulting tens of millions of sequence reads generated using this method are then assembled into contiguous sequences and ultimately scaffolds using a computational approach. In order to improve the accuracy of resulting contigs, reads are generated for multiple overlapping fragments. Referred to as coverage, this ensures that the genome is effectively over-sampled numerous times. This technique is repeated multiple times to improve coverage and increase the number of overlapping segments, thereby improving the integrity of the assembled genome by allowing contigs to be linked into scaffolds (as illustrated in Figure 1-12). A high multiple of coverage is therefore considered to be indicative of the quality of the genome sequence (Mardis, 2008). Some of the limitations of this technique are that it relies entirely on a bioinformatic method of assembly and that highly repetitive genomes remain fragmented (Commins et al. 2009). Multiple assembly algorithms have been developed with many more being written as new sequence data sets are being produced (Pop 2009). Several

programs are currently in place, such as Assemblathon where teams of different assemblers 'compete' against each other using the same data with the aim of improving the overall standard of assemblies (Baker 2012).

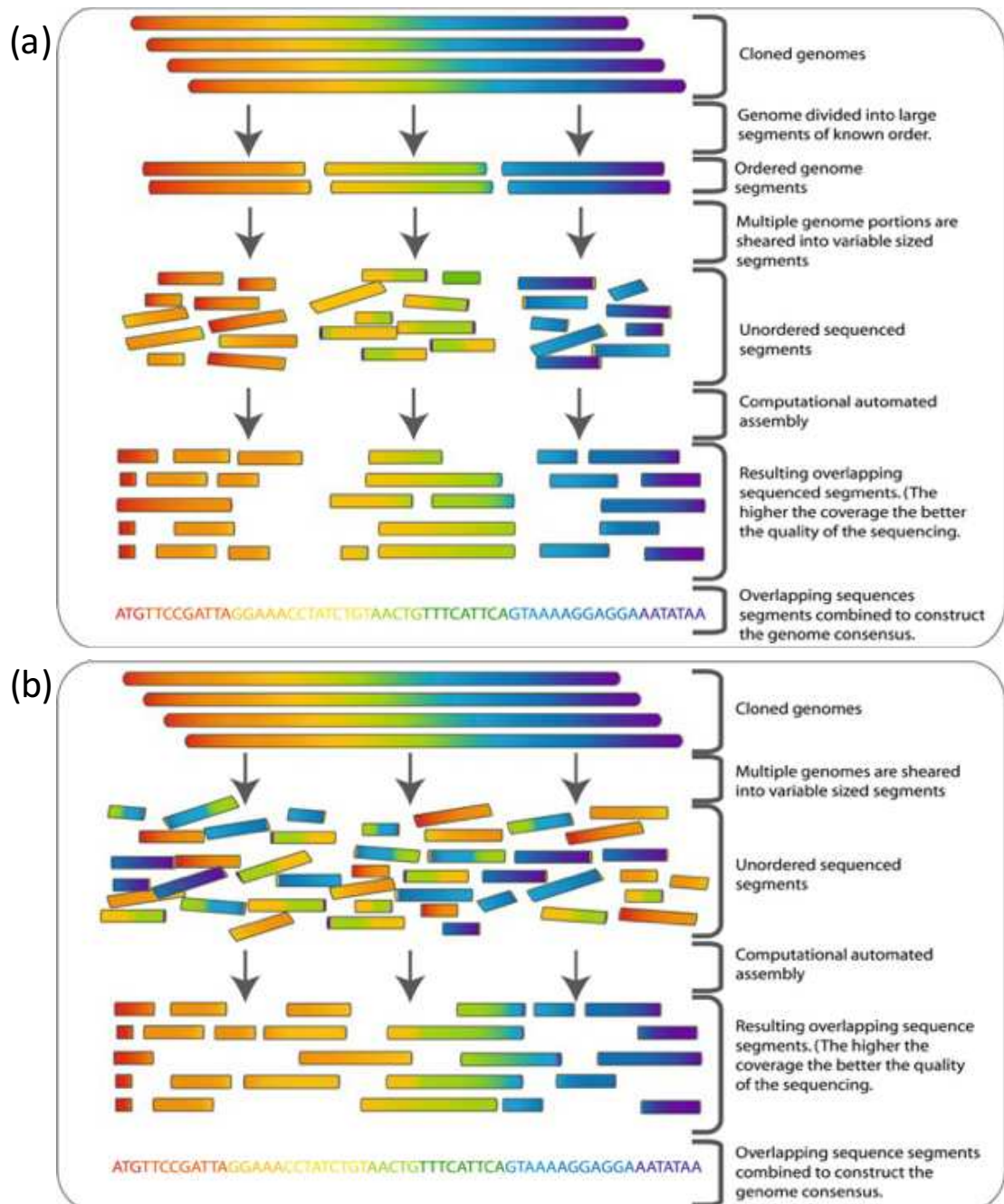


Figure 1-12: Schematic representation illustrating the differences between (a) BAC-by-BAC or hierarchical sequencing and (b) Shotgun sequencing (Commins et al. 2009).

Because this method does not use the 'map-first, sequence second' approach there is no indication as to where the scaffold is anchored or ordered on the chromosome. For genomes

that have a reference this is less of a problem than for *de novo* genomes, although both rely on some form of map in order to orientate and order the scaffolds (Sakai et al. 2015). In the absence of tools to facilitate detailed genome reconstruction such as linkage map and RH panels, bioinformatics programs can, through the use of algorithms, assist with mapping these scaffolds but there is still a requirement for physical mapping of the data to give an accurate portrayal of the genome and where the genes are localised.

#### **1.1.5.1 Genome Sequencing Challenges**

Accurate genome assemblies are fundamental to genome research, particularly for studying evolutionary relationships between species. The N50 values referred to in section 1.1.4.2 do not however take into account many other aspects of genome build quality, which can have a significant effect on the validity of any downstream investigation. For example, errors in the sequence read such as incorrect nucleotide substitutions, insertions and deletion errors can distort gene mapping and annotation analysis (Meader et al. 2010). Contigs can also be artificially inflated in size by over aggressive joining of reads. This can result in scaffolds containing a high degree of repetitive content, producing mis-assemblies and a misleading N50 value (Salzberg & Yorke 2005). Reducing costs and increasing throughput capabilities has led to a surge in the number of species being sequenced using NGS. The very nature of NGS however, means that the reads produced are short. Whereas first generation Sanger sequencing produces relatively long stretches of DNA sequence (around 1kb), second generation NGS produces millions of short reads which are sequenced multiple times in order to reduce error rates. Segmental duplications and large common repeats can therefore be difficult to place resulting in up to 20% of the genome being missed. In addition, analysis of genome sequences without an appropriate reference genome can mean that these missing or misassembled regions are unclassifiable and therefore have the potential to lead to biased outputs. In a comparative study using NGS sequenced human genomes compared to the reference human genome, Alkan and colleagues (2010) found that the fragmentation of the genome caused by short reads meant that only 56.3% of the genes had sufficient representation in the assembly when compared to the reference human genome (Alkan et al. 2010).

Earlier sequencing efforts such as that of the human genome used a large-insert BAC clone library (McPherson et al. 2001) but this approach is often considered too costly to create or maintain for new sequencing projects. Third-generation technologies, already described in the

previous section, aim to overcome some of these difficulties by increasing read length and library insert sizes.

Newer mapping technologies such as optical mapping methods (Teague et al. 2010) and the platforms BioNano (Mak et al. 2016) and Dovetail (Putnam et al. 2016) also aim to provide a long-term solution to these issues. The optical mapping system used by BioNano adds fluorescently tagged probes to 'nicked' restriction digest sites to fingerprint long DNA molecules which are then imaged. The resulting fingerprints are then assembled into larger optical maps, often spanning many megabases of a chromosome (Lee et al. 2016). The Dovetail method uses proprietary technology to generate long range mate pairs that span hundreds of kilobases (Putnam et al. 2016). Such approaches are however still in their infancy and each has its limitations. For example, BioNano is limited in its ability to link contigs across centromeres or large heterochromatin blocks thereby restricting the assembly to a sub-chromosomal level and Dovetail generates large super-scaffolds but these are, nonetheless, sub-chromosomal in size.

Using a combined optical mapping and third generation sequencing shows great promise for Improving genome scaffold size and structural resolution (Lee et al. 2016). In fact, a recent combined PacBio and BioNano approach performed by Pendleton and colleagues resulted in the generation of a highly contiguous *de novo* human genome with a contig N50 of 1.4Mb and a scaffold N50 of 31.1 (Pendleton et al. 2015). This combined method is illustrated in Figure 1-13.

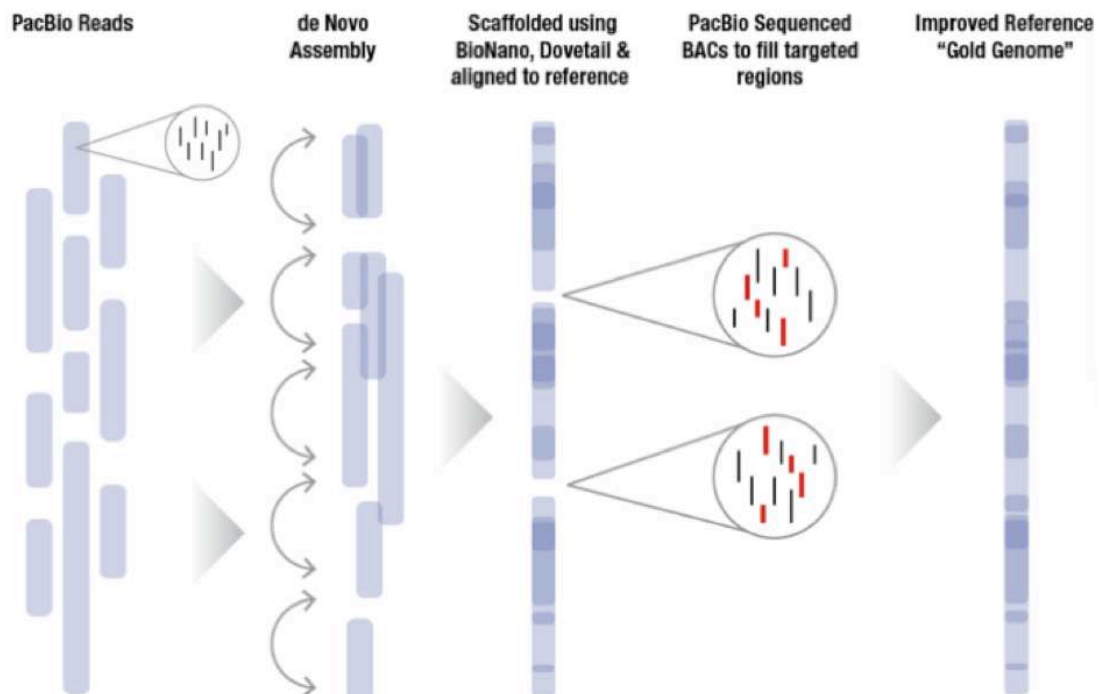


Figure 1-13: Combined PacBio sequencing and third generation mapping to improve sequence contiguity (dnanexus.com).

The current genome assembly status for a range of vertebrates is discussed below.

#### 1.1.5.2 Mammalian Genome Sequencing

Launched in 1990, the Human Genome Project was the first sequencing project of its scale in terms of ambition and collaboration. With the aim of sequencing all 3 billion base pairs of the human genome, a publicly funded initiative was set up using a BAC by BAC sequencing approach at the same time as a privately funded initiative was established which used a whole genome shotgun sequencing approach. The two projects led in early 2001 to the publication of the first sequenced human genomes (Lander et al. 2001; Venter et al. 2001) with coverage initially of around 90% of the euchromatic region of the genome of which around 3% was in finished form. By 2004, this coverage was increased to 98% of which 95% was in finished form (International Human Genome Sequencing 2004).

At around the same time, the mouse genome was sequenced, in part by virtue of its importance as a biomedical model, but also because despite the 75 million years since they shared a common ancestor it provided a crucial link for understanding the human genome (Waterston et al. 2002). Draft sequences of other mammalian species followed, including the chimpanzee (*Pan troglodytes*) (The Chimpanzee Sequencing and Analysis Consortium 2005), the rat (*Rattus*



*norvegicus*) (Gibbs et al. 2004) and the dog (*Canis lupis familiaris*) (Lindblad-Toh et al. 2005). The genomes of key agricultural species followed after with a draft sequence of the pig genome (*Sus scrofa domestica*) published in 2012 (Groenen et al. 2012) and of the cattle genome (*Bos taurus*) in 2009 (Elsik et al. 2009), both of which were sequenced using a combined BAC and whole-genome shotgun approach. Of the more than 5,000 extant mammalian species therefore, only 20 have genomes assembled to chromosome level. This small proportion of assembled genomes is also heavily biased to primates (Fang et al. 2014), artiodactyls (Larkin et al. 2012), carnivores (Murphy et al. 2005) and murid rodents (Murphy et al. 2005) with more than ten of the remaining 22 orders having no chromosome level assemblies at all. Generation of further *de-novo* mammalian assemblies is continuing at an extraordinary pace; these are at best, however, collections of scaffolds and require anchoring to chromosomes (for example, by FISH) to achieve full chromosomal level assembly.

### **1.1.5.3 Avian Genome Sequencing**

#### **1.1.5.3.1 The Chicken Genome**

The sequencing of the chicken (*Gallus gallus* - red junglefowl) genome in 2004 signalled a new dawn in avian genetics. As one of the 'big 10' genomes sequenced the chicken was an important choice for early sequencing (ICGSC 2004). Not only does chicken constitute 20% of the global meat market along with nearly the entire global egg production market. It is also a vital model for studying developmental biology, infectious disease (in particular viral diseases), immune system disorders, musculoskeletal disorders, cancer, Marek's disease and many more conditions that directly impact on human health (Brown et al. 2003). The first draft of the chicken genome was assembled using a whole-genome shotgun approach. Since then, the genome assembly has been enhanced with additional data sets, including genetic linkage maps (Groenen et al. 2009) and radiation hybrid maps (Morrison et al. 2007).

The current assembly, Galgal4 (Nov 2011) covers 96% of the predicted genome size (1.03Gb) and includes two linkage groups that are currently unassigned (around 1.77 Mb in size) as well as 14,093 unplaced scaffolds (Schmid et al. 2015). The chicken genome is nevertheless a work in progress, however, with sequence gaps still present, particularly for chromosome 16 and a near absence of sequence data for the smallest of the microchromosomes (GGA 29-31 and GGA 33-38). The W chromosome is also poorly assembled due to the extensive repetitive content found in this chromosome.

Annotation of the chicken genome has been based on gene homology with sequences from other species. This method has served well for defining protein-coding sequences, but genes that evolve quickly such as immune genes are harder to define over large evolutionary distances (Schmid et al. 2015). Despite this, the chicken genome appears to have significantly fewer protein-coding genes than other vertebrates, with 15,508 found in chicken (Cunningham et al. 2015) compared to 20,806 in humans. As well as appearing to have lost ancestral protein coding genes, the gene families themselves appear to have fewer members than other vertebrates (Hughes & Friedman, 2008). Significantly, 274 protein coding genes have been found to be absent in the chicken genome but present in most other vertebrate genomes including crocodiles, suggesting that their loss occurred after the split of dinosaurs and crocodiles (Lovell et al. 2014).

In 2004, the chicken also became the first, and to date only, avian species for which a karyotype was fully defined (Masabanda et al. 2004). At the time, chromosome paints or clones were generated to identify all chromosomes uniquely, however subsequent efforts to sequence from these clones proved unsuccessful (Griffin, personal communication). To date, the very smallest of the microchromosomes, the 'D group' (chromosomes 33–38) still do not have sequences associated with them in the genome assembly and the original probes from Masabanda et al. (2004) have since degraded (Griffin, personal communication). Reliable tools for detection of these chromosomes are therefore still not available.

#### **1.1.5.3.2 The Zebra Finch Genome**

In 2010, the zebra finch (*Taeniopygia guttata*) became the second avian species to have a sequenced genome. Belonging to the *Passeriformes* order (the largest of all avian orders, with over 5,000 identified species), the zebra finch is an important scientific model, in part due to its ability to communicate through learned vocalisation (shared only with the parrots and the hummingbirds) and its short generation turnover both of which make it a crucial model for understanding neurobiology (Clayton et al. 2009). The zebra finch genome was sequenced and assembled in the same manner as that of the chicken, whereby 1.0Gb of the 1.2Gb draft assembly was assigned to 33 chromosomes and three linkage groups using linkage maps and BAC mapping (Warren et al. 2010). 17,475 protein coding genes (since revised to 17,488 in the latest assembly – Ensembl taeGut3.2.4, Aug 2008) were predicted of which 57% were expressed in the forebrain of an adult zebra finch, consistent with the adaptations required for song and

memory (Warren et al. 2010). In contrast to the chicken genome where the major histocompatibility complex (MHC) is organised in a very compact manner with only 46 genes covering a region of 242 kb (Kaufman et al. 1999), sequencing and cytogenetic mapping of the zebra finch MHC suggests that it is in fact dispersed across 4 chromosomes in a manner that is similar to that seen in some mammalian lineages (Balakrishnan et al. 2010).

#### **1.1.5.3.3 The Turkey Genome**

Extending previous efforts to produce genetic linkage maps for the turkey (*Meleagris gallopavo*) (Harry et al. 2003; Reed et al. 2005; Reed et al. 2007) and in a departure from techniques used to sequence and assemble the chicken and zebra finch genomes, the turkey was the first avian species to be sequenced using next-generation sequencing (NGS). A combination of NGS platforms were used to sequence the genome (Roche 454 and Illumina GAI) producing a genome size of ~1.1 Gb of which 917 Mb was assigned to specific chromosomes using a BAC mapping approach (Dalloul et al. 2010). A total of 15,093 protein-coding genes were identified, subsequently revised to 14,123 in the latest assembly (Ensembl Turkey\_2.01, Sep 2010). Further analysis of the turkey genome by Zhang G. et al. (2014a) has since revealed a surprisingly high number of lineage specific rearrangements suggesting that there may in fact be some local misassemblies (Zhang G. et al. 2014a).

#### **1.1.5.3.4 The Duck Genome**

As a natural reservoir for influenza A viruses, the duck (*Anas platyrhynchos*) is an important model for understanding the pathogenesis of viruses and provides a unique insight into host immune responses. Of particular interest is the H5N1 virus, which has been seen to cross the species barrier to humans causing 622 infections (as of March 2013) with a fatality rate of 59% (Hulse-Post et al. 2005). The duck genome was sequenced using a whole-genome shotgun sequencing strategy with Illumina Genome Analyser sequencing technology, generating a draft assembly covering 1.1 Gb, with 15,634 protein-coding genes (Huang et al. 2013). At the same time radiation hybrid (RH) mapping was used to assign scaffolds to chromosomes using the chicken genome as a reference, whereby an algorithmic approach was initially used to locate the chicken homologs on the duck genome prior to RH alignment of scaffolds to duck chromosomes (Rao et al. 2012).

#### **1.1.5.3.5 Falcon Genomes**

Sequencing efforts were also extended to two falcon species, both of which, like all falcons, display extraordinary morphological, physiological and behavioural adaptations including fast flying speeds and visual acuity (see chapter 5) that allow them to be successful predators. The peregrine (*Falco peregrinus*) and Saker falcon (*Falco cherrug*) genomes were deep sequenced with an illumina HiSeq2000 with greater than 100-fold coverage. Assembly was performed using SOAPdenovo with genome sizes for both species estimated at 1.2Gb along with scaffold N50 values of 3.89 Mb for *F. peregrinus* and 4.15 Mb for *F. cherrug*. 16,263 genes were predicted for the peregrine falcon and 16,204 for the saker falcon of which around 92% were functionally annotated (Zhan et al. 2013).

#### **1.1.5.3.6 Additional Sequenced Avian Genomes**

The publication of further avian genomes followed with the addition of the budgerigar, *Melopsittacus undulatus* (Koren et al. 2012); the pigeon, *Columba livia*, (Shapiro et al. 2013); the collared flycatcher, *Ficedula albicollis* (Ellegren et al. 2012; Kawakawmi et al. 2014); the Puerto Rican amazon, *Amazona vittata* (Oleksyk et al. 2012); medium ground finch, *Geospiza fortis* (Zhang G. et al. 2012), the large ground finch, *Geospiza magnirostris* (Rands et al. 2013) and the canary, *Serinus canaria* (Frankl-Vilches et al. 2015). In 2014, the ostrich genome (*Struthio camelus*) was improved 5-fold with the use of optical mapping. This approach generated significantly larger scaffolds known as 'super-scaffolds' and enhanced the overall quality of the ostrich genome taking the initial N50 from 3.59 Mb to an N50 of 17.71 Mb (Zhang J. et al. 2015). The budgerigar (*Melopsittacus undulatus*) genome was the only one of the 45 that was assembled using a multiplatform (Illumina/GS-FLX/PacBio) approach. This genome was also enhanced using data from optical mapping experiments, in this case increasing the assembly's N50 scaffold size to around 14 Mb (Ganapathy et al. 2014).

### **1.1.6 Avian Genome Sequencing in the New Era**

The incorporation of NGS into genome sequencing efforts changed the avian genomic landscaped dramatically in 2014, with the publication of numerous avian genome sequences by Guoije Zhang and collaborators from the Avian Phylogenomics Group, taking the total number of sequenced avian genomes to around 60 at the time of writing. These newly published genome sequences, along with the previously published sequences represented all 32 neognath orders and two of the paleognath orders giving nearly entire coverage (92%) of the major extant avian

orders (sequenced species listed in Table 1-3, data available from (Zhang G. et al. 2014b)). Almost all of these genomes were sequenced and assembled using Illumina short reads, with varying degrees of coverage. A high (>50x) coverage approach was used for 20 of the bird species while a lower coverage of around 30x was used for 25 of the species as shown in Figure 1-14, thereby ensuring that the coverage was sufficient for extensive comparative analysis into bird macroevolution and the link between genetic variation and phenotypic diversity (Zhang G. et al. 2014a).

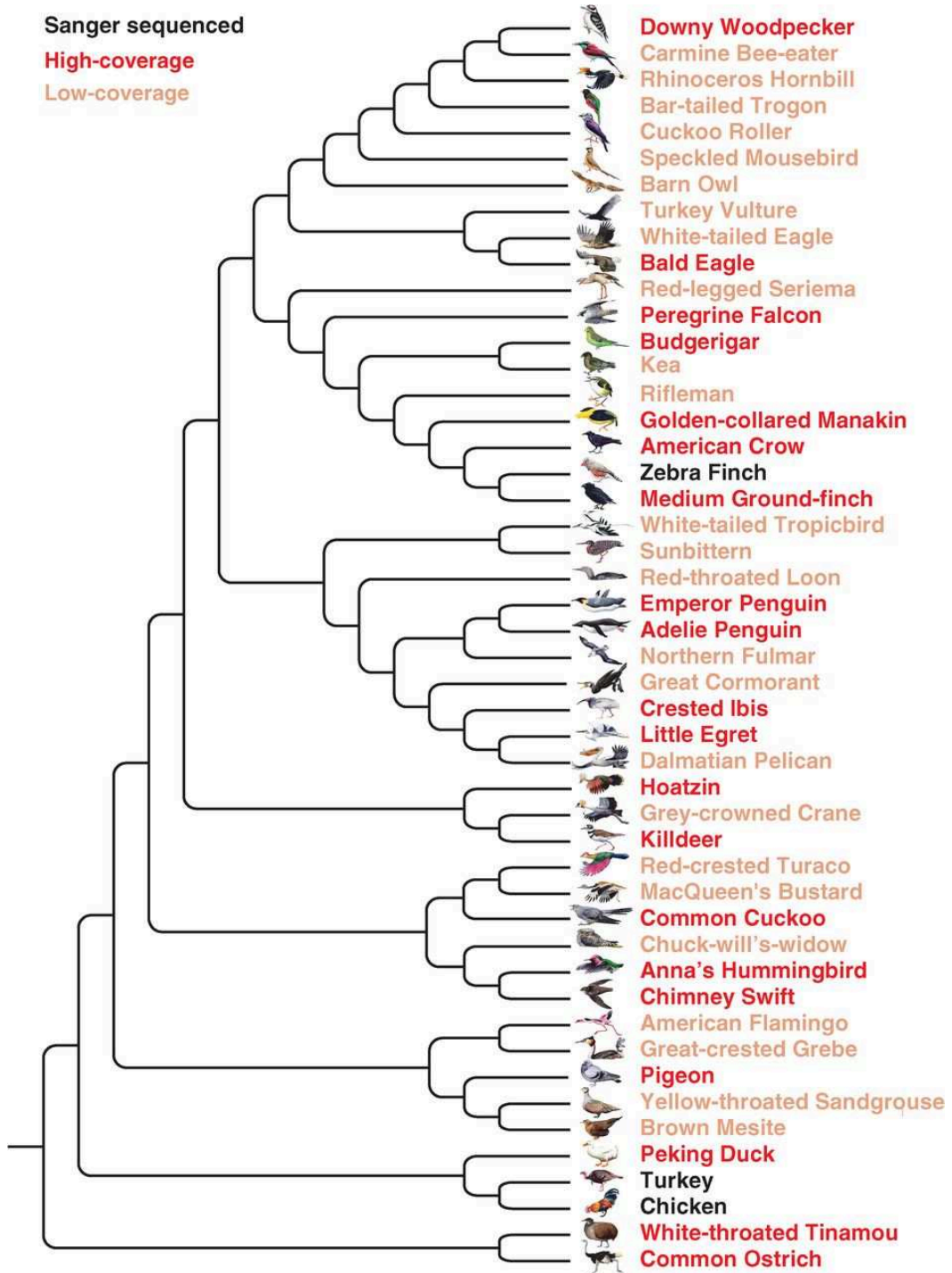


Figure 1-14: Sequence Coverage of the Avian Phylogenetic Tree (Zhang G. et al. 2014a).

Species	Common name	Genome Size	Number of Genes
<i>Acanthisitta chloris</i>	Rifleman	1.05Gb	14,596
<i>Anas platyrhynchos</i>	Peking duck	1.1Gb	16,521
<i>Antrastomus carolinensis</i>	Chuck-will's-widow	1.15Gb	14,676
<i>Apaloderma vittatum</i>	Bar-tailed trogon	1.08Gb	13,615
<i>Aptenodytes forsteri</i>	Emperor penguin	1.26Gb	16,070
<i>Balearica regulorum</i>	Grey-crowned crane	1.14Gb	14,173
<i>Buceros rhinoceros</i>	Rhinoceros hornbill	1.08Gb	13,873
<i>Calypte anna</i>	Anna's hummingbird	1.1Gb	16,000
<i>Cariama cristata</i>	Red-legged seriema	1.15Gb	14,216
<i>Cathartes aura</i>	Turkey vulture	1.17Gb	13,534
<i>Chaetura pelagica</i>	Chimney swift	1.1Gb	15,373
<i>Charadrius vociferus</i>	Killdeer	1.2Gb	16,856
<i>Chlamydotis macqueenii</i>	Macqueen's bustard	1.09Gb	13,582
<i>Colius striatus</i>	Speckled mousebird	1.08Gb	13,538
<i>Columba livia</i>	Pigeon	1.11Gb	16,652
<i>Corvus brachyrhynchos</i>	American crow	1.1Gb	16,562
<i>Cuculus canorus</i>	Common cuckoo	1.15Gb	15,889
<i>Egretta garzetta</i>	Little egret	1.2Gb	16,585
<i>Eurypyga helias</i>	Sunbittern	1.1Gb	13,974
<i>Falco peregrinus</i>	Peregrine falcon	1.18Gb	16,242
<i>Fulmarus glacialis</i>	Northern fulmar	1.14Gb	14,306
<i>Gallus gallus</i>	Chicken	1.05Gb	16,516
<i>Gavia stellata</i>	Red-throated loon	1.15Gb	13,454
<i>Geospiza fortis</i>	Medium ground finch	1.07Gb	16,286

Species	Common name	Genome Size	Number of Genes
<i>Haliaeetus albicilla</i>	White-tailed eagle	1.14Gb	13,831
<i>Haliaeetus leucocephalus</i>	Bald eagle	1.26Gb	16,526
<i>Leptosomus discolor</i>	Cuckoo-roller	1.15Gb	14,831
<i>Manacus vitellinus</i>	Golden-collared manakin	1.12Gb	15,285
<i>Meleagris gallopavo</i>	Turkey	1.04Gb	16,051
<i>Melopsittacus undulatus</i>	Budgerigar	1.1Gb	15,470
<i>Merops nubicus</i>	Carmine bee-eater	1.06Gb	13,467
<i>Mesitornis unicolor</i>	Brown mesite	1.1Gb	15,371
<i>Nestor notabilis</i>	Kea	1.14Gb	14,074
<i>Nipponia nippon</i>	Crested ibis	1.17Gb	16,756
<i>Ophisthocomus hoazin</i>	Hoatzin	1.14Gb	15,702
<i>Pelecanus crispus</i>	Dalmatian pelican	1.17Gb	14,813
<i>Phaethon lepturus</i>	White-tailed tropicbird	1.16Gb	14,970
<i>Phalacrocorax carbo</i>	Great cormorant	1.15Gb	13,479
<i>Phoenicopterus ruber</i>	American flamingo	1.14Gb	14,024
<i>Picoides pubescens</i>	Downy woodpecker	1.17Gb	15,576
<i>Podiceps cristatus</i>	Great-crested grebe	1.15Gb	13,913
<i>Pterocles gutturalis</i>	Yellow-throated sandgrouse	1.07Gb	13,867
<i>Pygoscelis adeliae</i>	Adélie penguin	1.23Gb	15,270
<i>Struthio camelus</i>	Common ostrich	1.23Gb	16,178
<i>Taeniopygia guttata</i>	Zebra finch	1.2Gb	17,471
<i>Tauraco erythrolophus</i>	Red-crested turaco	1.17Gb	15,435
<i>Tinamus guttatus</i>	White-throated tinamou	1.05Gb	15,773
<i>Tyto alba</i>	Barn owl	1.14Gb	13,613

**Table 1-3: Recently sequenced avian genomes with corresponding N50 sizes, genome sizes and total number of genes per species (Zhang G. et al. 2014b).**

#### **1.1.6.1.1 Genome 10k Project**

The upsurge in sequencing has led to the development of ambitious projects such as the Genome 10K Project (G10K) (Hausler et al. 2009). The main aim of this project is to sequence 10,000 vertebrate genomes, but also to standardise procedures for specimen collection and preparation as well as develop a uniform approach to genome assembly and alignment (Koepfli et al. 2015). The Avian Phylogenomics Consortium, inspired by the success of the 48 genomes project published in 2014, has since announced its 'Bird 10K' project under which it plans to generate draft genome sequences for all extant birds over the next five years (Zhang G. et al. 2015).

#### **1.1.6.1.2 Physical Genome Mapping**

As described above, recent NGS efforts rarely have assemblies at a similar level of 'chromosome level' integrity as those provided by traditional methods. One of the key aims of this thesis (see subsequent chapters) will address this issue. Inability of NGS to produce long enough contigs to cover whole chromosomes and unavailability of inexpensive mapping technologies to assemble complex genomes by chromosome invariably means that scaffolds require mapping to chromosomes. A recurring theme in this thesis is how not having a genome that is assembled to whole chromosome level limits the utility of current genome assemblies for critical aspects of evolutionary and applied genetics such as phenotype-to-genotype associations, gene and regulatory networks research and exploration of the mechanics of chromosome evolution. For this reason, cytogenetic mapping using fluorescence in situ hybridisation (FISH) of BAC clones can be a powerful technique that allows for reconciliation of this sequence data by directly visualising the regions of interest on the genome. FISH mapping was used for this purpose in mapping the draft sequence of the human genome where 7,600 clones were successfully mapped cytogenetically (Cheung et al. 2001), but has rarely been applied to many genomes since because of the technical difficulties associated with multiplexing FISH experiments.

Of the 50 or so avian genomes so far sequenced, all apart from the five assembled to chromosome level (chicken, turkey, duck, zebra finch and collared flycatcher) are constructed to the scaffold level meaning that there are at least 100-150 scaffolds per genome. Without being physically anchored to chromosomes, these scaffolds result in a genome that is essentially a 'bag of bits' from which little inference can be made about their role in chromosomal rearrangements.



In order to overcome some of the limitations of NGS outputs, Kim et al. (2013) developed the algorithm 'reference-assisted chromosome assembly' (RACA) as a means of ordering and orientating scaffolds into longer chromosomal fragments. The algorithm works by using a closely related reference genome with an outgroup genome to create predictions of chromosome organization of predicted chromosome fragments (PCFs) in a *de novo* sequenced species (Figure 1-15). Validation with PCR to detect 'chimeric' (misaligned) regions ensures accuracy of mapping without the need for physical or genetic maps. At best however, this method produces sub-chromosome sized fragments that require further verification and subsequent chromosome assembly. RACA applied to the Tibetan antelope (*Pantholops hodgsonii*) and blind mole rat (*Spalax galili*) genomes significantly improved continuities of these assemblies but they still contain more than one large PCF for most chromosomes (Kim et al. 2013; Fang et al. 2014). The authors also found that higher starting N50 values produced more robust assemblies, therefore reinforcing the notion that even using tools that overcome some of the limitations of NGS, an improvement in NGS read length is nevertheless a priority. However, this technique is still limited by its inability to place the PCFs correctly onto the physical chromosome meaning that the PCFs are still in effect 'super scaffolds'.

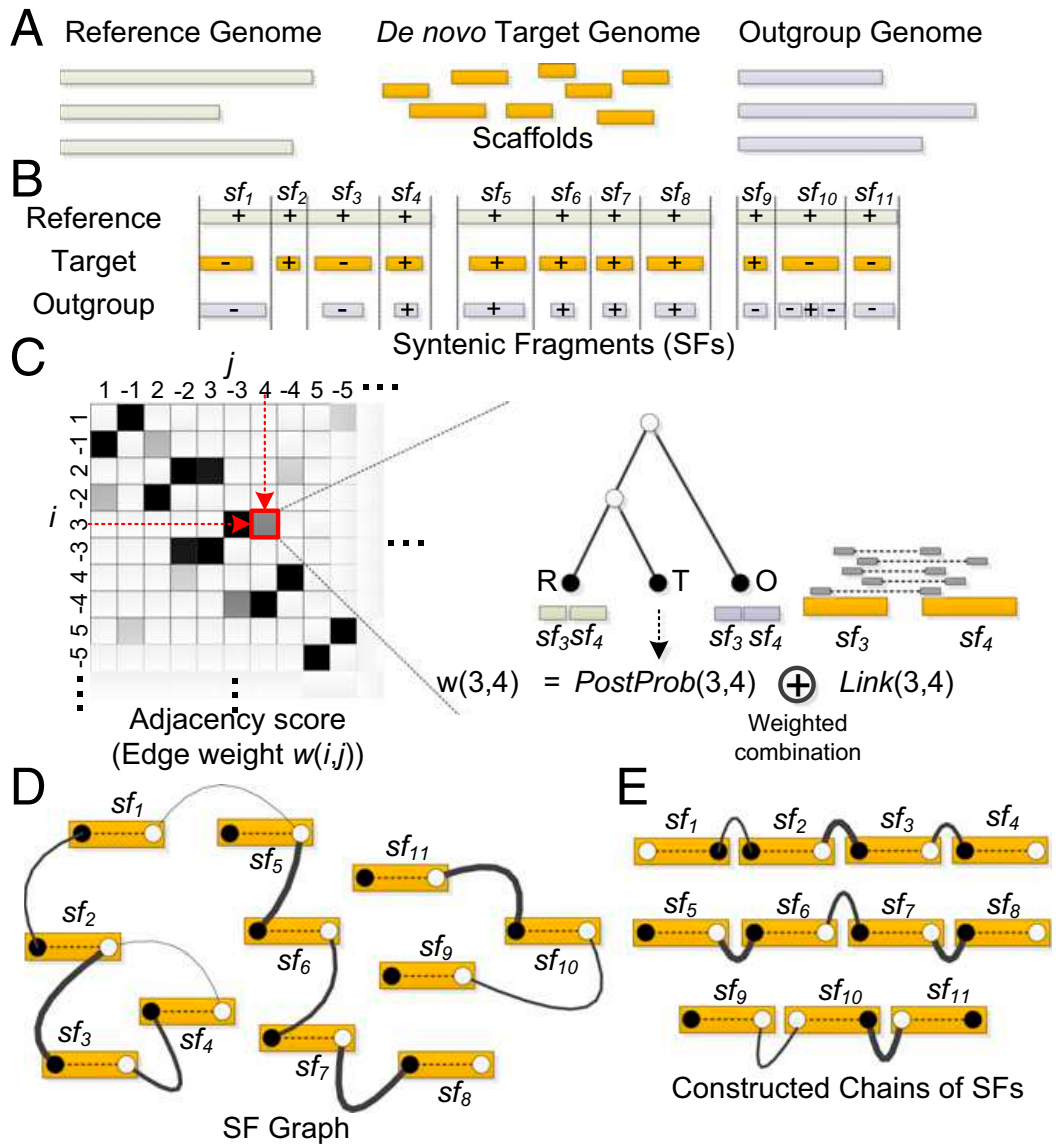


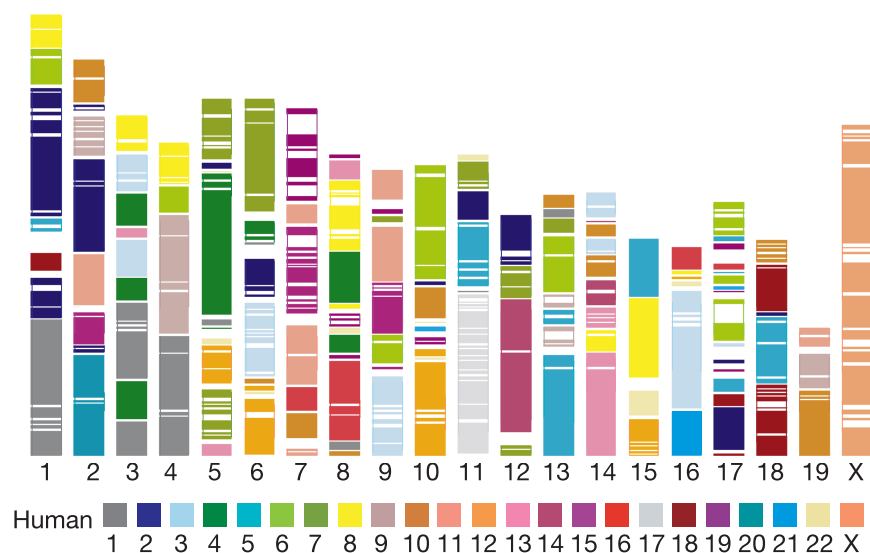
Figure 1-15: Overview of the RACA Algorithm.

(A) RACA requires a reference genome, a *de novo* sequenced target (in scaffolds), and an outgroup genome as input data. (B) Syntenic Fragments (SFs) are constructed by aligning the reference and target genome sequences. Pluses and minuses represent the orientations of the target and outgroup on the reference, and three groups of SFs represent regions from three reference chromosomes. (C) Adjacency scores are calculated for each pair of SFs. (D) The SF graph is calculated using the adjacencies. (E) Constructed chains of SFs are produced by the RACA algorithm (Kim et al. 2013).

### 1.1.7 Comparative Genomic Visualisation Tools at the Chromosome Level

Comparative genomics is the combination of genomic/cytogenetic data and evolutionary biology to address questions of genome structure, function and evolution between species. Whole genomes or parts of genomes are compared to those of other species to identify both similarities and differences between organisms - the assumption being that closely related organisms have a higher proportion of genomic similarity.

The starting point in comparative genomics is the comparison of a minimum of two genomes. Genomic sequences are aligned in order to identify orthologous regions from which genomic and molecular patterns can be inferred. At a sequence level, regions of conserved synteny are not just limited to functional DNA regions. In fact, comparison between the mouse and human genomes reveal that 99% of protein coding genes align with homologs in mouse but at the nucleotide level only around 40% of the human genome aligns with mouse. The remaining 60% of the genome is comprised of repetitive elements from each species that do not align with one another (Waterston et al. 2002). Despite a high degree of gene synteny between the two species, comparison of syntenic regions at a karyotypic level reveals a significant amount of gross chromosomal rearrangement as shown in Figure 1-16.

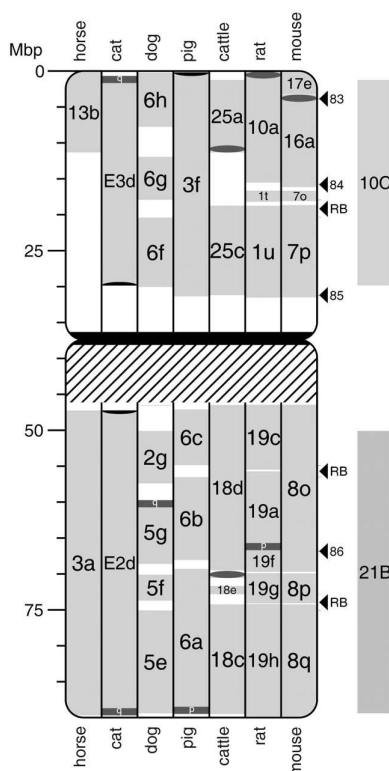


**Figure 1-16: Syntenic blocks between human and mouse superimposed on to the mouse genome, illustrating a significant level of genomic rearrangement (Waterston et al. 2002).**

The chromosomal rearrangements observed between humans and mouse occurred after the two species diverged from their common ancestor 75 mya. In order to elucidate within which

lineage they occurred and when, comparison with other closely related species is required (Ma 2011). Armed with this information, inferences about speciation events and species-specific phenotypic adaptations can be more readily made.

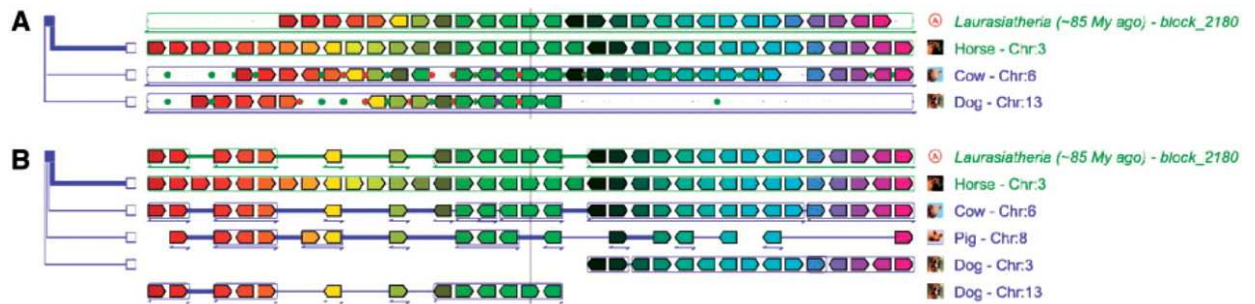
The abundance of sequenced genome data now available, whether at a scaffold or chromosomally assembled level, requires tools that enable visualisation of the genomes relative to each other. The ability to directly compare syntenic regions of multiple species is vital for the identification of conserved regions and evolutionary breakpoints and therefore the investigation of functional significance within them. Several tools have been developed for this purpose: at a sequence level, Evolution Highway (<http://evolutionhighway.ncsa.uiuc.edu>) allows for direct comparison of multiple sequence level genomes using one species as the reference genome, as illustrated in Figure 1-17 (Murphy et al. 2005).



**Figure 1-17: Screenshot of Evolution Highway, showing the alignments of multiple mammalian genomes against human chromosome 16 with the grey blocks representing HSBs and the chromosome number for each species listed against each block (Murphy et al. 2005).**

At a gene rather than sequence level, Genomicus (<http://dyogen.ens.fr/genomicus>) allows for the comparison of gene order of multiple species across a wide phylogenetic tree (Muffato et

al. 2010) as shown in Figure 1-18. In addition, ancestral reconstructions based on gene order adjacencies are possible using Genomicus.



**Figure 1-18:** Screenshot from Genomicus demonstrating how the gene PHOX2B compares to other closely related species with (A) showing PhyloView which illustrates the gene in a phylogenetic context and (B) showing AlignView which illustrates the PHOX2B and its orthologues in a greater number of species (Muffato et al. 2010).

Both tools are powerful means of visualising multiple, entire genomes allowing for comparative genome analysis at a high resolution.

### 1.1.8 Genome Reconstruction using Bioinformatics and FISH

#### 1.1.8.1 Ancestral Genome Reconstruction

The fundamental principle behind Darwin's theory of evolution is that all extant species share common ancestors at varying points of time (Darwin, 1859). In addition to facilitating inter-species comparisons, comparative genomics also provides a glimpse into the past by using genomic comparisons to make inferences about the common ancestor of the lineage of interest. At a morphological level, clues to the features of ancestral species have long been possible through the investigation of and interpretation of fossil evidence but now the relatively young field of paleogenomics is seeking to uncover shared ancestral features at a DNA level. The discipline of paleogenomics is focused on two primary areas of research; the first being the identification of ancestral features through the extraction and subsequent analysis of ancient DNA. The second area focuses on the analysis of genomic sequence DNA of extant species as a means of identifying ancestral features. The latter is possible due to the limited number of possible chromosomal rearrangement mechanisms that lead to genomic reshuffling, making it feasible to draw assumptions about rearrangements that may have occurred in each lineage and therefore make inferences about the ancestral state (Muffato & Roest Crolius 2008). Reconstruction of ancestral genomes firstly requires correct phylogenetic placement of the taxa

of interest. These taxa are represented as the tips of the phylogenetic tree with the region where they join (node) denoting their common ancestor. Once the phylogeny has been established synteny information is required in order to compare DNA organisation of related species to each other. This synteny data can be either be sequence based and therefore analysed using a computational approach or can be cytogenetic data that is investigated using cross species chromosome painting (described in section 1.1.2.2).

#### **1.1.8.2 Zoo-FISH Ancestral Reconstruction Methods**

Prior to the development of genome sequencing, early efforts to reconstruct ancestral karyotypes were restricted to using cytogenetic methods (zoo-FISH). Preliminary research focused on closely related mammals (Wienberg & Stanyon 1997) with a particular focus on the Boreoeutherian ancestor (Froenicke 2005). This reconstruction method has the advantage that data for in excess of 80 species of mammal is already available (without requiring a fully sequenced genome). However, whilst able to provide informative gross structural information this method is nevertheless limited; firstly by the evolutionary distance between target species (as described in section 1.1.2.2), secondly by the lack of precision in terms identifying intrachromosomal rearrangements (Wienberg 2004) and finally by the limitations of only being able to visualise regions greater than 4Mb (Froenicke et al. 2006).

#### **1.1.8.3 Bioinformatic Ancestral Reconstruction Methods**

An alternative approach to ancestral genome reconstruction is to analyse sequenced genomes using a bioinformatics approach. The ever-increasing number of assembled genomes has created this additional avenue of research thereby maximizing the time and cost already committed to genome sequencing. Robust tools that are able to decipher the most likely ancestral configurations are required for this type of analysis, particularly with the large number of genomes continually being sequenced. Increasing amounts and resolution of data require tools that have the capacity to perform ever more complex multi species comparisons at any one time (Avdeyev et al. 2016).

Genomic sequences may be analysed for ancestral reconstruction purposes at a range of resolutions: from the karyotype level to that of gene order and finally that of genomic sequence (Muffato & Roest Crollius 2008). At a karyotype level, the overall gross genomic structure can

be compared between species and inferences about gross genomic rearrangements can be made. At the gene order level, targeted loci can be investigated by examining orthologous genes across multiple species (Louis et al. 2013). Finally, at a sequence level, genome comparison enables the visualisation of homologous syntenic blocks (HSBs) and the corresponding breakpoints between these regions of conservation known as evolutionary breakpoint regions (EBRs). Analysis of functional significance within these regions is then possible along with the identification of rearrangements that have occurred in one or more species compared to their common ancestor (Larkin et al. 2009).

The three primary methods of ancestral reconstruction are based on either the principles of maximum parsimony, maximum likelihood or Bayesian inference. Briefly, the method of maximum parsimony requires that the minimum number of changes must have occurred between two points of evolution (Steel & Penny, 2000). Whilst technically the simplest method, it does not take into account the realities of selection pressures that cause evolutionary changes and therefore don't necessarily follow this simplistic pattern. The second of the three primary methods is that of maximum likelihood (Steel & Penny, 2000). This uses a statistical approach based on observed phenotypes in the taxa from which it can conclude the most likely ancestral state. The third of the methods is that of Bayesian inference which uses a probabilistic model for reconstruction (Ronquist & Huelsenbeck, 2003).

Reconstruction of ancestral genomes at this level requires a complex algorithmic approach. A series of different algorithms have been developed for this purpose, including MGR (Multiple Genome Rearrangement) developed by Bourque and Pevzner (2002) which uses a maximum parsimony model; inferCARs developed by Ma et al. (2006) and more recently inferCARsPRO (Ma 2010) and PMAG (Hu et al. 2014) which rely on gene adjacencies to reconstruct contiguous ancestral regions (CARs); and MGRA written by Alekseyev and Pevzner (2009) which relies on multiple breakpoint graphs and was recently updated to MGRA2 to take into account gene gains and losses as well as uneven genome sizes (Avdeyev et al. 2016). Each of these algorithms uses a slightly different method of ancestral reconstruction but fundamentally each method starts with comparing a number of genomes and generates the most likely common ancestral configuration by assimilating the closest species to a median ancestral state (Muffato & Roest Crollius 2008).

#### **1.1.8.4 Summary of Cytogenetic Methods used to Study Chromosomes**

Taken together, the means of studying chromosomes, karyotypes and chromosomal rearrangements extends from classical cytogenetics, molecular cytogenetics (FISH), cytogenomics (array-CGH), radiation hybrid (RH) mapping, genetic linkage mapping through to whole genome sequencing. The study of chromosomal rearrangement, both from an individual (clinical) perspective and from an evolutionary point of view is possible using all of these tools. In addition, by applying statistical models and bioinformatic tools, it becomes possible to reconstruct evolutionary events and determine the overall genome structures of common ancestors. The following section explores how this has been achieved (with an emphasis on the species of interest in this thesis) and, in particular how methodologies and mechanisms can apply both clinically and in an evolutionary context.

## **1.2 Chromosome rearrangements in medicine and evolution**

Chromosomal rearrangements can be studied in the context of changes that affect individuals (deviations from the norm) or changes that occur between species (comparative genomics). A central tenet of this thesis is to explore commonalities between the two, both from a technical (see above) and from a mechanistic (as follows) perspective. In humans, chromosomal abnormalities are seen in around 1 in 200 live born individuals, 1 in 20 stillbirths and around 1 in 2 spontaneous abortions (Hassold & Hunt, 2001). Recent evidence suggests that the majority of human embryos have some sort of chromosome abnormality, at least in low level mosaic form (Taylor et al. 2014). Constitutional abnormalities (i.e. affecting most cells in the conceptus) are generally lethal at early stages of embryogenesis resulting in early embryo loss, and are probably the leading cause of IVF failure (Munné, 2006). A small minority of abnormalities (such as trisomy 21) can result in a live birth but affected individuals often have a reduced lifespan and/or fertility. Balanced structural chromosomal abnormalities (i.e. ones that do not lead to a net gain or loss of DNA) usually result in a normal phenotype with the abnormality only becoming evident when fertility is evidently compromised (Stern et al. 1999).

Chromosome rearrangements can either cause or reinforce reproductive isolation, either reducing reproductive fitness by causing meiotic segregation errors in gametes, thereby reducing fertility or causing a mutation that is sufficiently deleterious that it becomes lost by natural selection. Unless the population size within which the mutation occurs is small then it is unlikely that these mutations will become fixed, although at times of enforced population



reduction (such as in response to natural disaster) it is more likely that genetic drift will promote fixation (Burt et al. 1999). Despite this, genome evolution is frequently investigated using quantitative methods at the level of nucleic acid and protein variation rather than focusing on the chromosomes themselves as a means for generating phenotypic variation (Larkin et al. 2009). This thesis is an exception.

### **1.2.1 Numerical chromosome abnormalities**

Numerical chromosome abnormalities, by definition, involve the gain or loss of whole chromosomes. They can be divided into polyploidy, where an entire extra set or sets of chromosomes are present, and aneuploidy, where there is a gain or loss of a whole chromosome and therefore an alteration to chromosome number in only part of the set.

#### **1.2.1.1 Polyploidy**

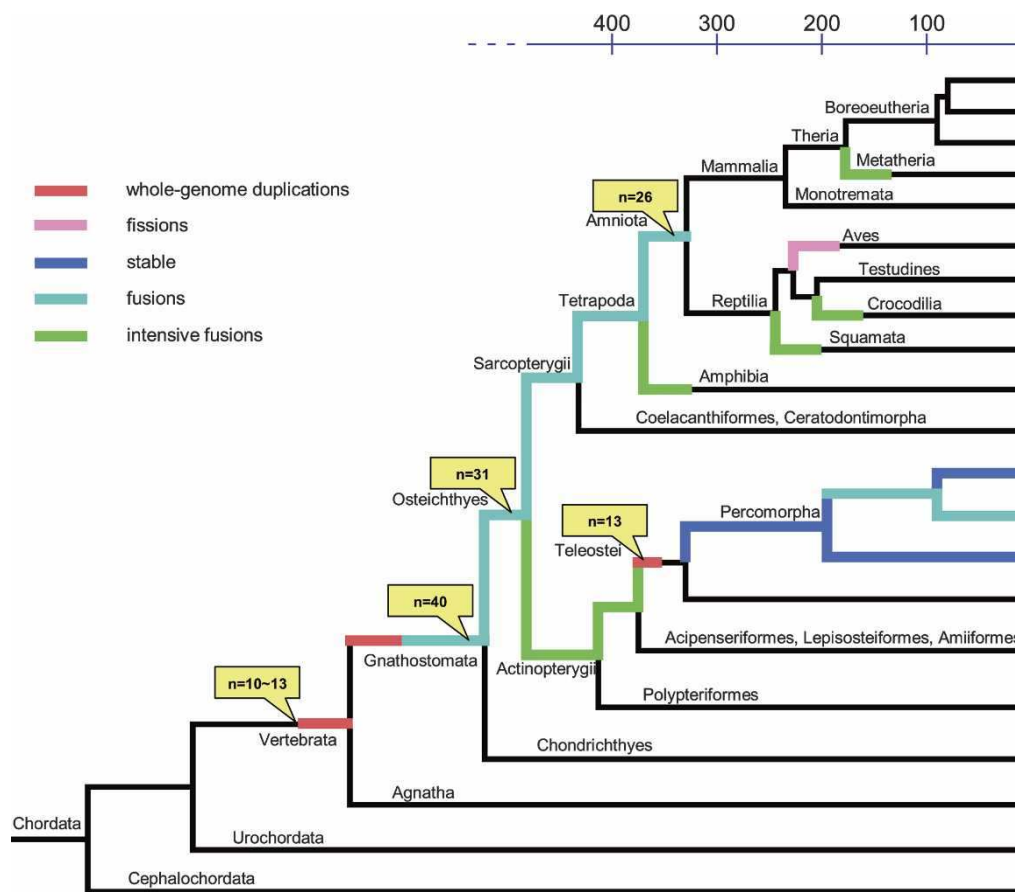
An organism exhibiting an extra set or sets of a genome has implications both clinically and in evolution. In humans, polyploidy is seen as either triploidy (3x) or tetraploidy (4x) and usually results in spontaneous abortion, causing around 1 in 6 first trimester pregnancy losses (Eiben et al. 1990). Occasionally foetuses survive to term although survival beyond birth is unlikely. Usually arising by polyspermy or endoreduplication, polyploidy is an important consideration in IVF (*in-vitro* fertilisation) where embryos are routinely screened for 3PN (pronuclei) fertilizations to reduce the likelihood of a triploid embryo being transferred.

In evolutionary terms, genome duplication events (known as paleopolyploidy) have occurred in several eukaryote lineages. Plants in particular, appear to have undergone multiple whole genome duplication events, in many cases followed by diploidisation. Wheat, for example can be genomically diploid as seen in wild wheat (*Triticum monococcum* –  $2n=14$ ), or tetraploid as is seen in durum wheat (*Triticum dicoccoides* –  $2n=4x=28$ ) and hexaploid as seen in bread wheat (*Triticum aestivum* –  $2n=6x=42$ ) (Simmonds 1976).

Paleopolyploidy occurs much less frequently in the animal kingdom than in plants. In a landmark study in the 1970's, Susumo Ohno hypothesised that two rounds of whole genome duplication must have occurred early in vertebrate evolution in order to produce the complexity seen in modern vertebrate genomes (Ohno 1970). Doubling (or tripling) of the genome in this manner has two important evolutionary consequences: firstly, redundant alleles are able to evolve freely

with little selection pressure, thereby creating genetic diversity; and secondly, alterations in chromosome number can lead to reproductive incompatibilities within a population and ultimately provide a basis for speciation.

The ancestral vertebrate, dating back to around 450 mya, appears to have had 12-13 pairs of chromosomes (Postlethwait 2000). Two rounds of whole genome duplication followed, resulting in a diploid number of 80 in the gnathostome (jawed vertebrate) ancestor. A series of fusions appear to have taken place in the teleost lineage (from which amniotes evolved) bringing the chromosome number down to 26 pairs around 350 mya, with little change in the teleost lineage ever since. Lineage specific fusions subsequently occurred in the ancestral genomes of reptiles, birds, amphibians and marsupials as shown in Figure 1-19 (Nakatani et al. 2007).



**Figure 1-19: Vertebrate phylogenetic tree illustrating whole genome duplications and lineage specific chromosome fission and fusion events that occurred across the vertebrates (adapted from Nakatani et al. 2007).**

### **1.2.1.2 Aneuploidy**

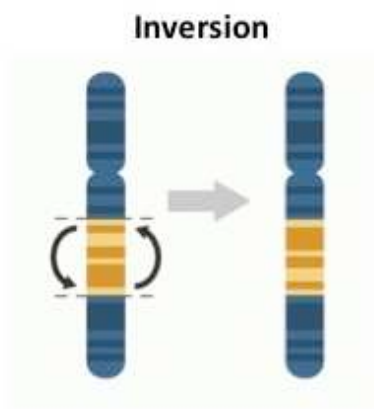
Aneuploidy refers to a state where one or more individual chromosomes has been lost or gained and is caused by non-disjunction events that occur during meiosis in the formation of gametes. In humans it is relatively common with around 5% of clinically recognised pregnancies, and 35% of spontaneous abortions exhibiting aneuploidy (Hassold & Hunt 2001). Aneuploidy can result in monosomies such as Turners syndrome (45 X0) or, more commonly trisomies such as trisomy 13 (Patau syndrome), trisomy 18 (Edwards syndrome) and trisomy 21 (Down syndrome) (Torres et al. 2008).

### **1.2.2 Structural Chromosome Abnormalities**

Structural alterations include chromosomal insertions, deletions, duplications and inversions as well as reciprocal translocations, all of which are initiated by erroneous repair from a double strand break (DSB). The faulty nature of these repairs can be either due to direct joining of incorrect DSBs or due to recombination with non-allelic homologous sequences (Schubert & Lysak 2011).

#### **1.2.2.1 Inversions**

Chromosomal inversions occur when two breaks in the same chromosome create a segment that then rotates 180 degrees and repairs in an inverted position (see Figure 1-20). Inversions can be 'pericentric' where the centromere is included in the inversion or 'paracentric' where it is restricted to the chromosome arm.



**Figure 1-20: Chromosomal inversion**

Despite inverting the gene order in the affected region, the phenotypic effect can be minimal, particularly in the case of paracentric inversions, which are seen more frequently than pericentric and often exhibited as polymorphisms within a species and fixed differences between species (Coyne et al. 1991). However, the rearrangement caused by an inversion can interfere with pairing of homologous chromosomes by creating inversion loops which impede meiosis (Griffin & Finch 2005) leading to the production of unbalanced gametes and subsequent reduced fertility or underdominance (Hoffman & Rieseberg, 2008). Inversions disrupt recombination to varying degrees according to the chromosomal region involved. From an evolutionary perspective, even if the effect of this disruption is comparatively small, the perpetuation of this rearrangement in a population will create an additional source of genetic variation over many generations (Kirkpatrick 2010). Examples of meiotic drive have been shown in heterozygotes carrying an inversion, where more than half of the gametes of the heterozygotes exhibit the inversion (Lyon 2003).

Until the advent of the genomics era, inversions were only detectable using classical or molecular cytogenetic methods. As the closest relative to humans, comparison of human and chimpanzee (*Pan troglodytes*) chromosomes, revealed only nine inversions using cytogenetic techniques (Yunis & Prakash, 1982). Comparison at a sequence level however, revealed 1,576 inversions (illustrated in Figure 1-21), all of which were distributed throughout the genome with no evidence of bias in regions prone to inversion (Feuk et al. 2005), although there is a strong likelihood that some of these may be due to assembly errors.

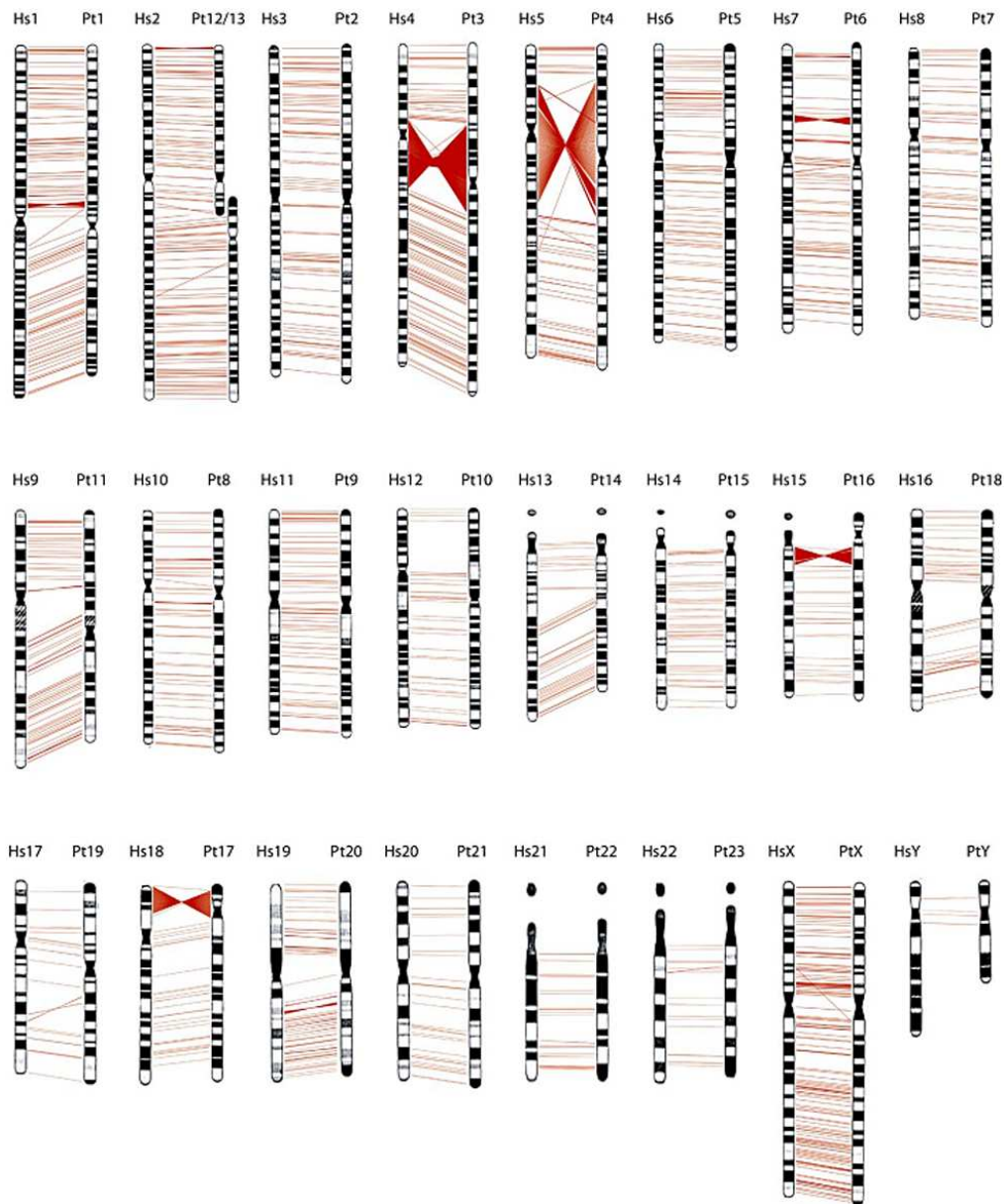


Figure 1-21: Distribution of chromosomal inversions between humans (on the left) and the chimpanzee homolog (on the right) inferred from sequence comparisons (Feuk et al. 2005).

### 1.2.2.2 Chromosomal Fissions, Fusions and Translocations

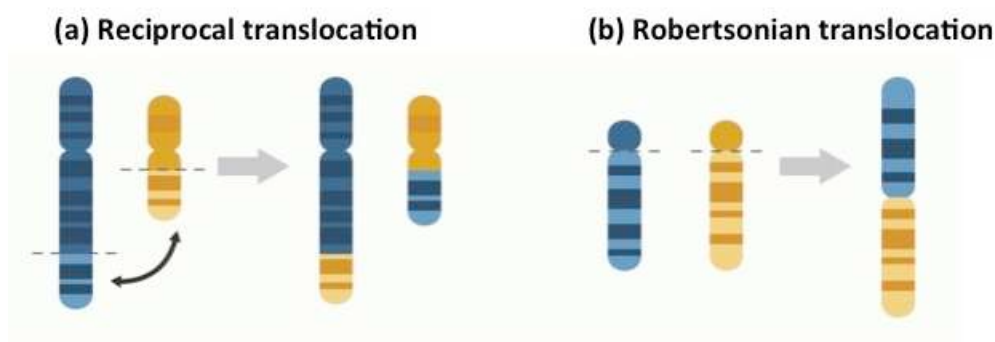
Karyotypic diversity is often explained by chromosomal fission and fusion. Using the example of chromosomal fission, it is thought that the ancestral mammalian karyotype was made up of a number of large mediocentric chromosomes which, in many species, through the process of chromosomal fission led to a karyotype consisting of a large number of small, acrocentric chromosomes (Todd 1970). A classic example of fission theory is that of human chromosomes 14 and 15 which were generated by the fission of an ancestral chromosome 14 (Ventura et al.

2003), the products of which then went on to become two telocentric chromosomes which developed their own telomeric regions to prevent further fusions (Giannuzzi et al. 2013).

In terms of chromosomal fusions, another example using the human karyotype is that of chromosome 2 which in this example is the result of an end-to-end fusion of two ancestral hominid chromosomes as evidenced by the presence of a vestigial centromere and interstitial telomeres (Ijdo et al. 1991). The specific translocation mechanisms are described below.

#### 1.2.2.2.1 Reciprocal Translocations

Reciprocal translocations occur when chromosomal regions are exchanged between two non-homologous chromosomes without any clear loss of genetic material (as shown in Figure 1-22). Carriers are typically phenotypically normal but often have a history of spontaneous abortions or have offspring with birth defects (Griffin & Finch 2005).



**Figure 1-22: Structural chromosome rearrangements. (a) Reciprocal translocation, illustrating exchange of genetic material between non-homologous chromosomes; (b) Robertsonian translocation.**

If the translocation has occurred within an important gene region in which expression is up- or down-regulated, the risk of cancer is greatly increased by potentially inactivating a tumour suppression gene or activating an oncogene (McLachlan & O'Bryan 2010). The translocation may also affect fertility by disrupting normal synapsis during meiosis due to the formation of a quadrivalent (as seen in Figure 1-23) which impedes the mechanics and timings of meiosis. The resulting chromosomal segregation leads to generation of unbalanced gametes, some of which are non-viable because they carry an incomplete set of genes (Shah et al. 2003).

Strictly speaking, most chromosomal fusions are technically reciprocal translocations as evidenced by the presence of only one centromere and infrequent evidence of interstitial

telomeric sequences. 'End to end fusions' are theoretically reciprocal translocations that have occurred in acrocentric or telocentric chromosomes where the exchanged region of each chromosome occurs in a very distal region of the chromosome, leaving a very small derivative product which is subsequently lost in meiosis (Schubert & Lysak 2011). In humans, the source of structural chromosome anomalies may be *de novo* (20%) or familial (80%) (Jacobs 1992). *De novo* structural abnormalities are overwhelmingly paternal in origin (84.4%) with the exception of Robertsonian translocations which are primarily maternal in origin (65%) (Chandley 1991).

#### **1.2.2.2.2 Robertsonian Translocations**

Robertsonian translocations are a product of the centric fusion of two non-homologous acrocentric chromosomes as shown in Figure 1-22. The product of which is the formation of one metacentric chromosome with two q-arms (long arms) and the loss of the p-arms (short arms) leading to a reduction in the overall chromosome number by one. Fertilised gametes from the carrier of a Robertsonian translocation may result in an embryo that is normal, a carrier, a full trisomy (e.g. familial Down's syndrome) or a complete monosomy. Monosomic zygotes from Robertsonian translocations are incompatible with life and most translocation trisomy conceptuses (with the exception of trisomy 21) result in first trimester loss (Scriven 2001).

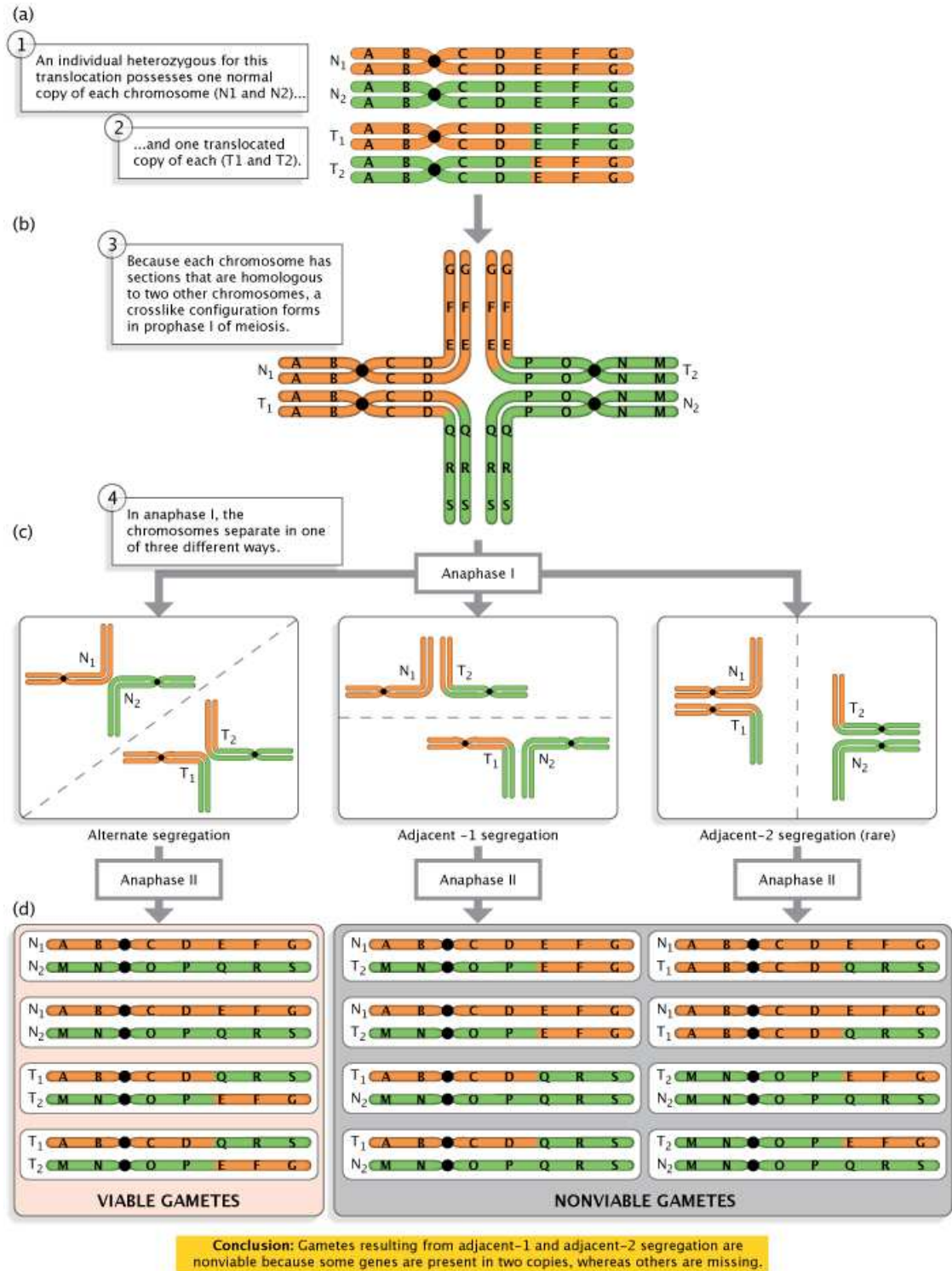


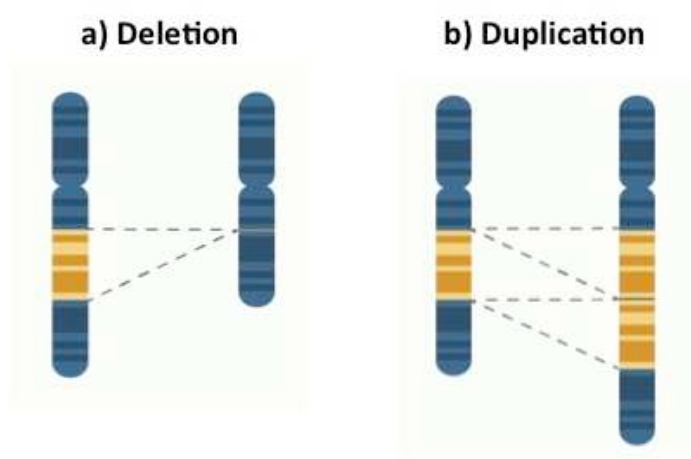
Figure 1-23: Meiosis in an individual carrying a reciprocal translocation illustrating formation of a quadrivalent, subsequently leading to incorrect segregation during anaphase which results in unbalanced gametes (O'Connor C, 2008).



### 1.2.2.3 *Indels (Insertions and Deletions), Duplications, CNVs and Other Structural Rearrangements*

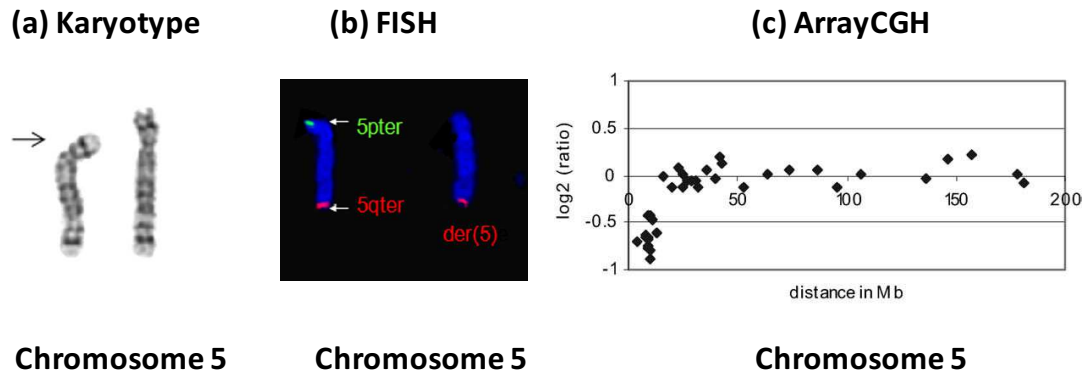
#### 1.2.2.3.1 Indels

Often described together as 'Indels' – Insertions and deletions can affect fertility and speciation to varying degrees. Insertions occur when additional regions are deleted from one chromosomal region and inserted into another as a consequence of pairing errors in meiosis (Feuk et al. 2006). The size of these insertions can vary widely and phenotypic consequences vary according to the genic region that has been disrupted. The loss of DNA sequences from the genome is described as a deletion (illustrated in Figure 1-24). The resulting phenotypic effect and viability depend on the size of the deletion and the region involved with larger deletions more likely to involve more genes and therefore have a more pronounced effect. A reduction in the number of genes in a region may have implications for gene dosage, which may in turn affect phenotype depending on the size and region involved.



**Figure 1-24: Structural Chromosome Rearrangements: (a) Deletion; (b) Duplication.**

A well characterised example of a deletion in humans is Cri du Chat syndrome. This is a rare medical condition that results in mental retardation, microcephaly and a characteristic cry. It is caused by a deletion in the p-arm of chromosome 5 (Lejeune et al. 1963) and can be diagnosed cytogenetically with a karyotype, with FISH and with array-CGH as illustrated in Figure 1-25.



**Figure 1-25:** Different methods of diagnosing Cri du Chat syndrome. (a) The G-banded karyotype on the left has an arrow marking the small deletion seen on the p-arm of chromosome 5. (b) The FISH image illustrates the normal chromosome on the left with two signals and the abnormal chromosome on the right with an absence of a p-arm (green) signal illustrating the deletion (Sun et al. 2014). (c) Array-CGH output illustrating the deletion of DNA at the p-arm of chromosome 5 (the X axis represents the distance in Mb from the p-terminus and the Y axis represents the hybridisation ratio (Rickman et al. 2005).

#### 1.2.2.3.2 Duplications

Duplications at a chromosomal level occur when an extra copy of chromosomal region is formed (as shown in Figure 1-24). This can be a result of meiotic pairing errors between misaligned chromosomes, often at sites with a large degree of repeat DNA. This type of rearrangement may have the effect of increasing the number of genes in the region affected thereby altering gene dosage. Duplications have long been considered to be an important factor in evolution, generating genetic diversity through the creation of duplicate genes that are under reduced selection pressure, thereby allowing mutations to occur more readily (Ohno 1970).

#### 1.2.2.3.3 Copy Number Variation

Copy number variation (CNV) is thought to contribute to phenotypic variation, disease and speciation, however until recently most studies focussed on the differences within a species (mostly humans and some farm animals). An example of CNVs in humans is that of variation in the *AMY1* gene which encodes for one type of the starch digesting enzyme amylase. In this case, variations are seen both between primates and humans (suggesting an ancestral origin) and between populations (Perry et al. 2007). From a disease perspective, CNVs in human chromosome 22q11.2 are linked with development of DiGeorge syndrome (Bittel et al. 2009; Hiroi et al. 2013). Among pigs, CNVs in the region containing the *KIT* gene are known to result in the white/patch coat seen in some European pig breeds (Pielberg et al. 2002; Paudel et al. 2013).

As described in section 1.1.3 above, array-CGH has also been used as a tool beyond clinical diagnosis with investigations into cross species genomic comparisons allowing for the interpretation of the role of CNVs in evolution between many species, including primates (Dumas et al. 2007) and birds (Griffin et al. 2008; Skinner et al. 2009; Völker et al. 2010; Skinner et al. 2014).

#### **1.2.2.3.4 Other Structural Rearrangements**

Other structural rearrangements include abnormalities such as isochromosomes and ring chromosomes. Isochromosomes occur when the two arms of a chromosome divide incorrectly at the centromere during mitosis or meiosis creating a chromosome with two arms that are mirror images of each other. The resulting chromosome has a partial trisomy of some genes and monosomy of others. An example of a disorder caused by this mechanism is Turner syndrome – where 15% of cases are caused by an isochromosome X (Santana et al. 1977). Ring chromosomes occur when breaks in the subtelomeric region of each arm of a chromosome result in a fusion or from the fusion of one broken chromosome end and the telomeric region of the other end of the chromosome (Sigurdardottir et al. 1999). Ring chromosomes have been reported for all human chromosomes but like most structural rearrangements, the regions involved determine the genes affected and therefore the severity of the phenotypic effect.

### 1.3 Karyotype Variation

As implied in the beginning of this introduction, diploid chromosome numbers can vary widely, both within and across taxonomic groups. A graphical representation of the variation in diploid chromosome number across a range of vertebrate taxa is shown in Figure 1-26, demonstrating the extensive range of chromosome numbers seen, particularly across mammals.

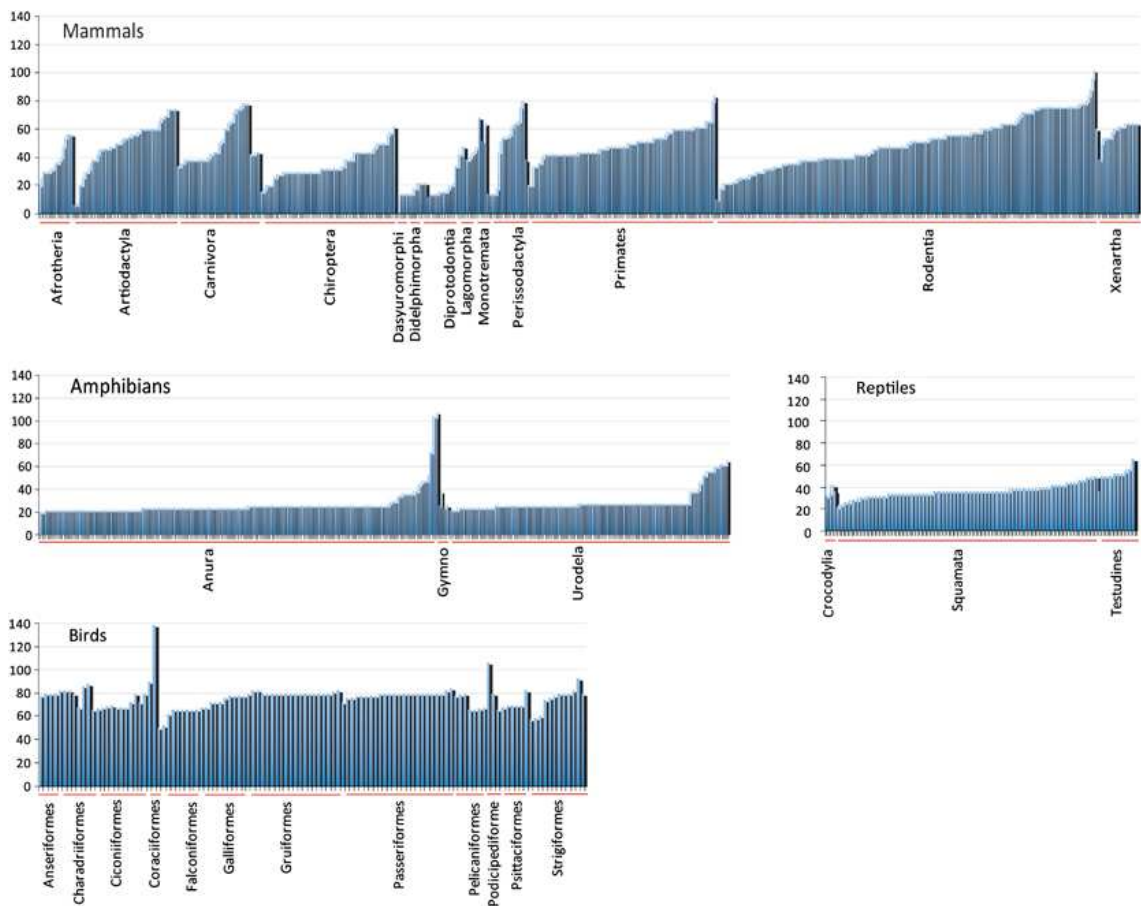


Figure 1-26: Chromosomal number variety in vertebrates. Diploid number is represented on the x-axis, with the y-axis indicating species in the different orders (Ruiz-Herrera et al. 2011).

#### 1.3.1 Mammalian Karyotype Diversity

In mammals, chromosome numbers can be as low as 6 (female)/7 (male) in the Indian muntjac (*Muntiacus muntjak*) and as high as 102 in the viscacha rat (*Tympanoctomys barrerae*) as illustrated in Figure 1-27 (Contreras et al. 1990). Evolved from an ancestral mammal with a diploid number of 46 (Ruiz-Herrera et al. 2011) investigations indicate that even within a genus, the karyotype number can vary enormously between species as seen in the muntjacs, where the Chinese muntjac (*Muntiacus reevesi*) has a diploid number of 46 compared to that of the Indian muntjac which (as described) has a diploid number of 6/7 (Wurster & Benirschke 1970). Within

the *Canidae* family, for example, there is great variety with some species such as the domestic dog exhibiting a karyotype of 38 acrocentric autosomes but the red fox displaying 16 pairs of metacentric autosomes. In this example it appears that almost all of the dog chromosomes have remained intact other than having fused at the centromere therefore forming metacentric chromosomes (Yang et al. 1999).

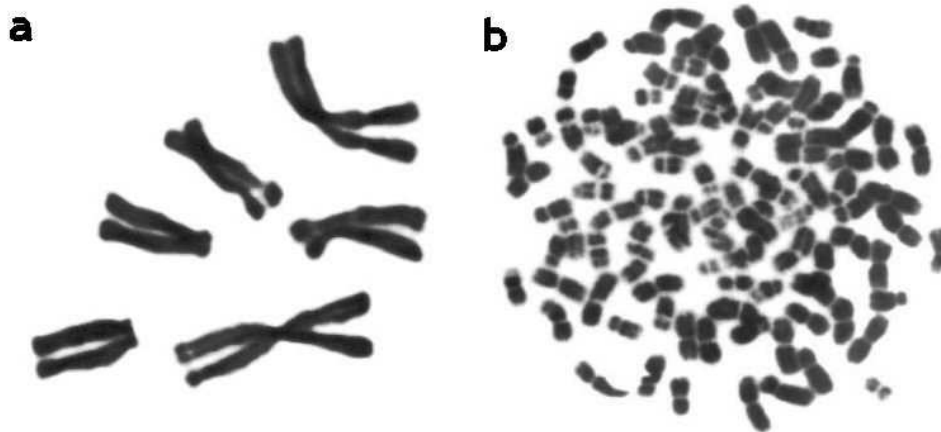


Figure 1-27: Variation in diploid chromosome number. a) Metaphase spread of the Indian Muntjac (*Muntiacus muntjak*) where  $2n=6/7$ . b) Metaphase spread of the Viscacha rat (*Tymanoctomys barrerae*) where  $2n=102$  (Graphodatsky et al. 2011).

### 1.3.2 Avian Karyotype Diversity

The highly distinctive, 'signature' avian karyotype is typically divided into around 10 macro chromosomes and around 30 multiple evenly sized, morphologically indistinguishable microchromosomes (Christidis 1990; Masabanda et al. 2004; Griffin et al. 2007), some of the better-characterised examples of which are illustrated in Figure 1-28. The morphological similarity of the microchromosomes and the sheer number of them makes it almost impossible to analyse with classic cytogenetic techniques such as karyotyping. In fact, to date although around 1,000 karyotypes have been published for a class that represents around 10,000 extant species, these are all partial karyotypes with only 5–10 pairs of chromosomes being easily identifiable. Rare exceptions to this 'avian style' karyotype include those with an unusually small diploid number such as the stone curlew (*Burhinus oedicnemus*;  $2n=42$ ), the beach thick knee (*Esacus magnirostris*;  $2n=40$ ) and several hornbills ( $2n=42$ ); and those with an uncommonly high diploid number such as the kingfishers and hoopoes (*Upupa epops*;  $2n > 120$ ) (Christidis 1990).

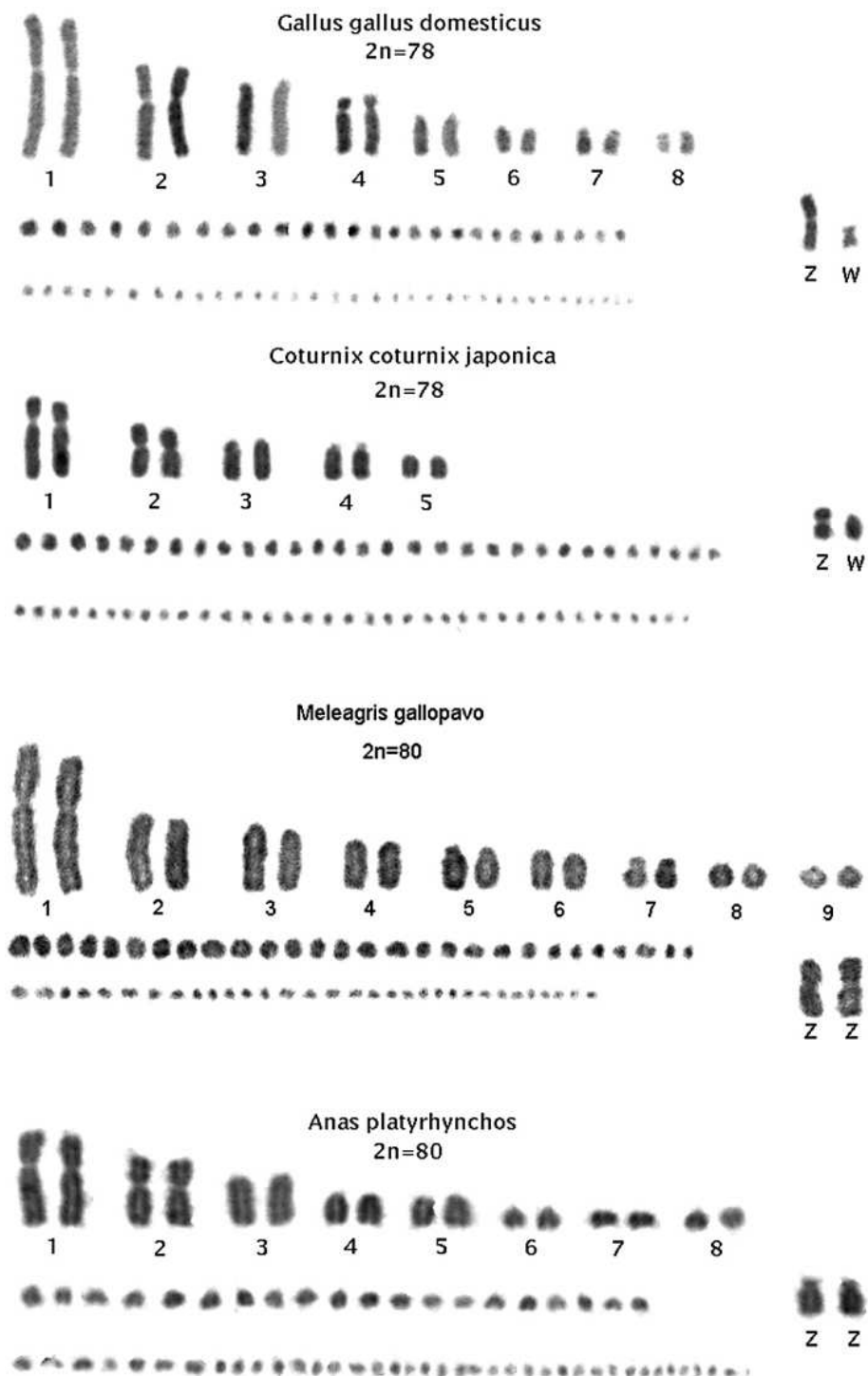


Figure 1-28: Typical avian karyotype organisation with few macrochromosomes and many, morphologically indistinguishable microchromosomes. From the top down, chicken, Japanese quail, turkey and duck are represented (Schmid et al. 2005).

### 1.3.3 Reptilian Karyotype Diversity

Both non-avian reptiles and birds generally display a similar karyotype pattern with a small number of macrochromosomes (up to 10 pairs) and many smaller morphologically indistinguishable microchromosomes, the exception being the crocodylians which lack microchromosomes and have an average diploid number of 30 (Cohen & Gans 1970), an example of which for the Nile crocodile (*Crocodylus niloticus*) is shown in Figure 1-29.



Figure 1-29: Giemsa stained karyotype of the Nile crocodile (*Crocodylus niloticus*;  $2n=32$ ) with chicken homology by chromosome number listed to the right of the chromosome (Kasai et al. 2012).

Snakes have a similar karyotype to that of birds but with many fewer microchromosomes, an example of which for the Burmese python (*Python morolus*) is shown in Figure 1-30. The average diploid number seen in snakes is  $2n=36$ , consisting of 8 pairs of macrochromosomes with the remainder being microchromosomes (Matsubara et al. 2006).

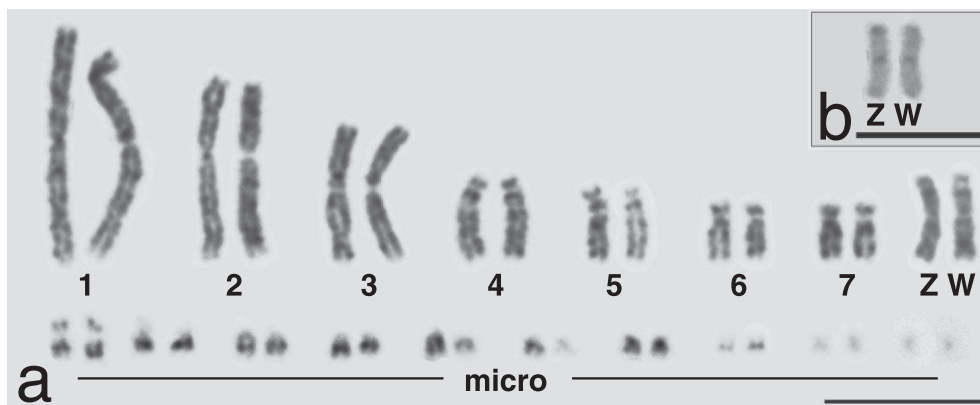


Figure 1-30: Karyotype of the Burmese python (*Python morolus*;  $2n=36$ ) a) Giemsa stained karyotype b) C-banded sex chromosomes (Matsubara et al. 2006).

Turtles also exhibit the macro and microchromosome karyotype structure as shown in Figure 1-31 with the most common diploid number being 50-52 among the *Emydidae* and 66 in the *Trionycidae* with evidence of a high level of conservation at the macrochromosomal level between turtles and birds (Matsuda et al. 2005).

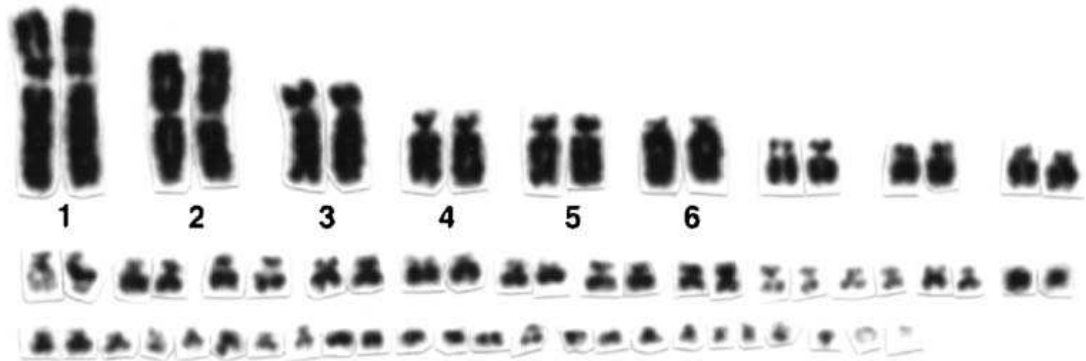


Figure 1-31: Giemsa stained karyotype of the Chinese soft-shelled turtle (*Pelodiscus sinensis*;  $2n=66$ ) (Matsuda et al. 2005).

## 1.4 Genome Structure Variation

In addition to the karyotypic diversity seen between mammals, reptiles and birds, there are also variations in genomic content. A summary of the key differences between birds and reptiles is described in this section.

### 1.4.1 Avian Genome Structure

The avian genome is characteristically small in size, being approximately one third of the size of a typical mammalian genome (ICGSC 2004; McQueen et al. 1998; Smith et al. 2000; Habermann et al. 2001) with little variance between species. Ninety per cent of avian genomes are between 1.0 to 1.6 Gb in size with a mean of 1.39 Gb. In mammals, the genome sizes are markedly larger with 90% of species having a genome size between 2.2 to 4.9 Gb with a mean of 3.36 Gb (Ellegren 2013). The larger genomes seen in mammals are primarily due to the presence of increased numbers of transposable elements (TEs) whereas the majority of avian genomes have lower levels of repeat elements than other tetrapod vertebrates with numbers ranging from 4-22% compared 34 to 52% in mammals (Böhne et al. 2008). The average total length of SINEs (short interspersed elements) in birds is also significantly lower than that of other reptiles at ~1.3 Mb



compared to 12.6 Mb in the alligator and 34.9 Mb in green sea turtle (*Chelonia mydas*), suggesting that reduction in SINEs occurred in avian lineage (Zhang G. et al. 2014a). The increased gene density seen in bird genomes (and also in the only flying mammals – bats) is due to the shortening of introns and a reduction in intergenic distance and the significantly shorter protein coding genes seen in birds (50% shorter than mammalian genes and 27% shorter than reptilian genes) and is speculated to be connected with the requirements of rapid gene regulation required for flight (Zhang G. et al. 2014a). In the chicken (representative karyologically of the majority of avian species), macrochromosomes range in size from 30 to 250Mb and microchromosomes are, on average, only 12Mb in length, the smallest being 3Mb (Pichugin et al. 2001). One remarkable feature observed in most avian karyotypes is the high diploid number (average  $2n=78$ ) and the presence of a large number of unusually small microchromosomes. At a molecular level, these microchromosomes are particularly unique in being extraordinarily GC-rich and gene-dense, whilst accounting for only 23% of the genome but 48% of the genes (McQueen et al. 1998; Smith et al. 2000; Habermann et al. 2001; Burt 2002). Burt (2002) proposed that the microchromosomes present in birds were established in the ancestral vertebrate karyotype 400 mya, which appears to be supported by the Nakatani 2007 study described in section 1.2.1.1 which found that many of the avian microchromosomes corresponded directly with gnathostome ancestor protochromosomes, inferring that the characteristically avian karyotype was established at an extraordinarily early stage of evolution.

#### **1.4.1.1 Interchromosomal Rearrangements in Birds**

A key feature, unique to birds, is the high level of karyotypic stability, with the majority of avian species having a karyotype that is very similar in terms of size and gross genomic structure to that of the chicken ( $2n=78$ ). Exceptions to this rule include the *Falconiformes* (falcons) and the *Psittaciformes* (parrots), both of which have reduced diploid numbers relative to chicken along with fewer microchromosomes suggesting evidence of chromosomal fusion. The use of chromosome paints derived from chicken (see section 1.1.2.3) has shown a high degree of conservation between the macrochromosomes, which supports the view that the avian genome is highly conserved, even across large phylogenetic distances. Technical difficulties creating microchromosomal paints have limited the scope of their use for this analysis (discussed in section 1.1.2.3). Preliminary work using a microchromosomal BAC – FISH approach performed in our lab found evidence of microchromosomes fusing to macrochromosomes in the gyrfalcon and the budgerigar, therefore beginning the process of using BACs to fill in gaps that have as yet been unassigned (Lithgow, O'Connor R.E. et al. 2014).

#### **1.4.1.2 Intrachromosomal Rearrangements in Birds**

Despite the apparently slow interchromosomal rearrangement rates seen in birds, it is apparent that the same does not apply to intrachromosomal rearrangements, which are seen considerably more frequently (Völker et al. 2010; Skinner & Griffin 2011). Comparison of the genomes of chicken, turkey and zebra finch and analysis using the Genalyzer tool (Choudhuri et al. 2004) revealed a high degree intrachromosomal rearrangement within the macrochromosomes, many of which were subsequently confirmed by FISH (fluorescence *in situ* hybridisation) (Völker et al. 2010; Skinner & Griffin 2011). Analysis of intrachromosomal rearrangements in the microchromosomes however is limited. To date, cross-species analysis of microchromosomes between birds has been limited to two main studies: the first by Rao and colleagues used their radiation hybrid assembled duck genome to compare the microchromosomes of the duck to those of the chicken (Rao et al. 2012). The second study was performed by our group (Lithgow, O'Connor R.E. et al. 2014) and found no interchromosomal rearrangements between the chicken, turkey and zebra finch microchromosomes but found multiple intrachromosomal changes.

#### **1.4.2 Genomic Structure of Reptiles in Comparison to Birds**

Despite the high levels of synteny between chicken and anole lizard microchromosomes, initial reports suggested that the high GC and low repeat content pattern evident in the avian microchromosomes is not seen in the lizard. In fact, the landmark anole lizard sequencing paper suggested that the lizard exhibits a remarkable level of GC homogeneity not seen in other birds or in fact mammals (Alföldi et al. 2011). This homogeneity was attributed to the existence of a proportionately higher number of transposable elements seen in larger genomes such as those of amphibians, non-avian reptiles and mammals. The presence of these repeat elements is also believed to lead to genomes that are more prone to inter and intrachromosomal rearrangements, resulting in a less bimodal distribution of chromosome size and a more even distribution of gene content (Burt 2002). Conversely, the low number of repeat elements found in the avian genome is consistent with the small genome size and karyotypic structure featuring gene and GC rich microchromosomes (ICGSC 2004). It originally appeared that the anole lizard is the exception however as the GC microchromosomal content in another lepidosaur, the Japanese four-striped rat snake (*Elaphe quadrivirgata*) is more consistent with that of birds (Matsubara et al. 2012). Alföldi and colleagues also argued that rather than just exhibiting a homogenous proportion of GC content between micro and macrochromosomes, they also found a complete lack of isochores in the anole genome. Isochores are long (>300kb) regions

found in amniote genomes which exhibit a pattern of either high GC content or low GC content (Bernardi, 2000) compared to the surrounding regions and therefore correspond to chromosomal bands (Saccone et al. 1993). Costantini and colleagues have recently disputed this claim, instead finding that within the macrochromosomes, isochores from the L2 and H1 families (GC poor and GC rich families respectively) are in fact present in the *Anolis* genome and importantly that the microchromosomes exhibit a higher density of H1 isochores, consistent with that seen in the chicken genome. Attributing the difference in results to a high proportion of gaps in the sequencing of the *Anolis* genome performed by Alföldi's group, Costantini's group conclude that the *Anolis* genome should not therefore be considered to differ from other vertebrates in this case (Costantini et al. 2016). It appears therefore that this low repeat genome with a GC bias towards the microchromosomes was established very early in the saurischian lineage and has hence been subject to extraordinarily low rates of change over this period of approximately 275 million years.

### 1.4.3 Sex Chromosome Evolution

Sex chromosomes in vertebrates are often differentiated from each other, as either male heterogamety as seen in mammals (XX female and XY male) or female heterogamety (ZW female and ZZ male), as seen in birds and most reptiles (Graves 2014). In mammals, the X and Y chromosome are markedly different in size with the Y chromosome often significantly smaller than the X and much richer in repetitive sequences (Waters et al. 2007). The Y is around 84 Mb in size with few genes, most of which are male specific. It has a small region of homology with the X (pseudoautosomal region, PAR) adjacent to where the sex-determining gene *SRY* is located (Skaletsky et al. 2003). The human X chromosome conversely is around 155 Mb in size with approximately 1,100 genes with multiple functions (Ross et al. 2005), it is also highly conserved between species with an almost identical gene order seen between human and horse (*Equus caballus*) (Raudsepp et al. 2004). It appears that the X and Y date back 180 mya to just before the eutherian and marsupial divergence (Cortez et al. 2014).

Unlike mammals, birds exhibit the highly conserved ZW sex chromosome system, with females being heterogametic (ZW) and males homogametic (ZZ). In all neognathae birds, the Z and W chromosomes are differentiated in terms of size with the W being largely heterochromatic, gene poor and significantly smaller than the Z. Ratites, however, have a W chromosome which is of a similar size to the Z and is homologous in its entirety with the exception (in the case of emus) of a small region near the centromere (Shetty et al. 1999). Despite the difference in size, it can be

inferred that that the ZW system was evident prior to the divergence of the Palaeognathae and Neognathae lineages (Deakin & Ezaz 2014). In chickens, the Z chromosome is 80Mb in size and contains around 1,000 genes making it far less gene dense than the autosomes. It also has 60% more interspersed repeats making it significantly distinct from the rest of the chicken genome (Bellott et al. 2010) but interestingly, similar (in terms of content structure) to the mammalian X (Graves 2014). Although morphologically resembling the XY system seen in mammals (with the exception of the ratites as described), the two systems exhibit no homology (Nanda et al. 1999) and in fact have completely independent origins with the avian Z chromosome instead sharing homology with human autosomes 5, 9 and 18 (Bellott et al. 2010) and the human X chromosome sharing homology with a block of the q-arm of chicken chromosome 1 and 20Mb of the p-arm of chicken chromosome 4 (a microchromosome in most birds) (Ross et al. 2005). There also appears to be no homology between the majority of reptile ZW and avian ZW systems, the snake for example has a ZW system but in this case, the homology is shared with chicken chromosome 2 not Z or W (Matsubara et al. 2006). The exception to this rule appears to be the gecko lizard (*Gekko hokouensis*), which does in fact share homology with the chicken Z chromosome (Kawai et al. 2009). The ZW system is not a feature that is consistent among all reptiles. Both XY and ZW sex chromosome systems are seen in lizards and turtles, with lizards showing the largest degree of variability among reptiles. Some reptiles such as crocodiles also exhibit temperature dependent sex determination (TSD) with the gekkonidae lizard family being particularly unusual in exhibiting all three methods of sex determination (O'Meally et al. 2012).

Unlike many other reptiles where sex determination is often temperature dependent, birds exhibit genetic sex determination (like mammals). Unlike mammals, however, the sex-determining gene is not SRY (which in fact lies on chicken chromosome 4) (Graves 2014). Instead, it has been suggested that the gene DMRT1 found on the Z chromosome may be the key to sex determination by using a dosage dependent system meaning that male determination requires two copies of the gene as found in ZZ males. DMRT1 has also been shown to be required for testis formation (Smith et al. 2009). There is still much debate as to what determines gender in birds with possible candidates including W specific genes that may determine ovarian function among other theories (Graves 2014). Improvements in the assembly of the Z chromosome achieved using a BAC by BAC approach (Bellott et al. 2010) along with work underway to improve the assembly of the W chromosome (Chen et al. 2012) will assist with resolving some of these questions. An overall comparison of genomic elements and sex determination systems and how they differ between amniotes is illustrated below in Figure 1-32.

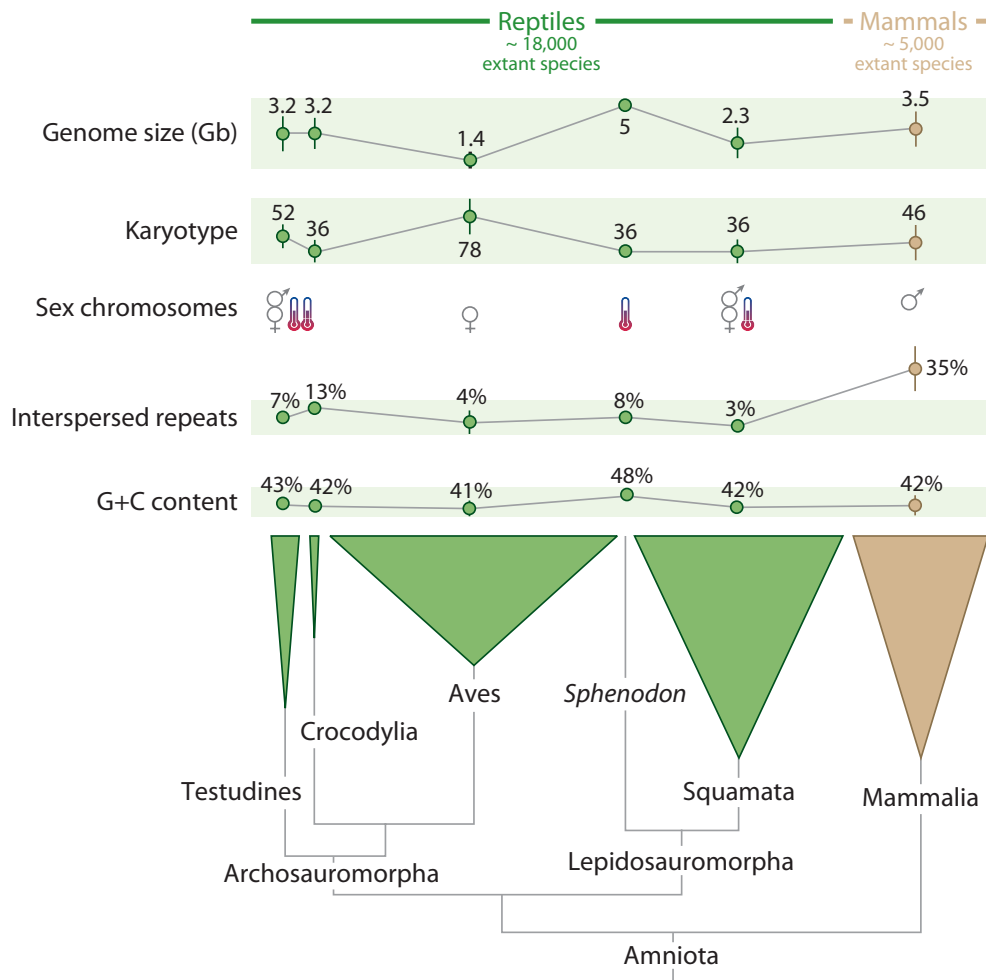


Figure 1-32: A comparison of genomic elements, genome size, karyotype number and sex determination systems in reptiles and mammals. The width of clades is illustrative of species diversity (Janes et al. 2010).

#### 1.4.4 Genome Structure in the Dinosaurs

Using osteocyte size as an indicator of genome size from fossilized bones, Organ et al. (2007) were able to identify a clear distinction between the small, characteristically avian genomes identified in the saurischian dinosaur lineage (that gave rise to modern birds) and the larger inferred genome seen in the ornithischian dinosaur lineage (that led to the lineage that includes stegosaurs and triceratops) that is comparable in size to those of modern crocodiles. They found that saurischian dinosaur genomes averaged around 1.78pg compared to ornithischian dinosaur genomes of about 2.49pg. Their findings suggest that much of the genomic adaptation widely considered to be an adaptation for the metabolic demands of flight in birds, such as small genome size and low repeat content, are features that evolved between 230 and 250 mya and, in this lineage, have changed little since (Organ et al. 2007). Despite claims in the scientific

literature (DeSalle et al. 1992; Cano et al. 1992) as well as fictional accounts such as 'Jurassic Park', studies have found that the preservation of DNA from dinosaurs or any other species (in amber or elsewhere) is unlikely to be achievable and therefore attempts to study the dinosaur genome directly are beyond the scope of possibility (Penney et al. 2013). Nonetheless using the tools described in previous sections it should be possible to reconstruct the overall dinosaur karyotype structure and this will therefore form part of the aims of this thesis.

## **1.5 Mechanisms of Genome and Chromosome Evolution**

High levels of genome conservation are not limited to closely related species. In fact, genome sequencing has revealed that conserved elements are present in species as evolutionary distant as humans and sea anemones. Specific blocks of linked genes common to both species have been identified, suggesting that the regions must be present in the eumatazoan (all animal groups with the exception of sponges and placozoa) ancestor 700 mya (Putnam et al. 2007).

### **1.5.1 Non-allelic Homologous Recombination**

Non-allelic homologous recombination (NAHR) at meiosis results in the separation of conserved segments of the genome and their subsequent fusion into different combinations. This then leads to the generation of gametes that can be both genetically balanced such as those seen in reciprocal translocations and those that are unbalanced (such as in Robertsonian translocations). Homologous sites that take part in chromosome rearrangements are often in repeat regions on the genome, most of which are derived from transposable elements (TEs) (Kidwell 2001). A low frequency of chromosomal rearrangement is widely considered to be due to a reduction in NAHR which in turn is considered to be due the low density of repeat elements and corresponding lack of substrates for recombination. The small number of repeats, segmental duplications and pseudogenes found in the avian genome previously reported by Burt and colleagues (Burt 2002; Burt et al. 1999) was also supported by evidence found using a bioinformatics approach by Skinner and Griffin (Skinner & Griffin 2011), and is consistent with the genomic structure found in birds. Equally, areas prone to recombination would therefore logically exhibit enrichment of CNVs and chromosomal rearrangements in these regions. The hypothesis would therefore follow that avian genomes with their characteristically low recombination rates should exhibit few CNVs and chromosomal breakage points (Völker et al. 2010).

In addition to NAHR as a mechanism for genomic rearrangement, non-homologous end-joining also leads to chromosomal rearrangement. Interestingly, breakpoint sites tend to lie near telomeres and centromeres, consistent with the formation of acrocentric chromosomes in some species and metacentric chromosomes in others. Whether chromosomal rearrangements occur because there is a specific benefit to the adapted new configuration is unclear although it is apparent that rearrangements between species can cause, or reinforce, reproductive isolation (Delneri et al. 2003; Noor et al. 2001; Rieseberg 2001).

### 1.5.2 Genetic Recombination

Genetic diversity is a consequence of and is maintained by both mutation and recombination. Understanding the mechanisms responsible for both is crucial for comprehending the patterns of variation seen across species and for understanding evolutionary patterns (Jensen-Seaman et al. 2004). In some animals, including humans, recombination 'hotspots' have been identified where a concentration of recombination events occurs (Arnheim et al. 2003). In the chicken genome specifically, GC-content and recombination appear to increase with decreasing chromosome size (ICGSC 2004), possibly explained by the cycle of 'biased gene conversion' which may increase the GC content as the recombination rate increases (Marais 2003). This finding is consistent with the significantly higher rate of recombination found in avian microchromosomes than mammalian chromosomes. Even within avian species there is a marked difference between the macro and microchromosomes with a median value of 2.8 cM/Mb in the macrochromosomes and 6.4 cM/Mb in the microchromosomes, compared to only 1-2 cM/Mb in mammals (ICGSC 2004). Given that at least one chiasma per bivalent is required at meiosis, a karyotype that retains multiple chromosomes will inevitably lead to a greater number of genetic variants in the gametes thereby increasing the recombination rate (Romanov et al. 2014).

Recombination rates were also studied in the zebra finch (*Taeniopygia guttata*) genome for variation by Backström et al. (2010) who found a lower rate (2 cM/Mb than that seen in chicken (revised figure based on inclusion of 10 microchromosomes not featured in the assembly). They also found a significant recombination bias towards the ends of the chromosomes, which resulted in the macrochromosomes effectively being a 'recombination desert' with 90% of the recombination being concentrated in 23% of the genome. Consistent with that found in the chicken genome, they found a positive correlation between recombination rate and GC content (Backström et al. 2010) suggesting that the zebra finch may also exhibit a recombination

'hotspot' pattern. Similar results were found in studies of the chicken genome, where a telomeric recombination bias and strong correlation was also found between high recombination rates and GC rich sequences (Groenen et al. 2009). A significant correlation between recombination rate and chromosomal breakpoints in the zebra finch has also been identified (Völker et al. 2010). Pairwise comparisons of avian genomes have facilitated the identification of evolutionary breakpoint regions (EBRs) (further described in section 1.5.3), allowing for a deeper analysis of genomic patterns associated with recombination.

In mammals, Larkin and colleagues found an increased rate of recombination between EBRs (Larkin et al. 2009). In a similar study, Romanov et al. (2014) tested the theory that higher recombination rates were found in EBRs between chicken and zebra finch. No association between regional recombination rate and EBRs in chicken was identified, and although a slightly elevated rate was identified in the zebra finch, this was not found to be statistically significant (Romanov et al. 2014). Interestingly, the recombination rate of the smallest human chromosome (HSA 22) is double that of the largest (HSA 1), suggesting that these features are evident in other species too and are in fact more obvious in the chicken genome due to the wide variation in chromosome size. In addition, the p-arm of chromosome 4 in the chicken exhibits an especially high recombination rate and GC content consistent with its origins as a microchromosome that fused in the lineage leading to chicken (Griffin et al. 2008).

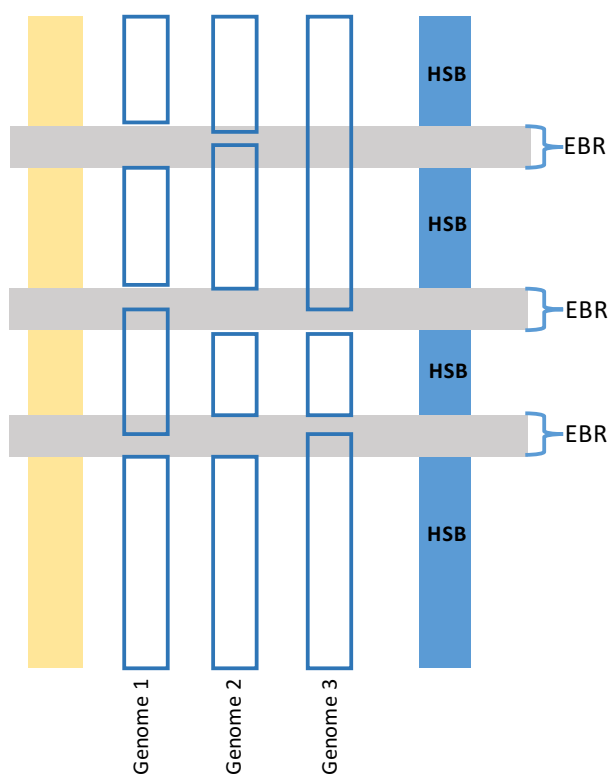
An alternative theory suggests that the underlying driver of chromosomal rearrangement is in fact the proximity of DNA regions to each other within chromatin and that the repetitive sequences in fact have little influence in this mechanism (Branco & Pombo 2006). A comparison of multiple mammalian and avian genomes shows that regardless of which theory underlies the mechanism there are regions of chromosomal stability - homologous synteny blocks (HSBs), between the evolutionary breakpoint regions (EBRs). The significance of these conserved HSBs in animal evolution is, however, still to be determined thanks to the paucity of available whole genome assemblies upon which meaningful comparisons can be made, despite evidence clearly showing a clear role for EBRs as 'hotspots' of chromosome evolution in mammals (Bailey et al. 2004), birds (Skinner & Griffin 2012), insects (Cáceres et al. 2001) and yeast (Gordon et al. 2009),

### **1.5.3 Homologous Synteny Blocks and Evolutionary Breakpoint Regions**

As described previously, on-going multiple genome sequencing projects are providing an extraordinary amount of data that allow for genome comparisons at a deeper level of resolution



than previously possible. Pairwise comparisons at a sequence level allow for the visualisation and identification of HSBs and EBRs as illustrated in Figure 1-33. The term HSB has more recently become routinely used as a means of defining regions of shared synteny between more than one species (sometimes known as msHSBs – multispecies HSBs). In general, however, HSBs are defined as regions of shared synteny where sequence and therefore gene order has been conserved among species and EBRs are defined as regions where chromosomal rearrangements have disrupted the gene order (Farré et al. 2011). More precisely EBRs are defined as the region between two HSBs demarcated by the boundaries at the end of each surrounding HSB (Murphy et al. 2005).



**Figure 1-33: Schematic representation of HSBs and EBRs (adapted from Farré et al. 2011).**

Extensive research focused on mammals has revealed differences in the type of genes that are situated within HSBs and flanking EBRs, with significant ontology signatures related to organismal development seen in mammalian msHSBs (Larkin et al. 2009) suggesting that there is an evolutionary requirement to keep these genes together. Within EBRs (frequently found in gene dense regions), lineage-specific biology and adaptive features were found to be enriched (Elsik et al. 2009; Larkin et al. 2009; Groenen et al. 2012). The assumption is that these ontology signatures are related to phenotypes specific to the lineages in which they occur in mammals. Similar research using the sequenced genomes of birds and reptiles by Farré and colleagues

revealed enrichment in archosaurian/testudine HSBs for Gene Ontology (GO) terms related to gene expression regulation and biosynthetic processes. Their analysis also found avian specific HSB enrichments for skeleton and limb development along with EBR GO enrichments for lineage specific phenotypic traits, examples of which include forebrain development in the budgerigar, potentially contributing to vocal learning development seen in parrots. Enrichments in pattern specification and regionalisation in the Adélie penguin may also reflect changes that have contributed to phenotypic adaptations that allow these penguins to swim and use less energy than other penguins (Farré et al. 2016).

#### **1.5.4 Chromosomal Breakpoints - Random or Non-random?**

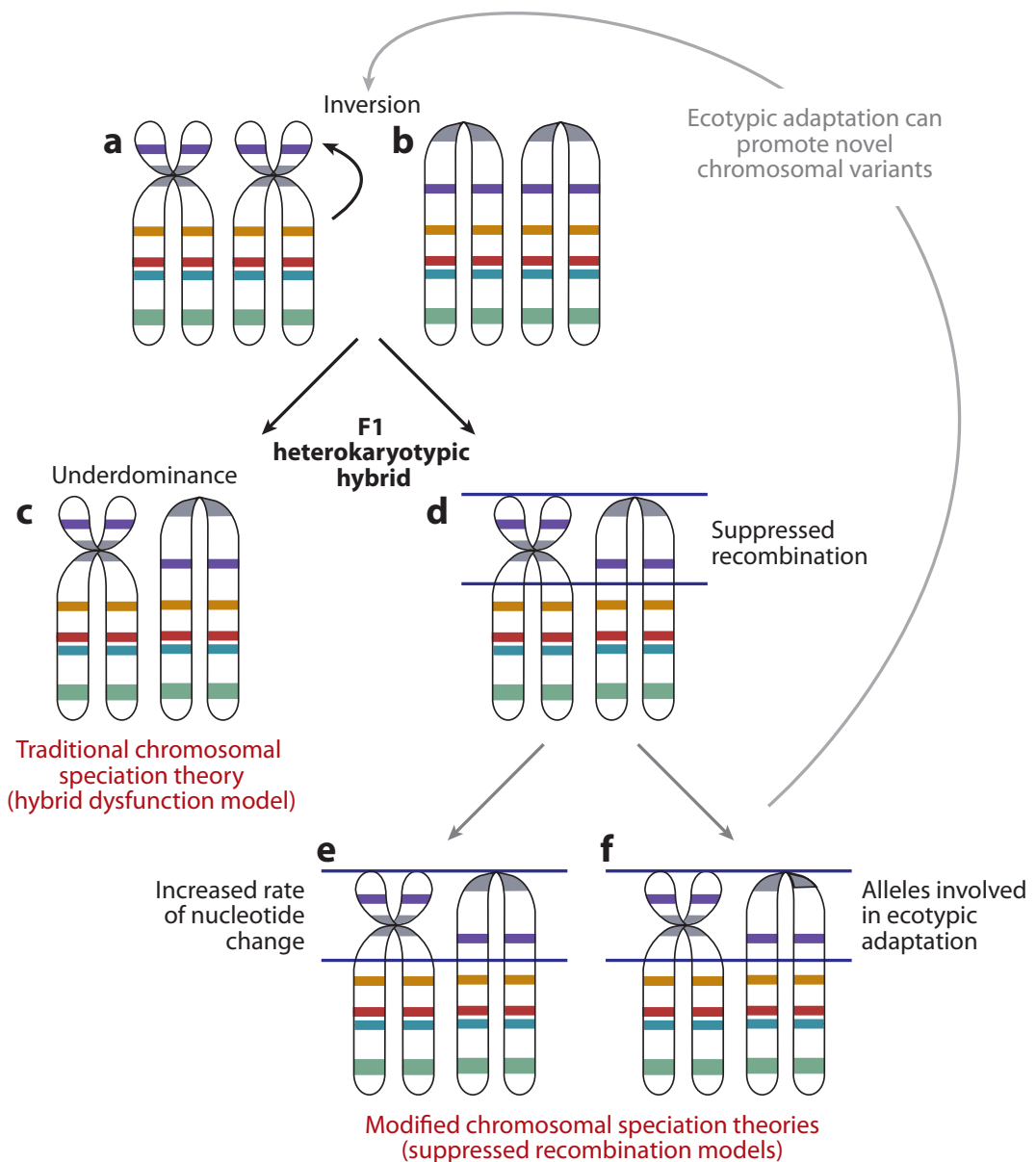
Chromosomal breakpoint reuse is defined as the appearance of the same breakpoint in more than one species that does not appear in the common ancestor. Whether these breakpoints occur at random genomic positions or at evolutionary 'hotspots' remains a subject of much debate. The random breakage model (RBM) proposed by Nadeau and Taylor (1984) examined the locations of around 80 homologous loci between the genomes of human and mouse and observed an independent and uniform distribution of rearrangement breakpoints. Studies completed since with increasing levels of resolution (Copeland et al. 1993; DeBry & Seldin 1996; Burt et al. 1999; Lander et al. 2001), went on to support this model ultimately leading to RBM becoming established as the default theory of chromosome evolution (Peng et al. 2006). In 2003 however, Pevzner and Tesler proposed an alternative model whereby the shuffling of gene order between human and mouse would require multiple closely located breakages inferring that there was some kind of clustering effect in the breakage pattern. They described these regions as 'fragile', proposing that these are more prone to rearrangements thus eliminating any random nature to their occurrence (Pevzner & Tesler 2003). Since then, sequence based and radiation hybrid maps of eight mammals have been analysed and found to support the 'hotspot' theory, even finding that around 20% of the breakpoint regions identified appeared to be reused (Murphy et al. 2005). The number of fragile sites in any given genome is still to be established, meaning that a hybrid model may be possible whereby a small number of fragile sites may be the starting position for any break but that the other end of the break may occur randomly (Peng et al. 2006).

### 1.5.5 Chromosomal Rearrangements and Speciation

Rearrangements at the chromosome level affect speciation in two primary ways; firstly, they can cause pairing errors in meiosis (this is seen in patients who are heterozygous for a rearrangement) and secondly they result in unbalanced gametes. Both of these causes disrupt the appropriate genetic material being passed onto the gametes and when they occur in gene rich regions, they can cause mutations or disrupt their position relative to regulatory regions. These can be lethal but can also lead to the evolution of lineage specific genes. In order for any chromosomal rearrangement to become fixed in the population, it must persist through to the gamete stage and subsequently contribute to the creation a viable heterokaryotypic zygote. Once it has reached reproductive maturity it must survive through generations of selection, ultimately emerging as a homokaryotypic feature. The presence of a fragile site is therefore just the beginning of a very long path to evolutionary fixation (Sankoff 2009).

The traditional chromosome speciation theory, also known as the hybrid dysfunction model (illustrated in Figure 1-34), uses three definitions of the impact chromosome rearrangements have on speciation: the first category encompasses those changes that are deleterious and therefore rapidly eliminated by natural selection, the second are those that result in population polymorphisms and the third category are those that cause 'underdominance' – reduced reproductive fitness in heterozygotes (White 1969). In this model, changes such as translocations, inversions etc. become fixed in a population leading to speciation. As described, rearrangements at the chromosomal level can occur in several different ways: fusions and fission of chromosomes, duplications or deletions of chromosomal segments, inversions of chromosome segments and finally translocations between non-homologous chromosomes. These can all lead to chromosomal changes that in individuals that are heterozygous for the change, will affect meiotic segregation, often resulting in unbalanced gametes (Rieseberg 2001). The resulting unbalanced gametes may be unviable and if they do make it to fertilisation, are more likely to result in zygotes that fail to develop to maturation. A major criticisms of the hybrid dysfunction model is that it does not allow for the possibility that a strongly underdominant rearrangement would be immediately removed from the population and a weakly underdominant one would not necessarily have a strong enough effect to cause isolation as well as making the assumption that heterozygotes are always less reproductively fit than the homozygotes (Brown & O'Neill 2010).

In contrast to the hybrid dysfunction model, the suppressed recombination speciation model (as shown in Figure 1-34) concentrates less on reduced reproductive fitness but instead is focused on reduced recombination levels which in turn lead to a partial reproductive barrier (Navarro & Barton 2003a; Rieseberg 2001; Ayala & Coluzzi 2005). The theory proposes that reduced recombination leads to an accumulation of alleles that contribute to reproductive isolation, adaptive differentiation and the build-up of gene divergence which lead to hybrid incompatibilities. These differences act as a 'genetic filter' between populations allowing for mutations to accumulate where they confer a local advantage. Statistically significant results that support this model have, however, proven difficult to verify (Strasburg et al. 2009; Noor & Bennett 2009; Navarro & Barton 2003b; Zhang J. et al. 2004).



**Figure 1-34: Speciation theory models illustrated using the example of an inversion. When (a) and (b) hybridise the F1 hybrid is heterokaryotypic. (c) illustrates the traditional speciation theory (hybrid dysfunction model) leading to underdominance. (d) illustrates the suppressed recombination model which states that suppressed recombination in the heterokaryotype will lead to (e) increased nucleotide rate change or (f) the capture of alleles allowing ecotypic adaption in that region (Brown & O'Neill, 2010).**

### 1.5.6 Homoplasy and Hemioplasy

Homoplasy is the term used to describe phenotypic similarities that have arisen due to evolutionary convergence rather than being inherited from a shared common ancestor. Distinguishing whether a trait is homoplastic in origin or due to homology is a key challenge in phylogenetic reconstruction (Avise & Robinson 2008). Homology itself can be defined as being

orthologous in origin i.e. genes originating from a single ancestral gene in the common ancestor of the genomes being compared or paralogous in origin i.e. originating from a gene duplication event (Koonin 2005). The advent of zoo-FISH has facilitated the reconstruction of phylogenies and ancestral karyotypes but some of the data produced using this approach has challenged previously established phylogenetic patterns (Robinson et al. 2008). In order to address these inconsistencies, Avise and Robinson defined the term 'hemiplasy' to describe situations where traits appear to be homoplastic but are in fact homologous in origin and are caused by discrepancies in gene tree-species tree analysis (Avise & Robinson 2008). Precise mapping of EBRs is an important step towards determining whether a trait is homoplastic or hemiplastic in origin (Skinner & Griffin 2011).

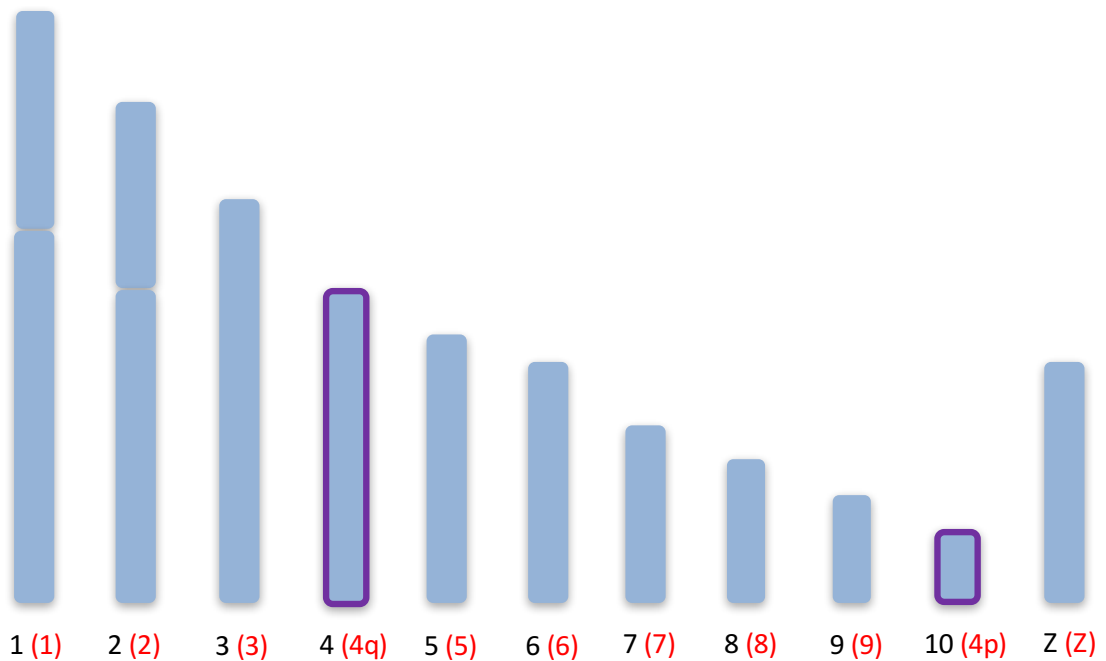
As mentioned in section 1.1.2.3 chromosome painting studies (zoo-FISH) have identified multiple avian species that exhibit a fusion of ancestral avian chromosomes 4 and 10 (listed in Table 1-4) and hence show examples of possible homoplasy.

Order	Common Name	Species Name	Author
Anseriformes	Greylag goose	<i>Anser anser</i>	Guttenbach et al. 2003
Galliformes	Chinese bamboo-partridge	<i>Bambusicola thoracica</i>	Shibusawa et al. 2004
Galliformes	Chinese quail	<i>Coturnix chinensis</i>	Shibusawa et al. 2004
Galliformes	Common peafowl	<i>Pavo cristatus</i>	Shibusawa et al. 2004
Galliformes	Chicken	<i>Gallus gallus</i>	Matsubara et al. 2004
Galliformes	Japanese quail	<i>Coturnix japonica</i>	Guttenbach et al. 2003 Shibusawa et al. 2004
Columbiformes	African collared dove	<i>Streptopelia roseogrisea</i>	Guttenbach et al. 2003

**Table 1-4: Avian species that exhibit a fusion on chromosome 4 as revealed by chromosome painting with chicken macrochromosomal paints.**

In the majority of other species analysed, including representatives of the *Anseriformes*, *Casuariiformes*, *Cathartiformes*, *Galliformes*, *Passeriformes*, *Psittaciformes*, *Rheiformes*, *Struthioinformes* and *Tinamiformes* order, the two chromosomes appear independently as illustrated in Figure 1-35. This example is of particular interest because the ancestral chromosome 4 (chicken chromosome 4 q-arm) is highly conserved with humans meaning that it must be present in the shared ancestor suggesting 310 million years of genome conservation (Chowdhary et al. 1998). In addition, the p-arm of chromosome 4 is orthologous to an ancestral

microchromosome, with FISH evidence of interstitial telomere signals next to the centromere. It also seems that this ancestral region has not lost the characteristically microchromosomal properties of high gene density and recombination rate (ICGSC 2004). The repeated pattern of this rearrangement across multiple species may of course be an example of homoplasy and therefore the result of numerous independent fusions or an interesting example of hemiplasy.



**Figure 1-35: Ancestral avian macrochromosome karyotype with chicken orthologues highlighted in red text (adapted from Griffin et al. 2007).**

Before considering the role of chromosomes in the evolution of birds, reptiles (and indeed dinosaurs) it is perhaps prudent to conclude this introduction with a consideration of the evolutionary events and phenotypic features of the species being studied.

## **1.6 Evolution and Genomics: Amniotes, Birds and Dinosaurs**

Any study of karyotypic (gross genomic) evolution needs to be considered in the context of the evolutionary events that occurred at the time and the phenotypic features of the species of interest.

### **1.6.1 Amniote Evolution**

Amniota, the clade that includes mammals, birds and non-avian reptiles is a remarkably diverse group of tetrapod vertebrates believed to have shared a common ancestor 325 mya during the Permian period (Shedlock & Edwards 2009). The amniotes diverged into the mammalian lineage (synapsids), which includes the monotremes and marsupials and the reptile lineage (diapsids). Reptilia is a clade that includes birds and all non-avian reptiles (reptiles other than birds and henceforth referred to as 'reptiles' for simplicity) with around 17,500 extant members (~10,500 of which are birds and ~7,000 reptiles). The crown-group of Diapsida is known as the Sauria and is subdivided into two groups (as illustrated in Figure 1-36). Lizards, snakes and tuataras form a monophyletic group (lepidosauromorpha) that is the sister group to either archosauromorpha (crocodilians, dinosaurs, pterosaurs) or archosauromorpha and testudinata (the relationship of turtles within amniota remains controversial) all of which shared a common ancestor 275 mya (Hedges et al. 2015). In terms of relationships between individual reptile groups, there is still some discussion, however, most authors agree upon the broad pattern of kinship within Sauria, which includes living birds, crocodiles, lizards, snakes, turtles, tuatara and all of the other extinct descendants from their common ancestor (Shedlock & Edwards 2009). Numbers appear to have remained relatively low in both the lepidosaurs and the archosauromorphs until the Permo-Triassic mass extinction event (PTME) devastated the synapsids around 251 mya (Benton & Twitchett 2003). Massive volcanic eruptions in the Siberian Trapps are thought to have initiated the conditions that created the PTME. These eruptions led to a prolonged period of global warming, creating anoxic conditions that devastated 80-90% of life on land and in the oceans (Benton & Twitchett 2003). The subsequent period of climate change led to increasingly arid conditions that marked out the beginning of the Triassic as a period of extraordinary ecological change. Estimates indicate that it took 10-15 million years before ocean reefs, forests and vertebrates were re-established after the PTME (Benton et al. 2013).

The stem-groups of lepidosauria and archosauria also include several extinct lineages that existed in the Triassic period including the rynchosaurs (Ezcurra et al. 2014). Of the lepidosaurs, the tuatara diverged 272 mya making them an extraordinarily ancient species and the only



extant example of its order, the rhynchocophelia (Rauhut et al. 2012). Assuming that the majority of recent molecular phylogenies of amniote interrelationships are correct, turtles (testudines) diverged from archosauromorphs first, with the origin of crown archosaurs (crocodiles, birds and their extinct relatives). Archosauria exhibits a basal split into the crocodile-line (Pseudosuchia or Crurotarsi) and bird-line (Ornithodira or Avemetatarsalia) clades. The dinosaurs, defined as the clade including *Triceratops*, *Passer* and all of the descendants of their common ancestor, are nested within Ornithodira and they include birds, which are nested within the theropod clade Maniraptora.

Reported divergence times of these major clades vary between different studies. Nonetheless the earliest dinosaur appeared about 235 mya (in the early Late Triassic period) and their divergence from non-dinosaur dinosauromorphs, by definition, was slightly earlier (in the Middle Triassic, approximately 240 mya). Pterosauria is the sister-group to Dinosauria and therefore diverged earlier, perhaps 245 mya, although an earlier origin is also possible. The ornithodiran/crurotarsan divergence occurred in the Lower Triassic period, or potentially earlier around the time of the Permo-Triassic boundary 252 mya although other studies date the crocodylian split from birds to 219 mya (Shedlock & Edwards 2009).

Prolonged debate has surrounded the phylogenetic relationship of turtles (testudines). The lack of temporal fenestrae in the skull led for many years to their assumed placement as a primitive anapsid, however molecular evidence now places the turtles as a sister group to the archosaurs (Chiari et al. 2012; Crawford et al. 2012; Shaffer et al. 2013). Again, debate surrounds the date the testudine divergence from the archosaurs as the earliest recorded testudine fossil (*Odontochelys semitestacea*) is late triassic in age (237-223 mya) (Benton et al. 2015; Nicholson et al. 2015) although stem turtle species (*Eunotosaurus africanus*) have been dated even further back to 260mya (Lyson et al. 2010). This therefore implies that the testudine archosauromorph divergence occurred around 260 mya during the Permian period. Specific divergence dates aside, the order within which each of these groups diverged from each other is widely accepted to be as represented in Figure 1-36.

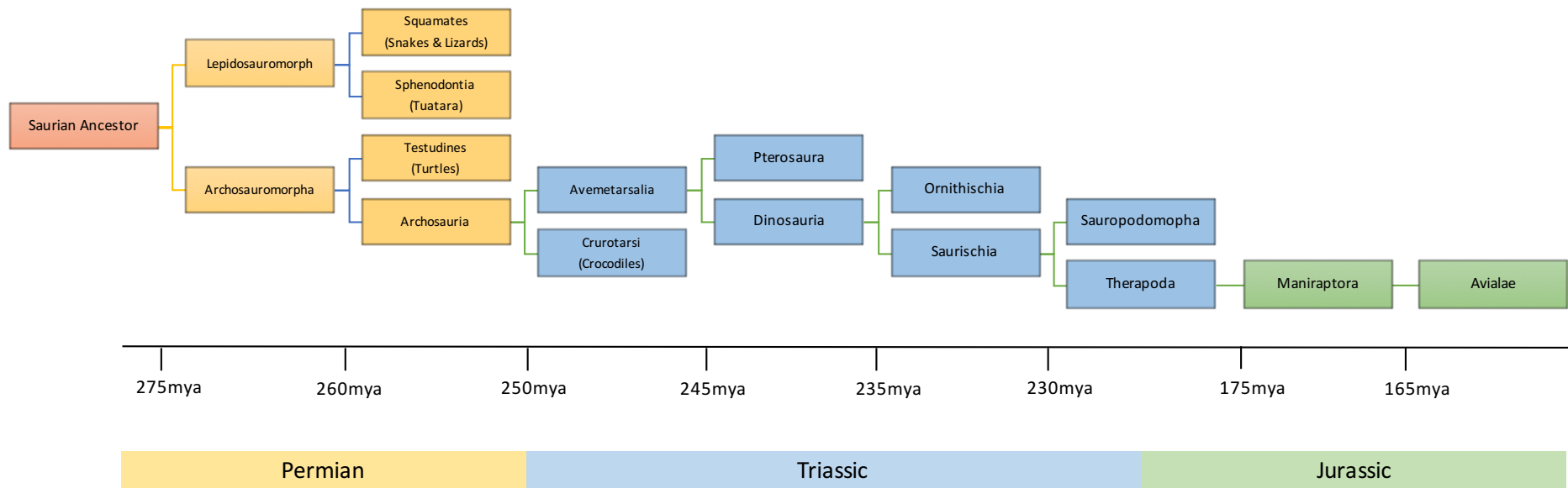


Figure 1-36: A time tree of amniotes (adapted from Shedlock and Edwards, 2009; Brusatte et al. 2011 and Benton et al. 2014).

### **1.6.2 Evolution of the Dinosaurs**

The group of animals most commonly known as dinosaurs typically refers to a group of small, bipedal archosaurs with long hind limbs and extended metatarsals that sit within the avemetatarsalian clade (Benton et al. 2013). Previous research dated the oldest unequivocal dinosaur fossils to 230 mya (Martinez et al. 2011), however recent fossil finds now indicate that the earliest dinosaurs may have appeared further back, around 240-245 mya (Nesbitt et al. 2013). In terms of species diversity and abundance, dinosaurs were still relatively low in number over the first 30 million years of their evolution, but by the mid Jurassic began to increase vastly in abundance, geographical spread and body size (Benton et al. 2014). The following 135 million years is remarkable for being a period not only for when dinosaurs were the dominant vertebrates but also for being a time when the dinosaurs displayed an extraordinary range of species diversity before finally being decimated by the Cretaceous-paleogene (K-Pg) extinction event 66 mya. Throughout this period, the dinosaurs survived further extinction events. These included the Carnian-Norian extinction event (CNEE) 228 mya that saw the end of the rhynchosaurs and dicynodonts (Brusatte et al. 2008) and the End-Triassic mass extinction event (ETME) 201 mya that devastated another group of archosauromorphs - the crurotarsans (the archosaurian crocodile lineage). The period after the ETME corresponded with a steady increase in dinosaur diversity and abundance arguing against the widely held belief that the release of an ecological niche by the extinction of their competitors led to a surge in dinosaur disparity (Brusatte et al. 2008). There are now over 1000 known species of dinosaur that appear in the fossil record with around 30 more being identified each year (Weishampel 2004), particularly in regions rich in newly discovered fossils such as China.

### **1.6.3 Speciation and Radiation of the Dinosaurs**

The extraordinary species diversity and abundance seen in the dinosaurs is often attributed to the eradication of competitor species that allowed the dinosaurs to flourish. However, it has also been suggested that these high levels of diversity and abundance reflect adaptations unique to dinosaurs that enabled them to survive through such harsh conditions while other species perished. For example, the extraordinary growth rates evidenced by bone growth patterns along with highly adapted respiration systems such as pneumatized bones (Farmer & Sanders 2010) and unidirectional respiration are both considered to be key features that enabled the dinosaurs to flourish (O'Connor P.M. & Claessens 2005). Interestingly, these very adaptations that may

have led to the success of the dinosaurs are also key features that contribute to the success of the descendants of dinosaurs – the birds. There is little doubt that modern birds are the descendants of dinosaurs with fossil evidence showing that both groups shared features such as feathers, oviparity, brooding behaviours and skeletal similarities (Varricchio et al. 2008), the question then arises as to whether dinosaurs also exhibited distinctively avian features at a genetic level.

#### 1.6.4 Avian Evolution from the Dinosaurs

Originating around 150 mya in the late Jurassic, birds (the living descendants of dinosaurs) evolved from a theropod lineage (Chiappe & Dyke 2006) at a time when the supercontinent Pangaea was separating into two landforms – Laurasia and Gondwana. The fossil *Archaeopteryx lithographica* (Figure 1-37) dating back to 150 mya and found in the 19th century in late Jurassic limestone in Germany (Meyer 1861) provides evidence of a transitional species between the dinosaurs and modern birds. Although previously considered to be the fossil representative of an early modern bird, features such as a bony tail and teeth rule *A. lithographica* out from being considered a true avian ancestor (Mayr et al. 2007).



**Figure 1-37: *Archaeopteryx lithographica* fossil showing clear evidence of feathered wings.**

As the oldest unambiguous fossil representative of Neornithes (modern birds) *Vegavis* is an aquatic bird classified within *Anseriformes* and most closely related to Anatidae – ducks, geese

and swans (Clarke et al. 2005). Dating back to ~67 mya, the discovery of this fossil supports the notion that representatives of modern birds were co-extant with non-avian dinosaurs prior to the Cretaceous-Paleogene (K-Pg) boundary 66mya (Clarke et al. 2005). The inherent difficulties in fossil dating due to geographic and depositional sampling bias has led to much controversy in the field of paleontology (Chiappe & Dyke, 2006) meaning that analyses at a genomic level are a useful complement to a fossil record that may imperfectly represent actual avian ancestors. Interestingly, the dinosaur ancestor of birds is generally considered to be bipedal, terrestrial, relatively small (small size being a pre-adaptation to flight) with limited flying ability, not dissimilar to the *Galliformes* (Witmer, 2002).

Until the publication in 2014 of a revised avian phylogeny by Jarvis and colleagues the timing of avian diversification has been a subject of much debate (Jarvis et al. 2014). The first avian divergence is now considered to have taken place around 100 mya when the Paleognathae (Ratites and Tinamous) diverged from the Neognathae (Galloanseres and Neoaves). Within the Paleognathae the Ratites and Tinamous then diverged 84 mya, while the Neognathae diverged into its stem lineages, the Galloanseres and Neoaves, 88 mya. The Galloansere divergence into the *Galliformes* (landfowl) and *Anseriformes* (waterfowl) occurred around the time of the K-Pg (see Figure 1-38). The major divergences of the Neoaves into Columbea and Passarea are now dated to before the K-Pg boundary (67-69 mya). The rest of the divergences within neoaves were largely complete at the ordinal level by 50 mya with the *Passeriformes* basal split estimated to be approximately 39 mya (Jarvis et al. 2014). The K-Pg event was another period of abrupt, mass global extinction and extreme climate change coinciding with the Chicxulub asteroid impact in Mexico (Schulte et al. 2010). It was a significant event for archaic birds (Ornithurae), of which the Neornithes are descendants. Recent fossil evidence points to a major radiation of advanced ornithurines occurring prior to the end of the Cretaceous period. The same group then suffered an abrupt extinction around the K-Pg event with their disappearance from the fossil evidence from the Paleogene period onwards (Longrich et al. 2011). Data from the Jarvis et al. analysis also suggests that the K-Pg transition period was one of rapid Neornithine speciation with 36 lineages radiating over a period of 10-15 million years. These revised dates challenge some of the previously held assumptions that Neornithine lineages diversified explosively after the K-Pg boundary rather than before as proposed by Jarvis and colleagues (2014).

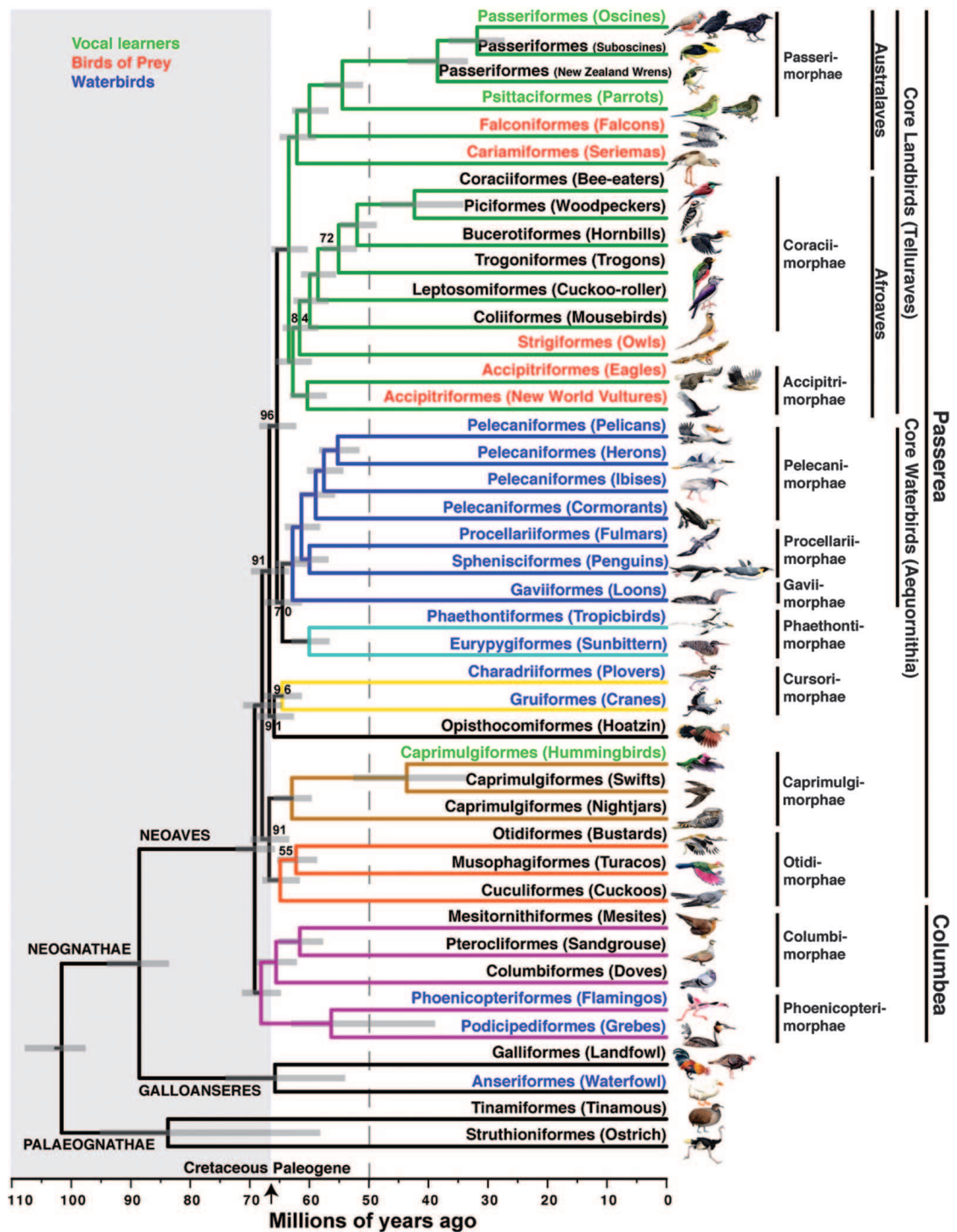


Figure 1-38: The revised avian phylogenetic tree illustrating the phylogenetic placement of avian taxa and radiation time points before and after the K-Pg extinction event 66 mya (Jarvis et al. 2014).

With around 10,500 extant species, birds are the most species rich of tetrapod vertebrates (Gill 2016). Modern birds belong to the class Aves and the subclass Neornithes. They are characterised by a combination of features not seen together in other vertebrates, such as homeothermy, flight (with the exception of penguins and ratites), oviparity, nesting, the

presence of a beak (and lack of teeth), a high metabolic rate, feathers and a lightweight skeleton. Occupying almost all terrestrial habitats, birds have adapted to a range of climate extremes, from inland Antarctica to the tropics, with the highest levels of species diversity seen in tropical regions (Weir and Schluter 2007).

The phenotypic diversity seen in birds is extraordinary, with sizes ranging from the bee hummingbird (*Mellisuga helenae*) at approximately 5cm in length to the ostrich (*Struthio camelus*), which stands over 2 metres tall. Birds have a high core body temperature (39-41°C), high blood glucose levels and energy expenditure levels that are five or more times higher than commonly seen in mammals. Comparison with similar sized mammals show the birds in fact tend to live longer despite the higher energy use (Holmes & Ottinger 2003). Birds are social, with varying degrees of communication complexity including the use of calls and song, and in some cases communicating with visual display. Birds can also be socially cooperative, exhibiting behaviours such as flocking and mobbing. Most birds also provide an extended period of parental care that is often shared between parents and/or with other birds.

Birds are critical to agriculture (both meat and eggs) and are also a model organism for studies of virology, immunology and developmental biology as well as being valuable companion animals for humans. From an evolutionary point of view, the Aves class is also a direct descendant of the dinosaur, sharing a common ancestor with mammals around 310 mya, making this class crucial to our understanding of evolution. In addition, approximately 1,200 extant birds are currently listed as endangered, with over 140 species having become extinct since the 16th century. Many of these extinctions are considered to be a result of anthropogenic climate and habitat change, in particular due to the introduction of alien species such as rats into island habitats (BirdLife International 2014). Further understanding of birds from an evolutionary point of view is therefore crucial as a means to understand vertebrate evolution and to protect current species from further risk.

### **1.6.5 Chromosomal Studies in the Context of Evolution**

From a chromosomal perspective therefore, birds (and the dinosaurs from which they evolved) remain remarkably understudied. Such studies act as an independent record of the actual substance of inheritance of living birds, genomic characters complementing a fossil record that may imperfectly represent actual neornithine predecessors (Romanov et al. 2014). As such, karyotype evolution studies may provide information on the ecological adaptations of avian



ancestors that the fossil record may never be able to establish unambiguously (Gatesy 2002). Several studies have endeavored to reconstruct ancestral karyotypes including a study from our lab by Romanov and colleagues (2014) to assemble the putative ancestral avian karyotype and a study by Uno et al. (2012) that attempted to establish the ancestral amniote karyotype. The Uno study found through gene mapping that the chicken and Chinese soft-shelled turtle (*Pelodiscus sinensis*) chromosomes are in fact true counterparts of each other suggesting that the characteristically avian karyotype has been conserved in this lineage for more than 250 million years since the lepidosaur divergence. The lepidosaurs conversely have exhibited a much larger degree of rearrangement as illustrated in Figure 1-39 (Uno et al. 2012).

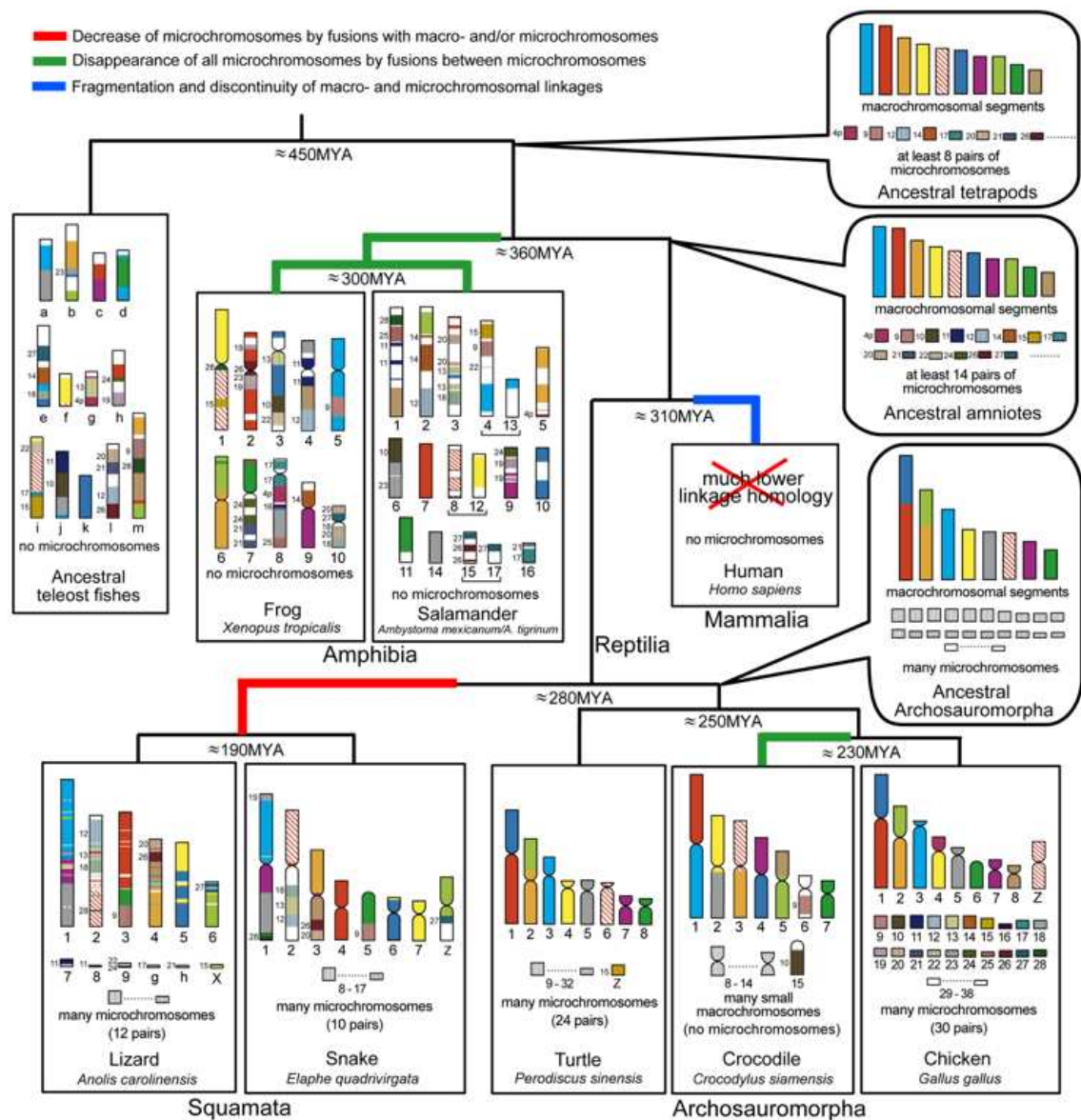


Figure 1-39: Schematic illustration of karyotype evolution among amniotes (Uno et al. 2012).



Taking all the above into account, it is clear that a number of areas remain under-explored in the field of genome-led chromosomal studies in non-human amniotes. As mentioned in section 1.1.1 molecular cytogenetic technology is not widespread in the detection of chromosomal rearrangements in individual animals such as pigs and cattle and therefore molecular cytogenetic tools derived from studies of whole genome sequences have great potential to advance 'clinical' cytogenetics for these animals. Similar tools used for this purpose can be used to investigate the hitherto undiscovered nature of chromosomal evolution of the avian microchromosomes at a molecular cytogenetic level. As mentioned in section 1.1.5.1, the shortfalls of NGS or TGS for chromosomal level assembly of newly sequenced species requires the intervention of molecular cytogenetic tools in conjunction with bioinformatics in order to reconstruct the whole genome structure and address the mechanisms of gross genomic evolution. Finally, the path of whole genome evolution from the saurian ancestor, through the dinosaur lineage to the avian ancestor and into extant birds remains under-discovered at the time of writing. With the above in mind, the purpose of this thesis is to address the shortfalls in our knowledge, while providing tools and technologies for further study through the pursuit of the following specific aims.

## 1.7 Specific Aims of this Thesis

**Specific aim 1.** To isolate sub-telomeric sequences (see section 1.1.2.1) from pig and cattle genome assemblies and develop a multiprobe device (see section 1.1.2.5) for the screening of both overt and subtle chromosome rearrangements in these species.

**Specific aim 2.** To apply the technology developed in specific aim 1 for the solving of previously intractable karyotypes and test the hypothesis that microchromosomal rearrangement is rare in avian evolution

**Specific aim 3.** To use the technology developed in specific aims 1 and 2 in conjunction with a bioinformatics approach developed by colleagues at the Royal Veterinary College, London to complete the cytogenetic mapping of scaffold based genome assemblies to full chromosomal level in two key (but karyotypically dissimilar) avian species (peregrine falcon and pigeon)

**Specific aim 4.** To use bioinformatic tools to re-create the overall genome structure (karyotype) of both Saurian and Avian ancestors and to retrace the gross evolutionary changes that occurred along the dinosaur lineage. To perform gene ontology analysis of homologous synteny blocks and evolutionary breakpoint regions (EBRs) of chromosomes (see section 1.5.3) to test the hypothesis that there is an enrichment for genes that correspond to known phenotypic characteristics of the species in question.

## **2 Materials and methods**

### **2.1 Chromosome Preparation**

#### **2.1.1 Fibroblast Culturing**

Avian tissues (trachea or early stage embryos) were derived from local suppliers. Sampling was reviewed and approved by the Animal Welfare and Ethics Review Board (AWERB) at the University of Kent.

##### **2.1.1.1 Media Preparation**

Standard fibroblasts were cultured in Alpha MEM (Fisher), supplemented with 10% Fetal Bovine Serum (Gibco) and 1% Pen-Strep-L-Glutamine (Sigma). For cells that required further supplementation (e.g. budgerigars) the serum was increased to 20%. Completed media was stored at 4°C until needed and was made up in a class II hood.

##### **2.1.1.2 Primary Culture Preparation – Enzyme Digestion**

Trachea was the preferred tissue for establishing cell cultures and was obtained either through biopsies provided by local veterinarians from deceased birds or from birds that were euthanized on site. Tissue was transferred into a sterile petri dish containing around 1ml Hanks Balanced Salt Solution (HBSS) (Gibco) with 1% Pen-Strep Fungizone (Gibco). Tissue was cut into 0.5mm<sup>3</sup> squares using a scalpel with a rocking motion so as not to tear the tissue. Macerated tissue was transferred to a 15ml falcon tube before adding around 1ml Liberase (Roche) made to a 0.125mg/ml working solution and incubated at 37°C for 3 to 6 hours. When tissue appeared to be digested, 4-5ml complete media was added to the tube and the entire solution was transferred to a T25 culture flask before incubating at 40°C (32°C for reptile cells) and 5% CO<sub>2</sub>. All steps were carried out aseptically in a class II laminar flow hood.

##### **2.1.1.3 Primary Culture Preparation – Embryonic Tissue**

Fertilised eggs were opened using a scalpel in a class II laminar airflow hood under sterile conditions. The embryo was extracted from the egg and placed in a petri dish containing around 2ml Phosphate Buffered Saline (PBS). The embryo was decapitated and the body tissue was macerated using a cell dissociation sieve before being passed through a 70µm filter into a 50ml falcon tube. 9ml complete media was added to the tube before transferring the solution to a T75 cell culture flask and incubating at 40°C with 5% CO<sub>2</sub>.

#### **2.1.1.4 Refreshing**

Flasks were refreshed every other day by removing spent media and replacing it with fresh complete media (made up as described in section 2.1.1.1) to a total volume of 5ml in a T25 flask and 10ml in a T75 flask.

#### **2.1.1.5 Passaging**

Flasks were passaged when cells appeared confluent by aspirating the media from the flask then rinsing with 1ml HBSS (2ml for T75) and discarding. 1ml (2ml in T75) pre-warmed 0.05% Trypsin-EDTA (Gibco) solution was added to the flask to coat the cells and the flask was placed on a hotplate at 37°C for 1-2 minutes to facilitate enzymatic detachment of cells. Flasks were checked under the microscope to ensure they had rounded up and detached and the flask was tapped firmly on the side to dislodge any attached cells. 9.5ml of complete media was added to the flask to resuspend the cells and 10ml was transferred to a T75 flask and the T25 was re-fed with 4.5ml complete media. When passaging from a T75 flask, 6ml of the cell suspension was added to a fresh T75 and a further 4ml complete media was added to both T75 flasks.

### **2.1.2 Chromosome Harvesting**

Flasks were selected for harvesting when confluent, exhibiting abundant mitotic doublets and when the flask was not required for passaging. 50µl colcemid (Gibco) at a concentration of 10.0µg/ml was added to each T25 flask (100µl added to a T75) for 1 hour. Cultures were transferred to a 15ml falcon tube and the flask rinsed with 1ml HBSS, which was added to the falcon tube. 1ml Trypsin-EDTA (Gibco) was added to the flask for 2 minutes at 37°C to detach cells. 2ml HBSS was added to rinse the flask of any remaining cells and then transferred to the same falcon tube. Samples were centrifuged for 10 minutes at 1,000 rpm, the supernatant was then discarded and the cell pellet was resuspended. Hypotonic treatment to swell the cells was performed by adding 5ml pre-warmed 0.075M KCl drop-wise and incubating at 37°C for 20 minutes. Three drops of fixative (3:1 methanol:acetic acid) were added to the cells while gently agitating and the solution was centrifuged for a further 10 minutes at 1,000rpm. The supernatant was discarded and the pellet resuspended using a Pasteur pipette. The resuspended pellet was drawn up into the Pasteur pipette, 5ml fix was then added to the tube and the cell suspension was released gently into the fix. The tube was centrifuged for 10 minutes at 1,000 rpm and the fix process repeated a further 2 times and samples were stored at -20°C.

### **2.1.3 Blood Culture and DNA Extraction**

#### **2.1.3.1 Blood (Lymphocyte) Culturing – Mammals**

Mammalian blood was taken via standard phlebotomy using heparinized tubes. 500µl of whole blood was added to 9.5ml pre-warmed PB Max karyotyping medium in a T25 culture flask and incubated (flask upright) at 37°C with 5% CO<sub>2</sub> for 72 hours. Flasks were then tapped to resuspend the red cell layer before adding 100µl colcemid (10µg/ml concentration) to the flask and incubating for 30 minutes at 37°C. The solution was transferred to 15ml falcons and then centrifuged for 5 minutes at 1,900 rpm. The supernatant was discarded and the pellet resuspended before adding 6ml 0.075M drop-wise with agitation followed by incubation at 37°C for a total period of 12 minutes in order to swell the cells osmotically. Fixative was added slowly down the side of the tube to a final volume of 14ml before inverting the tube to mix. Tubes were then centrifuged at 1,900rpm for 5 minutes before discarding the supernatant and re-suspending the cell pellet. Fixative was added drop-wise with agitation to 5ml before further centrifugation at 1,900rpm for 5 minutes followed by removal of supernatant and resuspension of the pellet. This fixative wash was repeated a further 3-4 times and samples were stored at -20°C.

#### **2.1.3.2 Blood (Lymphocyte) Culturing – Birds**

Culture medium made up of 217.5ml RPMI 1640 medium (Sigma), 25ml Chicken serum (Invitrogen), 5ml Penicillin-Streptomycin (Invitrogen), 2.5ml L-Glutamine (Invitrogen) and 25mg Concanavalin A type IV (Sigma) was mixed in a class II hood. Whole blood was collected using standard heparinized vacutainers. 3ml of Histopaque (Sigma) was put in a 15ml falcon tube and brought to room temperature before adding 2-3ml of blood on top to form a layer. Tubes were centrifuged at 400g for 30 minutes to form layers. The plasma layer was then removed and discarded and the opaque interphase layer was transferred to a clean 15ml falcon tube. 10ml of sterile PBS was added and mixed gently. Tubes were centrifuged at 250g for 10 minutes and the supernatant discarded. The cell pellet was resuspended in 5ml of sterile PBS and mixed gently before centrifuging at 250g for 10 minutes. The supernatant was discarded and the cells resuspended before transferring to a T25 flask containing 10ml complete medium and culturing for 72 hours at 40°C and 5% CO<sub>2</sub>. 50µl colcemid (concentration 5µg/ml) was added to the culture for 1 hour at 40°C before transferring the solution to a 15 ml falcon tube. Cells were centrifuged at 400g for 5 minutes and the supernatant discarded. KCl (0.075M) was added drop-wise with agitation to 6ml over a 15-minute period with incubation at 37°C. Fixative was added slowly

down the side of the tube to a final volume of 14ml before inverting the tube to mix. Tubes were then centrifuged at 400g for 5 minutes before discarding the supernatant and resuspending the cell pellet. Fixative was added drop-wise to 5ml before further centrifugation at 400g for 5 minutes followed by removal of supernatant and resuspension of the pellet. This fixative wash was repeated a further 3-4 times and samples were stored at -20°C.

### **2.1.3.3 Extraction of Genomic DNA**

DNA was extracted using the Qiagen DNEasy Blood and Tissue kit from tissue obtained by skin biopsy from various birds. The manufacturer guidelines were followed with the exception of an additional final elution step to increase the DNA yield, where the eluted DNA was put back through the same spin column and the final centrifugation was repeated.

## **2.2 Generation of Labelled FISH Probes**

### **2.2.1 Selection of BAC clones**

#### **2.2.1.1 BAC Selection – Subtelomeric Porcine, Bovine and Avian BACs**

BAC clones of approximately 150kb in size were selected using the Sscrofa Version 10.2 NCBI database ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) for porcine BACs. BACs in the subtelomeric region of the p-arm and q-arm of each chromosome were identified and ordered from both the PigE-BAC library (ARK-Genomics) and the CHORI-242 Porcine BAC library (BACPAC). BAC clones in the same chromosomal regions were identified from the Btau 4.6.1 NCBI database for cattle BACs and ordered from the CHORI-240 Bovine BAC library. Avian Subtelomeric BAC clones were selected using the same principle but using the *Gallus\_gallus* Version 2.1 NCBI database for Chicken BACs and *Taeniopygia guttata* Version 3.2.4 NCBI database for Zebra Finch BACs ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). BACs in the subtelomeric region of the p-arm and q-arm of each chromosome were identified and ordered from both CHORI-261 Chicken BAC library (BACPAC) and the Zebra Finch TGMCBa library (Wageningen).

#### **2.2.1.2 BAC Selection – Selected Avian BACs**

BACS for cross-species FISH testing were designed in collaboration with members of the Larkin lab based at the Royal Veterinary College, London. The specific criteria used to increase the likelihood of the BACs working across a diverse range of species included the proportion of

conserved sequence between species, the number of repetitive elements and the GC content of the BAC. The pipeline for detection of appropriate BACs consisted of the following steps: a) generation of multi-species alignments; b) detection of conserved elements (CEs) in different reference genomes; c) repeat-masking reference genomes; d) mapping BAC clones to the reference genome sequence with BAC whole sequence or sequenced BAC ends; e) calculations performed to identify the following elements of each BAC: the fraction of repetitive sequences in the reference genome, average nucleotide conservation score, average GC content of the BAC and separately of the CEs within the BAC, length of CEs, number of exons; f) selection of a training set of BACs to be tested on different species metaphases to test the criteria; g) building a statistical model based on the results of the training set. BACs were ordered from the two avian libraries referred to above in section 2.2.1.1 and each was labelled in both FITC and Texas Red (as described in section 2.2.3.2).

## **2.2.2 Preparation of BAC Clones**

### **2.2.2.1 LB Agar Preparation**

16g of LB Agar (Invitrogen) was added to 500ml of ddH<sub>2</sub>O, autoclaved at 120°C and left to cool to 50°C. 300µl of the antibiotic chloramphenicol (25mg/ml) (Sigma) was added to give a final concentration of 15µg/ml. Approximately 10ml of cooled agar solution was then poured into sterile plastic Petri dishes and left to set overnight before refrigeration at 4°C

### **2.2.2.2 LB Broth Preparation**

10g LB Broth (Sigma) was mixed with 500ml ddH<sub>2</sub>O and autoclaved at 120°C and left to cool to 50°C before adding 300µl of chloramphenicol to give a final concentration of 15µg/ml.

### **2.2.2.3 Plating out of BACs**

A sterile disposable pipette tip was inserted into each agar stab containing the BAC clone and transferred to separate 50ml falcon tubes containing 20ml of prepared LB broth and left to culture overnight in the shaker at 37°C at 140rpm. From these tubes, a sterile disposable pipette was used to streak 10µl of each sample onto agar plates and left to culture overnight at 37°C. On the following day, two colonies were picked from the agar plate with a sterile disposable tip and transferred to 15ml falcon tubes containing 5ml of LB Broth/glycerol solution (7% glycerol).

The tube containing the pipette tip was then left to culture overnight at 37°C in a shaker at 140rpm. 2ml of each culture was taken and stored at -20°C.

#### **2.2.2.4 BAC DNA Isolation**

A sterile disposable pipette was inserted into the glycerol stock of the BAC clone and used to streak a further agar plate. The agar plates were incubated upside-down at 37°C overnight. On the following day, the plates were washed with PBS and the colonies scraped with a Pasteur pipette. The resulting solution was transferred to a centrifuge tube and centrifuged at 8,000rpm for 3 minutes in readiness for isolation using the Qiagen Miniprep kit.

### **2.2.3 Amplification and Labelling BACs**

#### **2.2.3.1 Probe DNA Amplification**

Each DNA sample was pulse centrifuged prior to being analysed for concentration and absorbance ratio on a spectrophotometer (NanoDrop, ThermoScientific). GenomiPhi V.2 (GE Healthcare) sample buffer and reaction buffer were thawed on ice alongside probe samples. All reagents were pulse centrifuged and mixed. 3µl of each probe DNA was transferred into labelled 0.5ml tubes with 27µl Sample Buffer before mixing and pulse centrifuging. Samples were heated at 95°C for 3 minutes in the PCR block to denature the DNA and then immediately placed on ice.

Enzyme/Reaction buffer mix was prepared with the enzyme volume calculated at a ratio of 3µl x the number of tubes x 1.2, and the reaction buffer volume calculated at 9 x the volume of enzyme. Both were mixed together in a fresh 1.5ml tube and held on ice before adding 30µl to the cooled probe DNA, mixing and pulse centrifuging. All samples were then incubated at 30°C for 1.5 hours in a dry incubator before being returned to the PCR block for 10 minutes at 65°C. 60µl of MBG H<sub>2</sub>O was added, mixed and transferred to fresh 1.5ml tubes, followed by 12µl of sodium acetate/EDTA buffer (50 ml of 3M Sodium acetate (pH 8.0) mixed with 50 ml of 0.5M EDTA (pH 8.0)). 300µl of 100% ethanol was added and mixed gently by inversion and centrifuged for 15 minutes at 13,000rpm. The supernatant was removed to leave a pellet before 500µl of 70% ethanol was added and the solution centrifuged again at 13,000rpm for 2 minutes. The supernatant was discarded and the pellet pulsed in the centrifuge and any remaining ethanol was removed with a small pipette before leaving residual ethanol to evaporate with the tube lid open for 2-3 minutes. 60µl of 10mM Tris-HCl buffer (pH 8.0) was added to the pellet to resuspend and left overnight at 4°C before remixing and measuring the DNA concentration with



the nanodrop. Each probe DNA sample was then diluted with 10mM Tris-HCl buffer (Sigma) to a volume of 166.5µg/µl using each individual concentration measurement (by multiplying the DNA concentration by sample volume, dividing by the dilution required (166.5 µg/µl) and subtracting the sample volume) to achieve a sample DNA concentration of 2µg/µl per sample.

### **2.2.3.2 Nick Translation**

Probe mixes were made by transferring 12µl diluted DNA to a fresh 1.5ml tube and adding 10µl NT buffer (Cytocell), 10µl 10xDTT (Dithiothreitol), 8µl NucMixA (Cytocell), 1.5µl Texas Red-12-dUTP (Invitrogen) for the q-arm probes or 1.5µl FITC-Fluorescein-12-UTP (Roche) for p-arm probes (in the case of subtelomeric probes) was added along with 4µl DNA Polymerase 1 (Promega), 2.5µl DNase 1 (Roche) (at a concentration of 9.4U/ml which is then diluted to 1µl DNase 1 and 999µl MBG H<sub>2</sub>O and mixed by inverting) and 51µl MBG H<sub>2</sub>O on ice to reach a final volume of 100µl per sample. Samples were mixed, pulsed and incubated for 2 hours at 15°C in a water bath, followed by heat inactivation for 10 minutes at 65°C in the water bath. All cross species BACs were labelled in both colours and a selection were labelled in aqua using PromoFluor-532-aadUTP (Promokine) in the same quantities that were used for FITC and Texas Red.

### **2.2.3.3 Agarose Gel Preparation**

A 1.4% agarose gel was made by dissolving 0.42g agarose (Biogene) in 30ml 1xTBE (Invitrogen) The solution was boiled for 1 minute before adding 1µl SYBR Safe (Invitrogen) and pouring into a gel case and leaving to set. Once set, the gel case was filled with 0.5x TBE buffer (Invitrogen). 4µl loading buffer was mixed with 4µl DNA probe sample and 7µl of each mixed sample was loaded into the agarose gel wells alongside a 2µl 100BP DNA ladder (Biogene). The gel was run at 90 volts for 23 minutes and images captured on a transilluminator to check for smears under 500bp.

### **2.2.3.4 Probe Purification**

Probes were cleaned with QIAquick Nucleotide Removal Kit (Qiagen), by adding 1ml Buffer PNI (including isopropanol) to each 100µl probe and mixing. The solution was transferred to the quickspin column and centrifuged at 6,300rpm for 1 minute and flow-through discarded. The column was washed with 750µl of PE Buffer (containing ethanol) and centrifuged for 1 minute at 6,300rpm and flow-through discarded before centrifugation for a further minute at

13,000rpm. The column was then moved to a new 1.5ml centrifuge tube and 100µl MBG H<sub>2</sub>O was added to elute the DNA and left to stand for 5 minutes before centrifugation at 13,000rpm for a further minute. Columns were discarded and the purified probe stored at 4°C for in readiness for FISH.

## **2.3 Fluorescence *in situ* Hybridisation (FISH)**

### **2.3.1 Metaphase Slide Preparation and Hybridisation**

Stored chromosome samples were centrifuged at 1,000 rpm for 10 minutes and supernatant discarded before adding more fixative (3:1 methanol:acetic acid) until the solution was semi-opaque. 10µl of chromosome suspension was placed on each half of the slide and left to dry after which the slides were washed in 2xSSC (saline-sodium citrate) for 2 minutes, followed by dehydration in an ethanol series for 2 minutes in each of 70%, 85% and 100% ethanol. Probes were diluted in a formamide buffer with porcine Hybloc for pig probes, chicken Hybloc for avian probes and bovine Hybloc for cattle probes (Applied Genetics Laboratories) to a total volume of 10µl, of which 1.5µl was the FITC labelled probe, 1µl was the Texas red labelled probe, 1µl was Hybloc and the remainder was Hyb I (Cytocell) hybridisation buffer giving a probe concentration of 10ng/µl.

Once the slide was dry, 10µl of probe mix was pipetted onto a 22x22mm coverslip on the 37°C hotplate before placing the coverslip onto the slide and sealing with rubber cement. The slides were then placed on a 75°C hotplate for 2 minutes to simultaneously denature the probes and target DNA, then moved to a hybridisation chamber and incubated at 37°C overnight (or for 72 hours if cross-species) to allow probe DNA to anneal to the target DNA. Slides were then removed from the hybridisation chamber, cover slips discarded and the slide washed in 0.4xSSC for 2 minutes at 72°C (for same-species FISH) to remove any unbound probe before being placed in 2xSSC with Tween (0.05%) at room temperature for 30 seconds and drained. Cross-species FISH slides were only washed in the second of these two washes - 2xSSC with Tween (0.05%) at room temperature for 30 seconds, to reduce the risk of removing any weakly bound probe. 10µl DAPI (Vectorlab) was dropped on each half of the slide and a 24x50mm coverslip placed on top.

### 2.3.2 Multiprobe Device Preparation

Fluorescently labelled probes were diluted to a concentration of 10ng/ $\mu$ l in sterile distilled water along with Hybloc competitor DNA. Each probe combination contained a probe isolated from the distal p-arm (labelled in FITC) and distal q-arm (labelled in Texas Red) from a single chromosome. Where the chromosome is acrocentric, the most proximal sequence was isolated. For simplicity sake, these were individually assigned with the chromosome number followed by the letter p in green type and the letter q in red type, as indicated in Figure 2-1 for the porcine multiprobe device and Figure 2-2 for the avian multiprobe device.

	1pq	2pq	3pq	4pq	5pq	6pq	7pq	8pq
	9pq	10pq	11pq	12pq	13pq	14pq	15pq	16pq
	17pq	18pq	Xpq					

**Figure 2-1: Schematic representation of probe on the porcine multiprobe device with each square containing a FITC labelled p-arm probe and a Texas Red labelled q-arm probe for each chromosome.**

Each probe combination (e.g. 1pq) was air dried on to a square of the device in the orientations illustrated in Figures 2-1 and 2-2. The second part of the device is a glass slide subdivided into 24 squares which are designed to align to the 24 squares on the first part of the device. 2 $\mu$ l of metaphase suspension was dropped onto each square of the slide which was then put through the ethanol series as described in section 2.3.1. The individual probe combinations were then re-hydrated in 1 $\mu$ l of formamide based hybridisation buffer (Hyb I, Cytocell) by pipetting onto each square of the first part of the device containing the probe before aligning the squares of the device and the slide. Probe and target DNA were subsequently denatured on a 75°C hotplate for 5 minutes prior to hybridisation overnight in a dry hybridisation chamber floating in a 37°C water bath in order for probe DNA to anneal to the target DNA. Second day washes were performed as described above in section 2.3.1.

	10pq	11pq	12pq	13pq	14pq	15pq	16pq	17pq
	1,4,3	18pq	2,5,8	19pq	6,7,9	20pq	Z,W	21pq
	22pq	23pq	24pq	25pq	26pq	27pq	28pq	

**Figure 2-2:** Schematic representation of probe layout for the avian multiprobe device with each square containing a FITC labelled p-arm probe and a Texas Red labelled q-arm probe for each of the microchromosomes with the shaded squares in the middle row containing macrochromosome paints labelled in multiple colours as represented by the font colour in the diagram.

### 2.3.3 Microscopy

An Olympus BX-61 epifluorescence microscope equipped with a cooled CCD camera with DAPI, Aqua, Gold and Texas Red filters was used to visualise the chromosomes, which were then captured using SmartCapture 3 software (Digital Scientific UK).

### 2.3.4 Image Analysis - Karyotyping

SmartType software (Digital Scientific UK) was used for karyotyping purposes after being custom-adapted for porcine karyotyping according to the standard karyotype as established by the Committee for the Standardised Karyotype of the Domestic Pig (Gustavsson 1990) for porcine samples and according to the ISCNDB (Cribru et al. 2001) for bovine samples.

## 2.4 Physical Genome Mapping

The approaches listed below represent the stages that the Larkin lab at the Royal Veterinary College, London (RVC) performed in order to produce predicted chromosome fragments (PCFs) which were then mapped in our lab by FISH. Our method jointly devised at the University of Kent and RVC London involves: (1) the construction of predicted chromosome fragments (PCFs) from scaffold-based assemblies based on alignment of raw sequencing read data using the RACA algorithm (RVC); (2) PCR and computational verification of these PCFs (RVC); (3) development of a refined set of PCFs based on the previous verification set; (4) the use of a 'universal set' of BACs spread uniformly across the genome and designed to hybridise efficiently in

phylogenetically divergent species to anchor PCFs to chromosomes using fluorescence *in situ* hybridisation (zoo-FISH). I would like to express my thanks to Joana Damas, Dr Marta Farré and Dr Denis Larkin for performing this part of the project. For clarity, the steps performed at RVC are outlined below.

#### **2.4.1 Construction of PCFs using the RACA Algorithm (RVC)**

RACA (Kim et al. 2013) assemblies were generated for peregrine falcon (*Falco peregrinus*) and pigeon (*Columba livia*) genomes from the fragmented Illumina assemblies previously published (Shapiro et al. 2013; Zhan et al. 2013). For the peregrine falcon, the zebra finch chromosome level genome assembly was used as a closely related reference (divergence 62 mya) and the chicken genome assembly as the outgroup (divergence 96 mya). Colleagues at the RVC generated 113 PCFs with an N50 of 27.44 Mb using default RACA parameters. For the pigeon with a large phylogenetic distance from both zebra finch and chicken (>70 mya), the chicken was used as reference and zebra finch as the outgroup because: a) fewer pigeon scaffolds were split by RACA in this configuration and b) there is a higher degree of similarity between pigeon and chicken karyotypes (Derjusheva et al. 2004). This resulted in 150 pigeon PCFs with an N50 of 34.54 Mb. These initial RACA assemblies contained 72 (15.06%) and 78 (13.64%) scaffolds, in peregrine falcon and pigeon PCFs respectively, that were split by RACA due to insufficient read or comparative evidence to support their structures.

#### **2.4.2 Verification of PCFs (RVC)**

As indicated above, using default parameters RACA produced splits in 13-15% of the target genome scaffolds. To verify the structures of these scaffolds RVC colleagues attempted polymerase chain reaction (PCR) amplification across the split regions less than 6kb in the target species scaffolds. Of these, 41 (83.67%) and 58 (84.06%) resulted in positive PCR results with products of expected length in pigeon and peregrine falcon genomic DNA, respectively. For the split regions with negative PCR results they tested an alternative (RACA-suggested) order of the flanking syntenic fragments (SFs). Of these split regions, positive PCR results were obtained for two (50%) in peregrine and seven (100%) in pigeon, confirming the chimeric nature of the original scaffolds indicated by RACA. To estimate which of the remaining split regions greater than 6kb in size (36 in peregrine and 40 in pigeon PCFs) were likely to be chimeric they used a minimum physical read coverage across the SF joining regions for which PCR results were consistent with RACA predictions. In total, the number of scaffolds containing real structural

differences with the reference chromosomes that would be split by RACA was established to be 56.25% in the peregrine falcon and 42.86% in pigeon PCFs.

### **2.4.3 Creation of a Refined Set of Pigeon and Peregrine PCFs (RVC)**

A refined set of PCFs were created using adjusted coverage thresholds and in addition, those scaffolds confirmed by PCR were kept intact, but those that were shown to be chimeric and/or disagreeing with the cytogenetic map were split (see below) resulting in a total of 96 PCFs with N50 25.82 Mb for the peregrine falcon and 137 PCFs with N50 of 22.17 Mb for the pigeon, covering 97.17% and 95.86% of the original scaffold assemblies, respectively. The peregrine falcon RACA assembly contained six PCFs homologous to complete zebra finch chromosomes (TGU 4A, 9, 11, 14, 17 and 19) while five pigeon PCFs were homologous to complete chicken chromosomes (GGA 11, 13, 17, 22 and 25). Only 3.5% of the original scaffolds used by RACA were reported as chimeric in pigeon and 3.77% in the peregrine falcon final PCFs. The total accuracy for the RACA assembly was estimated as 84-86% for the peregrine falcon and 86-92% for the pigeon based on the ratio of the number of SFs to the number of scaffolds.

## **2.5 Ancestral Karyotype Reconstruction**

### **2.5.1 Avian Ancestor**

The following amniote genomes were selected in order to reconstruct the hypothetical avian ancestor: chicken (*Gallus gallus*), duck (*Anas platyrhynchos*), zebra finch (*Taeniopygia guttata*), ostrich (*Struthio camelus*), budgerigar (*Melopsittacus undulatus*) and turkey (*Meleagris gallopavo*) along with the anole lizard (*Anolis carolinensis*) included in the analysis as the outgroup. The seven species-specific msHSB sets served as MGRA inputs for individual genomes which then produced a series of contiguous ancestral regions (CARs) representing the most likely ancestral configuration for the avian ancestors.

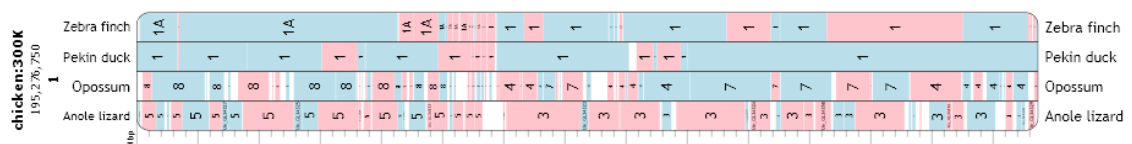
### **2.5.2 Saurian Ancestor**

The following amniote genomes assembled at the chromosomal level were selected in order to reconstruct the hypothetical saurian ancestor: chicken (*Gallus gallus*), duck (*Anas platyrhynchos*) and zebra finch (*Taeniopygia guttata*), along with green anole lizard (*Anolis carolinensis*) a reptilian clade representative, and opossum (*Monodelphis domestica*) as a basal mammal outgroup representative. In addition, the above three avian genomes were used to reconstruct

the avian ancestor for this data set, in this case using the lizard genome as the outgroup.

### 2.5.3 Multiple Genomes Alignment and Identification of HSBs and EBRs

The whole genome sequences of the species of interest were aligned and visualised using the interactive genome browser 'Evolution Highway' (Murphy et al. 2005; Farré et al. 2016; <http://evolutionhighway.ncsa.uiuc.edu>). Pairwise blocks of synteny were identified and displayed relative to chromosomes of the chicken, which served as the reference genome (ICGSC *Gallus gallus* 4.0). The visualised genome alignments as inferred from orthology maps for the five species were analysed side by side and by chicken chromosome. The start and end coordinates of the aligned orthologous regions observed in all the species compared were used to define multiple-species homologous synteny blocks (msHSBs; Figure 2-3) at the 300-Kb resolution. These msHSBs were assigned to and subsequently sorted in individual chromosomes in each species according to their location, orientation and sequential order. Coordinates of the msHSBs identified for the species being analysed were used to define evolutionary breakpoints (EBRs). Each EBR represented a chromosome segment between two adjacent msHSBs



**Figure 2-3: Example of Evolution Highway alignment output that produced 96 msHSBs in four (zebra finch, duck, opossum and anole Lizard) genomes as aligned against GGA1 totalling 137.47 Mb out of 195.28 Mb (or ~70%).**

### 2.5.4 Ancestral Saurian and Avian Karyotype Reconstruction

To reconstruct a putative avian ancestor, the Multiple Genomes Rearrangements and Ancestors tool version 2 (MGRA2) (Avdeyev et al. 2016; Alekseyev & Pevzner 2009) was used as follows: pairwise alignments for duck, zebra finch and anole lizard were visualised relative to the chicken as referred to above. The orthology map of the opossum was used as an input for the MGRA program and included in the analysis as the outgroup. The five species-specific msHSB sets served as MGRA2 inputs for individual genomes which then produced a series of contiguous ancestral regions (CARs) representing the most likely ancestral configuration for the species identified in both hypothetical saurian and avian ancestors.

## **2.5.5 Genome Rearrangement Analysis**

To reconstruct the chromosomal changes that occurred between the two sets of ancestral groups, two approaches were used. The first was a manual approach to identify the most parsimonious series of events that occurred from the ancestor to the extant species. The second approach required the use of the Multiple Genome Rearrangements (MGR) and Genome Rearrangements In Man and Mouse (GRIMM) tools ((Bourque & Pevzner 2002); <http://grimm.ucsd.edu/>). MGRA2 outputs served as MGR/GRIMM inputs to trace the most parsimonious scenarios for evolutionary changes in two scenarios: firstly, the intra- and interchromosomal rearrangements that might have occurred from the hypothetical saurian ancestor to the avian one and secondly, those rearrangements that may have occurred between the avian ancestor and the extant species.

## **2.5.6 Gene Ontology Analysis**

### **2.5.6.1 Microchromosome Analysis (section 6.4.2)**

Ensembl gene ID data and gene name for each microchromosome was extracted from the BioMart Ensembl Genes 75 Database (Kinsella et al. 2011) using galGal4 as the dataset. In order to eliminate any significant results arising through the presence of multiple copies of genes in the same family being present on the same chromosome, gene families were reduced to a single representative member. Downloaded gene IDs and gene names were uploaded into DAVID (Dennis et al. 2003) using Ensembl Gene ID as the list identifier and subsequently analysed using the Functional Annotation Clustering tool. Cluster data from each microchromosome gene list output was downloaded into Microsoft Excel and filtered using an enrichment score of 1.3 and above and a P value less than 0.05 to edit the list for clusters considered to be significant. BioMart (Ensembl) derived gene names for each microchromosome were also uploaded into GOEAST (Zheng and Wang, 2008) using *Gallus gallus* as the reference. Batch-gene analysis was performed by GOEAST, and enriched GO term outputs were filtered for those with a P value less than 0.05 considered to be significant. The GO results obtained from GOEAST were downloaded into Microsoft Excel and presented with graphic files created directly from GOEAST for each microchromosome where results were available. Finally, in order to correct for multiple sampling error, an FDR threshold of 0.05 was used.



#### **2.5.6.2 *Dinosaur HSB and EBR Analysis (section 6.4.5)***

Gene lists for msHSBs and EBRs were extracted from the Ensembl BioMart data system (Kinsella et al. 2011) using Galgal4 as the dataset. Since human genes are best annotated, the gene lists derived from chicken were matched to orthologous human genes and filtered for homology type and orthology confidence, leaving only those genes that were one-to-one orthologues and had the maximum orthology confidence.

Background gene lists were also generated using the gene lists of all chicken–human orthologues. The first background gene list tested covered all chicken chromosomes and the second list only included results for 19 of the chicken chromosomes where the msHSBs and EBRs were found. In addition, in order to test whether genes with low gene identity matches affected the GO analysis, thresholds of 70, 60 and 50% homology for the orthologue gene lists were set and the resulting GO outputs compared. Based on these tests, the 70% gene identity threshold was selected for generating the msHSBs/EBRs gene lists, and the 19-chromosome list was used for the background GO analysis list.

Gene lists were used as inputs for the web-based functional annotation tool DAVID (Dennis et al. 2003) using Ensembl Gene ID as the list identifier and subsequently analysed using the Functional Annotation Clustering tool. Cluster data from each gene list output was downloaded into Microsoft Excel and filtered using an enrichment score of 1.3 and above and a P value less than 0.05 to edit the list for clusters considered to be significant. Finally, in order to correct for multiple sampling error, an FDR threshold of 0.05 was used

### **3 Specific aim 1: To isolate sub-telomeric sequences from pig and cattle genome assemblies and develop a multiprobe device for the screening of both overt and subtle chromosome rearrangements in these species.**

#### **3.1 Background**

The domestic pig (*Sus scrofa*) is a eutherian mammal and member of the artiodactyl order. It shared a common ancestor with humans around 79 to 87 mya (Kumar and Hedges, 1998) and has been domesticated since around 7000 BC (Giuffra et al. 2000). It provides 43% of meat consumed worldwide making it the leading source of meat protein globally (United States Department of Agriculture 2015). Purebred boars selected for their genetic merit are used at the top (nucleus) level of the breeding pyramid meaning that any fertility problems in these animals could significantly reduce litter sizes throughout the breeding population. This ultimately leads to a reduction in food production and higher environmental costs per mating animal, issues that are perpetuated further through an increasing emphasis on artificial insemination (AI) (Rodríguez-Gil & Estrada 2013).

Semen used in AI preparations is routinely assessed for parameters that are considered to be indicative of fertility such as sperm concentration, morphology and motility. Evidence suggests that these parameters are in fact, not reliable indicators of prolificacy (Gadea 2005). Indeed, the primary identification of boars that exhibit hypoprolificacy is deduced from both litter sizes and 'non-return rates', i.e. the proportion of sows/gilts served by that boar that return to heat (i.e. fail to conceive) after 21 days. With a gestation length of 115 days and an average litter size of 12 piglets, each sow can produce around 23 slaughter pigs per year assuming there are no fertility problems (The BPEX Yearbook 2014). In addition, fertility is assessed using farrowing rates, which indicate how many litters are produced against how many sows were originally served (ideally >85% (Gadea et al. 2004)). The mating of hypoprolific boars into the sow population can have a significant effect on non-return rates and litter sizes, in some cases reducing the number of piglets in a litter by up to 50%. In order to prevent the perpetuation of reduced fertility, the identification and elimination of hypoprolific boars from the breeding population is a priority, particularly given rising global populations and increasing demand for meat products.

Balanced chromosomal rearrangements described in section 1.2.2.2.1, occur frequently in pigs and are seen in as many as 0.47% of AI boars awaiting service (Ducos et al. 2007). Over 130 reciprocal translocations have been identified with chromosomes 1,7, 14 and 15 the most frequently involved (Rothschild & Ruvinsky 2011). Reciprocal translocations adversely affect reproductive performance in pigs by causing a reduction in litter size due to high mortality among early embryos. Approximately 50% of boars exhibiting hypoprolificacy are reciprocal translocation carriers, even though they have a normal phenotype and semen parameters (Rodríguez et al. 2010). Balanced translocations are considered to be the primary reason for hypoprolificacy in pigs due to the generation of unbalanced gametes and subsequent partially aneuploid conceptuses that lead to early loss of zygotes and ultimately litters that are 25-50% smaller than would be expected (Gustavsson 1990; Pinton et al. 2000).

Among cattle, the most commonly seen structural chromosomal rearrangements are Robertsonian translocations with the 1/29 being seen most frequently of the 44 that have been identified (Garrick and Ruvinsky, 2014). In one 15-year study of the Italian breeding population 7.1% animals were identified as carrying a Robertsonian translocation (Ducos et al. 2008). Reciprocal translocations have been reported in cattle, although much less frequently, with the same Italian study reporting a rate of 0.03% (Ducos et al. 2008). A recent study by De Lorenzi et al. (2012) suggested that the frequency of reciprocal translocations is grossly underreported, largely due to the inherent difficulties in detecting these rearrangements using routine cytogenetics (De Lorenzi et al. 2012). Techniques beyond the cytogenetic resolution of detection are also therefore essential for screening the cattle breeding population.

As described in section 1.1.2.5, technology has been developed in which cryptic translocations can be identified using a FISH strategy that involves 24 individual hybridisations on a single slide. In humans, this approach has been used extensively in clinical cytogenetics (Knight et al. 1996; Horsley et al. 1998; Ravnán et al. 2006; Dawson et al. 2002). The purpose of this study was to develop these investigations further to generate a panel of equivalent porcine BACs, extending on the Knight et al. study to develop a porcine version of the human system by isolating the most distal probes identified directly from the pig genome assembly. The aim was to employ a strategy that would significantly increase the speed and accuracy of boar translocation screening ultimately leading to the identification and removal of hypoprolific boars from the breeding population and to make preliminary steps towards extending this method of screening to cattle.

## **3.2 Specific aims**

With the above background in mind, the specific aims of this chapter were as follows:

- Specific aim 1a: To develop a practicable and commercially viable system for the screening of chromosome translocations in domestic male mammals by classical approaches
- Specific aim 1b: To isolate sub-telomeric chromosome identifier probes as tools for chromosome translocation detection in pigs and, in so doing test the hypothesis that genome assembly information in the porcine assembly accurately represents chromosomal position in the sub-telomeric regions
- Specific aim 1c: To develop a means of screening for porcine cryptic translocations for the whole karyotype in a single experiment
- Specific aim 1d: To apply the above system in the screening of sub-fertile boars and, in particular test the hypothesis that some sub-fertile boars have translocations that can be detected by this system but not by karyotyping alone
- Specific aim 1e: To test the hypothesis that translocation screening technology developed for boars can also be applied to other mammals of agricultural importance such as cattle

## **3.3 Materials and Methods**

### **3.3.1 Blood culture**

Porcine blood samples were provided (in heparin tubes) by the breeding companies referring their pigs for screening (ACMC, JSR Genetics, Danish Agricultural Food Council, TOPIGS Norsvin and Klasse Ki). Cattle blood for initial testing was provided by collaborators at Paragon Vets. Blood was cultured and metaphase suspensions harvested as described in section 2.1.3.1.

### **3.3.2 BAC Selection and FISH**

BAC clone selection and generation of labelled FISH probes was performed as described in section 2.2. FISH and multiprobe hybridisation was performed as described in section 2.3.2.

### 3.4 Results

#### 3.4.1 Specific aim 1a: To develop a practicable and commercially viable system for the screening of chromosome translocations in domestic male mammals by classical approaches

A routine karyotyping service was established for five different pig breeding companies, two of which are based in the UK: JSR Genetics and ACMC, and three which are based in continental Europe: The Danish Agricultural Food Council (Denmark), Klasse Ki (Holland) and TOPIGS Norsvin (Norway). Using the methods described in section 2.3.4, karyotypes were successfully produced for a total of 161 boars from different breeding populations with an average of 10 karyotypes created per boar. Twelve translocation carriers of six different reciprocal translocations were identified using this approach with no abnormalities identified in the remainder (as shown in Table 3-1).

<b>Results from boar karyotyping</b>	
<b>Normal boars:</b>	149
<b>Translocations found:</b>	12
<b>RCP 1:2</b>	1
<b>RCP 3:9</b>	1
<b>RCP 7:10</b>	7
<b>RCP 7:12</b>	1
<b>RCP 13:15</b>	1
<b>RCP 16:17</b>	1
<b>Total boars tested:</b>	161

**Table 3-1: Summary results from boar karyotyping for translocation screening.**

A large number of translocations (7), were identified between chromosomes 7 and 10 (an example of which is illustrated in Figure 3-1), primarily due to the identification of a translocation in one boar leading to the priority screening of closely related boars.

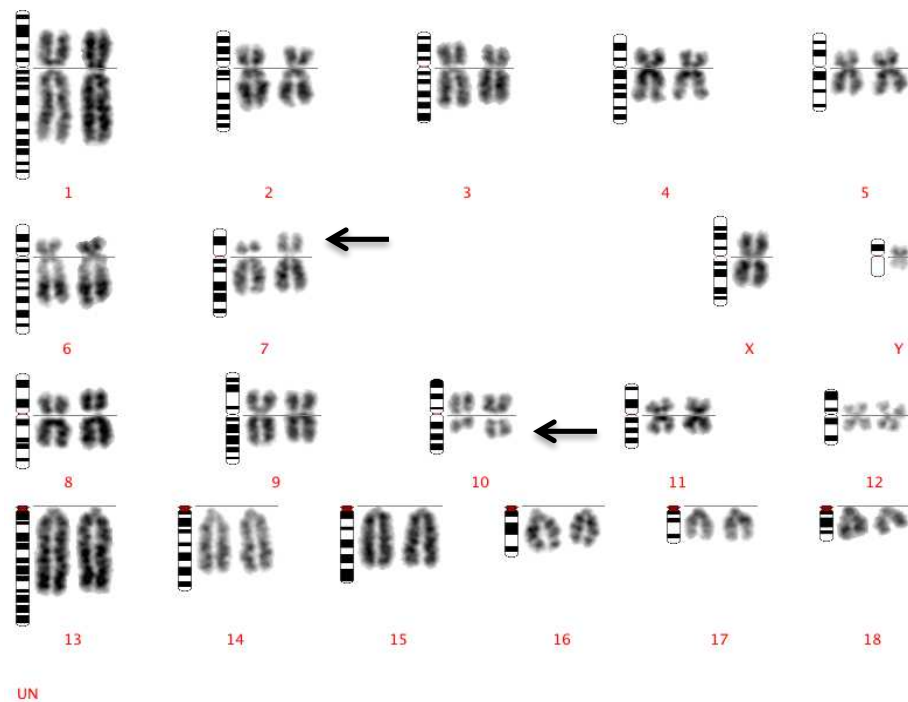
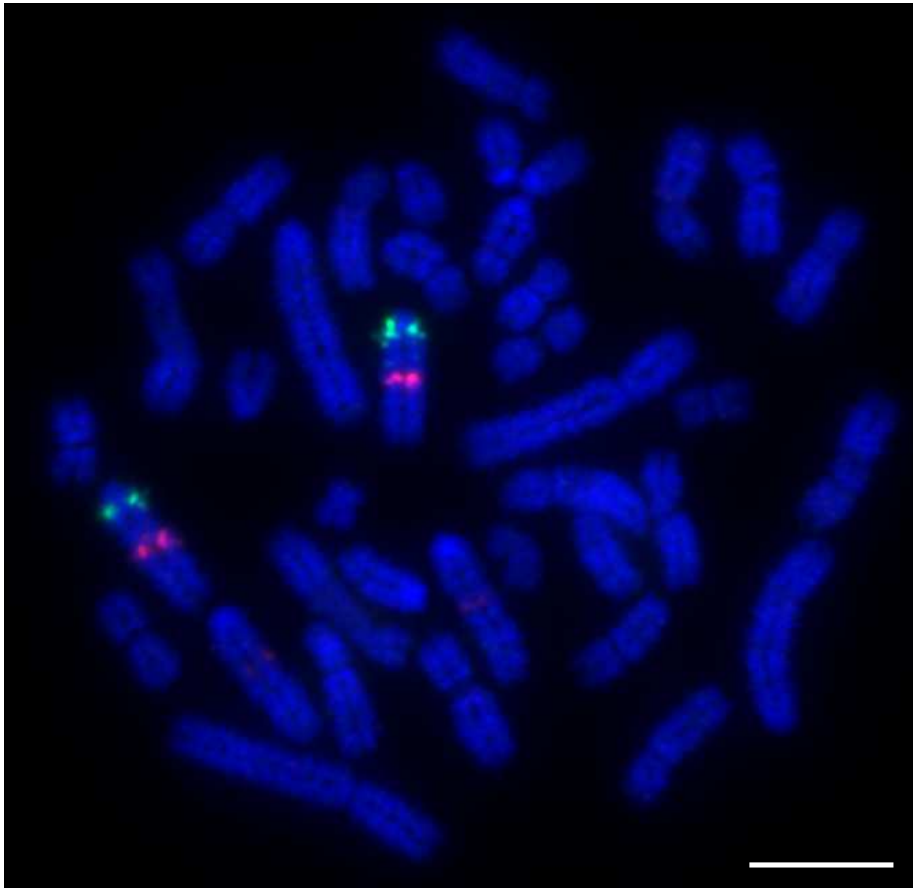


Figure 3-1: Standard DAPI banded karyotype of boar carrying a 7:10 RCP. The affected chromosomes are highlighted with arrows.

#### 3.4.2 Specific Aim 1b: To isolate sub-telomeric chromosome identifier probes as tools for chromosome translocation detection in pigs and, in so doing test the hypothesis that genome assembly information in the porcine assembly accurately represents chromosomal position in the sub-telomeric regions

A total of 82 BACs were tested, of which 45 BACs mapped correctly and 37 did not map as anticipated. All FITC labelled probes mapped to the expected locus at or near the p-terminus of the chromosome with the exception of the first attempt for a BAC (PigE-134L21) for the p-arm of chromosome 1 (which actually mapped to chromosome 8), along with a p-arm BAC for chromosome 10 (PigE-231H10) which mapped to chromosome 3 and three BACs originally assigned to chromosome 9p, which mapped elsewhere in the karyotype. After selecting alternative BACs however, bright green signals were observed at the appropriate end of the chromosome. Surprisingly, of the 51 probes that were originally assigned to the q-terminus of the chromosome, while displaying bright red signals, 32 mapped to a place in the genome other than that which was predicted. Of these, 24 clones (75%) mapped to the correct chromosome, but not to the q-terminus. An example is given in Figure 3-2 for chromosome 15 with the

anticipated mapping illustrated in a screenshot from NCBI clone finder shown in Figure 3-3 and the full list of incorrectly mapped BACs identified through this project given in Table 3-2.



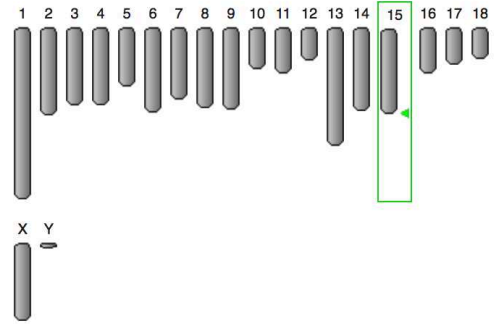
**Figure 3-2: Clone ID PigE-108N22 labelled in Texas Red which should map to the distal end of SSC15 but appears halfway along this acrocentric chromosome. The FITC labeled probe mapped correctly. Scale bar 10 $\mu$ m.**

**PigE-108N22**

(ID:22741930)

**Clone Summary**

**Clone name aliases:** bT108n22  
**Organism:** Sus scrofa  
**Library name:** [The ARK-GENOMICS pig BAC library \(PigE\)](#)  
**Library type:** genomic



**Sscrofa10.2**

◀ PigE-108N22: Selected sequence-based clone placement

Genome View | Clone Placements | Associated Sequences | Distributors

Genome sequence View for the Clone record

**Sequence-Based Placements** ⓘ

Assembly	Asm. unit	Chr	Seq. ID	Start	End	Length	Method	Conc.	Confidence	Placed by
Sscrofa10.2	Primary Assembly	15	<a href="#">NC_010457.4</a>	156,720,525	156,845,791	125,267	end-seq	Y	Unique	NCBI

Figure 3-3: Screenshot from NCBI clone finder illustrating where BAC PigE-108N22 was expected to map on SSC15.

An unexpected finding of this study therefore was that that probes assigned to the q-arm were frequently incorrectly mapped, with the majority of probes mapping to the correct chromosome but the incorrect locus. Correctly mapping q-arm probes were eventually assigned by choosing BACs (using an *in silico* approach) that were assigned to larger, fully mapped contigs closest to the q-terminus.



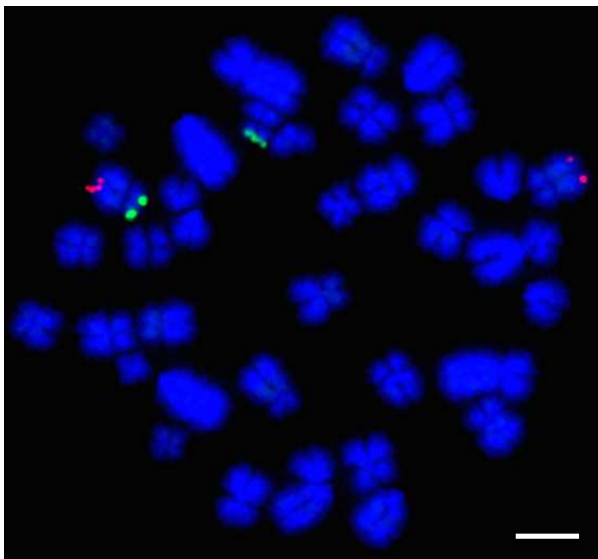
Number	Arm	Clone Name	FISH Assignment	Same Chromosome?
1	p	PigE-134L21	8 p-arm	No
1	q	CH242-137C1	10 centromere	No
1	q	CH242-35I10	Multiple	No
1	q	CH242-83P21	7 centromere	No
2	q	CH242-188K23	2 centromere	Yes
2	q	CH242-230M23	2 centromere	Yes
2	q	CH242-441A1	2 centromere	Yes
2	q	PigE-117G14	2 p-arm	Yes
3	q	CH242-265K24	3 p-arm	Yes
3	q	PigE-221G14	3 p-arm	Yes
3	q	PigE-264D16	3 p-arm	Yes
5	q	CH242-133F9	5 p-arm	Yes
5	q	CH242-288F8	5 p-arm	Yes
5	q	PigE-127K14	5 p-arm	Yes
5	q	PigE-178M22	5 p-arm	Yes
7	q	CH242-272F22	7 centromere	Yes
7	q	CH242-518F14	7 centromere	Yes
7	q	PigE-208I10	3 q-arm	No
7	q	PigE-230H8	7 centromere	Yes

Number	Arm	Clone Name	FISH Assignment	Same Chromosome?
7	q	PigE-75E21	7 mid q-arm	Yes
9	p	CH242-215O14	9 centromere	Yes
9	p	CH242-44O5	9 centromere	Yes
9	p	CH242-178L4	9 centromere	Yes
10	p	PigE-231H10	3 p-arm	No
10	q	CH242-237D22	10 centromere	Yes
10	q	CH242-36D16	10 q-arm + extra signal	Yes
10	q	PigE-60N24	1 centromere	No
11	q	PigE-199B10	11 p-arm	Yes
11	q	PigE-232N19	11 p-arm	Yes
15	q	PigE-108N22	15 mid q-arm	Yes
16	q	CH242-4G9	16 p-arm	Yes
16	q	PigE-124C22	16 p-arm	Yes
16	q	PigE-173H6	16 p-arm	Yes
17	q	PigE-112L22	10 centromere	No
18	q	PigE-141I21	6 p-arm	No
X	q	CH242-447L20	X p-arm	Yes
X	q	PigE-214O4	13 centromere	No

Table 3-2: Incorrectly mapped porcine BACs.

### **3.4.3 Specific Aim 1c: To develop a means of screening for porcine cryptic translocations for the whole karyotype in a single experiment**

Once the full complement of mapped and optimised porcine probes was established, they were applied to the multi probe device as described in methods section 2.3.2. The device enabled visualisation of bright, punctate signals (one green, one red) for each chromosome. The multiprobe device was subsequently tested on 20 chromosomally normal preparations and each translocation carrier in order to confirm the cytogenetic diagnosis. The device confirmed the diagnosis of translocations between chromosomes 1 and 2; 3 and 9, 3 and 7, 3 and 14, 7 and 10 (Figure 3-4); 7 and 12; 9 and 13; 13 and 15.



**Figure 3-4: Labelled probes for SSC7 illustrating a reciprocal translocation between SSC7 and SSC10. Scale bar 10µm.**

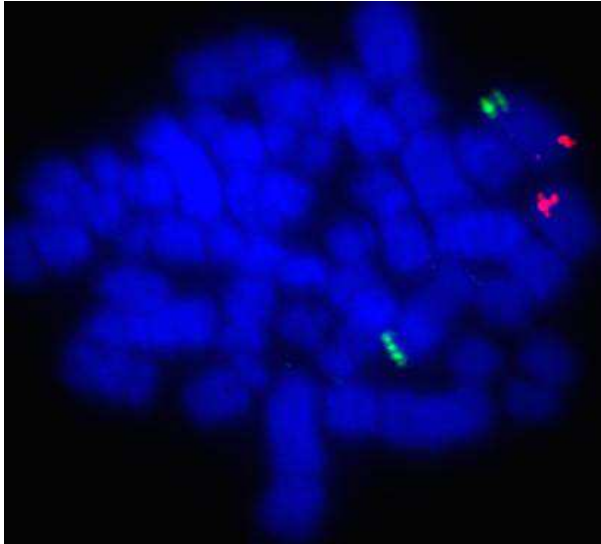
In addition, no abnormalities were seen in the other preparations. A full list of subtelomeric BACs that give bright signals on the appropriate chromosome arm is provided in Table 3-3.

Chromosome	Arm	Clone Name	Chromosome	Arm	Clone Name
1	p	CH242-248F13	10	q	CH242-517L16
1	q	CH242-151E10	11	p	PigE-211E21
2	p	PigE-8G19	11	q	CH242-239O11
2	q	CH242-294F6	12	p	PigE-253K5
3	p	PigE-168G22	12	q	PigE-124G15
3	q	CH242-315N8	13	P	PigE-197C11
4	p	PigE-131J18	13	q	PigE-179J15
4	q	PigE-85G21	14	p	PigE-137C12
5	p	PigE-74P10	14	q	PigE-167E18
5	q	CH242-63B20	15	p	PigE-90C11
6	p	PigE-238J17	15	q	CH242-170N3
6	q	CH242-510F2	16	p	PigE-149F10
7	p	PigE-52L22	16	q	CH242-42L16
7	q	CH242-103I13	17	p	CH242-70L7
8	p	PigE-2N1	17	q	CH242-243H19
8	q	PigE-118B21	18	p	PigE-253N22
9	p	CH242-65G4	18	q	PigE-202I11
9	q	CH242-411M8	X	p	CH242-19N1
10	p	CH242-451I23	X	q	CH242-305A15

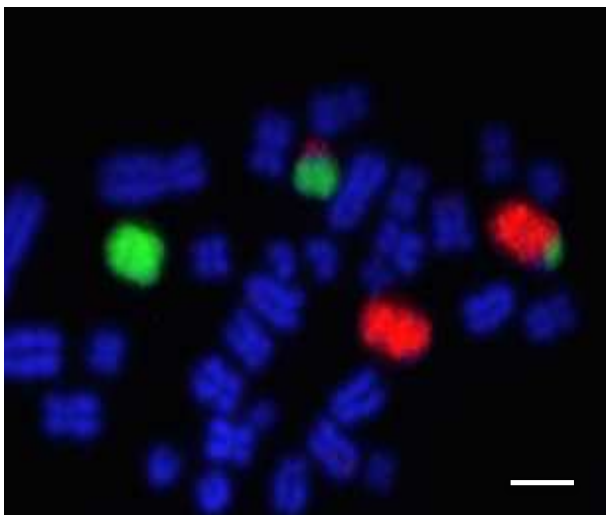
**Table 3-3: Correctly mapped BACs for each porcine chromosome arm.**

#### **3.4.4 Specific Aim 1d: To apply the above system in the screening of sub-fertile boars and, in particular test the hypothesis that some sub-fertile boars have translocations that can be detected by this system but not by karyotyping alone**

A further boar that had previously been identified as karyotypically normal was re-referred for analysis using the Multiprobe device. Breeding records indicated a history of reduced litter sizes i.e. 6.5 born alive and 0.1 born dead with a farrowing rate of 67% compared to the average rates from other boars at the same farm of 11.3 born alive and 0.6 born dead and a farrowing rate of 74%. The Multiprobe device revealed a chromosome translocation between chromosomes 5 and 6 that was missed by classical karyotyping (shown in Figure 3-5). Further analysis with chromosome painting for porcine chromosomes 5 and 6 on this boar revealed a cryptic translocation with the distal portions of the two chromosomes exchanged (Figure 3-6).



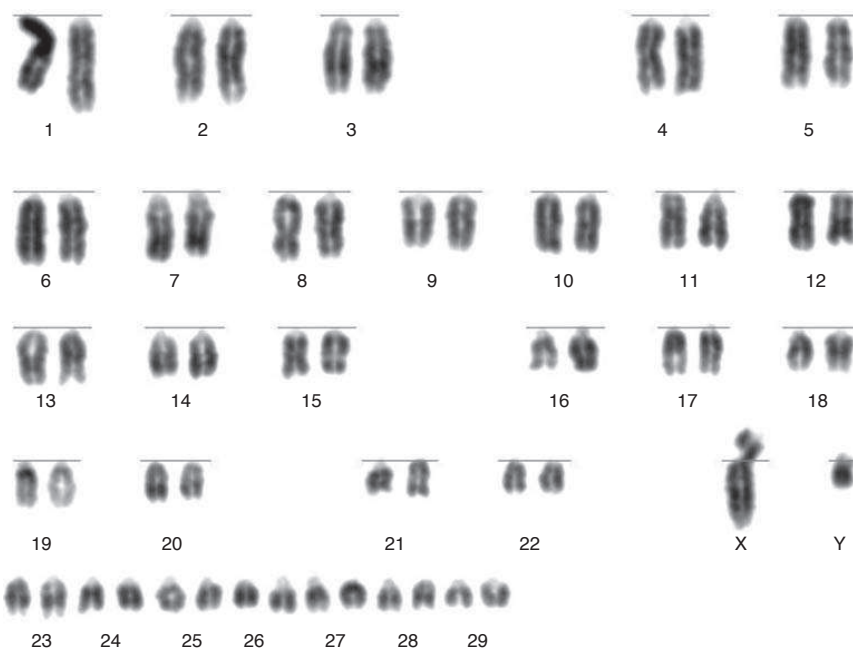
**Figure 3-5: Labelled BAC clones for SSC5 (p-arm labelled in FITC and q-arm labelled in Texas Red) showing a translocation between chromosome 5 and 6 despite suboptimal chromosome preparation. Scale bar 10 $\mu$ m.**



**Figure 3-6: Chromosome paints for SSC5 (Texas Red) and SSC6 (FITC) illustrating the cryptic translocation that had been previously undetectable from the karyotype. Scale bar 10 $\mu$ m.**

**3.4.5 Specific Aim 1e: To test the hypothesis that translocation screening technology developed for boars can also be applied to other mammals of agricultural importance such as cattle**

Prior to selecting BACs for subtelomeric chromosome screening, karyotype analysis of a bull was performed as illustrated in Figure 3-7. The *Bos taurus* karyotype has a diploid chromosome number of 60 of which all autosomes are acrocentric and of a similar size meaning that producing an accurate karyotype is challenging.



**Figure 3-7: DAPI banded cattle karyotype (*Bos taurus*) where 2n=60 (Larkin et al. 2014).**

In order to create a set of cattle specific subtelomeric FISH probes, BAC clones were selected from the CHORI-240 library using the same criteria as defined for selection of the porcine BACs, described in methods section 2.2.1.1. Signals were successfully produced for all BACs, however a small proportion (6 BAC clones: 11p: CH240—205F19, 12p: CH240—3K6, 18p: CH240—448G16, 20q: CH240—6H13, 24p: CH240—219O21, 25q: CH240—98G11) did not co-localise on the same chromosome and were therefore replaced with BACs that did map correctly.

Chrom	Arm	Clone Name	Span
1	p	CH240-321O2	179,965
	q	CH240-96M6	187,920
2	p	CH240-457J20	198,157
	q	CH240-227E16	179,789
3	p	CH240-154A5	174,225
	q	CH240-302G6	190,291
4	p	CH240-416O20	170,609
	q	CH240-193F3	179,112
5	p	CH240-326L8	188,525
	q	CH240-248M21	163,993
6	p	CH240-324B6	180,970
	q	CH240-5F18	184,848
7	p	CH240-415D2	182,547
	q	CH240-276L16	168,781
8	p	CH240-443K7	175,465
	q	CH240-241A18	176,318
9	p	CH240-25A3	177,086
	q	CH240-298I24	172,331
10	p	CH240-421B11	166,378
	q	CH240-325F16	179,292
11	p	CH240-314K5	165,445
	q	CH240-344O3	183,795
12	p	CH240-261C16	164,440
	q	CH240-262C4	165,223
13	p	CH240-461F6	188,788
	q	CH240-471M8	178,736
14	p	CH240-319C15	181,738
	q	CH240-240M1	178,587
15	p	CH240-225A24	151,902
	q	CH240-386C2	168,728
16	p	CH240-139M7	166,377
	q	CH240-315I10	186,228
17	p	CH240-267P22	176,654
	q	CH240-313I20	182,729
18	p	CH240-14C14	163,878
	q	CH240-436N22	179,260
19	p	CH240-349G17	169,018
	q	CH240-390C5	180,283
20	p	CH240-394L14	182,595
	q	CH240-339K22	183,557
21	p	CH240-301D14	163,699
	q	CH240-62O23	176,169
22	p	CH240-426O23	182,818
	q	CH240-313B20	173,299
23	p	CH240-102P19	179,615
	q	CH240-374G6	174,942
24	p	CH240-382F1	171,530
	q	CH240-19L13	171,917
25	p	CH240-198J4	186,545
	q	CH240-379D22	163,818
26	p	CH240-428I10	181,997
	q	CH240-389H1	176,691
27	p	CH240-7G11	184,155
	q	CH240-352M8	184,694
28	p	CH240-313L4	181,707
	q	CH240-63D12	183,932
29	p	CH240-367D17	179,713
	q	CH240-257F23	188,054
X	p	CH240-121E1	176,736
	q	CH240-472J20	186,872

**Table 3-4: Correctly mapped subtelomeric cattle BACs by chromosome from the CHORI-240 library.**

The definitive list of correctly mapping BACs is listed in Table 3-4. An example of two BACs that successfully mapped and hybridised to cattle chromosome suspensions for chromosome 24 is illustrated in Figure 3-8.

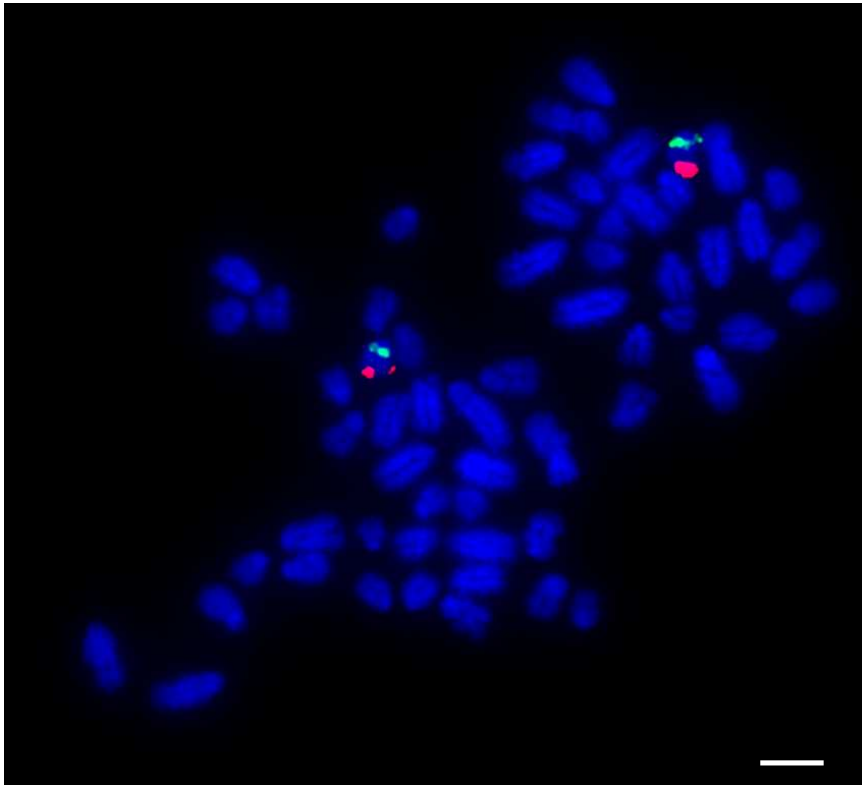


Figure 3-8: Subtelomeric cattle probes for BTA24 p-arm labelled in FITC (CH240-382F1) and q-arm labelled in Texas Red (CH240-19L13). Scale bar 10 $\mu$ m.

A further chicken specific device was developed which is discussed in the next chapter.

## 3.5 Discussion

### 3.5.1 Chromosome translocation detection using FISH

Results from this study provide proof of principle of an approach that can be used successfully to diagnose chromosomal translocations that directly impact fertility in pigs at a resolution previously difficult to achieve by standard karyotyping. It was also extended to cattle and chicken. There are three advantages of using this approach over classical karyotyping: The first is that it detects more cryptic translocations than standard karyotyping otherwise would. The boar with a 5:6 reciprocal translocation described in section 3.4.4 is an example. Indeed, the fact that a previously undetected cryptic translocation was identified would suggest that the actual number of translocations in the boar breeding population may in fact be significantly higher than previously reported. It is possible that these karyotypically cryptic and unreported translocations are seen more frequently than expected but that the routine use of multiple inseminations per sow may be diluting the effect on the farrowing rates. The boar with a cryptic

translocation in this study had a significantly reduced farrowing rate and interestingly also had a significantly lower 'born dead' rate suggesting that the translocation in this case results in early embryo loss. It would appear that the production of unbalanced gametes caused by the translocation in question results in embryos that are not compatible with early life causing early embryo mortality in a pattern that is also seen in humans (Tempest & Simpson 2010). In humans, reciprocal translocations arise more frequently *de-novo* rather than from being inherited from a carrier parent (Tempest & Simpson 2010). It would therefore be reasonable to suggest that the same pattern of familial inheritance applies to pigs and other animals. The *de novo* nature of these translocations supports the theory that all boars awaiting service should be screened chromosomally to reduce the risk of using a hypoprolific animal for breeding purposes. In fact, despite over 130 reciprocal translocations being reported in the literature, to date this is the first reported translocation to have occurred between chromosomes 5 and 6 suggesting that this fits that category (Rothschild & Ruvinsky 2011). Secondly, as in this case, when preparations are sub-optimal, this approach provides necessary 'back-up' to ensure accurate diagnosis. That is, provided FISH signals are clear, confident diagnosis can be made on a single metaphase, regardless of the length of the chromosomes.

The use of this method of translocation screening also permits analysis by individuals who are less well trained in karyotype analysis. Although several laboratories have pioneered animal cytogenetics for the purposes of AI boar (and bull) screening there are fewer now than in previous decades despite the continuing need to continue screening in this manner (Ducos et al. 2008). Nonetheless, it should be made clear that specialist cytogenetic skills are still required to make chromosome preparations reliably in the lab and to perform overall analyses. The scheme developed here should therefore be considered an adjunct to classical cytogenetics, not a replacement for it.

### **3.5.2 Genome Assembly Integrity**

A second outcome of this study was the revelation that a large number of BACs isolated from current published pig genome assembly (Sscrofa10.2, Groenen et al. 2012) mapped incorrectly. That is, those that were predicted to map to the q-terminus of a particular chromosome mapped elsewhere on the same chromosome. The high level of mapping errors found in this study led to further investigation of the clone placement with members of the Swine Genome Sequencing Consortium. It became evident that the problem was the result of errors in the way in which parts of the draft pig genome sequence were assembled. Specifically, analysis of the BAC



sequences revealed that the high error rate was due to misplacement of some of the smaller fingerprint contigs (FPCs) within which the BAC was located. These small FPCs did not have full sequence and orientation data when the genome was assembled and it appears that these small contigs were added to the end of the list of contigs for each chromosome. This resulted in the sequences from the BACs in these poorly mapped contigs being randomly assigned to the end of the relevant chromosomes, which explains why the error rate was particularly high among BACs chosen to map to the subtelomeric q-arm region. The analysis of Warr et al. (2015) identified limitations of this current assembly at a more granular level.

The genome assembly errors found throughout the course of this project highlight the need for caution when choosing BACs for this purpose in any genome, particularly as this genome is considered to be one of the better assembled ones. A revised reference pig genome assembly is now being developed by scientists at The Roslin Institute, USDA Meat Animal Research Center and elsewhere (Smith et al. 2016) and is more than four hundred times more contiguous than the current Sscrofa10.2 assembly (A.L. Archibald, personal communication). In addition, the cytogenetic mapping information generated here will be used for checking the chromosomal assignments and long range order and orientation for this new pig reference genome assembly. Data generated through the work presented here will therefore contribute to the next published pig genome. With the rapid expansion in the number of newly sequenced animal genomes being published, along with corresponding BAC libraries for many, the possibility of assembly errors should be an important consideration for future similar studies.

### **3.5.3 Extension of the FISH screening method to other agricultural species**

As demonstrated in 3.4.5, the application of these subtelomeric FISH probes for translocation screening is not limited to pig breeding. Artificial insemination is also widely used in cattle breeding with a high premium placed on bull semen of superior genetic merit. With sufficient alterations (i.e. incorporating cattle subtelomeric BACs) the method has successfully been adapted for cattle and, could in the future, be adapted for other species. In addition, the increasingly widespread use of embryo transfers in cattle would indicate that both the cow and the bull should be screened for chromosomal translocations. In fact, the cattle karyotype is more difficult to analyse reliably because of a diploid number of 60, largely made up of similar sized acrocentric chromosomes as illustrated in Figure 3-7. The cattle karyotype therefore lends itself to the use of a FISH based screening approach such as is described here, as does the largely acrocentric sheep karyotype ( $2n=54$ ). Beyond screening for fertility related translocations, there

is nothing to stop this high throughput approach being applied to the analysis of chromosomes for any species. Indeed, in the next chapter it's use on chicken (and other avian) chromosomes is described – in this case as a means of identifying chromosomal changes at an evolutionary level.

### **3.6 Conclusions**

Now that a full set of porcine and bovine subtelomeric probes has been isolated and applied in the manner described, screening efficiency can be improved by allowing the analysis of the full chromosomal complement on one slide. Given the nature of translocations and their impact on fertility in pigs, the simple, rapid identification of (cryptic or otherwise) translocations will facilitate the detection and subsequent removal of affected animals from the breeding population at an early stage. This has the potential to lead to long-term improved productivity, delivering meat products in a more cost-effective and environmentally friendly way to a growing population. The widespread use of artificial insemination and the large market for superior boar and bull semen being sold to both small and large-scale pig breeding operations suggests that improvements in productivity impact not just the large commercial breeders but also the smaller farmers where reduced wastage may be more critical. Lessons regarding genome assembly learnt from this exercise would suggest that a cautionary approach be taken when identifying BACs for this purpose and that a combined *in-silico* and experimental (wet lab) approach is crucial in the development of similar tools. Ultimately, the FISH based translocation screening technique developed in this study is a powerful and reliable approach to screening with great potential to be adapted to other species as illustrated here and in the next chapter.

## **4 Specific aim 2: To apply the technology developed in specific aim 1 for the solving of previously intractable karyotypes and test the hypothesis that microchromosomal rearrangement is rare in avian evolution**

### **4.1 Background**

Defining the avian karyotype is notoriously difficult, largely due to the presence of a large number of morphologically indistinguishable microchromosomes seen in the majority of bird species (see section 1.3.2). Classification of avian macrochromosomes (up to chromosome 9) is possible using classical cytogenetic techniques such as karyotyping but beyond this size of chromosome it is near impossible, hence the publication of partial rather than full avian karyotypes. Even at the macrochromosomal level, chromosome banding can be difficult to identify, thereby making a robust analysis of cross-species homology difficult and unreliable (Griffin et al. 2007). The development of chromosome paints derived from amplification and fluorescence labelling of chicken macrochromosomes (Griffin et al. 1999) has improved on this limited resolution and has led to the publication of zoo-FISH analysis on over 55 avian species from 15 different orders as described in section 1.1.2.3. These studies are however restricted to analysis of the macrochromosomes. Some success has been achieved using microchromosomal paints (Lithgow, O'Connor R.E. et al. 2014) but again this is limited, largely due to the inability to separate each microchromosome at the flow sorting level, meaning that individual chromosome paints cannot be generated. A degree of success using a cross-species BAC mapping approach has been reported, although this is limited to closely related species, with 70% success rates reported using chicken BACs on turkey (*Meleagris gallopavo*) (Griffin et al. 2008) which reduces to under 40% when tested on duck (*Anas platyrhynchos*) (Skinner et al. 2009) although marginally higher rates between chicken and duck have been reported elsewhere (Fillon et al. 2007).

As described in section 1.4, the apparent highly conserved nature of microchromosomes (Völker et al. 2010; Griffin et al. 2008; Skinner et al. 2009) has proven difficult to investigate using either classical or molecular cytogenetic methods. The purpose of this chapter was then to develop tools that provide a means of reliably analysing avian karyotypes of multiple avian species at a higher cytogenetic resolution than previously achieved, and at a wider phylogenetic distance than previously possible. Results generated using these tools will permit a more detailed

investigation into inter and intrachromosomal rearrangements that have occurred between species, providing insights into the mechanisms of evolution and speciation.

## 4.2 Specific aims

With the above background in mind, the specific aims of this chapter were as follows:

- Specific aim 2a: To develop a means of identifying all sequenced avian chromosomes in a single experiment by molecular cytogenetics
- Specific aim 2b: To test the above system on closely related avian species to identify patterns of inter chromosomal rearrangement during avian evolution
- Specific aim 2c: To develop a means of improving the success of cross species FISH using BAC clones
- Specific aim 2d: To adapt the system described in specific aim 2a using technology developed in aim 2c so that it can be applied to all bird karyotypes
- Specific aim 2e: To test the hypothesis that micro chromosomal rearrangement is a rare event, compared to other vertebrate species, during avian genome evolution

## 4.3 Materials and Methods

### 4.3.1.1 *Chromosome preparation*

Chromosome preparations were derived from cells grown from skin biopsies, trachea, embryonic fibroblasts and lymphocyte culture as described in section 2.1.2.

### 4.3.1.2 *BAC Selection*

In order to improve the rate of successful hybridisation in distantly related avian species, the criteria for selecting BACs described in section 2.2.1.2 was designed in partnership with collaborators in the Larkin lab based at the Royal Veterinary College, London. These criteria included the following parameters: sequence conservation between species, low repeat content and GC content of the BAC being between 40-60% (described hereafter as 'selected' BACs). 'Non-selected' BACs were chosen using the criteria described in section 2.2.1.1.

### 4.3.1.3 *BAC Selection and FISH*

Generation of labelled FISH probes was performed as described in section 2.2. FISH and multiprobe hybridisation was performed as described in section 2.3.2.

## 4.4 Results

### 4.4.1 Specific Aim 2a: To develop a means of identifying all sequenced avian chromosomes in a single experiment by molecular cytogenetics

As the best characterised avian species with the largest BAC library, the chicken (*Gallus gallus*-GGA) was selected as the reference species to study. To facilitate analysis of the microchromosomes, BACs located in the subtelomeric regions of the p and q-arms of chromosomes 10-28 were selected as listed in Table 4-1. Co-hybridisation of p and q-arm BACs for each chromosome was performed to verify correct mapping of the BACs, producing bright punctate signals for each of the chromosomes, an example of which is shown for GGA chromosome 12 in Figure 4-1.

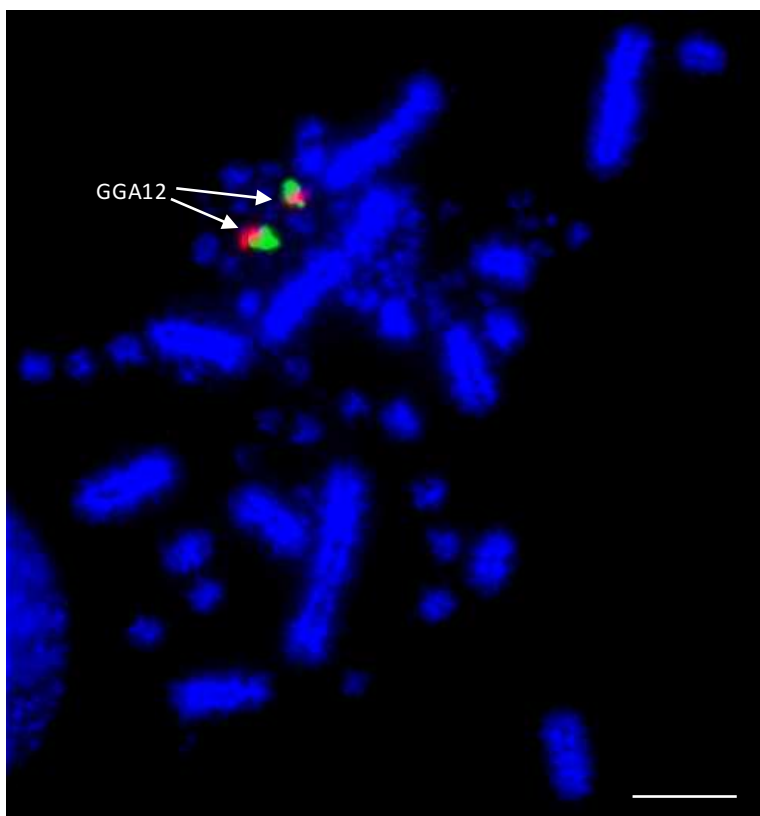


Figure 4-1: Example of dual FISH results for Chicken (*Gallus gallus* – GGA) chromosome 12 to confirm correct mapping. The p-arm BAC (CH261-88K1) is labelled with FITC and the q-arm BAC (CH261-152H14) is labelled with Texas Red. Scale bar 10µm.

Chrom	Arm	GGA Start Position	GGA End Position	Clone Name	Size
10	p	257,559	435,710	CH261-94C12	178,152
	q	19,689,227	19,878,620	CH261-118E15	189,394
11	p	297,412	487,723	CH261-28F3	190,312
	q	207,654,262	20,963,698	CH261-89P18	189,505
12	p	372,960	564,981	CH261-88K1	192,022
	q	19,943,504	20,143,412	CH261-152H14	186,044
13	p	639,659	830,204	CH261-71N1	190,546
	q	17,531,818	17,705,631	CH261-58H5	173,814
14	p	106,806	274,491	CH261-121M10	167,686
	q	14,836,141	15,050,135	CH261-17L7	213,995
15	p	37,719	230,572	CH261-131E4	192,854
	q	12,766,676	12,953,903	CH261-40D6	187,228
16	p	26,112	177,988	CH261-97F21	151,877
	q	26,068	226,529	CH261-96L12	200,462
17	p	180,149	368,825	CH261-72P11	188,677
	q	10,201,570	10,372,603	CH261-69M11	171,034
18	p	159,671	345,409	CH261-67N15	185,739
	q	10,715,138	10,889,646	CH261-72B18	174,509
19	p	120,682	300,545	CH261-167A1	179,864
	q	9,757,250	9,923,027	CH261-189E4	165,778
20	p	101,638	271,770	CH261-124A24	170,133
	q	13,731,632	13,930,041	CH261-10L6	198,410
21	p	190,095	394,600	CH261-10A18	204,506
	q	6,676,005	6,808,856	CH261-49L18	132,852
22	p	16,255	200,733	CH261-30D24	184,479
	q	3,718,853	3,905,054	CH261-49B2	186,202
23	p	6,821	200,984	CH261-191G17	194,164
	q	5,713,201	5,883,262	CH261-90K11	170,062
24	p	262,499	457,108	CH261-154L15	194,610
	q	5,801,069	6,000,425	CH261-154H17	199,357
25	p	1,053,207	1,220,181	CH261-82G24	166,975
	q	1,540,100	1,719,066	CH261-169N16	178,967
26	p	269,996	448,389	CH261-50J5	178,394
	q	4,211,425	4,409,618	CH261-40C14	198,194
27	p	112,338	254,681	CH261-167J20	142,344
	q	4,662,705	4,830,235	CH261-100E5	167,531
28	p	77,131	247,634	CH261-16I3	170,504
	q	457,773	653,378	CH261-101C8	195,606

**Table 4-1: Chicken (*Gallus gallus* - GGA) BACs selected from the CHORI-261 library for each chicken microchromosome with start and end positions in the chicken genome and the size of the BAC listed.**

Subsequent application of the labelled probes to the multiprobe device (see section 1.1.2.5) along with previously generated macrochromosome paints (Griffin et al. 1999) in the orientation illustrated in section 2.3.2 enabled fast efficient hybridisation of both macrochromosome paints and microchromosome BACs to chicken metaphases in one experiment, therefore allowing for analysis of all sequenced chromosomes on one slide.

#### 4.4.2 Specific Aim 2b: To test the above system on closely related avian species to identify patterns of inter chromosomal rearrangement during avian evolution

In order to assess whether gross chromosomal rearrangements occurred within the *Galliformes* order, the aforementioned device was tested on a variety of species including the peafowl (*Pavo cristatus*), sand partridge (*Ammoperdix heyi*), Chinese quail (*Coturnix chinensis*), the Japanese quail (*Coturnix japonica*) and the turkey (*Meleagris gallopavo*). Almost all of the probes hybridised well to each species with bright signals evident for each chromosome, with a few exceptions as listed in Table 4-2. No apparent microchromosome rearrangement was found in any of these species with all probes hybridising in the same pattern that is seen in the chicken. Paints for chicken macrochromosomes 2 (FITC), 5 (Texas Red) and 8 (Aqua) tested on the peafowl revealed a fusion in the region homologous to GGA chromosome 8 as shown in Figure 4-2.

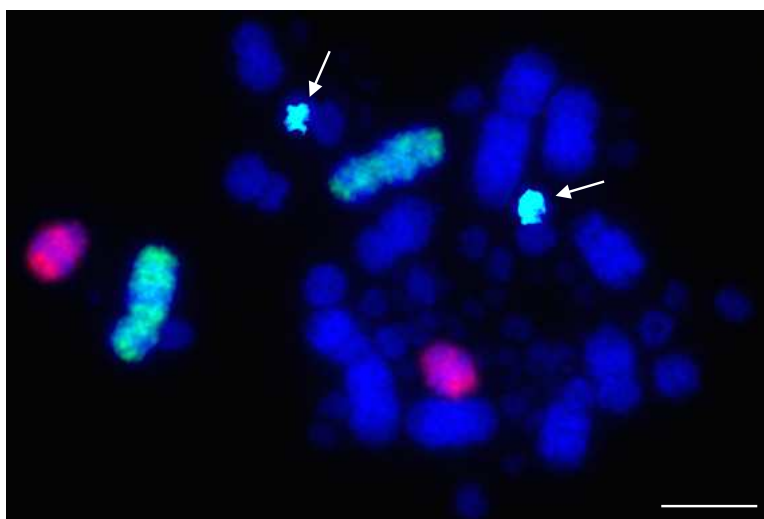


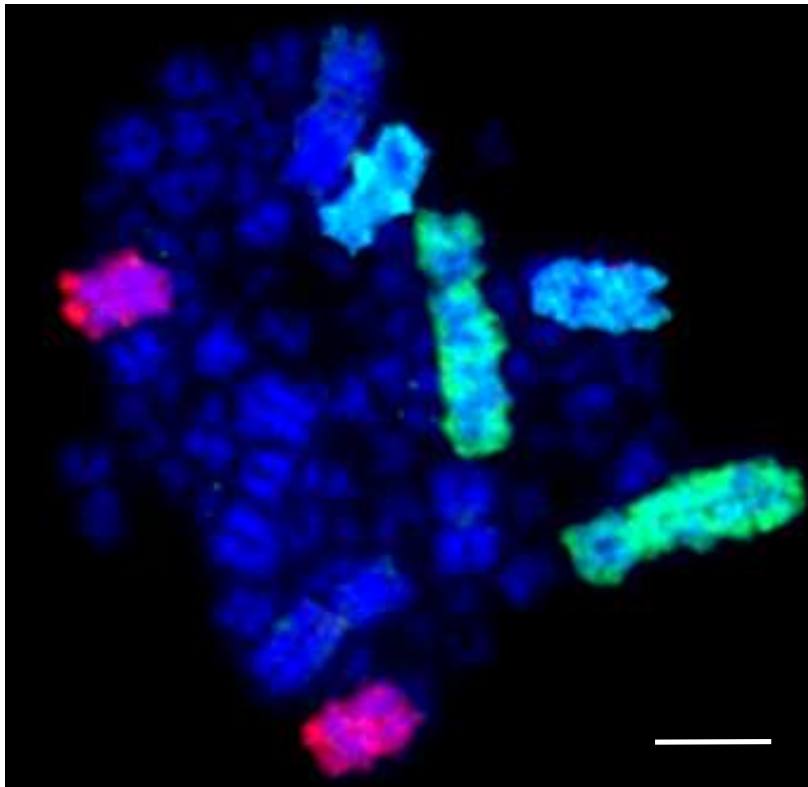
Figure 4-2: Chicken (*Gallus gallus* - GGA) macrochromosome paints for chromosomes 2 (FITC), 5 (Texas Red) and 8 (aqua) tested on the peafowl (*Pavo cristatus*) illustrating a fusion revealed by the aqua paint where the paint only covers around half of the chromosome. Scale bar 10µm.

BAC Clone Name	GGA Chr	Japanese Quail	Chinese Quail	Sand Partridge	Peafowl	Turkey
CH261-118E15	10	✓	✓	✓	✓	✓
CH261-94C12	10	✓	✓	✓	✓	✓
CH261-28F3	11	✓	✓	✓	✓	No
CH261-89P18	11	✓	✓	✓	✓	No
CH261-152H14	12	✓	✓	✓	✓	✓
CH261-88K1	12	✓	✓	✓	✓	✓
CH261-58H5	13	✓	✓	✓	✓	✓
CH261-71N1	13	✓	✓	✓	✓	✓
CH261-121M10	14	✓	No	✓	✓	No
CH261-17L7	14	✓	✓	✓	✓	No
CH261-131E4	15	✓	✓	✓	✓	✓
CH261-40D6	15	✓	✓	✓	✓	✓
CH261-97F21	16	✓	✓	✓	✓	✓
CH261-96L12	16	✓	✓	✓	✓	✓
CH261-69M11	17	✓	✓	✓	No	✓
CH261-72P11	17	✓	No	✓	No	✓
CH261-67N15	18	✓	✓	✓	✓	✓
CH261-72B18	18	✓	✓	✓	✓	✓
CH261-167A1	19	✓	✓	✓	✓	No
CH261-189E4	19	✓	✓	✓	✓	No
CH261-10L6	20	✓	✓	✓	✓	✓
CH261-124A24	20	✓	✓	✓	✓	✓
CH261-10A18	21	✓	✓	✓	✓	✓
CH261-49L18	21	✓	✓	✓	✓	✓
CH261-30D24	22	✓	✓	✓	✓	✓
CH261-49B2	22	✓	✓	✓	✓	✓
CH261-191G17	23	✓	✓	✓	✓	✓
CH261-90K11	23	✓	✓	✓	✓	✓
CH261-154H17	24	✓	✓	✓	✓	✓
CH261-154L15	24	✓	✓	✓	✓	✓
CH261-169N16	25	✓	✓	✓	✓	✓
CH261-82G24	25	✓	✓	✓	✓	✓
CH261-40C14	26	✓	✓	✓	✓	✓
CH261-50J5	26	✓	✓	✓	✓	✓
CH261-100E5	27	✓	✓	✓	✓	✓
CH261-167J20	27	✓	✓	✓	✓	✓
CH261-101C8	28	✓	✓	✓	✓	✓
CH261-16I3	28	✓	✓	✓	✓	✓

**Table 4-2: Successful cross species hybridisations of chicken microchromosome BACs within the *Galliformes* order. Clear signals are denoted by an '✓' and 'No' where no signal was evident.**



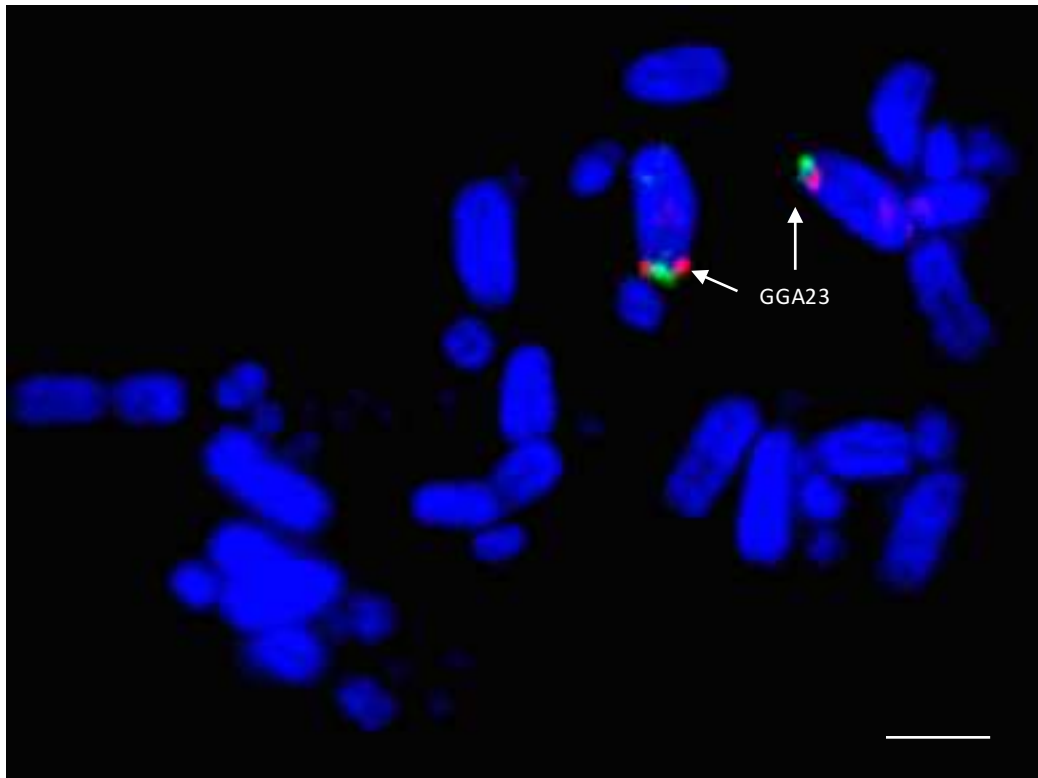
Results for the rest of the macrochromosome paints were consistent with that previously published with the (Japanese quail (Guttenbach et al. 2003; Shibusawa et al. 2004), Chinese quail (Shibusawa et al. 2004) shown in Figure 4-3 , the peafowl (Shibusawa et al. 2004) and the sand partridge exhibiting the chromosome 4 fusion seen in chicken. The results also supported previous studies on turkey illustrating the hybridisation of the chromosome 4 paint to a macro and a microchromosome and the chromosome 2 paint hybridising to 2 macrochromosomes (Shibusawa et al. 2004; Griffin et al. 2008).



**Figure 4-3: Macrochromosome GGA paints 1 (FITC), 3 (Aqua) and 4 (Texas Red) hybridised to Chinese quail chromosomes illustrating the fused chromosome four. Scale bar 10µm.**

To extend the analysis of microchromosomal rearrangements across species, the device was also tested on a variety of other species outside of the *Galliformes*. These included the houbara (*Chlamydotis undulata*), peregrine falcon (*Falco peregrinus*), gyr falcon (*Falco rusticolus*), budgerigar (*Melopsittacus undulatus*), pigeon (*Columba livia*), zebra finch (*Taeniopygia guttata*) and Pekin duck (*Anas platyrhynchos*). Very low rates of successful hybridisation were seen among this set of birds, although a small selection of BACs appeared to work well across all species tested. In particular, the BACs for chromosome 18 and 23 produced clear signals and in the case of the two falcons tested, the chromosome 23 BACs illustrated a fusion between a

microchromosome and a macrochromosome as shown in Figure 4-4. The full list of results for these probes tested on this set of birds is shown in Table 4-3.



**Figure 4-4: Chicken microchromosome probes (GGA23p - CH261-191G17 and GGA23q-arm - CH261-90K11) revealing a fusion to a macrochromosome in the peregrine falcon (*Falco peregrinus*- FPE). Scale bar 10 $\mu$ m.**

BAC Clone Name	GGA Chr	Houbara	Duck	Peregrine	Zebra Finch	Pigeon
CH261-118E15	10	No	No	√	No	No
CH261-94C12	10	No	No	No	No	No
CH261-28F3	11	No	No	No	No	No
CH261-89P18	11	No	No	No	No	No
CH261-152H14	12	No	No	No	No	No
CH261-88K1	12	No	No	No	No	No
CH261-58H5	13	No	No	No	No	No
CH261-71N1	13	No	No	No	No	No
CH261-121M10	14	No	No	No	No	No
CH261-17L7	14	No	No	No	No	No
CH261-131E4	15	No	No	No	No	No
CH261-40D6	15	No	√	√	No	√
CH261-97F21	16	No	No	No	No	No
CH261-96L12	16	No	No	No	No	No
CH261-69M11	17	No	No	√	No	No
CH261-72P11	17	No	No	No	No	No
CH261-67N15	18	√	No	√	√	No
CH261-72B18	18	√	No	√	√	√
CH261-167A1	19	√	No	No	No	No
CH261-189E4	19	√	No	No	No	No
CH261-10L6	20	No	No	No	No	No
CH261-124A24	20	No	No	No	No	No
CH261-10A18	21	No	No	√	No	No
CH261-49L18	21	No	No	No	No	No
CH261-30D24	22	No	No	No	No	No
CH261-49B2	22	No	No	No	No	No
CH261-191G17	23	√	√	√	√	√
CH261-90K11	23	√	√	√	√	√
CH261-154H17	24	No	No	√	No	No
CH261-154L15	24	No	No	No	No	No
CH261-169N16	25	No	No	No	No	No
CH261-82G24	25	No	No	No	No	No
CH261-40C14	26	No	No	No	No	No
CH261-50J5	26	No	No	No	No	No
CH261-100E5	27	No	√	√	No	No
CH261-167J20	27	No	No	No	No	No
CH261-101C8	28	No	No	√	No	√
CH261-16I3	28	No	No	No	No	No

**Table 4-3: Cross species hybridisation results for GGA subtelomeric microchromosome BACs on avian species tested beyond the *Galliformes* order.**

#### 4.4.3 Specific Aim 2c: To develop a means of improving the success of cross species FISH using BAC clones

In order to increase the cross species hybridisation success rate in birds, data generated from previous cross-species experimental work was used to isolate key genomic features that may contribute to the likely success of a labelled BAC working cross-species. Using the subtelomeric chicken BACs developed in section 4.4.1 and an additional set developed under the same criteria but with zebra finch derived BACs (see appendix), a group of 103 BAC clones (53 chicken and 50 zebra finch) selected purely based on their physical location in the reference genome were tested on a panel of five diverse avian species: chicken (*Gallus gallus*), turkey (*Meleagris gallopavo*), pigeon (*Columba livia*), peregrine falcon (*Falco peregrinus*) and zebra finch (*Taeniopygia guttata*). Of these, the success rate varied considerably from only 21% (chicken BACs hybridised to zebra finch chromosomes), to 92% (zebra finch BACs hybridised to peregrine falcon chromosomes). Based on these results a statistical approach was used by collaborators in the RVC lab (outlined in section 4.3.1.2) to identify BACs with a high likelihood of hybridising on multiple avian species. The additional 123 BACs (chosen from zebra finch and chicken BAC libraries) tested using this approach produced significantly higher success rates ranging from 71% to 100% as listed in Table 4-4 and Table 4-5.

In total, 226 BACs were tested on chicken, turkey, pigeon, peregrine falcon and zebra finch. Results for the chicken derived BAC clones are shown in Table 4-4.

	Chicken BAC Clones					
	Non-selected		Selected		Total	
	Number of signals present	% Success	Number of signals present	% Success	Total	% Success
<b>Chicken</b>	53	100%	99	100%	152	100%
<b>Turkey</b>	47	89%	99	100%	146	96%
<b>Pigeon</b>	14	26%	91	92%	105	69%
<b>Peregrine falcon</b>	25	47%	93	94%	118	78%
<b>Zebra finch</b>	11	21%	90	91%	101	66%
<b>Number of BACs</b>	53		99		152	

**Table 4-4: Successful hybridisation results across 5 avian species using selected and non-selected BACs derived from the CHORI chicken BAC library.**

As described, the set of BACs was extended to include those from the Zebra finch BAC library (TGMCBa) to give greater genome coverage, results for the zebra finch derived BACs are shown in Table 4-5.

	Zebra Finch BAC Clones					
	Non-selected		Selected		Total	
	Number of signals present	% Success	Number of signals present	% Success	Total	% Success
Chicken	29	58%	18	75%	47	64%
Turkey	27	54%	20	83%	47	64%
Pigeon	34	68%	17	71%	51	69%
Peregrine falcon	46	92%	22	92%	68	92%
Zebra finch	50	100%	24	100%	74	100%
Number of BACs	50		24		74	

Table 4-5: Successful hybridisation results across 5 avian species using selected and non-selected BACs derived from the TGMCBa Zebra finch BAC library

Overall success rates using the selected BACs were significantly higher than those that were selected by physical position only, as illustrated in Table 4-6. Results using the selected BACs were also more consistent between species, with an average of 93% of the BACs producing a signal from a range between 88 and 97%. The non-selected BACs had higher success rates when used on closely related species but results were more significantly affected by phylogenetic distance between species with rates reducing to an average of 65% but across a range of 47 to 80%.

	Selected BACs		Non-Selected	
	Number of signals present	% Success	Number of signals present	% Success
Chicken	117	95%	82	80%
Turkey	119	97%	74	72%
Pigeon	108	88%	48	47%
Peregrine falcon	115	93%	71	69%
Zebra finch	114	93%	61	59%
Total no. of BACs	123		103	

Table 4-6: Comparison of overall FISH success rates for selected versus non-selected BACs across 5 avian species.

Of the selected BACs, 94 of the 123 (76%) produced a clear signal on all 5 species and 112 (91%) worked on 4 of the 5 species compared to 28 of the 103 (27%) non-selected BACs working on all 5 species and 48 of the 103 (47%) non-selected BACs working on 4 of the 5 species.

For BACs to successfully hybridise across a wide phylogenetic distance (>73 my), the criteria identified to be important were a) the ability to align  $\geq 93\%$  of the BAC DNA sequence to other avian genomes and b) the BAC needed to either contain at least one conserved element (CE)  $\geq 300$  bp or if not long CE was present, the BAC had to contain only short repetitive elements (<1290 bp) and CEs of at least 3 bp long.

#### 4.4.4 Specific Aim 2d: To adapt the system described in specific aim 2a using technology developed in specific aim 2c so that it can be applied to all bird karyotypes

The full set of BACs analysed and tested in section 4.4.3 were edited down to those that hybridised successfully across all species tested. From this list, two BACs were chosen for each reference (chicken) chromosome at either end in the most distal regions (where possible). Those BACs with the highest overall conservation score according to the selection criteria described in section 4.3.1.2 were also preferentially chosen. This edited set of BACs, listed in Table 4-8 was subsequently tested on a range of 18 different species, listed below in Table 4-7:

Order	Common name	Species Name
<i>Anseriformes</i>	Duck	<i>Anas platyrhynchos</i>
<i>Charadriiformes</i>	Eurasian Woodcock	<i>Scolopax rusticola</i>
<i>Columbiformes</i>	Pigeon	<i>Columba livia</i>
<i>Columbiformes</i>	Collared Dove	<i>Streptopelia decaocto</i>
<i>Falconiformes</i>	Peregrine falcon	<i>Falco peregrinus</i>
<i>Falconiformes</i>	Saker falcon	<i>Falco cherrug</i>
<i>Galliformes</i>	Turkey	<i>Meleagris gallopavo</i>
<i>Galliformes</i>	Japanese Quail	<i>Coturnix japonica</i>
<i>Galliformes</i>	Guinea Fowl	<i>Numidea meleagris</i>
<i>Otidiformes</i>	Houbara	<i>Chlamydotis undulata</i>
<i>Passeriformes</i>	Blackbird	<i>Turdus merula</i>
<i>Passeriformes</i>	Canary	<i>Serinus canaria</i>
<i>Psittaciformes</i>	Budgie	<i>Melopsittacus undulatus</i>
<i>Psittaciformes</i>	Cockatiel	<i>Nymphus hollandicus</i>
<i>Psittaciformes</i>	Kakariki	<i>Cyanoramphus novaezelandia</i>
<i>Strigiformes</i>	Pharaoh eagle-owl	<i>Bubo ascalaphus</i>
<i>Struthioniformes</i>	Ostrich	<i>Struthio camelus</i>

Table 4-7: List of avian species tested using two selected BACs per reference (chicken) microchromosome.

Clear punctate signals were achieved for all microchromosome BACs for each species with the exception of the two BACs for chicken chromosome 25 which when tested on the blackbird and the canary, did not produce a signal. Clear signals were achieved for all macrochromosome BACs with a few individual species specific exceptions.

Clone Name	GGA Chr	Mean Score	BAC Size
CH261-89C18	1	0.08	171,359
CH261-98G4		0.1	182,677
CH261-169N6	2	0.24	195,579
CH261-44D16		0.36	205,011
TGMCBA-295P5	3	0.19	116,726
CH261-169K18		0.24	140,443
CH261-83E1	4	0.17	171,741
CH261-89P6		0.28	141,721
CH261-49B22	5	0.28	193,278
CH261-78F13		0.15	186,801
TGMCBA-382J4	6	0.14	125,515
CH261-49F3		0.29	112,713
CH261-56K7	7	0.2	176,360
CH261-180H18		0.33	232,888
CH261-107D8	8	0.16	225,011
TGMCBA-252A4		0.33	154,862
CH261-183N19	9	0.11	173,995
CH261-187M16		0.16	188,641
CH261-115G24	10	0.13	221,139
CH261-71G18		0.2	218,444
CH261-121N21	11	0.37	248,767
CH261-154H1		0.14	214,730
CH261-60P3	12	0.27	142,783
CH261-4M5		0.13	192,350
CH261-115I12	13	0.17	193,411
TGMCBA-321B13		0.21	149,734
CH261-122H14	14	0.27	198,117
CH261-69D20		0.15	187,173

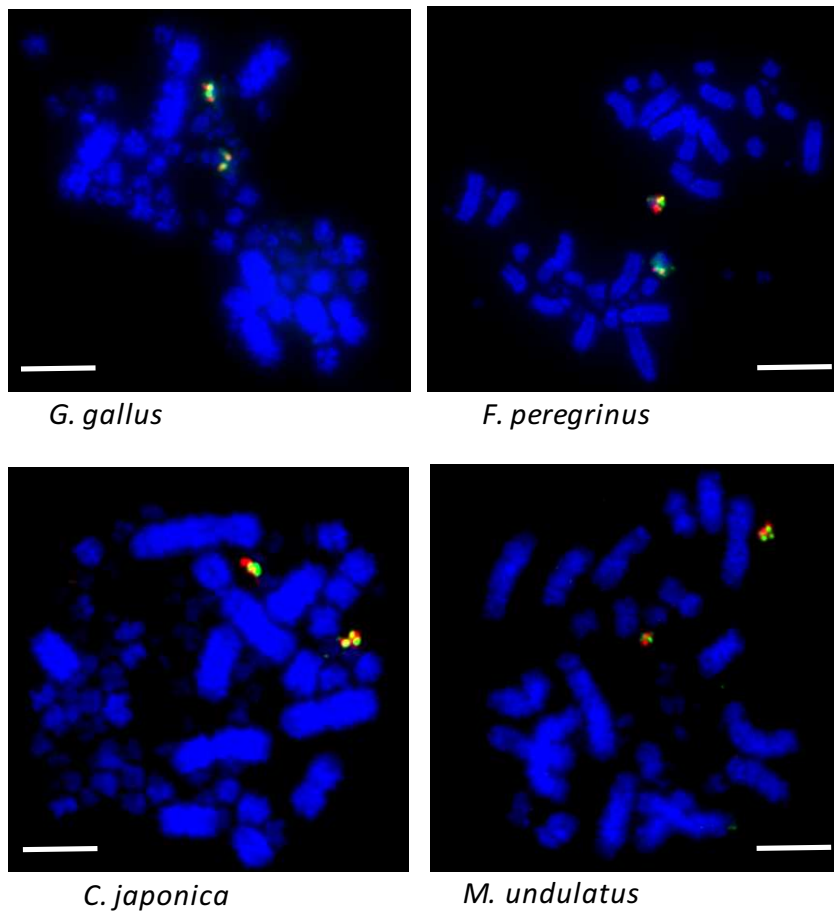
Clone Name	GGA Chr	Mean Score	BAC Size
CH261-90P23	15	0.31	185,386
TGMCBA-266G23		0.15	154,862
TGMCBA-375I5	17	0.21	129,289
CH261-42P16		0.28	171,737
CH261-60N6	18	0.17	232,938
CH261-72B18		0.19	172,245
CH261-10F1	19	0.33	140,529
CH261-50H12		0.23	151,852
TGMCBA-250E3	20	0.1	148,859
TGMCBA-341F20		0.22	127,681
CH261-83I20	21	0.11	194,106
CH261-122K8		0.38	181,181
CH261-40J9	22	0.27	177,235
CH261-18G17		0.15	214,459
CH261-191G17	23	0.16	217,609
CH261-90K11		0.19	162,094
CH261-103F4	24	0.17	151,952
CH261-65O4		0.21	151,636
CH261-59C21	25	0.06	155,051
CH261-127K7		0.07	124,418
CH261-186M13	26	0.16	176,643
CH261-170L23		0.18	194,856
CH261-66M16	27	0.14	177,166
CH261-28L10		0.16	216,399
CH261-64A15	28	0.14	169,997
CH261-72A10		0.22	210,049
CH261-129A16	Z	0.21	225,875
CH261-133M4		0.23	192,220

**Table 4-8: BACs chosen for multiple cross species analysis. Two BACs were chosen for each chromosome using the chicken genome as the reference.**



**4.4.5 Specific Aim 2e: To test the hypothesis that microchromosomal rearrangement is a rare event, compared to other vertebrate species, during avian genome evolution**

Using the set of cross species BACs defined in section 4.4.3 specifically to investigate the microchromosomes permitted analysis at a higher resolution than previously achieved. Two BACs were selected from each of the sequenced chicken microchromosomes (from GGA10 to GGA28, with the exception of GGA16, listed in Table 4-8) and dual FISH was performed on a total of 18 avian species. In all species tested, regions homologous to chicken chromosomes 20, 22, 24, 25, 26, 27 appear to have remained intact as entire microchromosomes with no evidence of chromosomal fusion. Figure 4-5 shows representative images for chicken chromosome 24 tested on multiple species with the BACs illustrating that this chromosome appears to remain intact as a microchromosome in all species tested with no sign of interchromosomal rearrangement.



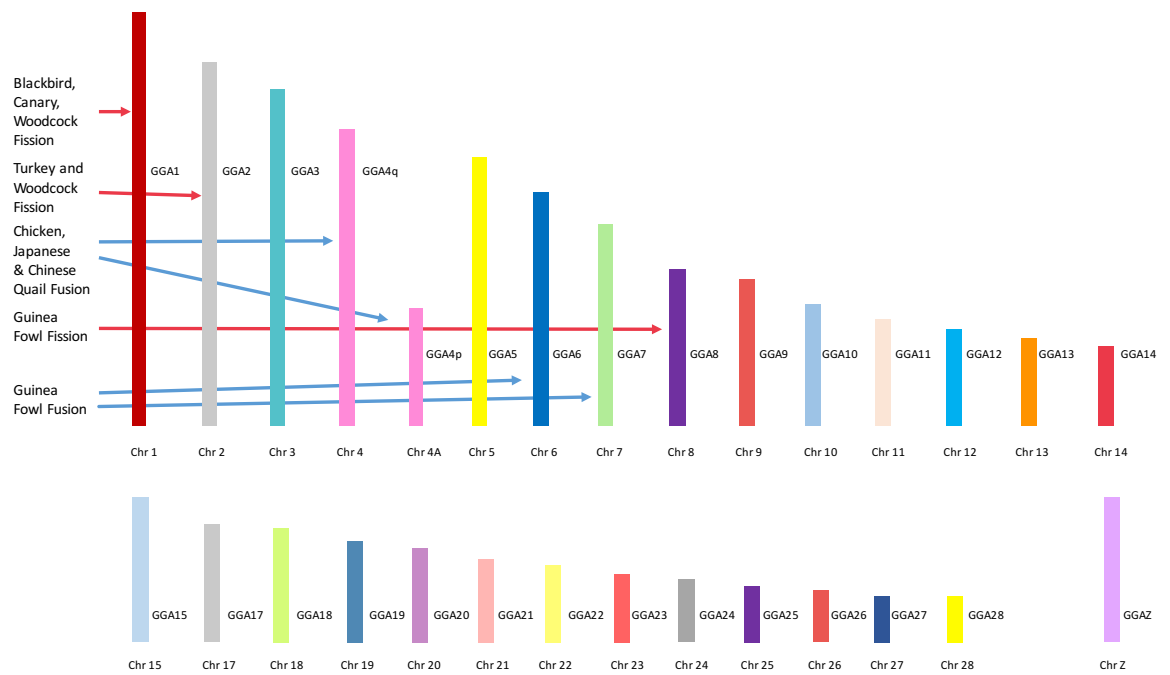
**Figure 4-5: Microchromosomal conservation observed across a wide range of avian species as revealed by testing BACs from chicken chromosome 24 (CH261-103F4 FITC and CH261-65O4 Texas Red). Scale bar 10µm.**

Results from the species within each order tested are outlined in the sections below:

**4.4.5.1 Species with no interchromosomal rearrangement between microchromosomes**

No apparent changes from the ancestral microchromosomes appear to have occurred among the representatives tested here from the following orders: *Galliformes*, *Anseriformes*, *Charadriiformes*, *Columbiformes*, *Otidiformes*, *Passeriformes*, *Strigiformes* and the *Struthioniformes*. The microchromosomes of each bird remain conserved in the same pattern exhibited in the chicken with BACs hybridising consistently together across all species tested despite there being over 100 million years since they diverged. No apparent interchromosomal rearrangement was therefore evident.

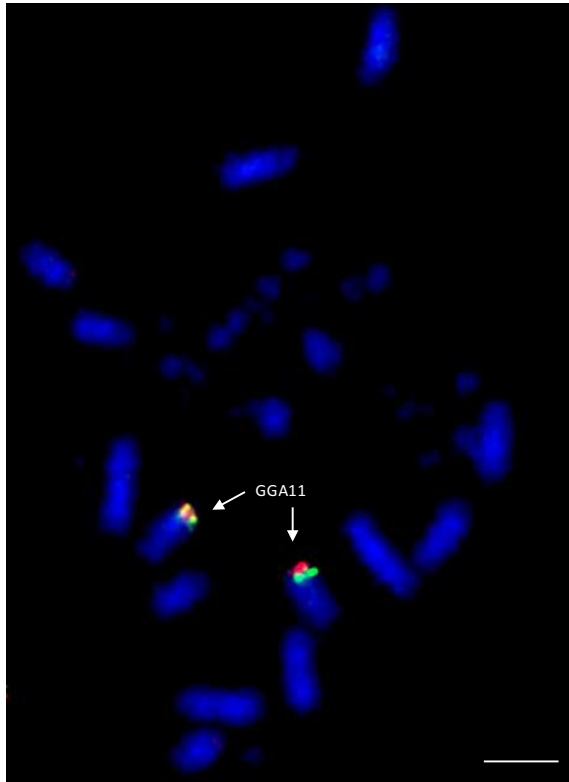
A representative ideogram for the whole karyotype (including macrochromosomes) is illustrated in Figure 4-6 for 12 of the 18 species tested here (duck, pigeon, ostrich, turkey, houbara, owl, collared dove, Eurasian woodcock, Japanese quail, Chinese quail, guinea fowl, blackbird and canary) illustrating microchromosomal conservation across all 12 species. Also indicated are the macrochromosomal rearrangements that have been previously published through chromosome painting studies (Shibushawa et al. 2002; Shibushawa et al. 2004; Guttenbach et al. 2003; Griffin et al. 2008; Nishida et al. 2007) and confirmed here. The only results that deviated from previous studies was an additional fission observed in the chicken homolog of chromosome 8 in the guinea fowl and two previously unreported fissions evident in the Eurasian woodcock.



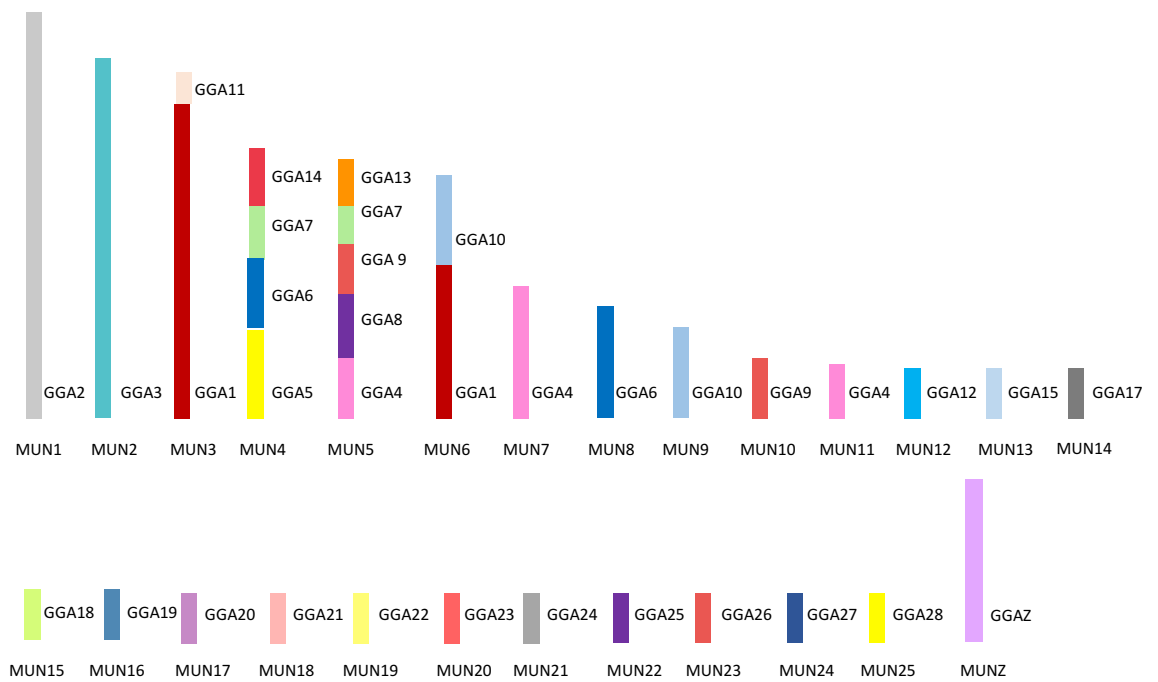
**Figure 4-6: Ideogram for 12 of the 18 species tested here (duck, pigeon, ostrich, turkey, houbara, pharaoh eagle owl, collared dove, Eurasian woodcock, Japanese quail, Chinese quail, guinea fowl, blackbird and canary), illustrating microchromosomal conservation across all 12 species plus representation of the macrochromosomal rearrangements previously published and confirmed here (indicated with a red arrow for a fission and blue for a fusion).**

#### 4.4.5.2 *Psittaciformes*

Among the *Psittaciformes*, the only lineage specific rearrangement that does not appear in the any of the other avian species tested involves the homolog of chicken chromosome 11. Overall, interchromosomal rearrangements were detected for the homologs for GGA10, 11 and 14. All three of the species tested: the kakariki (*Cyanoramphus novaezelandia*), the cockatiel (*Nymphus hollandicus*) and the budgerigar (*Melopsittacus undulatus*) exhibit a fusion of each of these homologs to macrochromosomes, as illustrated in Figure 4-7 where the homolog of GGA11 is shown fused to a macrochromosome in the cockatiel. In addition, there is a fusion of the GGA13 homolog seen in the budgerigar but not the other parrots tested here. The overall karyotypic rearrangement for the budgerigar is shown in the ideogram illustrated in Figure 4-8. Macrochromosomal rearrangements are based on those previously established through chromosome painting studies by Nanda and colleagues (2007) and confirmed here.

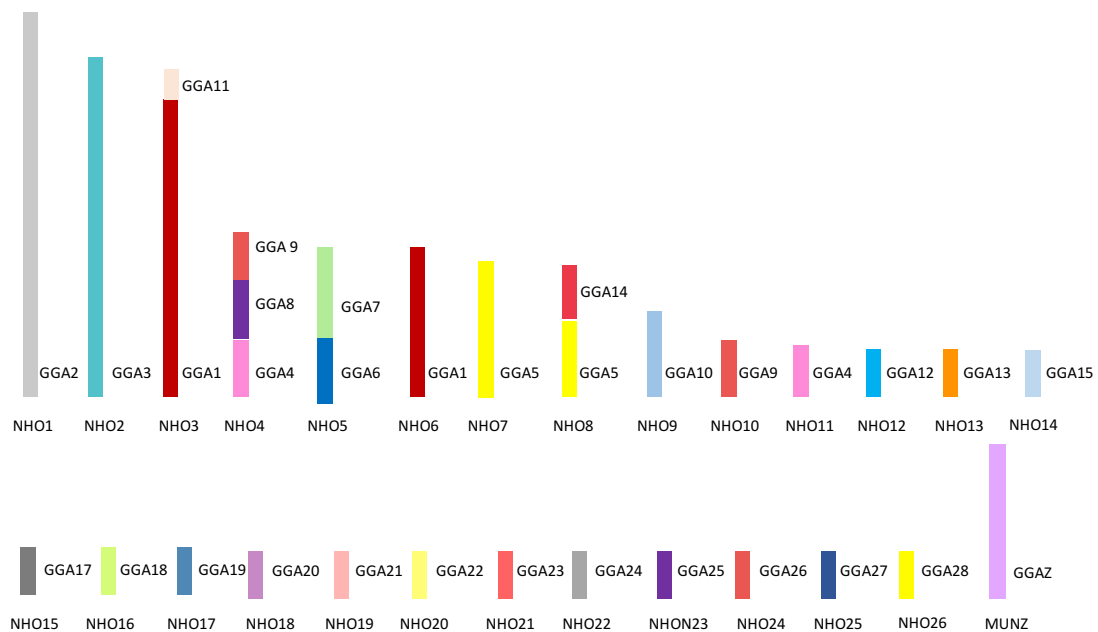


**Figure 4-7:** Hybridisation of GGA11 BACs (CH261-121N21-Fitc and CH261-154H1-Texas red) to cockatiel (*Nymphus hollandicus*) metaphases illustrating fusion of ancestral microchromosome to a macrochromosome. Scale bar 10µm.



**Figure 4-8:** Ideogram representing the budgerigar (*Melopsittacus undulatus* -MUN) karyotype, illustrating a high degree of interchromosomal rearrangement in the macrochromosomes and a high degree of microchromosomal conservation with the exception of 3 microchromosome homologs (GGA10, 11, 13 and 14).

Figure 4-9 shows the overall karyotypic structure of the cockatiel and illustrates that despite broadly similar patterns of rearrangement there are fewer rearrangements that have occurred between the macrochromosomes when compared to the budgerigar. The kakariki karyotype appears most similar to the budgerigar but requires further mapping to confirm the overall structure as there are no previously published studies for this species.

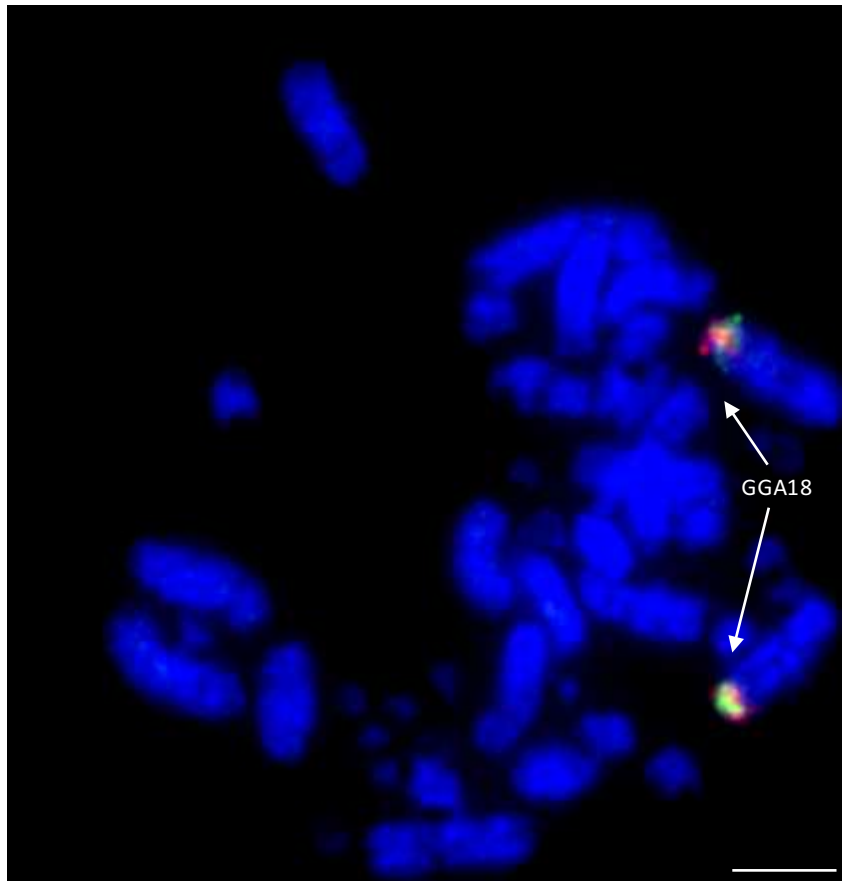


**Figure 4-9: Ideogram representing karyotypic structure of the cockatiel (*Nymphus hollandicus* - NHO) illustrating an overall structure that is broadly in alignment with the budgerigar but with one less microchromosome fusion and fewer interchromosomal rearrangements in the macrochromosomes.**

#### 4.4.5.3 Falconiformes

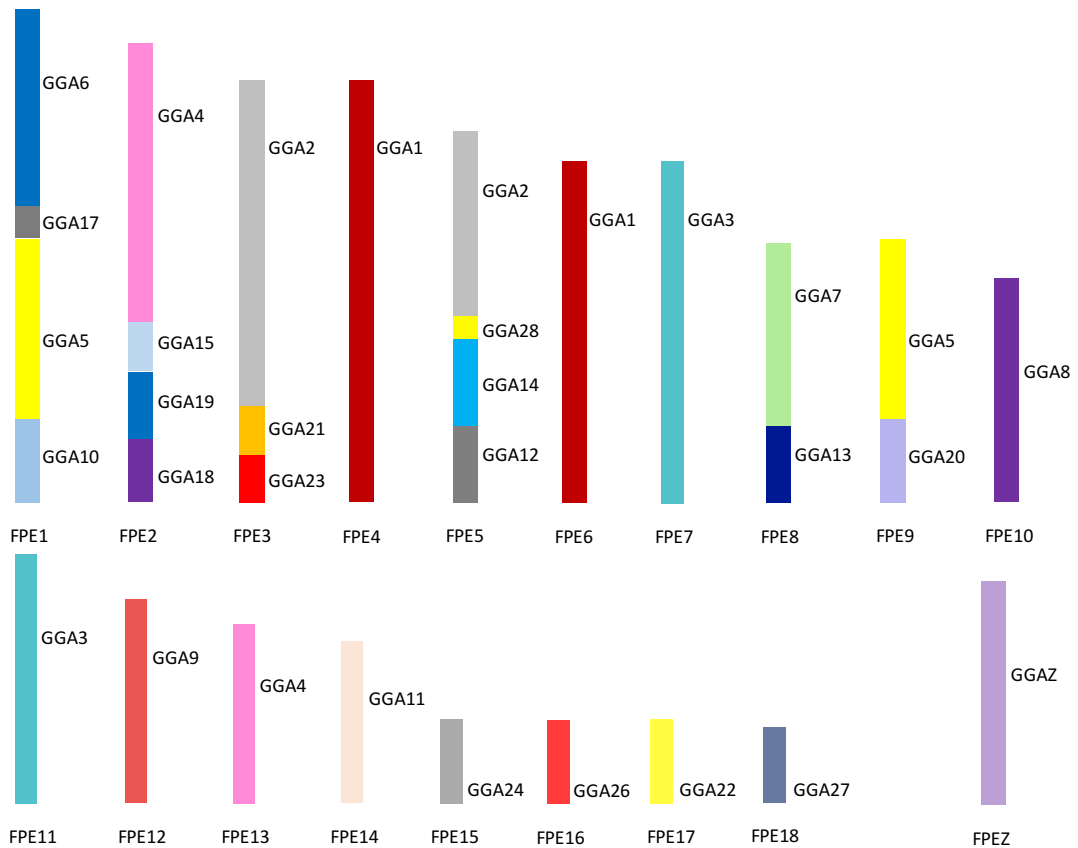
Among the *Falconiformes* extensive rearrangement appears to have taken place with regions homologous to GGA microchromosomes 10, 12, 13, 14, 15, 17, 18, 19, 21, 23 and 28 fused to GGA macrochromosome regions. An example of which is illustrated in Figure 4-10 where GGA18 homologs are fused to a macrochromosome in the peregrine falcon. Lineage specific rearrangements were apparent with no evidence of chicken chromosome homologs 15, 18, 19, 23 and 28 being rearranged in any of the other (non-falcon related) species tested. Interestingly, 15, 18 and 19 appear to have fused together as one chromosome (to chicken homolog 4) in all falcon species tested while 23 and 28 have both fused to the homolog of chicken chromosome 2 which at some point (either prior to or post fusion) has split in to two chromosomes. Both of the falcon species tested (peregrine and saker) appear to exhibit the same pattern of rearrangement (with the exception of peregrine chromosome 1 for which there is a centric

fusion) suggesting that any lineage specific rearrangements rapidly became fixed in the population with little interchromosomal rearrangement since. In addition, there appears to be no interchromosomal rearrangement between each pair of BACs tested suggesting that these regions of DNA are highly conserved and not prone to breakage.



**Figure 4-10: Hybridisation of GGA18 BACs (CH261-60N6-Fitc and CH261-72B18-Texas red) to peregrine falcon (*Falco peregrinus*) metaphases illustrating fusion of ancestral microchromosome to a macrochromosome. Scale bar 10µm.**

Figure 4-11 illustrates the overall karyotype structure of the peregrine falcon tested showing extensive interchromosomal rearrangement between the micro and the macrochromosomes. The karyotype of the saker falcon (also tested here) follows the same pattern with the exception of peregrine chromosome 1. In the saker falcon this chromosome has split into two chromosomes with the breakpoint occurring within the region of the GGA5 homolog. This suggest that this is a fission that has occurred after the falcon karyotype was formed rather than a later peregrine specific fusion. The peregrine falcon genome is covered in more depth in chapter 6.



**Figure 4-11: Ideogram representing overall karyotypic structure of the peregrine falcon (*Falco peregrinus* - FPE) illustrating an extensive amount of interchromosomal rearrangement throughout the karyotype.**

## 4.5 Discussion

### 4.5.1 Characterisation of all sequenced avian chromosomes in one experiment

As described in section 1.1.2.3, most of the avian comparative genomics studies performed at a chromosomal level have been limited to investigating the macrochromosomes (summarised in section 1.1.2.3, Table 1-1). Little success has been achieved with analysis of the microchromosomes because of the technical limitations in generating chromosome paints. The development of BAC libraries as a product of genome sequencing has however, enabled the approach developed here wherein BACs have been selected that successfully hybridise to chicken microchromosomes. Therefore, the analysis of the (sequenced) microchromosomes combined with the well characterised macrochromosomal chicken paints allows examination of the entire chicken genome. In addition, the incorporation of these paints and BACs into one device facilitates rapid detection of these probes in one experiment, thereby reducing set up and microscopy time.

#### 4.5.2 Cross-species FISH Improvements

Beyond the chicken genome, the results described in section 4.4.2 have shown that among the closely related *Galliformes* a high degree of success is achievable using the tools developed in section 1.4.1. The results generated using this tool confirmed previous findings by Shibusawa et al. (2004), Guttenbach et al. (2003) and Griffin et al. (2008) of the pattern of macrochromosomal rearrangements seen in the species tested. The high degree of success in hybridising chicken BACs to other *Galliformes* species can largely be attributed to the relatively recent divergence times of approximately 35-40 million years between these birds as evidenced by the reduced success rate when testing these probes on species that are phylogenetically further apart. Chromosome paints have been widely tested with a great deal of success on more distantly related species (see Table 1-1), however success using BACs has been limited in comparison with success rates rarely achieving higher than 40% of BACs producing a signal (Skinner et al. 2009). Use of the device developed in section 4.4.1 on more phylogenetically distant birds, quickly revealed a markedly reduced level of success as shown in Table 4-3 suggesting that another approach was required. Interestingly, some of these BACs, for example those that relate to chicken chromosome 23 (CH261-90K11 and CH261-191G17), were found to work consistently across multiple species suggesting that elements unique to these BACs improved the likelihood of successful hybridisation.

Identification of the genomic features unique to these high success BACs (as described in section 4.3.1.2) using a bioinformatics approach (in collaboration with the RVC lab) has led to a refinement in the methods used to select BACs designed to hybridise across multiple species. This has resulted in an improvement in hybridisation rates by several orders of magnitude. Extending this set of testable BACs to a pool around 230 BACs, all of which are tested on a core of five species of phylogenetically distant birds has resulted in the development of accurate prediction models as to the likelihood of success with each BAC on a multitude of species. In addition, generating and testing such a large number of BACs has resulted in there being at least two 'cross-species' BACs per reference (chicken) chromosome available along with providing an even level of coverage across the entire reference genome. Whilst the emphasis in this study was analysis of the microchromosomes, the set generated here includes multiple BACs for each of the macrochromosomes. For example, there are 20 BACs that have worked well on all four species for chicken chromosome 1 and an average of 9 BACs for each of chromosomes 2 to 9. The majority of these BACs are also evenly spaced along the reference chromosome therefore creating the possibility of future fine scale macrochromosome mapping across multiple species,



potentially using multiple colours as described in section 1.1.2.6, allowing for the investigation of intrachromosomal rearrangements at a resolution previously not possible.

Another benefit of generating a large set of BACs in this manner is that it has enabled the development of a refined, edited 'panel' of FISH probes that work across multiple species. Selection of the BACs used for this panel was based firstly on successful hybridisation across all five of the core species and secondly on their position in the most distal regions of each chromosome in the reference genome. This then provides a consistent anchor point from which to compare species in order to track chromosomal rearrangements over time. As this study has shown, rather than being limited to comparing multi-species chromosomal rearrangements within a specific order, this approach allows for comparison across an entire class (and to some degree even beyond this – this is discussed further in chapter 6).

#### **4.5.3 Interchromosomal rearrangement during avian evolution**

Avian cytogenetic studies to date suggest that avian genomes remain remarkably conserved in terms of chromosome number and that interchromosomal changes are relatively rare. When they occur, they tend to recur with several examples of homoplasmy (Griffin et al. 2007; Skinner & Griffin 2012) such as that of chromosome four as described in section 1.5.6. Cross-species FISH using chromosome paints has been extremely useful for characterising large scale interchromosomal rearrangements such as those seen in the *Falconiformes* and *Psittaciformes* (Nanda et al. 2007; Nishida et al. 2008), however without the ability to identify individual microchromosomes reliably, significant regions of these rearranged chromosomes are left without assignment. In this study, the use of microchromosomal BACs unique to each chromosome has improved the resolution of these changes dramatically. The individual regions of microchromosome homology that have fused to other chromosomes can now be reliably identified, filling in some of these otherwise intractable gaps. Analysis of the *Falconiformes* for example reveals that the majority of microchromosomes have fused to macrochromosomes with some chromosomes exhibiting as many as four chromosomes fusing to create one macrochromosome. Initial work using the BACs selected in section 4.4.1 revealed a series of microchromosome fusions to macrochromosomes in the gyrfalcon and budgerigar. Using a set of microchromosomal pooled paints (that were assigned to microchromosomes using BACs) we (Lithgow, O'Connor R.E. et al. 2014) were able to infer that the homologs of chicken microchromosomes 11 and 14 were fused to macrochromosomes in the budgerigar and that the chicken homolog of microchromosome 14 and one of the microchromosomes 16, 17 or 18 were

fused to macrochromosomes in the gyrfalcon (Lithgow, O'Connor R.E. et al. 2014). The results described in this chapter are a significant step on from these rudimentary assignments.

#### 4.5.4 Microchromosome Conservation

Microchromosomal rearrangement has long been considered to occur rarely compared to other chromosomal rearrangements in birds. These highly gene-dense chromosomes are thought to have changed very little throughout the last 100 million years of avian evolution (Ellegren 2013), with a high degree of conservation potentially dating back even further to the ancestral vertebrate 400mya (Burt 2002; Nakatani et al. 2007). Prior to the work presented in this thesis however, there has been little cytogenetic evidence to support this degree of conservation. What evidence is available, has been focused largely on closely related and karyologically similar species such as the chicken and the duck (Fillon et al. 2007; Skinner et al. 2009). The tools generated in this study have enabled analysis across avian representatives from ten different orders, all of which share a common ancestor over 100 mya. These results clearly illustrate an extraordinary level of microchromosome conservation with 2 out of the 10 orders exhibiting no change from the microchromosomal pattern exhibited in the chicken. From a microchromosomal point of view, these results support the hypothesis proposed in our previous study (Romanov et al. 2014) that the avian ancestor most closely resembled the chicken. Of the avian species that did exhibit microchromosomal rearrangements, the two representatives of the *Falconiformes* tested here (the saker and peregrine falcon) share the same pattern of fusion from which it can be inferred that early on in the evolution of this order, there was a large degree of rearrangement that became fixed in the population. It may therefore be that there is some biological advantage to this karyotypic structure for these birds, perhaps due to the high metabolic demands required by birds of prey. Of the other highly rearranged order, the *Psittaciformes*, it would appear that the microchromosomal fusions exhibited in each of the species tested here are not consistent with each other, suggesting that karyotypic evolution has continued from their common ancestor and that there are species specific rearrangements between the birds.

In all of these cases however, it appears that there is a bias towards the microchromosomes remaining as discrete units even when fused into highly complex karyotypes such as those of the falcons and the parrots. Interestingly, this same pattern is also evident in the chicken, where the p-arm of chromosome four is a microchromosome in the majority of other species, therefore suggesting that it is an ancestral microchromosome. In the chicken however, despite fusing to a

macrochromosome, it remains intact, even retaining all of its uniquely microchromosomal sequence characteristics such as high GC and gene content (ICGSC et al. 2004).

Even taking into account these lineage specific rearrangements, there appear to be four microchromosomes that across all birds tested, remain conserved as microchromosomes with no signs of apparent fusion. In the chicken, these are four of the smallest sequenced chromosomes with sizes ranging from 4 Mb to 6 Mb. Further sequence analysis may reveal signature features of these chromosomes that may indicate a biological reason as to why these chromosomes are left intact. If there is any correlation with the size of the chromosomes and their lack of rearrangement, then this would suggest that the very smallest 'D-group' chicken microchromosomes are also less prone to chromosomal fusion. In fact, when the two rearranged lineages are removed from this study, all other species analysed exhibit the same pattern of conserved microchromosomal arrangement, which given that the parrots and the falcons are karyotypically the exception rather than the rule among avian species, illustrates quite how profound this level of genome conservation really is.

## **4.6 Conclusions**

There are fundamental biological questions relating to the significance of genomic regions prone to chromosomal breakage (EBRs) and regions of conserved synteny (HSBs). Why, how often and what advantage these rearrangements (or lack of rearrangements) confer are crucial to understanding patterns of evolution and speciation. Given the extraordinary phenotypic diversity seen in birds and the highly conserved nature of their genomes, the tools developed in this chapter have started the process of answering some of these questions. The next obvious step in using this approach would be to extend the diversity of the avian species tested using these tools in order to assess whether these microchromosomal patterns are consistent across all orders. In addition, preliminary work presented in chapter 6 where these BACs have been used to identify chromosomal conservation and rearrangement in reptilian species illustrates that these tools can also be used in other, even more phylogenetically distant species. Finally, once sequences have been assigned to the smallest D-group chromosomes, a similar BAC based approach can be employed to establish whether these highly elusive microchromosomes also exhibit similar patterns to those identified in this chapter.

**5 Specific aim 3: To use the technology developed in specific aims 1 and 2 in conjunction with a bioinformatics approach developed by colleagues at the Royal Veterinary College, London to complete the cytogenetic mapping of scaffold based genome assemblies to full chromosomal level in two key (but karyotypically dissimilar) avian species (peregrine falcon and pigeon)**

**5.1 Background**

The ability to sequence *de novo* genomes cheaply and efficiently has led to multiple new genome projects including those that have gone beyond species that are medical models or of agricultural relevance (e.g. Hu et al. 2009; Groenen et al. 2012). Attention is now turning to the sequencing and assembly of numerous animals representing most families and orders (Zhang et al. 2014a; Koepfli et al. 2015). As mentioned in section 1.1.5, the ultimate aim of any *de novo* genome sequence assembly is to produce a chromosomal length array of sequence for each chromosome, from p- to q- terminus (hereafter referred to as a 'chromosome level' assembly). As detailed in section 1.1.5.1 however, genomes sequenced using next generation sequencing (NGS) technology result in shorter reads than previously achieved by Sanger sequencing (Larkin et al. 2012). This is because a) NGS does not typically generate contigs long enough to cover chromosomes completely and in an error-free manner; and b) there are few inexpensive mapping technologies to upgrade NGS assemblies to chromosome level. Even the best assembled NGS genomes are represented as sub chromosomal size 'scaffolds' that still require physical mapping to the chromosomes. Newer technologies such as optical mapping (Teague et al. 2010) including Dovetail (Putnam et al. 2016); BioNano (Mak et al. 2016) and PacBio long read sequencing (Rhoads & Au, 2015) provide a long-term solution to this problem (see section 1.1.5.1), however these still have difficulties crossing centromeres and large heterochromatin blocks (in the case of BioNano) or require hundreds of micrograms of high molecular weight DNA often not available for endangered or smaller animals (in the case of PacBio). Synteny-based bioinformatic approaches, such as RACA (Kim et al. 2013) – see section 1.1.6.1.2 - have been developed to predict near chromosome-sized fragments (PCFs) for a *de novo* NGS genome. However, size limitations apply here too meaning that these also require further mapping in order to place them correctly on the chromosome (Kim et al. 2013; Fang et al. 2014). FISH mapping of these PCFs therefore provides a practical solution to this problem.

The extent to which evolutionary and applied genomics studies can be performed is limited without these chromosome-level assemblies. Such studies are crucial for performing genome-assisted selection and breeding of agricultural species (Andersson and Georges 2004). They have already been established for cattle, pig, sheep, chicken, turkey and duck, but are not available for equivalent species that are important in other societies (e.g. camels, yaks, buffalo, ostrich, quail) or those bred for conservation reasons (e.g. falcons). In addition, chromosome-level assembly is also crucial for addressing fundamental biological questions relating to overall karyotype evolution and reproductive isolation (Lewin et al. 2009). With all of this in mind, the purpose of this chapter was to develop a novel, integrated and inexpensive approach to upgrade existing scaffold-based genome assemblies to a chromosomal level. As a proof of principle two newly sequenced avian genomes (the pigeon and the peregrine falcon) were selected for the reasons described below.

The first species, the pigeon (*Columba livia*) has a typical avian karyotype ( $2n=80$ ) similar to those previously assembled such as chicken, turkey and zebra finch (ICGSC et al. 2004; Dalloul et al. 2010; Warren et al. 2010). An early example of avian domestication (Driscoll et al. 2009), the pigeon is one of the most phenotypically diverse species of bird. Reared for food and for sport, the number of catalogued extant breeds in Europe is thought to be around 350 (Price, 2002), all of which are descended from a single ancestor (Darwin 1868). Exhibiting a considerably wider range of phenotypic diversity than any other of the domesticated birds, variation in colour, pattern, head crest, body shape, feathers, tails, vocalisation and flight display variations make them attractive to pigeon enthusiasts many of whom are interested in selective breeding of these traits (Price 2002; Stringham et al. 2012). In fact, much of Darwin's work on natural selection was derived from studying the diversity and selective breeding seen among pigeons (Darwin, 1868). Multiple studies have been performed in order to identify the genetic basis for this variety as well as to investigate the ancestry of pigeon populations (e.g. Shapiro et al. 2013; Stringham et al. 2012). Relatively few studies, however, have focused on the overall karyotypic structure of the pigeon genome.

The second species studied, the peregrine falcon (*Falco peregrinus*), populates almost all global niches and is second only in global distribution to the pigeon - its major source of prey (Ratcliffe, 1993; Drewitt & Dixon, 2008). The ability to travel at speeds of over 200 miles per hour and its enhanced visual acuity make the peregrine falcon a formidable predator (Tucker, 1998). A prolonged period at risk of extinction due to persecution around the time of the second world

war and secondary poisoning from organochlorine pesticides (such as DDT) in the 1950s and 60s (Ferguson-Lees and Christie 2001) led to their placement on the CITES list of endangered species. Since then, the recovery of peregrine falcons from the brink of extinction is considered to be one of the most successful species conservation stories (Crick & Ratcliffe, 1995). For these reasons, many population genetic studies have been performed, with an emphasis on molecular markers (such as microsatellites) for the identification of parentage and in order to trace the genetic relationships between birds (Nesje et al. 2000). Chromosomal studies have however been relatively limited, with only one key chromosome painting study to date which revealed an atypical avian karyotype where  $2n=50$  (Nishida et al. 2008).

As prime examples for sequencing by the International Avian Phylogenomics Consortium (Shapiro et al. 2013; Zhan et al. 2013; Zhang et al. 2014a), these NGS sequenced genomes are also obvious cases for completion to a chromosomal level. The aim of this chapter was to play a significant role (specifically the molecular cytogenetic arm) in the development of a novel, inexpensive approach to upgrade the pigeon and peregrine falcon genomes to a chromosome level. The newly developed method described in this chapter takes advantage of genome alignment algorithms developed by colleagues at the Royal Veterinary College, London (RVC) from which scaffold linking outputs are verified by PCR (performed at RVC), the subsequent results of which were physically mapped directly to chromosomes (my efforts). The set of scaffold-anchoring universal BAC clones developed here are also applicable to the genomes of all other birds (see previous chapter).

## 5.2 Specific Aims

The purpose of this chapter was to apply a panel of BAC clones designed to hybridise in phylogenetically distant avian species in order to confirm placement of and link PCFs to chromosomes. As previously discussed, mapping of BACs selected solely for cytogenetic position gave way to an iterative process in which colleagues at the RVC selected BACs on the basis of sequence conservation parameters. FISH results were then fed back to the RVC to refine these parameters. With this in mind, the specific aims of this chapter were:

- Specific aim 3a: To use the approach for BAC selection outlined in chapter 4 to select a larger panel (over 200) that map to PCFs of pigeon and peregrine genomes

- Specific aim 3b: To apply the above BAC set to pigeon (*Columba livia*) chromosomes to upgrade the scaffold-based assembly to chromosome level and map the intrachromosomal differences between pigeon and the reference genome
- Specific aim 3c: To apply the above BAC set to peregrine falcon (*Falco peregrinus*) chromosomes to upgrade the scaffold-based assembly to chromosome level and map the evolutionary changes that led to the gross genomic rearrangements that occurred in this species
- Specific aim 3d: To supply raw data from mapping PCFs in both species to RVC for uploading to the Evolution Highway browser

## **5.3 Materials and methods**

### **5.3.1 Bioinformatic Approach to Upgrading Genome Assemblies**

Predicted Chromosome Fragments (PCFs) produced by the multispecies RACA alignments performed by the Larkin group at the RVC were generated as described in section 2.4. The PCFs produced by this method were used to guide the physical mapping of the pigeon and peregrine falcon genomes using FISH. In order to test the validity of the RACA algorithm, BACs identified as being located within a PCF were systematically mapped with others predicted to be in the same PCF. Once the complete set of PCFs had been confirmed, each BAC was then tested with BACs from other verified PCFs in order to assemble the gross genomic structure of the karyotype and identify orientation of the PCFs. The approaches listed in section 2.4 represent the stages that the RVC team performed in order to produce PCFs which were then mapped in our lab by FISH.

### **5.3.2 BAC Selection and FISH**

BAC clone selection and generation of labelled FISH probes was performed as described in section 2.2. FISH was performed as described in section 2.3.

## 5.4 Results

### 5.4.1 Specific Aim 3a: To use the approach for BAC selection outlined in chapter 4 to select a larger panel (over 200) that map to PCFs of pigeon and peregrine genomes

As a result of joint efforts described in the previous chapter a total of 226 BACs were selected. Of these, successful hybridisation rates were achieved with 156 BACs in the pigeon (69%) and 186 BACs in the peregrine falcon (82%) with signals produced in both species for all of the reference chromosomes.

#### 5.4.1.1 Physical Assignment of Pigeon and Peregrine Falcon PCFs to Chromosomes using FISH

To assign PCFs to their appropriate place in the genome, BAC clones from the panel described above and assigned to PCFs based on RACA results were successfully hybridised to peregrine falcon (186 clones) and pigeon (156 clones) chromosomes. The 59 PCFs anchored to the peregrine falcon chromosomes represented 1.03Gb of its genome sequence (87% of the cumulative scaffold length). Of these, 723.71Mb were oriented on the chromosomes (see Table 5-1). The pigeon chromosome assembly consisted of 0.91Gb in 60 pigeon PCFs representing 82% of the combined scaffold length. Of these 687.59Mb were oriented on chromosomes (Table 5-1).

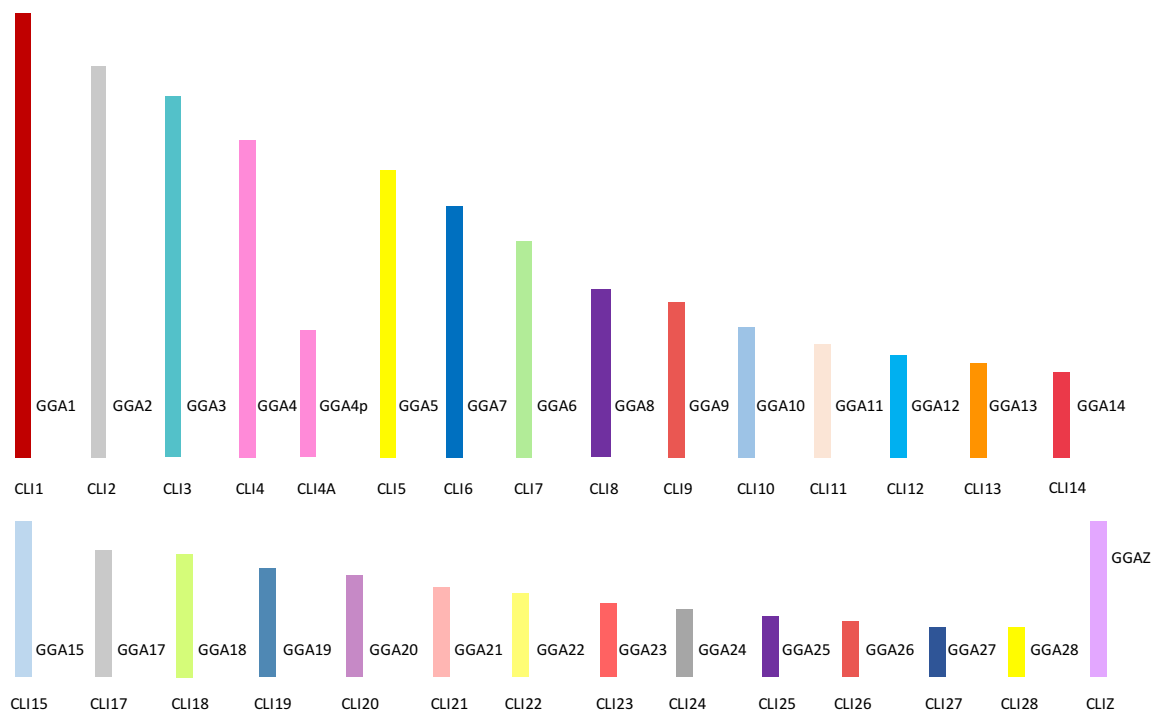
Genome Statistics	Peregrine	Pigeon
Genome size (Gb)	1.17	1.1
N50 (Mb)	3.94	3.15
Number of scaffolds	723	1081
Successfully hybridised BAC clones	186	156
Number of PCFs placed on chromosomes	59	60
Combined PCF length (Gb)	1.03	0.91
PCF assembly coverage (%)	90.03	85.23
Scaffold assembly coverage (%)	87.48	81.7
Number of oriented PCFs	26	26
Combined length (Mb)	723.71	687.59

Table 5-1: Statistics for the chromosome level assemblies of peregrine falcon (*Falco peregrinus*) and the pigeon (*Columba livia*).



**5.4.2 Specific aim 3b: To apply the above BAC set to pigeon (*Columba livia*) chromosomes to upgrade the scaffold-based assembly to chromosome level and map the intrachromosomal differences between pigeon and the reference genome**

The standard avian karyotype (2n=80) was confirmed, with each chromosome (where sequence data was available) having an appropriate chicken and zebra finch homolog as illustrated in Figure 5-1. Compared to chicken the only interchromosomal rearrangement identified was the ancestral configuration of GGA4 found as two separate chromosomes in pigeon and other birds (Derjusheva et al. 2004; Hansmann et al. 2009). Nonetheless 70 intrachromosomal rearrangements in the pigeon lineage were identified compared to chicken, 55 of which were found within scaffolds and 15 between PCFs (see Table 5-2).

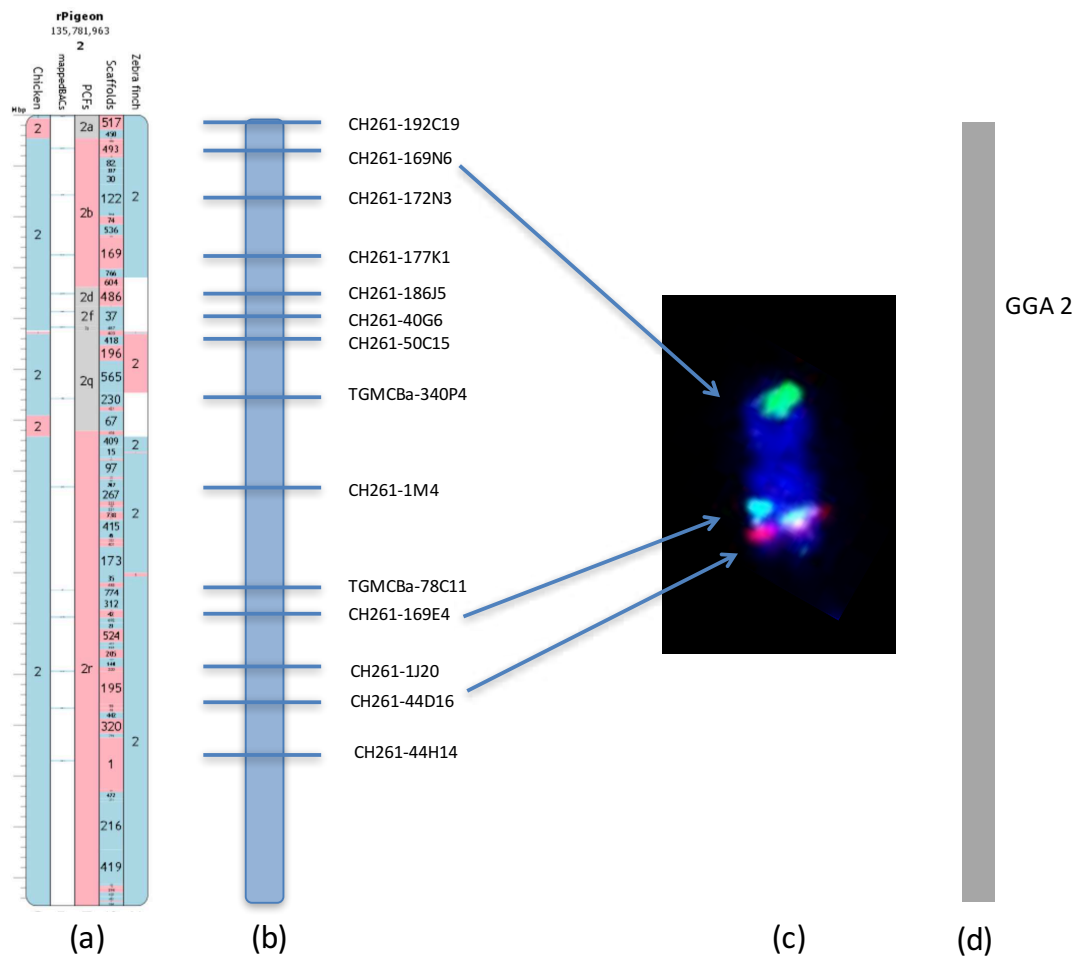


**Figure 5-1: Ideogram representation of pigeon (CLI) chromosomes and their chicken (GGA) homologs.**

Detected lineage-specific rearrangements	Pigeon
Number of chromosome fusions	0
Number of chromosome fissions	0
Number of intrachromosomal events	70
Within scaffolds	55
Between PCFs	15

**Table 5-2: Lineage specific rearrangements detected in the pigeon (*Columba livia*) when compared to the chicken genome.**

An example of PCF mapping for pigeon chromosome 2 (CLI2) is illustrated in Figure 5-2. In this example, one BAC mapping to PCF 2a (CH261-169N6 labelled in FITC) is shown co-localising with two BACs for PCF 2c (CH261-169E4 labelled in aqua and CH261-44D16 labeled in Texas red) where the orientation of PCFs relative to each other can be seen along with the orientation of BACs within PCF 2c.

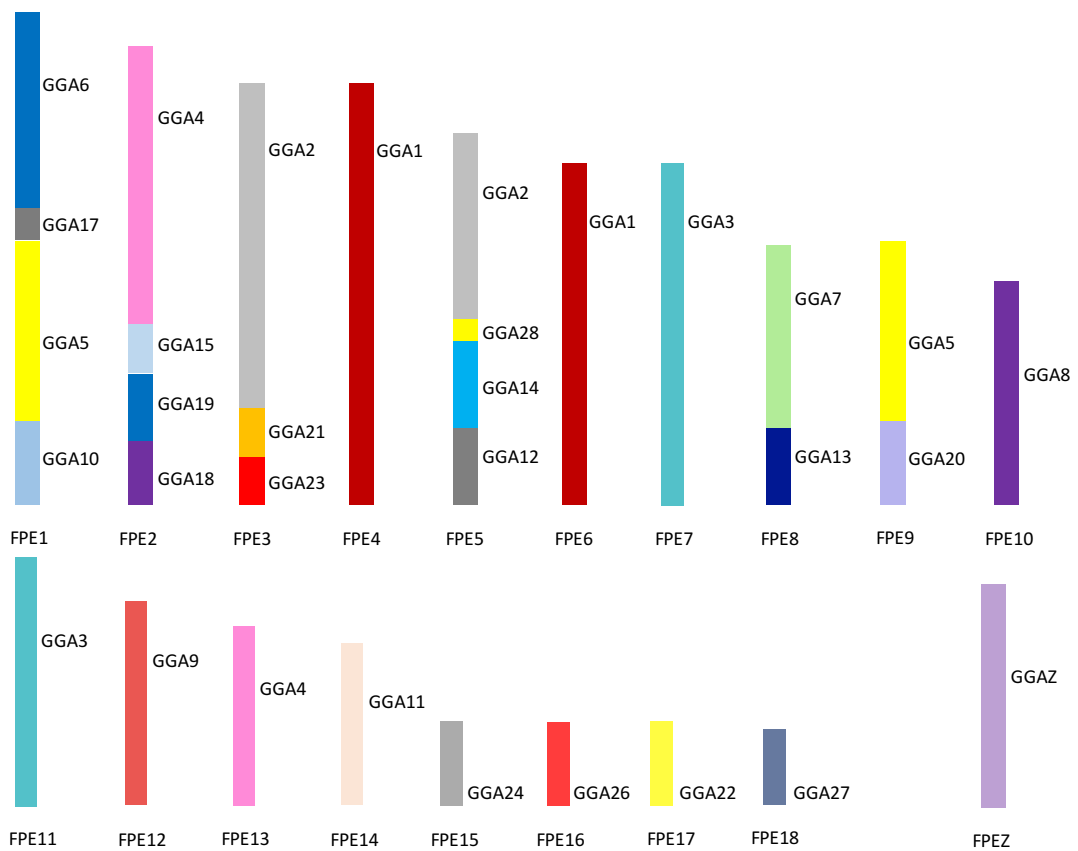


**Figure 5-2: Cytogenetic and PCF mapping of pigeon chromosome 2 (CLI2) using FISH; (a) Evolution highway alignment of zebra finch, chicken and pigeon genomes alongside the PCFs produced by RACA and the BACs that map in this region; (b) Cytogenetic map of BACs in the correct orientation on CLI2; (c) Physical mapping of BACs to CLI2 using FISH; (d) Ideogram illustration of GGA homolog to CLI2.**

A full list of BACs and their genomic coordinates in the pigeon genome is given in the appendix section supplementary table 1 (S1).

**5.4.3 Specific aim 3c: To apply the above BAC set to peregrine falcon (*Falco peregrinus*) chromosomes to upgrade the scaffold-based assembly to chromosome level and map the evolutionary changes that led to the gross genomic rearrangements that occurred in this species**

Homology between the chicken and the peregrine falcon was established for all sequenced chromosomes with the exception of GGA16 and GGA25 (Figure 5-3). In total, 13 peregrine-specific fusions and 4 fissions were detected when compared to chicken (Table 5-3).



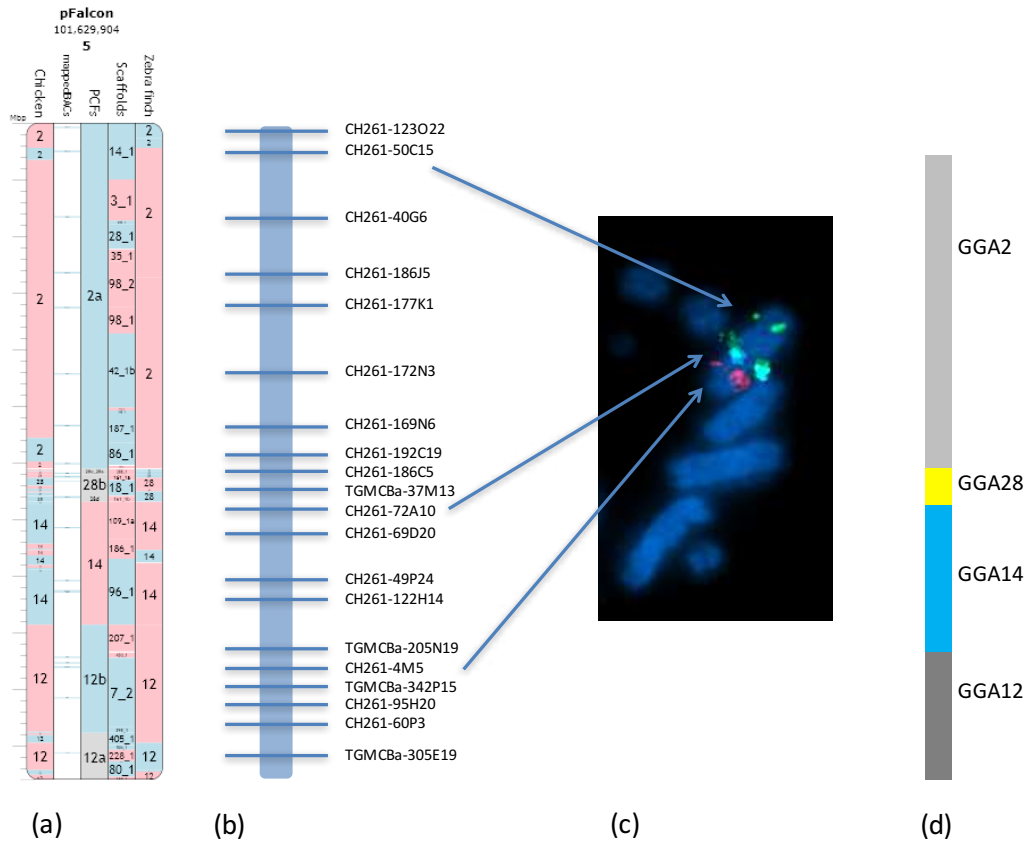
**Figure 5-3: Ideogram representation of peregrine falcon (FPE) chromosomes and their chicken (GGA) homologs.**

Detected lineage-specific rearrangements	Peregrine
Number of chromosome fusions	13
Number of chromosome fissions	4
Number of intrachromosomal events	68
Within scaffolds	51
Between PCFs	17

**Table 5-3: Lineage specific rearrangements detected in the peregrine falcon (*Falco peregrinus*) when compared to the chicken genome.**

Each of the largest chicken macrochromosome homologs (GGA1 to GGA5) were split across two peregrine falcon chromosomes. The peregrine falcon GGA1 and GGA3 counterparts were represented as two entire chromosomes each (FPE4 and FPE6, FPE7 and FPE11, respectively) with no additional fusions. GGA2 was split across FPE3 and FPE5, both of which exhibited additional fusions of microchromosomes with GGA21 and 23 fused in FPE3 and GGA12, 14 and 28 fused in FPE5. Consistent with the pigeon assembly results (and the majority of birds), GGA4 was found to be split across two peregrine falcon chromosomes (FPE2 and FPE13), the former of which exhibited 3 additional microchromosomal fusions (GGA15, 18 and 19). Both GGA6 and GGA7 homologs were found as single blocks fused with other chicken chromosome material within peregrine falcon chromosomes FPE1 and FPE8 respectively. Among the other chicken macrochromosomes, only GGA8 and GGA9 were represented as individual peregrine falcon chromosomes (FPE10 and FPE12, respectively).

Of the 17 mapped chicken microchromosomes, 11 were fused with other chromosomes. Five of the chicken microchromosomes (GGA11, 22, 24, 26, 27) were found as single microchromosomes. A total of 68 intrachromosomal rearrangements were detected in the peregrine falcon when compared with the chicken and zebra finch genomes. Of these, 51 were found within scaffolds and 17 between PCFs. An example of the FISH mapping of PCFs is illustrated in Figure 5-4 where the alignment of genomes on Evolution Highway is shown with the corresponding cytogenetic map established through FISH mapping of the set of 226 BACs. A three-colour FISH image demonstrates the orientation of BACs by FISH.



**Figure 5-4: Cytogetic and PCF mapping of peregrine falcon chromosome 5 (FPE5) using FISH; (a) Evolution highway alignment of zebra finch, chicken and peregrine genomes alongside the PCFs produced by RACA and the BACs that map in this region; (b) Cytogetic map of BACs in the correct orientation on FPE5; (c) Physical mapping of BACs to FPE5 using FISH; (d) Ideogram illustration of GGA homologs to FPE5.**

A full list of BACs and their coordinates in the peregrine falcon genome is given in the appendix, Table S2.

#### 5.4.4 Specific Aim 3d: To supply raw data from mapping PCFs in both species to RVC for uploading to the Evolution Highway browser

Comparative visualisations of both newly assembled genomes were uploaded to the Evolution Highway comparative chromosome browser under the reference genome name 'pFalcon' and 'rPigeon' (<http://eh-demo.ncsa.uiuc.edu/birds-test/#/SynBlocks>) thanks to the efforts of the RVC lab - an example of which is illustrated in Figure 5-5. Peregrine falcon chromosomes 1-13 and Z were named according to (Nishida et al. 2008), and chromosomes 14-19 were numbered by decreasing combined length of the assigned PCFs. Pigeon chromosomes 1 to 9 and Z were named according to Hansmann and colleagues (2009) with the remaining chromosome names assigned according to chicken homologs.

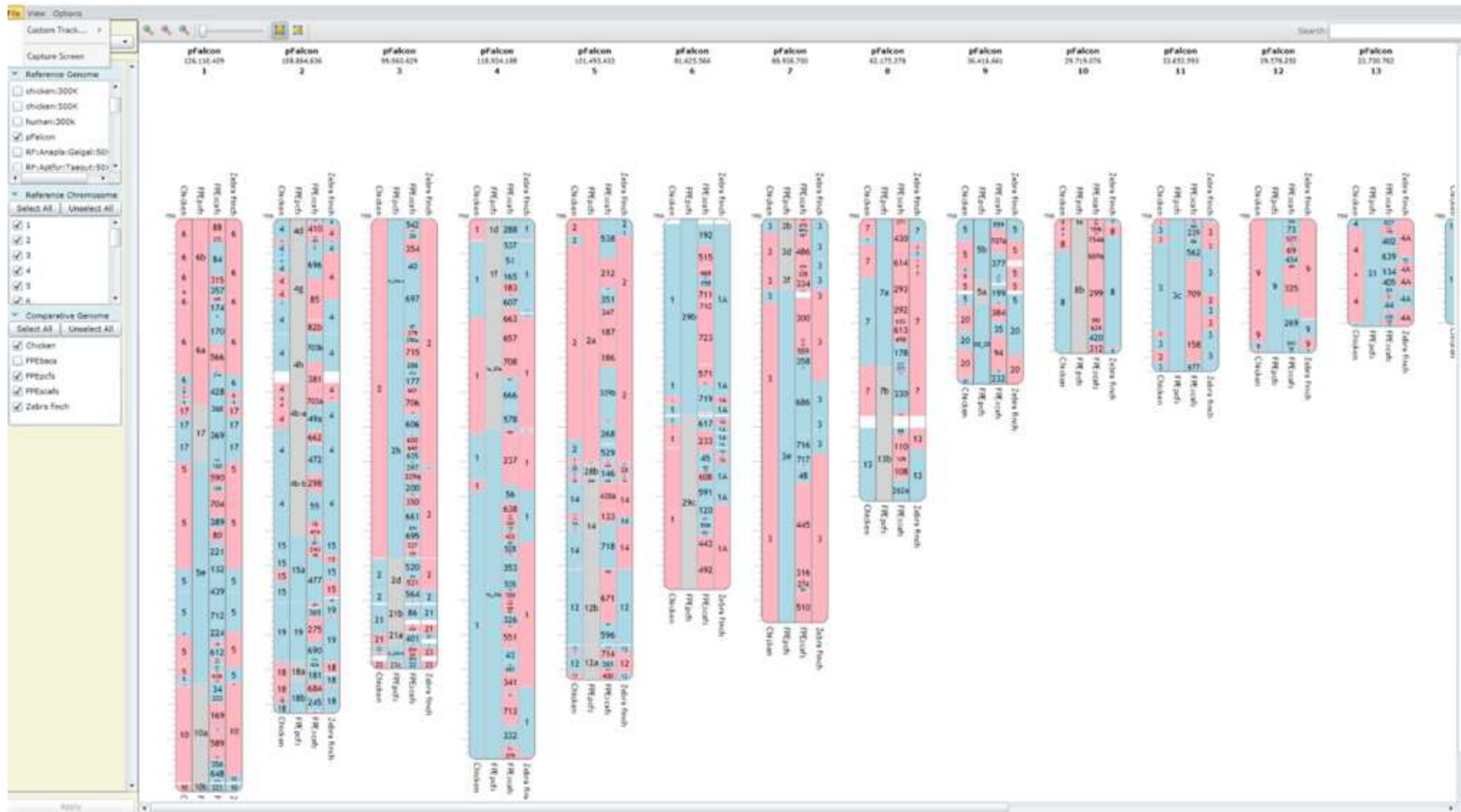


Figure 5-5: Screenshot from Evolution Highway illustrating the uploaded assembled peregrine falcon genome for chromosomes 1 to 13 (<http://eh-demo.ncsa.uiuc.edu/birds-test/#/SynBlocks>).

## 5.5 Discussion

### 5.5.1 Generation of Chromosomal Level Genome Assemblies

Increasing numbers of newly sequenced genomes require tools that enable inexpensive, efficient chromosome level mapping for the reasons already described. In close collaboration with the RVC, results presented in this chapter provide proof of principle that the tools developed here can play a significant part in the generation of chromosome-level assemblies for two previously published but highly fragmented Illumina sequenced genomes – the pigeon and the peregrine falcon (Shapiro et al. 2013; Zhan et al. 2013). The approach and the probe set developed here are also applicable to other bird species with efforts already underway to map the ostrich and the budgerigar genomes. The pigeon and peregrine falcon assemblies now have >80% of their genomes placed on chromosomes making them highly comparable to genomes assembled using Sanger sequencing techniques and high-density physical or genetic mapping (Lewin et al. 2009). The method described here is less expensive and requires fewer resources than traditional approaches, in part thanks to the ability to generate predicted chromosome fragments of a sub-chromosomal size using comparative genome information only. The subsequent use of BAC probes designed to work equally well on a large number of highly diverged avian species creates a resource for physical mapping that is transferrable to all avian species.

The ability to select BAC clones for this method only recently became a possibility thanks to progress achieved by the avian genome sequencing consortium (Zhang et al. 2014a). Based on an approach using cattle BACs on other artiodactyl species described by Larkin et al. (2003), the alignment of multiple avian genomes by colleagues at the RVC allowed the identification of sequence-based features that would significantly improve the chances of successful cross-species hybridisation (Larkin et al. 2003). As a result, we now have a highly efficient universal panel of BAC clones with the potential to hybridise to any avian chromosome preparation. In fact, preliminary results show that the probes developed here have worked well on turtle (*Trachemys scripta* – red-eared slider) and anole lizard (*Anolis carolinensis*) chromosomes (see chapter 6). As described in the chapters 3 and 4, application of the BAC set now routinely involves multiple hybridisations on octochrome and multiprobe devices (8 and 24 hybridisations) meaning that up to 72 BACs (when three colours are used) can be hybridised in a single experiment, making the process increasingly high throughput. To the best of my

knowledge this is the first example of high throughput FISH for single loci being applied for widespread comparative genomics.

### 5.5.2 Avian Interchromosomal Rearrangements

Classical cytogenetic and molecular cytogenetic (zoo-FISH) studies, (see also chapter 4) indicate that most bird genomes remain remarkably conserved in terms of their chromosome number (in 60-70% of species  $2n \sim 80$ ) and that interchromosomal changes are relatively rare, but that when they do occur, they tend to be lineage specific e.g. in *Psittaciformes* (parrots), *Falconiformes* (falcons) and *Sphenisciformes* (penguins) (Griffin et al. 2007; Schmid et al. 2015 and previous chapter). Indeed, this study is the first detailed reconstruction of a highly rearranged avian karyotype (peregrine falcon) and demonstrates that fusion is the most common mechanism of change. There was no evidence of reciprocal translocation and indeed all microchromosomes remained 'intact', even when fused to larger chromosomes. Recently we (Romanov et al. 2014) suggested possible mechanisms why, with relatively rare exceptions, avian genomes remain evolutionarily stable interchromosomally and why microchromosomes represent blocks of conserved synteny. This is examined further in the general discussion.

### 5.5.3 The Pigeon Genome

Despite its importance to human populations, relatively few studies have focused on the overall genomic structure of the pigeon. Using chicken macrochromosome paints hybridised to pigeon chromosomes, Derjusheva et al. (2004) identified a similar karyotypic pattern to that of the chicken (with the exception of GGA4 which is represented as two chromosomes in the pigeon, as confirmed by this study). The chromosome painting approach however, has two limitations: firstly, microchromosomes are not distinguishable individually due to the limitations of flow sorting chromosomes of this size from which to generate chromosome paints (Lithgow, O'Connor R.E. et al. 2014). Secondly, chromosome paints for macro or microchromosomes do not provide the resolution required to identify intrachromosomal rearrangements. A further study by Hansmann et al. (2009) attempted to rectify the first of these issues by using chromosome paints generated from the highly fused stone curlew karyotype (*Burhinus oedicnemus*;  $2n=42$ ). This approach is also limited due to the large number of fusions in the stone curlew, thereby making it difficult to differentiate the microchromosomes from each other. To the best of my knowledge, the study presented here is the first to characterise all



sequenced pigeon microchromosomes and to identify intrachromosomal rearrangements within the macrochromosomes.

#### **5.5.4 The Peregrine Falcon Genome**

Among the falcons, studies of karyotype structure have been similarly few. Only one zoo-FISH study (Nishida et al. 2008) has attempted to characterise the overall genome structure, finding the limited success common to most zoo-FISH studies. Results presented in this chapter illustrate the positions of ancestral microchromosomes demonstrating that fusion is the most common mechanism of interchromosomal rearrangement i.e. there was no evidence of reciprocal translocation and all microchromosomes remained intact, albeit fused to larger chromosomes. With around two thirds of birds exhibiting a remarkably conserved diploid number of around 80 with similar gross overall karyotypic structure (Ellegren, 2013), falcons are one of the few significant exceptions to this rule. The method described here for characterising the peregrine falcon genome may now act as a proof of principle for comparison of other highly rearranged avian karyotypes. Until now, the only available reference genomes were the chicken and the zebra finch, both of which unlike *Falconiformes* and *Psittaciformes* exhibit the highly conserved 'typical' avian karyotype, at least interchromosomally. Using chromosome paints, Nishida et al. (2008) found a very similar pattern of rearrangement in the common kestrel (*Falco tinnunculus*) ( $2n=52$ ), but a less similar pattern in the merlin (*Falco columbarius*) ( $2n=40$ ).

The chromosomally assembled peregrine falcon genome presented in this chapter now provides an additional reference genome from which further analysis of these and other similar species can be performed at a resolution previously not possible. In fact, assembly to chromosomal level of the saker falcon genome using RACA and PCF mapping with FISH has already begun - in this case, the peregrine falcon assembly is now the reference genome. Further clarity on these patterns of rearrangement by mapping increasing numbers of species will allow biological questions to be asked as to whether this degree of rearrangement occurred early in this order and became fixed in the population or whether rearrangement is ongoing. It may also be that there is some biological advantage to this karyotypic structure for these birds, perhaps due to the high metabolic demands required by birds of prey.

## 5.6 Conclusions

By combining comparative sequence analysis, targeted PCR and high-throughput molecular cytogenetics this work presents proof of principle for an approach that is theoretically applicable to any animal genome as a cost-effective means of transforming fragmented scaffold-level assemblies to chromosomal level. The significant fractions of the scaffold genome assemblies assigned to chromosomes for both species were 87% for peregrine falcon and 82% for pigeon, comparable to those previously reported for genomes assembled by traditional means (83-86%). In addition, the N50 of each genome was improved seven-fold and a series of intra and interchromosomal rearrangements that were previously undetectable were identified.

The results described here are, to the best of my knowledge, the first project of its kind describing assembly at a chromosome-level using NGS scaffold based assemblies as a starting point and therefore represents a step-change in the mapping of genome assemblies. This method permits comparative genome research at a higher resolution than previously possible thereby facilitating research into avian karyotype evolution and the role of chromosome rearrangements in adaptation and phenotypic diversity in birds.

**6 Specific aim 4: To use bioinformatic tools to re-create the overall genome structure of both Saurian and Avian ancestors and to retrace the gross evolutionary changes that occurred along the dinosaur lineage. To perform gene ontology analysis of homologous synteny blocks and evolutionary breakpoint regions (EBRs) of chromosomes to test the hypothesis that there is an enrichment for genes that correspond to known phenotypic characteristics of the species in question.**

### **6.1 Background**

The underlying basis of the work presented in this thesis (see general introduction) is that chromosome rearrangements can cause reproductive problems in individuals (discussed in chapter 3) and can cause or reinforce reproductive isolation between species (subsequent chapters). Often leading to compromised meiotic pairing, chromosome rearrangements may ultimately lead to reduced reproductive fitness in hybrid offspring. A reduced level of genetic recombination in rearranged regions is also thought to promote the accumulation of incompatibility loci, subsequently exacerbating a reduction in fitness (Delneri et al. 2003; Noor et al. 2001) – see section 1.5.

Karyotype evolution in birds (particularly at the microchromosome level) is relatively understudied. As already discussed, this is largely due to the fragmented nature of avian karyotypes. Results presented in the two previous chapters have gone some way to solving this problem by defining microchromosomes, identifying intrachromosomal rearrangements and constructing genome assemblies. The work described in these chapters adds significantly to the ~70 papers that have used chicken macrochromosome paints applied to the metaphases of other birds (see section 1.1.2.3) in addition to the 1,000 or so birds that have been partially karyotyped (Christidis. 1990). The underlying message from these karyotyping and chromosome painting studies is that the ancestral pattern of macrochromosomal organisation has remained largely unaltered in the majority of species. Results presented in chapter 4 and 5 illustrate that this pattern extends to the microchromosomes too, with the exception of some species. Both past and current (this thesis) work has provided insight into these rare exceptions that have

wholesale significant chromosome rearrangements e.g. *Psittaciformes* and *Falconiformes* (Nanda et al. 2007; Nishida et al. 2008).

Using individual BAC clones, studies into the turkey, duck and zebra finch (Griffin et al. 2008; Skinner et al. 2009; Völker et al. 2010) as well as the results presented in chapter 5 provide a low-resolution appraisal of intrachromosomal rearrangements. Zoo-FISH alone however is limited in its ability to identify the molecular coordinates of evolutionary breakpoints. The availability of chromosome level assembled genomes (chicken, duck, zebra finch, turkey, flycatcher and as a result of this thesis, pigeon and peregrine falcon) allows comparative genomics to be performed at a much higher resolution. David Burt and co-workers (Burt et al. 1999) were the first to use bioinformatics to determine cross-species analysis of whole avian chromosomes at a genomic level (chicken-human). The landmark chicken genome sequence paper (ICGSC et al. 2004) also established conserved synteny between the chicken and human genomes. In the twelve following years however conserved synteny comparisons have only been made between the chromosomes of two or three bird species (Völker et al, 2010; Rao et al. 2012; Skinner & Griffin 2011; Lithgow, O'Connor R.E. et al. 2014). Enriching the data by mapping chromosome level *de novo* avian genomes will lead to much deeper comparative cytogenetic analysis in birds and an increasingly robust means of investigating functional significance of HSBs and EBRs

Most EBR and HSB studies have focused on mammals (as covered in section 1.5.3). Analysis of other animals, such as birds, is a priority in order to establish whether these same patterns are seen in other groups or whether mammals are the exception. Of the studies performed on mammals, EBRs tend to appear in gene-dense regions (Larkin et al. 2009) with enrichment for zinc finger protein genes, genes associated with environmental stimulus response. These 'EBR genes' appear to be related to biological features specific to individual lineages (Larkin et al. 2009; Elsik et al. 2009; Groenen et al. 2012). A pattern of EBR reuse is also evident with some regions of the genome being particularly prone to chromosomal breakage (Sankoff 1999; Stankiewicz & Lupski. 2002). In fact, among birds (chicken, turkey and zebra finch) it appears that breakpoint re-use occurs more often than is seen in mammals (Skinner & Griffin 2012; Lithgow, O'Connor R.E. et al. 2014) with previous data produced from this lab (comparing chicken and zebra finch) that suggests a key role for recombination-based mechanisms in the generation of chromosome rearrangements (Völker et al. 2010). As described in section 1.5.3, Larkin and colleagues argue that the presence of HSBs across multiple species is the result of a

selective advantage to keeping particular combinations of genes together (Larkin et al. 2009) with evidence of gene ontology enrichment for terms related to organismal development and the central nervous system although some authors refute the notion that these proximity patterns occur or that there is any adaptive significance when they do (e.g. Singer et al. 2005; Sémon et al. 2006). A detailed bioinformatic analysis of the chromosomal differences and changes that have occurred during bird chromosome evolution is therefore essential.

The purpose of the first part of this chapter was to use six of the best chromosomally assembled avian genomes (Zhang et al. 2014a) and the only non-avian reptile species (*Anolis carolinensis*) with a sufficiently complete chromosomal level genome assembly (Alföldi et al. 2011) to reconstruct the ancestral avian karyotype and infer the evolutionary events that led to the karyotypes seen in extant birds. Based on the above approach (reported in Romanov et al. 2014), reconstruction of the series of gross genomic rearrangements that most likely occurred from the lepidosaur divergence 275 mya (Shedlock and Edwards, 2009) to the ancestral bird 100 mya forms the second part of this chapter. This period coincides with the Saurischia-Theropoda-Maniraptora-Avialae lineage and the age of the dinosaurs.

The group of animals most commonly known as dinosaurs typically refers to archosaurs with hind limbs held erect beneath the body, specifically *Triceratops*, Passerea and their relatives. While it is now accepted that birds are the living descendants of dinosaurs and that there is no clear distinction between the two groups, in common parlance, and for the purposes of this study dinosaurs are referred to as being phylogenetically distinct from their archosaurian relatives, the testudines (turtles), crocodylians and birds. The Dinosauria have captured the imagination of scientists, the creative arts and the general public since the earliest fossil discoveries. The Michel Crichton novel/Steven Spielberg movie Jurassic Park (and its sequels, most recently in 2015) popularised not only the organismal study of dinosaurs but also a popular interest in their genomics and molecular biology. To the best of my knowledge however there have been few academic studies that have made reasonable extrapolations about the nature of the dinosaur genome and none that have reconstructed the major events of genome reorganization that accompanied their evolution.

As described in section 1.1.8.2, ancestral karyotype reconstruction has largely been performed using cross species chromosome painting, which, while useful, is limited by the evolutionary distance between target species. Despite the difficulties hybridising *in-situ* across these

distances, FISH using paints derived from chicken chromosomes have been successfully hybridised to the chromosomes of turtles (which diverged from the archosaurs 260 mya) illustrating a remarkable degree of homology between them and birds suggesting a very early origin of the 'avian style' genome (Matsuda et al. 2005). Further regions of chromosome homology have also been identified using chicken paints on: crocodiles and lizards (Kasai et al. 2012; Pokorná et al. 2011; Pokorná et al. 2012) suggesting that these regions were present in their common ancestor ~250 mya. Whilst providing a fascinating insight into the ancestral genome, the significance of what lay within these regions and why these regions have remained conserved is limited at this level of resolution. Specifically, there are few reports of probes for individual loci hybridising effectively across long evolutionary distances. Part of the second half of this chapter therefore was to use the tools generated in chapter 4 and 5 to establish whether, despite the extraordinarily long periods since they shared a common ancestor, the highly conserved nature of the selected BACs would allow cross species hybridisation on turtle and lizard (specifically, red-eared slider turtle and anole lizard) chromosomes, results of which would provide clues as to the nature of the gross karyotypic rearrangement of avian and non avian dinosaurs.

Finally, to the best of my knowledge, analysis of any functionally significant gene ontology terms within EBRs and HSBs during avian and non-avian dinosaur evolution has yet to be investigated.

## 6.2 Specific Aims

With the above background in mind, the specific aims of this chapter were as follows:

- Specific aim 4a: To use bioinformatic tools to recreate the most likely ancestral avian karyotype through analysis of 6 avian genomes and to infer the evolution of chromosomal rearrangements from the divergence of the avian ancestor to extant avian karyotypes
- Specific aim 4b: To use gene ontology (GO) tools to assess whether avian micro-chromosomes were enriched for GO terms
- Specific aim 4c: To use the cytogenetic tools (sequence conserved BACs) developed in chapter four to investigate genome conservation between avian and non-avian reptiles
- Specific aim 4d: To use a similar approach to specific aim 4a to infer the arrangement of saurian ancestor chromosomes and trace the evolution of chromosomal rearrangements from the divergence of the saurian ancestor, to the appearance of the common avian ancestor and from there onwards to the extant chicken lineage
- Specific aim 4e: To analyse gene ontology terms in the HSBs and EBRs of this dinosaur lineage to test the hypothesis that these relate to phenotypic characteristics associated with the dinosaurs.

## 6.3 Materials and Methods

BAC clones selected from the set described in section 4.4.4 were generated, labelled and FISH was performed as described in section 2.3. Frozen fibroblast cells for the anole lizard (*Anolis carolinensis*) and red-eared slider (*Trachemys scripta*) were kindly provided by Professor Malcolm Ferguson-Smith (Cambridge University).

Bioinformatic analysis was performed as described in section 2.5.

## 6.4 Results

### 6.4.1 Specific Aim 4a: To use bioinformatic tools to recreate the most likely ancestral avian karyotype through analysis of 6 avian genomes and to infer the evolution of chromosomal rearrangements from the divergence of the avian ancestor to extant avian karyotypes

#### 6.4.1.1 Identification of Multispecies HSBs and EBRs

Results from this study originated from the alignment of the six best avian genomes that were assembled at a chromosomal (chicken, turkey, duck and zebra finch) or near-chromosomal level (ostrich and budgerigar) against one outgroup (anole lizard). Pairwise alignments of the genomes allowed for the visualisation of multi species HSBs (and their orientation in each genome) as well as the identification of EBRs between HSBs. Evolution Highway screenshots for the six avian species and the lizard outgroup compared to chicken chromosomes 5 and 11 are illustrated in Figure 6-1 illustrating the means of aligning genomes against each other.

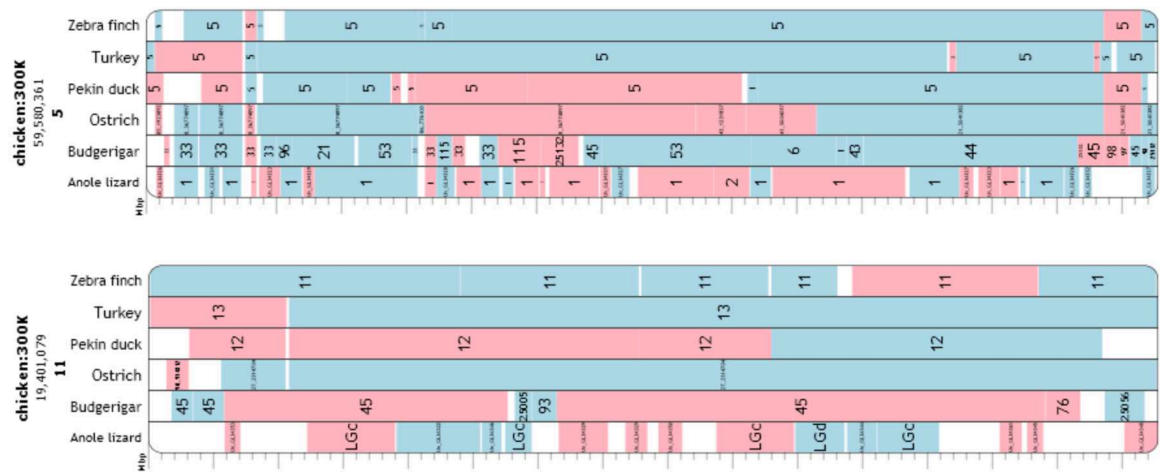


Figure 6-1: Evolution Highway screenshot of 5 avian genomes and the outgroup (anole lizard) aligned to chromosomes 5 and 11 of the reference genome - the chicken. Pink blocks are indicative of an inversion compared to the reference (chicken) genome and numbers in the blocks reflect the chromosome number of that species or the scaffold number in the budgerigar and the ostrich.



**6.4.1.2 MGRA Reconstruction of Ancestral Karyotypes and Chromosomal Changes Leading to Arrangements in 6 Extant Species**

Data derived in section 6.4.1.1 was analysed using MGRA (with help from Dr Michael Romanov) in order to reconstruct ancestral chromosomes 1-5 for all birds and chromosomes 6-28+Z for neognathae. Figure 6-2 indicates the most likely path of chromosomal rearrangement between ancestral chromosome 5 and its homologs in the extant species investigated. Figure 6-3 illustrates the changes that occurred in the homologs of neognathae ancestor chromosome 11. Using the lizard as the outgroup meant that coverage was limited beyond chromosome 6, so for chromosomes 6 and smaller (plus Z – the anole lizard uses XY sex determination), the ostrich was used as the outgroup. In the case of chromosome 11 presented here, the avian ancestral rearrangement was inferred due to the identical patterns seen in the ostrich and the chicken.

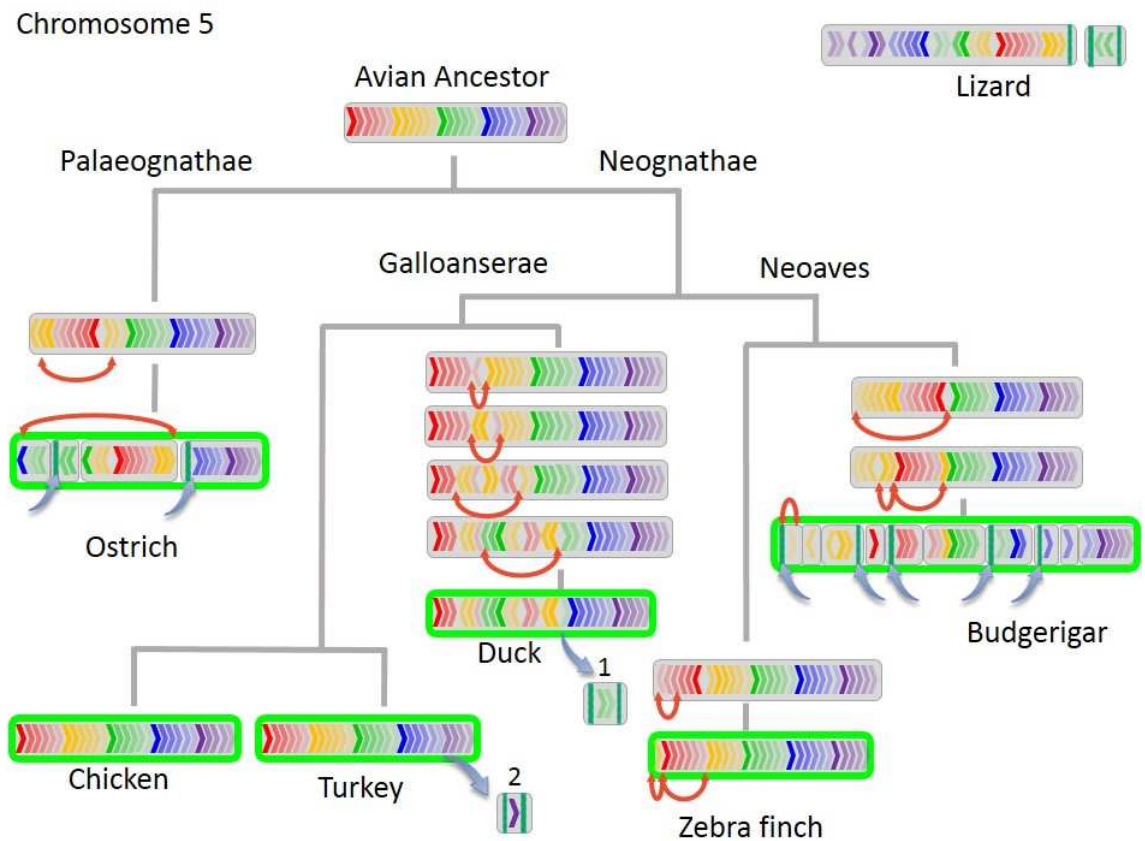
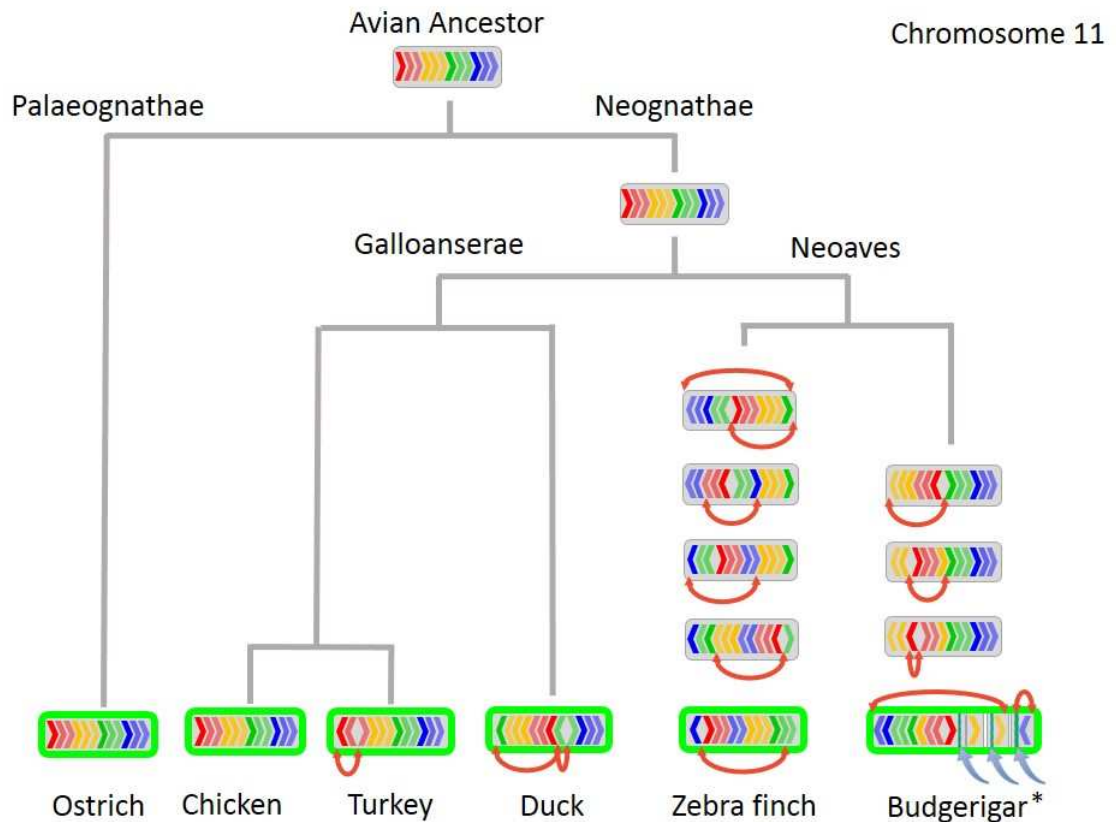


Figure 6-2: Avian ancestor chromosome 5 and the likely rearrangements that have occurred along each lineage to the extant bird. Rainbow patterned arrows within the chromosomes represent the HSBs, red curved arrows indicate chromosome inversions, blue arrows indicate chromosome translocations, green outline indicates previous chromosome painting results (anole lizard arrangement also indicated).



**Figure 6-3: Avian ancestor chromosome 11 and the likely rearrangements that have occurred along each lineage to the extant bird. As the arrangement for ostrich and Neognathae ancestors were the same, the avian ancestor could be derived (unlike for other chromosomes smaller than 5). \* in Budgerigar, FISH indicates fusion to a larger chromosome. Colour scheme and pattern consistent with Figure 6-2.**

The overall analysis suggests that, of the 6 species, the chicken lineage underwent the least number of intrachromosomal rearrangements (in other words, the chicken was most similar to the common ancestor). A disproportionately high number of rearrangements appear to have occurred in the turkey lineage since the divergence from chicken 30 mya, 19 of which were on chromosome 1, it appears likely however, that this is a consequence of assembly errors rather than actual rearrangements (Zhang et al. 2014a). These results also suggest that the ostrich lineage underwent 44 intrachromosomal changes on chromosomes 1-5 since the divergence from the common avian ancestor (approximately 100 mya), and the duck underwent 28 changes since the *Galliformes-Anseriformes* divergence (~65 mya). A faster rate of change was seen in the zebra finch and the budgerigar lineages, 41 in the former and 39 in the latter since the *Passeriformes-Psittaciformes* divergence (~54 mya). For the homologs of chromosomes 6-28 + Z, the absence of meaningful data from the lizard outgroup, meant that the analysis was focused on the neognathae (using the ostrich as an outgroup). The pattern here was consistent with the results for chromosomes 1-5, where the fewest number of changes appear to have occurred

between the chicken lineage and the avian ancestor and the greatest rate of change appears to have occurred in the budgerigar since the *Passeriformes-Psittaciformes* divergence 54 mya (68 for zebra finch and 79 for budgerigar).

Species	Ostrich	Chicken	Turkey	Duck	Zebra finch	Budgerigar
Number of interchromosomal rearrangements (as determined by FISH) from avian ancestor	0	1	1	0	2	8
Number of interchromosomal changes (determined using bioinformatics) from avian ancestor	26	1	5	1	2	40
Number of intrachromosomal rearrangements from avian ancestor – chromosomes 1-5 (excluding 4p)	44	22	46	40	54	52
Number of intrachromosomal rearrangements from Neognathae ancestor – chromosomes 6-28 + 4p + Z	N/A	25	32	49	71	82

**Table 6-1: Total numbers of inter- and intrachromosomal rearrangements in each species since their divergence from the avian ancestor 100 mya.**

Intrachromosomal events identified here are most parsimoniously explained by a series of inversions whereas the interchromosomal rearrangements are likely to have occurred due to a series of translocations. The number of changes and their timescales (and therefore rates of change) are presented in Figure 6-4 (for avian ancestor chromosomes 1-5) and Figure 6-5, for the neognathae (chromosomes 6-28 plus Z).

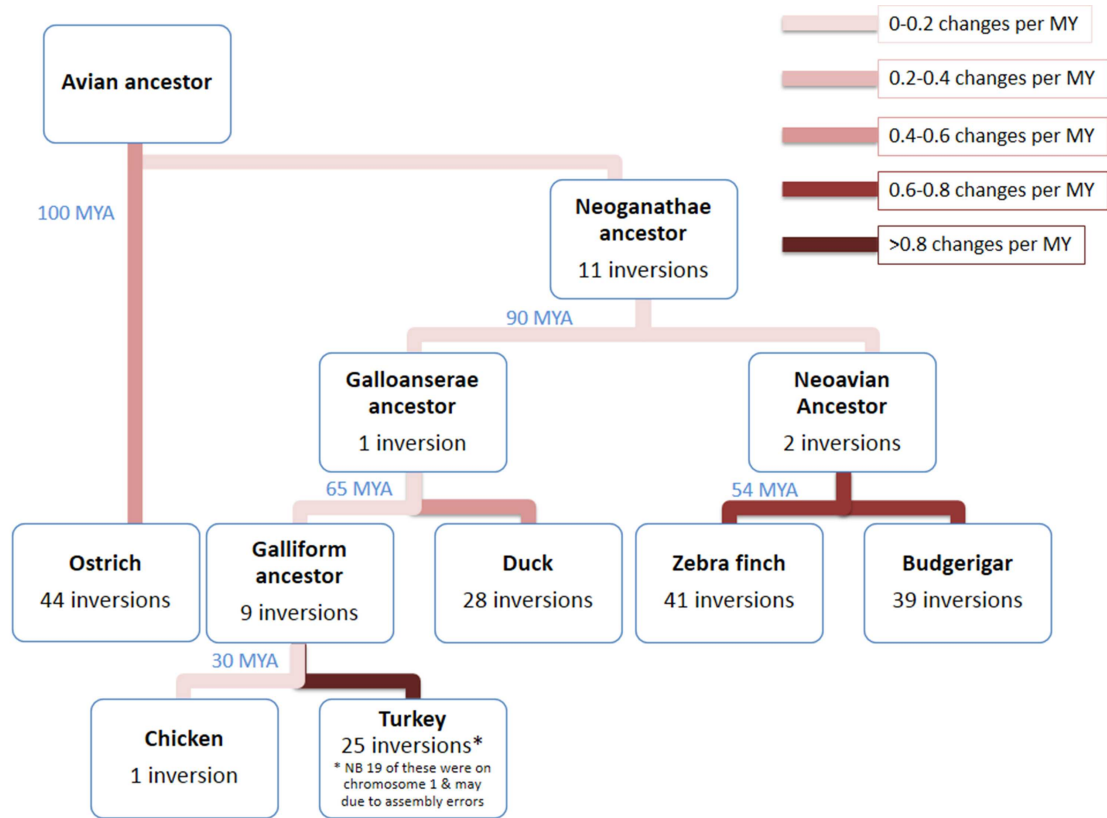


Figure 6-4: Number of chromosomal inversions that most parsimoniously explain the patterns seen in the 6 extant species as they diverged from the ancestor for chromosomes 1-5 (using the lizard outgroup). The greatest rates of change were seen in zebra finch and budgerigar. The phylogenetic is tree based on Jarvis et al. (2014).

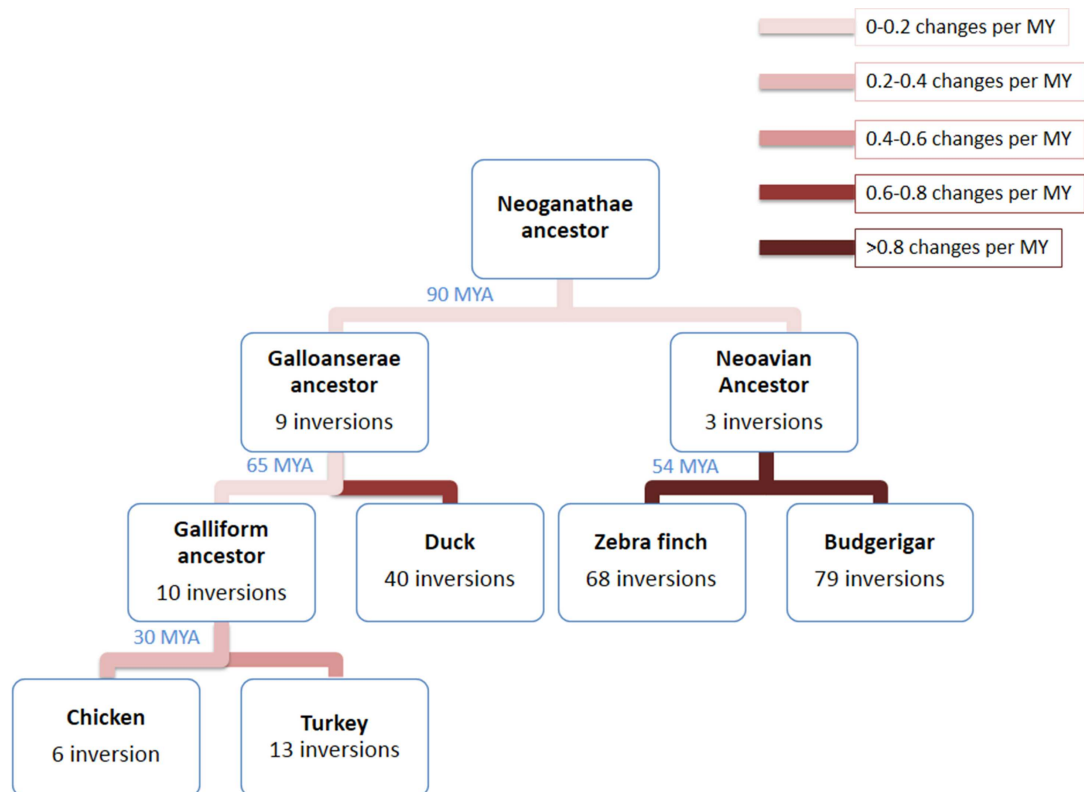


Figure 6-5: Number of chromosomal inversions that most parsimoniously explain the patterns seen in the 5 extant species as they diverged from the ancestor for chromosomes 6-28 + Z (using the ostrich as an outgroup).

The greatest rates of change were seen in zebra finch and budgerigar. The phylogenetic tree is based on Jarvis et al. (2014).

#### 6.4.2 Specific aim 4b: Gene Ontology Analysis of Avian microchromosomes

Given the previously hypothesised (but now illustrated – see chapter 4) conserved nature of microchromosomes this led to the hypothesis that each microchromosome was enriched for functionally enriched gene ontology (GO) terms for which there was an evolutionary advantage to remain clustered as a single entity during meiosis. Using the GO functional annotation tool DAVID all microchromosomes were analysed individually using Ensembl gene ID's obtained from Biomart. The DAVID outputs were then filtered by reducing the output to one gene representative per family prior to filtering for an enrichment score above 1.3 and a p-value less than 0.05.

GGA chromosome 11 showed significant results for metabolic processing, chromosomes 12 and 13 indicated involvement in nucleotide binding and chromosomes 14 and 17 showed a high density of genes involved in the WD40 complex. Chromosome 15 had significant results for both organelle binding and phosphorylation. As anticipated chromosome 16 produced significant

results for immune response genes. Chromosome 24 indicated involvement in lipid metabolism and chromosome 26 had significant results for sensory stimulation. Finally, chromosome 28 showed results for nucleotide binding, insulin signalling and killer cell mediated toxicity. Chromosomes 18, 19, 20, 21, 22 23 and 25 produced no significant results. However, further editing of this data using a false discovery rate (FDR) of 0.05 reduced the significant results down to those that related only to the immune related genes on chromosome 16.

#### **6.4.3 Specific Aim 4c: To use the cytogenetic tools (sequence conserved BACs) developed in chapter 4 to investigate genome conservation between avian and non-avian reptiles**

As described in section 4.4.3, chicken and zebra finch BACs were selected for cross species FISH according to a range of criteria including the proportion of conserved elements shared across multiple avian species. From this set of BACs, an edited panel was selected with the highest conservation scores and the highest degree of success hybridising across evolutionarily distant species. Two BACs were then selected per reference (chicken) chromosome to produce a 'panel' of BACs to test across all avian species. The reference chromosomes of interest analysed here were GGA10-GGA28 (with the exception of GGA16). Due to the high degree of apparent genome conservation observed between avian and reptilian species, this set of BACs was tested on chromosome suspensions from the anole lizard (*Anolis carolinensis*) and the red-eared slider turtle (*Trachemys scripta*).

##### **6.4.3.1 Overall Hybridisation Success Rates between Birds and Reptiles using FISH**

In total, 28 of the 34 (82%) avian microchromosome (GGA10-GGA28, not GGA16) FISH probes produced a signal in *T. scripta*. Of the microchromosome probes tested on *A. carolinensis*, 17 of the 36 (47%) microchromosome probes produced a signal. At least one signal per chicken microchromosome was achieved in the turtle and signals for all but three chicken microchromosomes were seen in the lizard. Overall results by chicken chromosome are presented in Table 6-2.

Clone Name	GGA Chr	Turtle	Lizard
CH261-115G24	10	✓	✓
CH261-71G18		✓	No
CH261-121N21	11	✓	✓
CH261-154H1		✓	No
CH261-60P3	12	Proto	No
CH261-4M5		Proto	Proto
CH261-115I12	13	No	No
TGMCBA-321B13		Proto	Proto
CH261-122H14	14	✓	No
CH261-69D20		✓	No
CH261-90P23	15	✓	✓
TGMCBA-266G23		✓	✓
TGMCBA-375I5	17	✓	✓
CH261-42P16		✓	No
CH261-60N6	18	✓	No
CH261-72B18		✓	No
CH261-10F1	19	✓	✓
CH261-50H12		✓	✓

Clone Name	GGA Chr	Turtle	Lizard
TGMCBA-250E3	20	No	No
TGMCBA-341F20		No	No
CH261-83I20	21	No	No
CH261-122K8		✓	✓
CH261-40J9	22	✓	✓
CH261-18G17		✓	✓
CH261-191G17	23	No	✓
CH261-90K11		✓	✓
CH261-103F4	24	✓	✓
CH261-65O4		✓	✓
CH261-59C21	25	No	No
CH261-127K7		✓	No
CH261-186M13	26	Proto	No
CH261-170L23		Proto	Proto
CH261-66M16	27	No	No
CH261-28L10		✓	No
CH261-64A15	28	✓	No
CH261-72A10		✓	No

**Table 6-2: Hybridisation success by GGA chromosome using GGA BACs on *Anolis carolinensis* (anole lizard) and *Trachemys scripta* (red-eared slider turtle). Clear signals are denoted by an '✓' and 'No' where no signal was evident; 'Proto' denotes GGA microchromosome BACs fused to a macrochromosome.**

#### **6.4.3.2 Microchromosome Conservation between Birds and Reptiles**

An extraordinary degree of conservation between the microchromosomes of birds and both reptiles investigated here was observed with apparent homology between 8 microchromosomes: GGA10, 11, 15, 17, 19, 21, 23 and 24 – all of which appear to be conserved as microchromosomes with no evidence of fusion or fission. In addition, homology as intact microchromosomes was observed between GGA14, 18, 25, 27 and 28 in the turtle. These results therefore suggest that GGA24 is conserved as a microchromosome across all avian and reptile species tested (including the *Psittaciformes* and *Falconiformes*) as illustrated in Figure 6-6; inferring that this microchromosome was fixed in the saurian ancestor; while GGA25 and 27 are conserved across all of the avian species and the turtle species tested here suggesting that these microchromosomes were fixed in the archosauromorph ancestor of turtles and birds.

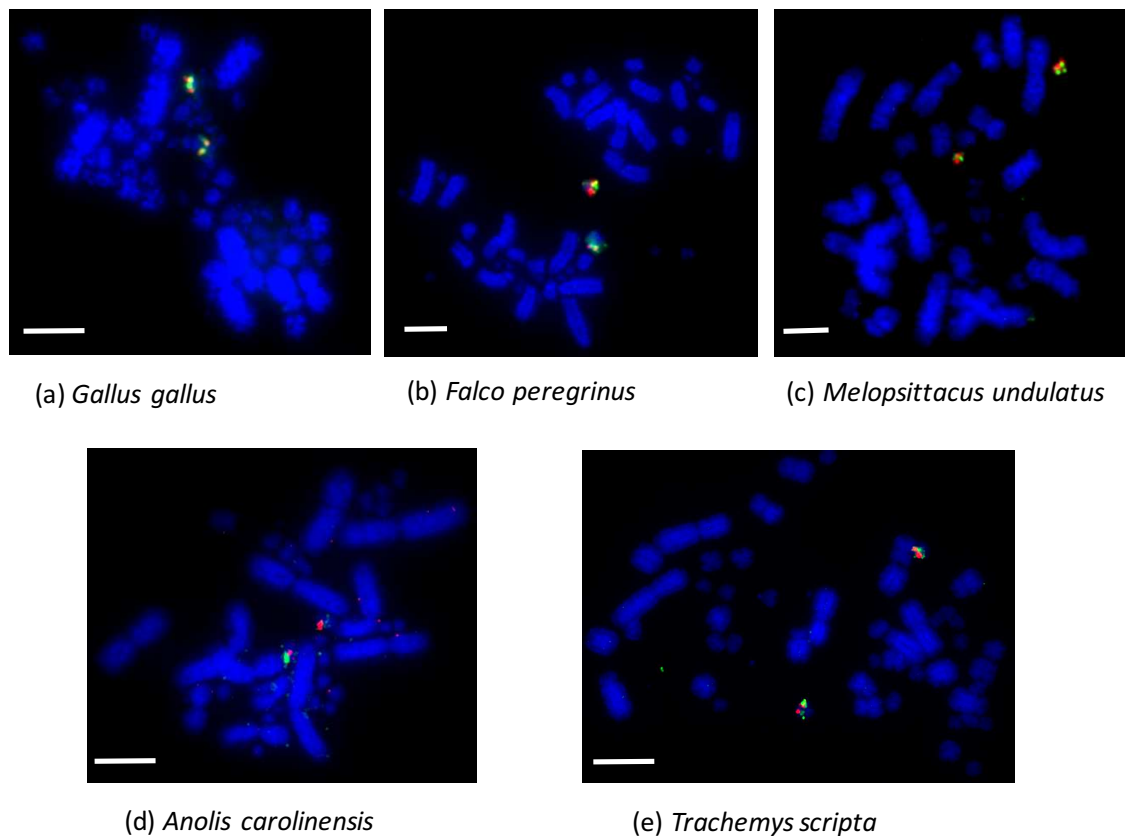
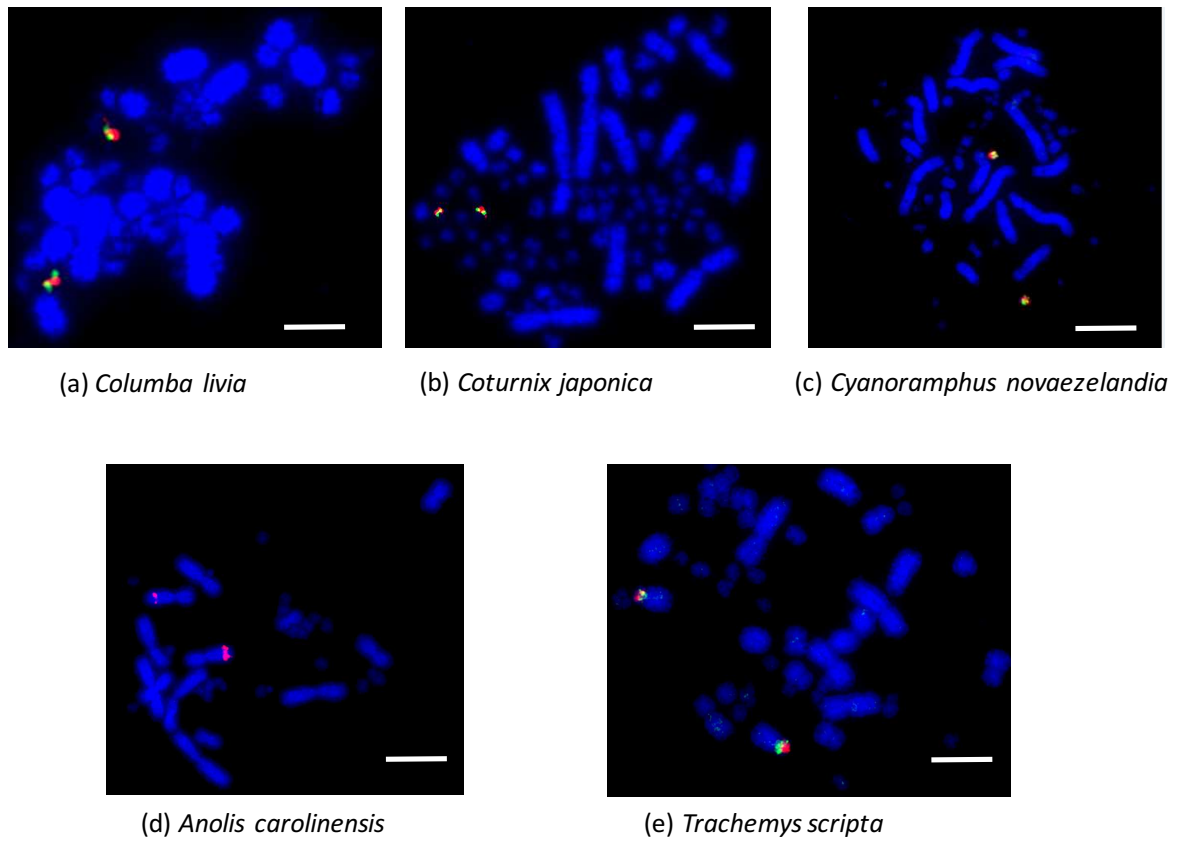


Figure 6-6: Dual colour FISH of GGA24 probes hybridised to 3 phylogenetically distant avian species (a) *Gallus gallus*, (b) *Falco peregrinus*, (c) *Melopsittacus undulatus* and two reptile species (d) *Anolis carolinensis* and (e) *Trachemys scripta*, illustrating an extraordinary degree of genome conservation. Scale bar 10µm.

#### 6.4.3.3 'Protomicrochromosome' Evidence in Reptiles

The lower diploid number seen in both lizards and turtles generally and in the species tested here (*T.scripta* 2n=50 and *A.carolinensis* 2n=38) would suggest that there would need to be a degree of chromosomal fission among these species for descendants along this lineage to reach a higher diploid number as is seen in most birds where ~2n=80. Results generated using the avian FISH probes revealed a minimum of three cases (GGA12, 13 and 26) where chicken microchromosome homologs appeared fused to macrochromosomes in both the lizard and turtle species suggesting that these were in fact 'protomicrochromosomes' that had subsequently split in the avian lineage to form microchromosomes. The homolog of GGA26 is also one of the microchromosomes that appears to remain intact across all avian species as illustrated in Figure 6-7 where the probes hybridise to a macrochromosome in both reptile species but are intact as a microchromosome across all species tested.





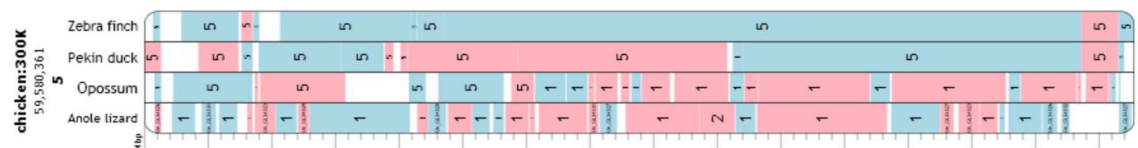
**Figure 6-7: Dual colour FISH of GGA26 illustrating a microchromosomal pattern in three representative avian species (a) *Columba livia*, (b) *Coturnix japonica*, (c) *Cyanoamphus novaezelandia* and the position of the same BACs hybridising to macrochromosomes of (d) *Anolis carolinensis* and (e) *Trachemys scripta* suggesting a macrochromosomal origin of the ancestral avian microchromosome 26. Scale bar 10µm.**

Finally, one of the pair of GGA22 probes hybridised to the middle of a macrochromosome in the turtle, but to a microchromosome in the lizard suggesting that this is a turtle specific fusion that has occurred after the turtles diverged from their common archosaurian ancestor with the birds.

**6.4.4 Specific aim 4d: To use a similar approach to specific aim 4a to infer the arrangement of saurian ancestor chromosomes and trace the evolution of chromosomal rearrangements from the divergence of the saurian ancestor, to the appearance of the common avian ancestor and from there onwards to the extant chicken lineage**

**6.4.4.1 Identification of Multispecies HSBs and EBRs**

Using a similar approach to that described in section 6.4.1.1, results from this study were generated from the alignment of the three best avian genomes that were assembled at a chromosomal level (chicken, duck and zebra finch) along with the best assembled reptile genome available assembled to a partial chromosomal level (*Anolis carolinensis*) against one mammalian outgroup (opossum – *Monodelphis domestica*). Pairwise alignments of the genomes allowed for the visualisation of multispecies HSBs - msHSBs (and their orientation in each genome) as well as the identification of EBRs between these msHSBs. An Evolution Highway screenshot for the three avian species, the lizard and the opossum outgroup compared to chicken chromosome 5 is shown in Figure 6-8, illustrating the means of aligning genomes against each other.



**Figure 6-8: Screenshot of pairwise alignments for zebra finch, duck, opossum and anole lizard aligned against chicken chromosome 5 using Evolution Highway.**

**6.4.4.2 msHSB Identification and Coverage**

Pairwise multiple genome alignments of the five species, including the chicken reference, resulted in a total of 397 msHSBs. These were distributed across 19 of the 28 sequenced chromosomes available on Evolution Highway: GGA1–GGA9, GGA11–GGA13, GGA15, GGA18, GGA20, GGA24, GGA26, GGA27, and GGAZ. The 397 msHSBs were also dispersed on 19 duck chromosomes, 21 zebra finch chromosomes, 10 lizard chromosomes, and 8 opossum chromosomes. Despite, reduced genome coverage due to the lizard chromosome assembly stopping at lizard chromosome 6, good coverage of the chicken genome was achieved with msHSBs covering 53% of the above 19 chicken chromosomes and 49% of the 28 chromosomes.

The proportion of coverage for each chicken chromosome is represented graphically in Figure 6-9.

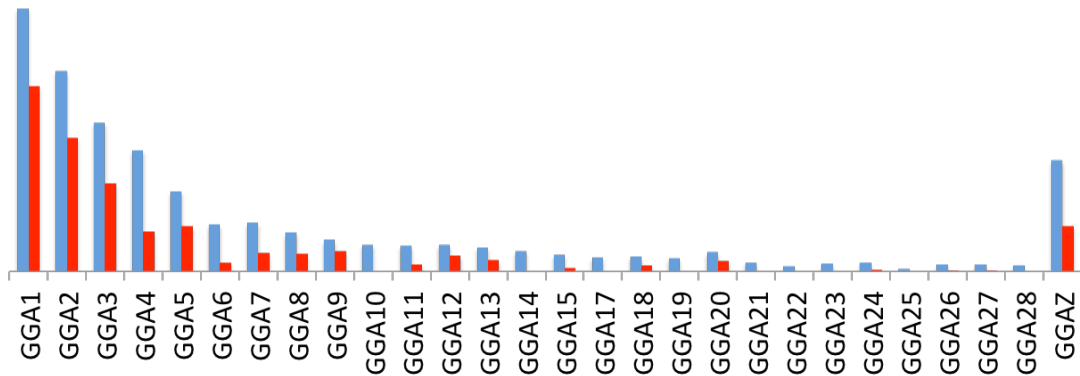


Figure 6-9: HSB coverage relative to the chicken genome where blue bars represent GGA chromosome length and the red bars represent HSB coverage of each chromosome.

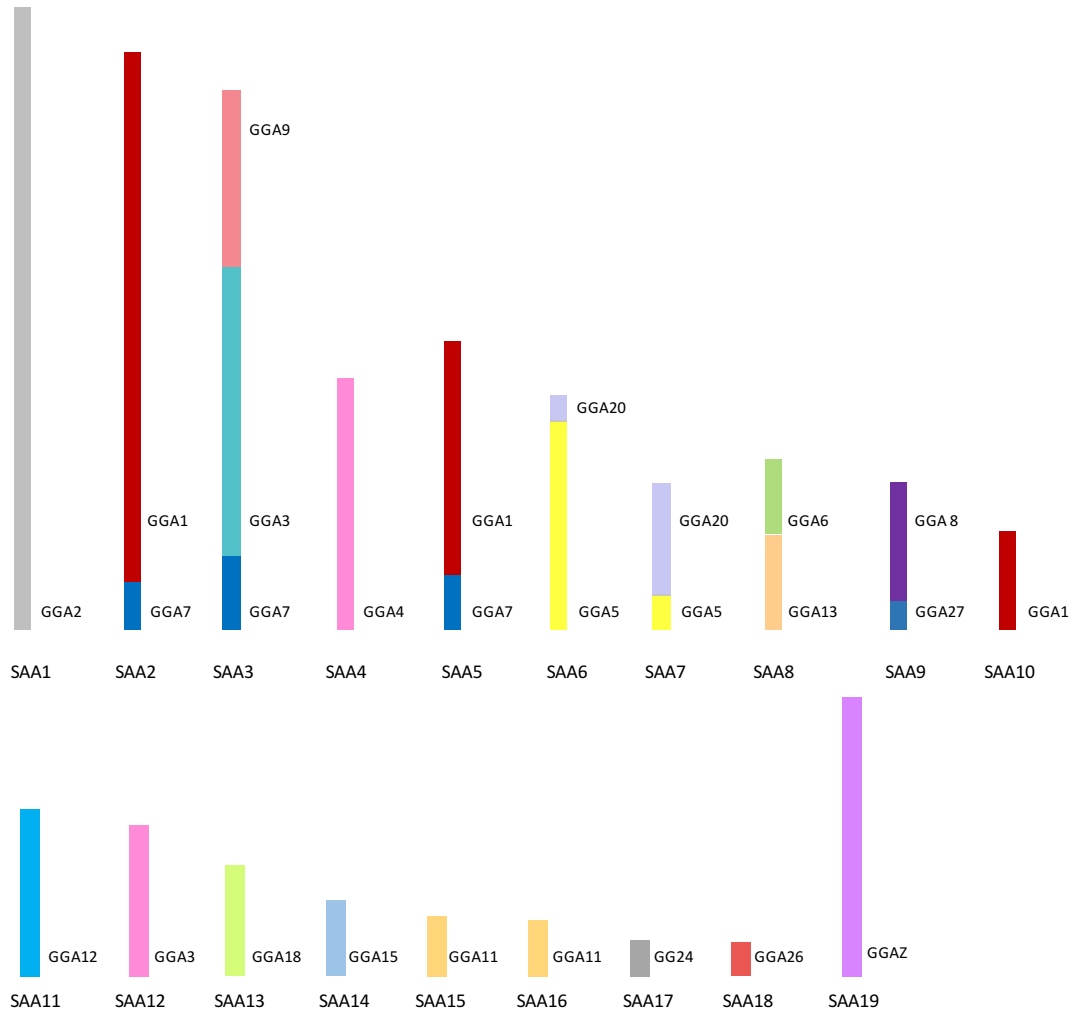
#### 6.4.4.3 Reconstruction of Saurian Ancestor using MGRA2

Data derived in section 6.4.4, was used to produce a series of contiguous ancestral regions (CARs) representing the most likely ancestral karyotype of the saurian ancestor (ancestor of archosauromorphs and lepidosaurs) that diverged from the mammalian lineage 275 mya using the MGRA2 algorithm. MGRA2 outputs generated 397 msHSBs spread across 19 CARs (the number of msHSBs per CAR varied between 2 and 59). Chicken homologs aligned to the CARs are illustrated in Figure 6-10. The overall CAR sizes produced by MGRA2 are listed in Table 6-3.

Saurian Ancestor CAR	CAR Size (bp)	Saurian Ancestor CAR	CAR Size (bp)
1	93,013,653	11	11,567,190
2	84,004,751	12	8,839,163
3	76,908,731	13	4,789,024
4	29,706,497	14	2,652,299
5	43,492,374	15	2,198,148
6	32,219,224	16	3,197,780
7	8,762,780	17	1,164,820
8	14,486,888	18	785,330
9	13,844,044	19 (Z)	33,692,620
10	12,612,535	<b>Total</b>	<b>477,937,851</b>

Table 6-3: Size in base pairs (bp) of each saurian ancestral CAR.

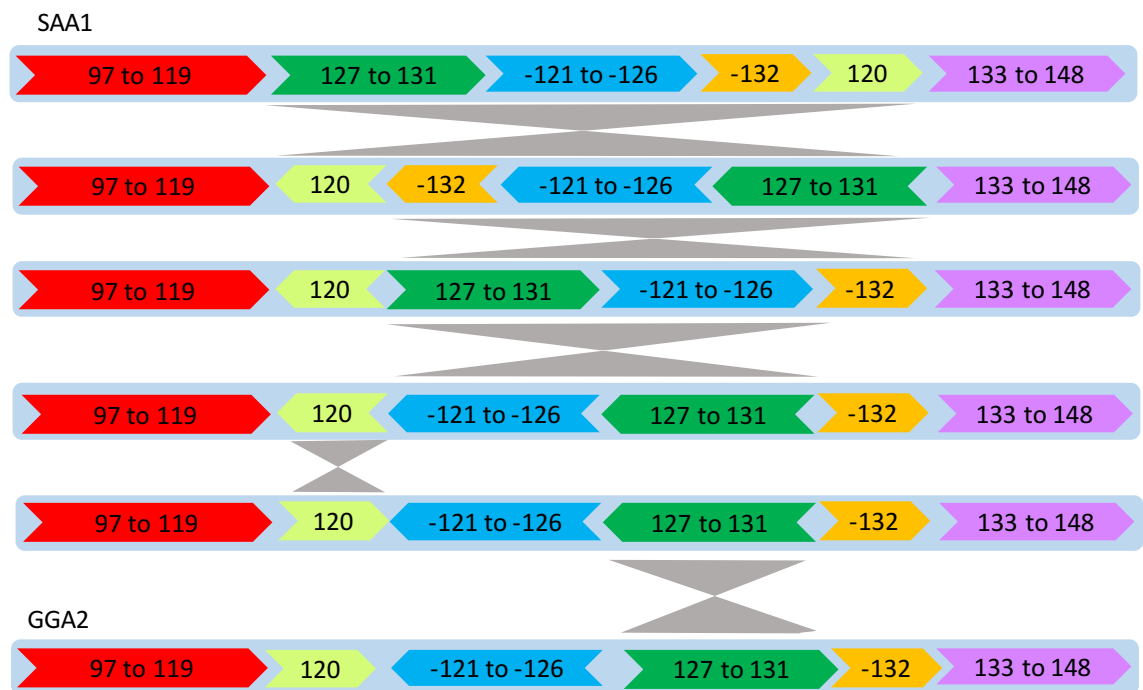
CARs were numbered according to the convention used for naming avian chromosomes with each CAR assigned the prefix SAA for 'Saurian ancestor'. The reconstructed chromosomes were numbered according to size or convention. i.e. the GGA homolog for the Z chromosome was numbered SAA19 as this was the last to be assigned despite being one of the largest chromosomes and to avoid confusion the highly conserved (across mammals and avian species) chromosome 4 was also labelled SAA4 despite being of medium size.



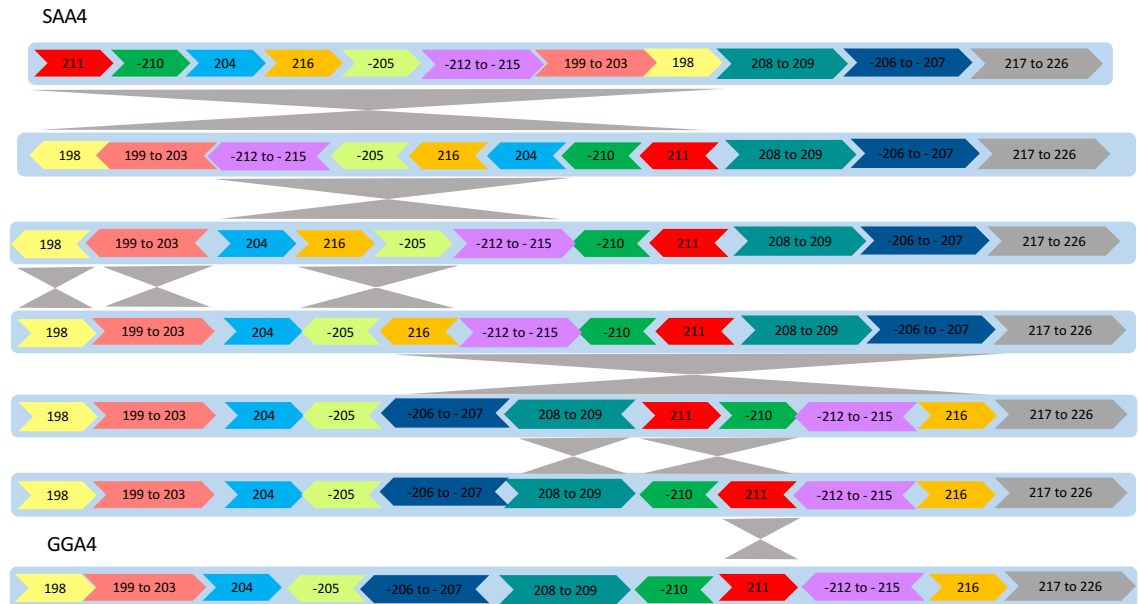
**Figure 6-10: Ideogram of saurian ancestor CARs (SAA) derived from MGRA2 with their GGA homologs illustrating interchromosomal rearrangement of GGA homologs relative to hypothetical ancestor.**

**6.4.4.4 Reconstruction of the Inter and Intrachromosomal Rearrangements that occurred between the Saurian Ancestor and the Chicken Genome**

Reconstructed CARs derived from MGRA2 were subsequently aligned against the extant genomes. The rearrangements between the saurian ancestor and each bird were then modelled using the model of maximum parsimony. A total of 49 inversions were identified between the saurian ancestor and the chicken genome along with 10 interchromosomal changes. Rearrangements are listed in Table 6-4 by ancestral chromosome and an example of the intrachromosomal rearrangements that occurred between SAA1 and GGA2 is shown in Figure 6-11. In addition, Figure 6-12 illustrates the changes that have occurred intrachromosomally between the saurian ancestor chromosome 4 and chicken chromosome 4. In this example, (as shown in Table 6-4) all of the changes appear to have occurred between the avian ancestor and the chicken.



**Figure 6-11: Intrachromosomal rearrangements between the saurian ancestor chromosome 1 (at the top of the image) and chicken chromosome 2 illustrating 5 inversions that have occurred between the two. The numbers in each block represent the msHSBs that have stayed together as a larger block of synteny without evidence of intra or interchromosomal rearrangement and a minus symbol indicates a change of direction.**



**Figure 6-12: Intrachromosomal rearrangements between the saurian ancestor chromosome 4 (at the top of the image) and chicken chromosome 4 illustrating 9 inversions that have occurred between the two.**

SAA Chr	Chromosomal rearrangements between:			
	Saurian ancestor to the avian ancestor		Saurian ancestor to the extant chicken	
	Interchromosomal	Intrachromosomal	Interchromosomal	Intrachromosomal
1	0	3	0	4
2	0	1	1	3
3	2	3	2	8
4	0	0	0	9
5	0	1	0	2
6	1	1	1	1
7	1	0	1	3
8	1	0	1	3
9	1	0	1	3
10	0	0	0	0
11	0	0	0	2
12	1	0	1	1
13	0	0	0	1
14	0	0	0	0
15	1	0	1	0
16	1	0	1	0
17	0	0	0	0
18	0	0	0	0
19	0	3	0	9
<b>Total</b>	<b>9</b>	<b>12</b>	<b>10</b>	<b>49</b>

**Table 6-4: Total number of inter and intrachromosomal rearrangements between the saurian ancestor and the avian ancestor and the saurian ancestor and the extant chicken genome.**

Of the interchromosomal rearrangements that were identified and listed in Table 6-4, a reciprocal translocation was identified between SAA6 and SAA7 which is illustrated below in Figure 6-13. Comparison of the data generated for the saurian ancestor and the avian ancestor revealed that almost all of the interchromosomal rearrangements (9 in total) had occurred prior to the avian ancestor. In terms of intrachromosomal events, it appears that between the saurian and the avian ancestor a total of 12 inversions occurred – 3 of which were on the ancestral chromosome from which the avian Z chromosome is derived.



**Figure 6-13:** Inter and intrachromosomal rearrangements between saurian ancestor chromosomes 6 and 7 resulting in a reciprocal translocation occurring between the two (highlighted in yellow) resulting in chromosomes 5 and 20 in the chicken.

Between the avian ancestor and the extant chicken genome a fusion occurred to form chromosome 1 in the chicken and a reciprocal translocation occurred between avian ancestor CARs that went onto become GGA chromosomes 1 and 7. The majority of rearrangements between these two ancestors were in fact intrachromosomal with a total of 37 inversions that appear to have occurred between the avian ancestor and the extant chicken genome. A summary of the total number of rearrangements between each ancestor is shown in Table 6-5.

	Chromosomal rearrangements between ancestral points	
	Interchromosomal	Intrachromosomal
Saurian Ancestor to Avian Ancestor	9	12
Avian Ancestor to Chicken	1	37
<b>Total</b>	10	49

**Table 6-5:** Number of the type of chromosomal rearrangement that occurred between the saurian ancestor, the avian ancestor and the extant chicken genome.

The rest of the rearrangements identified between the saurian ancestor and the extant chicken genome are illustrated in Figures S1 to S7 in the appendix.



**6.4.5 Specific aim 4e: To analyse gene ontology terms in the HSBs and EBRs of this dinosaur lineage to test the hypothesis that these relate to phenotypic characteristics associated with the dinosaurs**

The DAVID functional annotation tool was used in order to test the hypothesis that functional enrichment of GO terms within the msHSBs and EBRs identified in section 6.4.4.1 relates to established phenotypic characteristics associated with dinosaurs. Specifically, gene ontology of chicken-human one-to-one orthologous gene sets within the msHSBs found to be conserved in birds, lizards and opossum were analysed. Different scenarios were tested, the first of which was to test whether there was a difference in output when comparing GO terms analysed against a background list generated from the entire chicken genome or just those chicken genomes represented in the data set. The next scenario was to test whether the addition of flanking regions of 300kb to each EBR would alter the GO terms produced.

Within the msHSBs, significant enrichments were observed for GO terms relevant to transmembrane transport (symport) and signaling. Other msHSB-specific GO term enrichments appear to be related to synapse/neurotransmitter transport, nucleoside metabolism and use, cell morphogenesis and cytoskeleton, and sensory organ development. In the EBRs, DAVID outputs resulted in genes within these regions are enriched with GO terms for sequence variants affecting diversity of adult human height and nucleocytoplasmic transport. Other EBR-related GO term enrichments appear to be associated with nucleotide/ATP/metal binding, nucleoplasm part, chromatin/histone modification, RNA processing and splicing, and Golgi apparatus structure and transport. Table 6-6 illustrates the genes within the enrichment analysis that had some association with body size and their contemporary ontology terms at the time of writing,

Ensembl gene ID	Gene symbol	Gene name	HSA locus	Chicken ortholog	Gene Summary	Gene Function (related to body size)	Reference
ENSG00000111605	<i>CPSF6</i>	cleavage and polyadenylation specific factor 6	12q15	GGA1	Encodes for protein that is one subunit of a cleavage factor required for 3' RNA cleavage and polyadenylation processing.	Many sequence variants affecting diversity of human height, close to height associated SNP rs11177669	Gudbjartsson et al. 2008
ENSG00000166225	<i>FRS2</i>	fibroblast growth factor receptor substrate 2	12q15	GGA1	Targets signalling molecules to the plasma membrane in response to FGF stimulation to link receptor activation with MAPK and other signalling pathways essential for cell growth and differentiation	Essential for controlling process of skeletal development by mediating cellular responses; many sequence variants affecting diversity of human height, close to height associated SNP rs11177669	Su et al. 2014; Gudbjartsson et al. 2008
ENSG00000127337	<i>YEATS4</i>	YEATS domain containing 4	12q15	GGA1	Encodes for protein found in the nucleoli likely to represent a transcription factor thought to be required for RNA transcription. This gene has been shown to be amplified in tumours.	Candidate gene for short stature in humans and macrocephaly; many sequence variants affecting diversity of human height, close to height associated SNP rs11177669	Takenouchi et al. 2012; Gudbjartsson et al. 2008
ENSG00000166226	<i>CCT2</i>	chaperonin containing TCP1 subunit 2	12q15	GGA1	Encodes for protein that is a member of the chaperonin containing TCP1 complex (CCT), also known as the TCP1 ring complex (TRiC).	Associated with primordial short stature in humans; many sequence variants affecting diversity of human height, close to height associated SNP rs11177669	Hanson et al.2014; Gudbjartsson et al. 2008
ENSG00000198812	<i>LRR10</i>	leucine rich repeat containing 10	12q15	GGA1	Protein Coding gene with GO annotations related to actin binding and alpha-actinin binding	Many sequence variants affecting diversity of human height, close to height associated SNP rs11177669	Gudbjartsson et al. 2008
ENSG00000137177	<i>KIF13A</i>	kinesin family member 13A	6p22.3	GGA2	Encodes a member of the kinesin family of microtubule-based motor proteins that function in the positioning of endosomes	Many sequence variants affecting diversity of human height, close to height associated SNP rs12199222	Gudbjartsson et al. 2008
ENSG00000008988	<i>RPS20</i>	ribosomal protein S20	8q12.1	GGA2	Encodes a ribosomal protein that is a component of the 40S subunit.	Triggers stabilization and/or activation of p53; activation of p53 causes reduced growth and decreased body size; Genome-wide association study for birth weight in Nellore cattle points to orthologous genes affecting human and bovine height. Region also overlaps QTLs for birth weight, mature height, carcass weight, stature; many sequence variants affecting diversity of human height, close to height associated SNP rs10958476	Utsunomiya et al. 2013; Gudbjartsson et al. 2008
ENSG00000115137	<i>DNAJC27</i>	DnaJ heat shock protein family (Hsp40) member C27	2p23.3	GGA3	Protein Coding gene with GO annotations related to GTP binding and GTPase activity.	Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability	Tatton-Brown et al. 2014

ENSG00000138031	<i>ADCY3</i>	adenylate cyclase 3	2p23.3	GGA3	Encodes for adenylyl cyclase 3 which catalyses the formation of the secondary messenger cyclic adenosine monophosphate (cAMP). Widely expressed in various human tissues and may be involved in a number of physiological and pathophysiological metabolic processes.	Association with height-adjusted BMI; many sequence variants affecting diversity of human height, close to height associated SNP rs6733301	Stergiakouli et al. 2014; Gudbjartsson et al. 2008
ENSG00000119772	<i>DNMT3A</i>	DNA methyltransferase 3 alpha	2p23.3	GGA3	Encodes a protein that belongs to the serine/threonine protein kinase family, plays a role in regulation of fluid balance in the intestine.	Related to overgrowth syndrome and intellectual disability; many sequence variants affecting diversity of human height, close to height associated SNP rs6733301	Tatton-Brown et al. 2014; Gudbjartsson et al. 2008
ENSG00000196284	<i>SUPT3H</i>	SPT3 homolog, SAGA and STAGA complex component	6p21.1	GGA3	Protein Coding gene with pathways related to validated targets of C-MYC transcriptional activation and Chromatin organization.	Associated with height in Korean populations; many sequence variants affecting diversity of human height, close to height associated SNP rs9395066	Kim et al. 2010; Gudbjartsson et al. 2008
ENSG00000085382	<i>HACE1</i>	HECT domain-containing E3 ubiquitin ligase 1	6q16.3	GGA3	Encodes for protein involved in specific tagging of target proteins, leading to their subcellular localization or proteasomal degradation. The protein is a potential tumour suppressor and is involved in the pathophysiology of several tumours.	Related to neurodevelopment disorder with spasticity and abnormal gait; many sequence variants affecting diversity of human height, close to height associated SNP rs314268	Hollstein et al. 2015; Gudbjartsson et al. 2008
ENSG00000187772	<i>LIN28B</i>	lin-28 homolog B	6q16.3	GGA3	Encodes protein that belongs to the lin-28 family. Highly expressed in testis, fetal liver, placenta, and in primary human tumours and cancer cell lines.	Distinct Variants in gene influence Growth in Height from Birth to Adulthood; many sequence variants affecting diversity of human height, close to height associated SNP rs314263	Widén et al. 2010; Gudbjartsson et al. 2008
ENSG00000132429	<i>POPDC3</i>	popeye domain-containing protein 3	6q21	GGA3	Encodes a member of the POP family of proteins. Expressed in cardiac and skeletal muscle and may play important role during development.	Expressed in cardiac and skeletal muscle and may play important role in these tissues during development; many sequence variants affecting diversity of human height, close to height associated SNP rs314263	Refseq 2008; Gudbjartsson et al. 2008
ENSG00000112276	<i>BVES</i>	blood vessel epicardial substance	6q21	GGA3	Encodes a member of the POP family of proteins. Gene expressed in cardiac and skeletal muscle and may play an important role in development of these tissues.	This gene is expressed in cardiac and skeletal muscle and may play an important role in development of these tissues; many sequence variants affecting diversity of human height, close to height associated SNP rs314263	Refseq 2010; Gudbjartsson et al. 2008
ENSG00000198945	<i>L3MBTL3</i>	l(3)mbt-like 3 (Drosophila)	6q23.1	GGA3	Protein coding gene	Variants in this gene affect birth length and height in children; many sequence variants affecting diversity of human height, close to height associated SNP rs6899976	Paternoster et al. 2011; Gudbjartsson et al. 2008

ENSG00000181690	<i>PLAG1</i>	PLAG1 zinc finger	8q12.1	GGA3	Encodes a zinc finger protein with 2 putative nuclear localization signals. Developmentally regulated, and shown to be consistently rearranged in pleomorphic adenomas of the salivary glands.	Regulates several growth factors, including Igf2, a key regulator of body size; many sequence variants affecting diversity of human height, close to height associated SNP rs10958476	Karim et al. 2011; Gudbjartsson et al. 2008
ENSG00000138669	<i>PRKG2</i>	protein kinase, cGMP-dependent, type II	4q21.21	GGA4	Encodes a protein that belongs to the serine/threonine kinase family of proteins and plays a role in the regulation of fluid balance in the intestine.	Associated with copy number variants in patients with short stature; many sequence variants affecting diversity of human height, close to height associated SNP rs710841	van Duyvenvoorde et al. 2014; Gudbjartsson et al. 2008
ENSG00000138670	<i>RASGEF1B</i>	RasGEF domain family member 1B	4q21.22	GGA4	DNA methyltransferase. Protein localizes to cytoplasm and nucleus, expression is developmentally regulated	Identified as a promising candidate gene for short stature in Japanese populations; associated with growth restriction and development delay; many sequence variants affecting diversity of human height, close to height associated SNP rs710841	Komlósi et al. 2015; Harada et al. 2002; Friedman et al. 2006; Bonnet et al. 2010; Gudbjartsson et al. 2008
ENSG00000181195	<i>PENK</i>	proenkephalin	8q12.1	GGA4	Encodes a preproprotein that is processed to generate multiple products including opioids that modulate the perception of pain.	Genome-wide association study for birth weight in Nellore cattle points to orthologous genes affecting human and bovine height. Region also overlaps QTLs for birth weight, mature height, carcass weight, stature; many sequence variants affecting diversity of human height, close to height associated SNP rs10958476	Utsunomiya et al. 2013; Gudbjartsson et al. 2008
ENSG00000197959	<i>DNM3</i>	dynamamin 3	1q24.3	GGA8	Encodes for a member of guanosine triphosphate (GTP)-binding proteins involved in vesicular transport.	Associated with prenatal growth deficiency and a height associated SNP, rs678962	Burkhardt et al. 2011; Gudbjartsson et al. 2008
ENSG00000144535	<i>DIS3L2</i>	DIS3 like 3'-5' exoribonuclease 2	2q37.1	GGA9	Encodes for protein similar in sequence to 3'/5' exonucleolytic subunits of the RNA exosome responsible for degrading various RNA substrates.	Truncated gene related to patients both with short stature and skeletal overgrowth; GWAS of northwestern Europeans involves signaling pathway in the etiology of human height variation; many sequence variants affecting diversity of human height, close to height associated SNP rs749052	Tassano et al. 2013; Estrada et al. 2009; Gudbjartsson et al. 2008
ENSG00000144524	<i>COPS7B</i>	COP9 signalosome subunit 7B	2q37.1	GGA9	Protein Coding gene - related pathways are DNA Double-Strand Break Repair and Transcription-Coupled Nucleotide Excision Repair	GWAS many sequence variants affecting diversity of human height; many sequence variants affecting diversity of human height, close to height associated SNP rs749052	Gudbjartsson et al. 2008
ENSG00000156973	<i>PDE6D</i>	phosphodiesterase 6D	2q37.1	GGA9	Encodes for a key enzyme in the phototransduction cascade and thought to bind to proteins that target cilia.	Gene associated with Joubert Syndrome 22. Clinical features include growth abnormalities, small-for-dates baby, intrauterine growth retardation; many sequence variants affecting diversity of human height, close to height associated SNP rs749052	Thomas et al. 2014; Gudbjartsson et al. 2008

ENSG00000164615	<i>CAMLG</i>	calcium modulating ligand	5q31.1	GGA13	CAMLG is a membrane protein that binds to cyclophilin B and causes an influx of calcium.	Possible candidate for size selection in cattle; many sequence variants affecting diversity of human height, close to height associated SNP rs31198	Randhawa et al. 2015 ; Gudbjartsson et al. 2008
ENSG00000145833	<i>DDX46</i>	DEAD-box helicase 46	5q31.1	GGA13	Encodes member of the DEAD box protein family (putative RNA helicases). Implicated in cellular processes involving alteration of RNA secondary structure. Some members of this family are believed to be involved in embryogenesis, spermatogenesis, and cellular growth and division.	Possible candidate for size selection in cattle; many sequence variants affecting diversity of human height, close to height associated SNP rs31198	Randhawa et al. 2015 ; Gudbjartsson et al. 2008
ENSG00000128266	<i>GNAZ</i>	G protein subunit alpha z	22q11.22	GGA15	Encodes for protein that intermediates signal transduction in pertussis toxin-insensitive systems. May play a role in maintaining the ionic balance of cochlear fluids.	Many sequence variants affecting diversity of human height, close to height associated SNP rs5751614	Gudbjartsson et al. 2008
ENSG00000186716	<i>BCR</i>	BCR, RhoGEF and GTPase activating protein	22q11.23	GGA15	Breakpoint withing BCR related to Philadelphia chromosome and Chronic myeloid leukemia	Forms a fusion product with fibroblast growth factor receptor 1 - causes Chronic Myeloid Leukemia associated with short stature; many sequence variants affecting diversity of human height, close to height associated SNP rs5751614	Kossiva et al. 2010; Gudbjartsson et al. 2008
ENSG00000176390	<i>CRLF3</i>	cytokine receptor like factor 3	17q11.2	GGA18	Encodes a cytokine receptor-like factor that may negatively regulate cell cycle progression at the G0/G1 phase	Associated in European-American height cohort; many sequence variants affecting diversity of human height, close to height associated SNP rs3760318	Zhao et al. 2010; Gudbjartsson et al. 2008
ENSG00000153933	<i>DGKE</i>	diacylglycerol kinase epsilon	17q22	GGA18	DGKE gene encodes diacylglycerol kinase-epsilon, an intracellular lipid kinase that phosphorylates diacylglycerol (DAG) to phosphatidic acid.	Associated with height; many sequence variants affecting diversity of human height, close to height associated SNP rs4794665	Hirschhorn and Lettre 2009; Gudbjartsson et al. 2008
ENSG00000125966	<i>MMP24</i>	matrix metalloproteinase 24	20q11.22	GGA20	Protein Coding gene.with pathways related to degradation of the extracellular matrix.	GWAS Association with height variation; many sequence variants affecting diversity of human height, close to height associated SNP rs4911494	Gadelha Pereira Fontenele et al. 2015; Gudbjartsson et al. 2008

**Table 6-6: Sequence variants affecting diversity of height identified as having significant enrichments in the EBRs of the saurian karyotype (adapted from Gudbjartsson et al. 2008).**

## 6.5 Discussion

### 6.5.1 Reconstruction of Ancestral Genomes

The results presented here represent the most detailed reconstruction of reptilian (avian and non-avian) comparative cytogenetics and gross genomic evolution to date. They provide a more detailed path of karyotypic change evolution than could be achieved by zoo-FISH analysis alone and demonstrate proof of principle from which further studies (in the same animal group or in others) can ensue. Central to the analysis (both avian specific and dinosaur) was a highly interactive avian genome dataset from the Evolution Highway comparative chromosome browser developed by Denis Larkin's lab that can be used to compare the chromosome organisation of multiple species. Indeed, the ability to align genomes of multiple species against one another using Evolution Highway and similar approaches demonstrated in the comparison of mammalian species (Murphy et al. 2005) were the basis for which the bioinformatic approach described here was performed. The ultimate aim for Evolution Highway is that, in chromosomes for all avian species uploaded, HSBs will be displayed with reference to the chromosome number, as is currently the case for turkey, zebra finch, duck and collared flycatcher, rather than by scaffold. Most genomes (avian and mammalian) are however currently assembled in scaffold form, and are therefore somewhat fragmented. As discussed in section 1.1.4, there are several strategies through which this might be facilitated: (a) by improving scaffold sizes e.g. through optical mapping and PacBio (although these still fall short of full chromosome level assembly); (b) by linkage to radiation hybrid (RH) maps such as was achieved for duck (Rao et al. 2012); (c) by association with known genetic (linkage) or other physical maps (e.g. Wallis et al. 2004); (d) by use of algorithms (such as RACA) to order and orient scaffolds into longer chromosomal fragments (e) by systematic FISH mapping to chromosomes of orthologous clones derived from the individual scaffolds. The latter was demonstrated in turtle and lizard metaphases in this chapter and in other avian species in the two previous chapters. Given the efforts that have been made to sequence the genomes of birds, mammals and reptiles recently by current technologies (Zhang et al. 2014a) it is unclear how many of them will be re-sequenced using these newer technologies given funding and logistical constraints.

Previous iterations of the data modelled in this chapter to reconstruct the saurian ancestor were run with the scaffold based genome of the boa constrictor snake (*Boa constrictor*) instead of the lizard. Despite providing greater coverage of the genome overall, this analysis was discarded in favour of the lizard based data due to the highly fragmented nature of the snake genome. The

high degree of fragmentation in scaffold based genomes makes it difficult to distinguish 'real' EBRs from scaffold breaks, thereby reducing the validity of any subsequent GO analysis of both EBRs and HSBs. Ancestral genome reconstruction tools, largely designed for mammalian chromosome level genomes (Jarvis 2016) are also unable to reliably manage scaffold level genomes, hence the pragmatic decision taken here to focus on the best assembled reptile genome (the lizard) despite the lack of chromosomal level assembly for the microchromosomes.

### **6.5.2 Ostrich Genome Inconsistencies**

Early cytogenetic studies suggest that, for the majority of avian species, karyotypic patterns are broadly similar to one another (Christidis 1990; Griffin et al. 2007; Völker et al. 2010; Skinner & Griffin 2012). The question of whether the conserved interchromosomal synteny reported for the macrochromosomes applies to the microchromosomes has previously been beyond the resolution of contemporary methodology. This thesis, in the last three chapters, is the first to classify intermicrochromosomal rearrangements in any species. Evidence is provided that interchromosomal rearrangements are nonetheless rare, except in cases (around 1/3 of species) where we already knew that karyotypes were highly rearranged (Christidis 1990; Griffin et al. 2007). The ostrich is apparently the exception however according to results presented here. Avian-typical patterns in ostrich, rhea and emu have been shown with comparative chromosome painting data (Shetty et al. 1999; Guttenbach et al. 2003; Nishida et al. 2007) and in the previous chapter for ostrich. In this chapter however, this idea is challenged. While not analysis that I performed personally, the data generated here produced assembly reads from single ostrich chromosomes that spanned more than one chicken chromosome on 26 occasions, indicating the presence of 26 ostrich interchromosomal rearrangements compared to the avian ancestor. The extent to which this represents assembly errors or genuine chromosome rearrangements is still an open question, however the technology developed in three of the chapters of this thesis should be able to resolve this through design and hybridisation of conserved BACs targeted to the purportedly translocated regions.

### **6.5.3 Chicken is Most Closely Related to the Dinosaur Ancestor**

From the current analysis it appears that the chicken has undergone the fewest chromosomal changes compared to the ancestor of the species studied. There are similarities between this study and one performed by Zhou and colleagues who examined sex chromosome evolution (Zhou et al. 2014). The data presented here demonstrates that compared to the avian ancestor,

chicken autosomes 1–5 have rearranged the least while Zhou's group conclude that avian ancestral sex chromosome organisation is closer to that of the Paleognathae (ostrich and emu) in that they demonstrate less degradation and a closer synteny to the lizard. It only was only possible to examine the Z chromosome in the Neognathae in the study presented in this chapter meaning that further studies will be required to identify patterns of sex chromosome and autosome preservation among the different lineages.

A further interesting question also arises of whether the chicken, having undergone the fewest chromosomal changes, also has the least adaptive changes compared to the avian ancestor. In other words, was the avian ancestor more like a chicken than most other birds? Dinosaurs along the theropod lineage (the ancestors of modern birds) were bipedal, terrestrial and relatively small (small size being an immediate pre-adaptation to flight) with limited flying ability, similar to the *Galliformes* (Witmer 2002). As described in section 1.6.4, the oldest fossil representative of neornithes (modern birds) is widely considered to be *Vegavis* – a likely representative of the *Galloansere* (Chiappe & Dyke 2006; Clarke et al. 2005). However, as already discussed, inherent biases in the fossil record may distort the assumptions that can be made about some of these early ancestors. For this reason, additional data points provided by studies such as those presented here into genomic characteristics of these ancestors may complement an otherwise imperfect fossil record.

#### **6.5.4 The Saurian Ancestor**

Comparison of chromosome number in most reptiles and outgroups such as amphibians and mammals suggest that the overall chromosome number of the saurian ancestor was around 40, half of which were likely to be microchromosomes (Organ et al. 2008). This is not dissimilar in number to the average of 20-24 microchromosomes found in most extant non-avian reptiles (lizards and snakes) (Beçak et al. 1964; Alföldi et al. 2011). As described in section 1.4, Burt (2002) proposed that most microchromosomes were already present in the common dinosaur ancestor that gave rise to birds. Burt argues that this ancestor probably had the small genome size characteristic of birds and a karyotype of around  $2n=60$  (20 pairs of microchromosomes) with chromosomal fission being the mechanism that created the remainder of the avian microchromosomes (Burt 2002). From a cytogenetic point of view, this hypothesis is supported by the high degree of similarity seen between the turtle and the chicken karyotype. Zoo-FISH studies using chicken macrochromosome paints on the chromosomes of the Chinese soft-shelled turtle (*Pelodiscus sinensis*) ( $2n=66$ ) (Matsuda et al. 2005) and red-eared slider



(*Trachemys scripta*) ( $2n=50$ ) (Kasai et al. 2012) suggest that the characteristically avian karyotype was more or less structurally in place 260 mya with little interchromosomal change since, other than a series of fissions in the avian lineage to reach the average diploid number of 80 seen in birds. Cytogenetic mapping of the painted turtle (*Chrysemys picta*) by Badenhorst and colleagues revealed a higher degree of interchromosomal rearrangement than previously assumed, including the fusions of GGA12, 13 and 26 to macrochromosomes as identified in this chapter as well as the interchromosomal rearrangement of GGA22 to the peri-centromeric region of a macrochromosome. The overall turtle pattern was still however highly syntenic with the karyotype of the chicken (Badenhorst et al. 2015). It is highly probable then that the common ancestor of birds and turtles (around 260 mya – see discussion on this issue in section 1.6.1) had a similar karyotype. Whether this karyotypic structure was in place in the common ancestor of the archosauromorphs and the lepidosaurs 275 mya has, until now, been difficult to investigate due to the limitations of zoo-FISH across this degree of phylogenetic distance.

Interestingly, cytogenetic data from this chapter and chapter 4 reveal that across the anole lizard, the red-eared slider turtle and all avian species tested, there is one microchromosome homolog (GGA24) which appears to remain intact with no evidence of fusion or fission over the last 275 million years. When the lineage specific fusions of the *Falconiformes* and the *Psittaciformes* are excluded, the results are more striking with evidence of microchromosome conservation of 9 GGA homologs, with a further 3 that appear to be 'proto-microchromosomes' in the lizard and turtle which are likely to have been subject to a fission in the avian lineage. Using the FISH data alone, would suggest that the ancestral saurian genome probably had a diploid number of around 44, more than half of which were microchromosomes consistent with that hypothesized by Organ et al. (2008).

The overall structure of the archosauromorph ancestor was therefore likely to be close to that of the Chinese painted turtle (*Chrysemys picta*) with an average of 8-9 macrochromosomes and 24-25 pairs of microchromosomes ( $2n \sim 64-68$ ). The smaller number of overall chromosomes ( $2n \sim 30$ ) and absence of microchromosomes in the crocodylians suggests a large degree of chromosomal fusion in the 23 species that represent this group. Recent sequencing of the anole lizard genome including assembly to a chromosomal level for chromosomes 1 to 6 (Alföldi et al. 2011) facilitated the bioinformatics approach to ancestral genome reconstruction presented here, however full ancestral reconstruction using this method was limited due to the lack of chromosome level assembly for the anole microchromosomes. Remarkably, Alföldi and

colleagues found a direct syntenic correlation at a sequence level between the microchromosomes of *A. carolinensis* and the chicken genome with all but one anole microchromosome corresponding to a single chicken microchromosome. Given that the anole genome contains 12 pairs of microchromosomes compared to the 28-30 seen in most avian species, they suggest that these 12 microchromosomes may have arisen in the reptile ancestor with remaining avian microchromosomes being derived in the bird lineage (Alföldi et al. 2011). The results presented in this thesis support this theory with evidence already of at least three 'proto-microchromosomes' identified in the anole lizard using FISH.

### **6.5.5 Genome Evolution from 275 mya to 100 mya**

The primary mechanisms for chromosomal rearrangement in this lineage appear to have been fissions and intrachromosomal rearrangements (inversions). Results presented in section 6.4.4.4 illustrate that the majority of the interchromosomal rearrangements appear to have occurred between the saurian and the avian ancestor 100 mya. From this point onwards it appears that the main mechanism of change was intrachromosomal. Given the high degree of conservation revealed between birds and turtles through chromosome painting (Matsuda et al. 2005) and through the results presented in this chapter, it is reasonable to make the assumption that the majority of the interchromosomal rearrangements actually occurred prior to the testudine divergence 260 mya. This would suggest that the characteristic avian style karyotype was established at very early stage of theropod evolution and was therefore a feature evident in the dinosaur lineage and probably also in the pterosaurs. It appears therefore that since the lepidosaurs diverged from the archosauromorphs 275 mya, the archosauromorph genome has remained remarkably constrained by size, possibly due to the later requirements of flight (as would certainly be the case for the pterosaurs - this is discussed further in the general discussion). The crocodylians seem to be the exception to this pattern, perhaps due to fewer environmental constraints to genome size which allowed the accumulation of repetitive elements, thereby facilitating the high degree of fusion seen in the crocodile karyotype. The small genome size of saurischian dinosaurs identified through osteocyte size by Organ et al. (2007) supports the notion that this has been a long established feature in this lineage. The larger genome size seen in the lepidosaurs and the testudines (average 3.2Gb), which reduces in the avian lineage to an average of 1.4Gb, along with a correspondingly reduced level of interspersed repeats from 7% to 4% (Janes et al, 2010), suggests that there was a reduction in size after the turtle divergence 260mya. The period post turtle divergence corresponds to that of the dinosaur radiation when phenotypic adaptations such as the pneumatisation of bones

(Farmer & Sanders 2010) and a move to a bipedal stance coincided with a time of extraordinary speciation (Benton et al. 2014). Survival through multiple extinction events when most other reptilian and mammalian species were decimated also suggests some inherent advantages to this group of animals, perhaps due to or aided by this highly compact genome. Certainly the extraordinary phenotypic diversity seen among birds with over 10,000 extant species corresponds with this highly unique karyotype.

Results from this study suggest that the most parsimonious path from saurian to archosauromorph ancestor was via a series of 9 interchromosomal rearrangements of which there was a net result of one fusion (bringing the ancestral chromosome number to 20), and 12 intrachromosomal rearrangements. The lack of full chromosomal level genome coverage of any reptile hinders the generation of a full bioinformatics appraisal of the ancestral karyotype at this phylogenetic distance, however the further data provided by the FISH suggests that there were at least 14 microchromosomes present at this stage and an additional 3 microchromosomes that were ancestrally fused to a macrochromosome; 10 of these 17 microchromosomes are not covered by the bioinformatic ancestral approach. Taken together then, with the 20 ancestral chromosomes generated bioinformatically gives an overall chromosome number of 30 pairs, increasing the overall chromosome number to  $2n=60$ , much like that of the chromosome number of most turtles (e.g. *Pelodiscus sinensis*-  $2n=66$ ).

As previously described, there is evidence that some groups exhibit extensive interchromosomal rearrangement such as the crocodiles, (and much later) the parrots and the falcons it is likely that the overall structure probably stayed generally the same from this point onward other than a series of ~10 fissions that must have occurred to bring the overall diploid number from  $2n=60$  to  $2n=80$  (but with most chromosomes staying intact). In this study, three of these have been identified – macrochromosomes that then fissioned to become GGA 12, 13 and 26. The results also suggest that at least 12 intrachromosomal rearrangements occurred in this period suggesting that the overwhelming mechanism of change at this time point and at the level detectable here was interchromosomal. It appears then that at this point the overall karyotype structure became fixed with the predominant mechanism of change being through intrachromosomal rearrangement with evidence of at least a further 26 inversions occurring to form the extant chicken genome.

### 6.5.6 Gene Ontology Analysis

As mentioned in section 1.5.3 gene ontology analysis of HSBs and EBRs provide clues to the biological significance of evolutionary changes within specific lineages leading to the phenotypes characteristic of those species. In an effort to discover the biological significance behind these changes, Farré et al. (2016) analysed the GO terms present in EBRs that are associated with specific avian adaptive features in individual species, finding (among other results) significant enrichment categories for forebrain development in the budgerigar (consistent with vocal-learning) along with enrichments for avian specific features in the HSBs.

In this study, GO analysis performed on the HSBs present from the saurian ancestor through to the avian ancestor suggested significant enrichments within HSBs for cellular functions including transmembrane transport, cell morphogenesis and neurotransmitter transport. Given that HSBs are generally considered to be enriched for GO terms related to phenotypic features that remain constant (Larkin et al. 2009), the results presented here are consistent with this hypothesis. Realistically however, transmembrane and neurotransmitter transport as well as cell morphogenesis are generic features that are difficult to attach any biological significance of these being saurian, archosaurian, dinosaurian or avian.

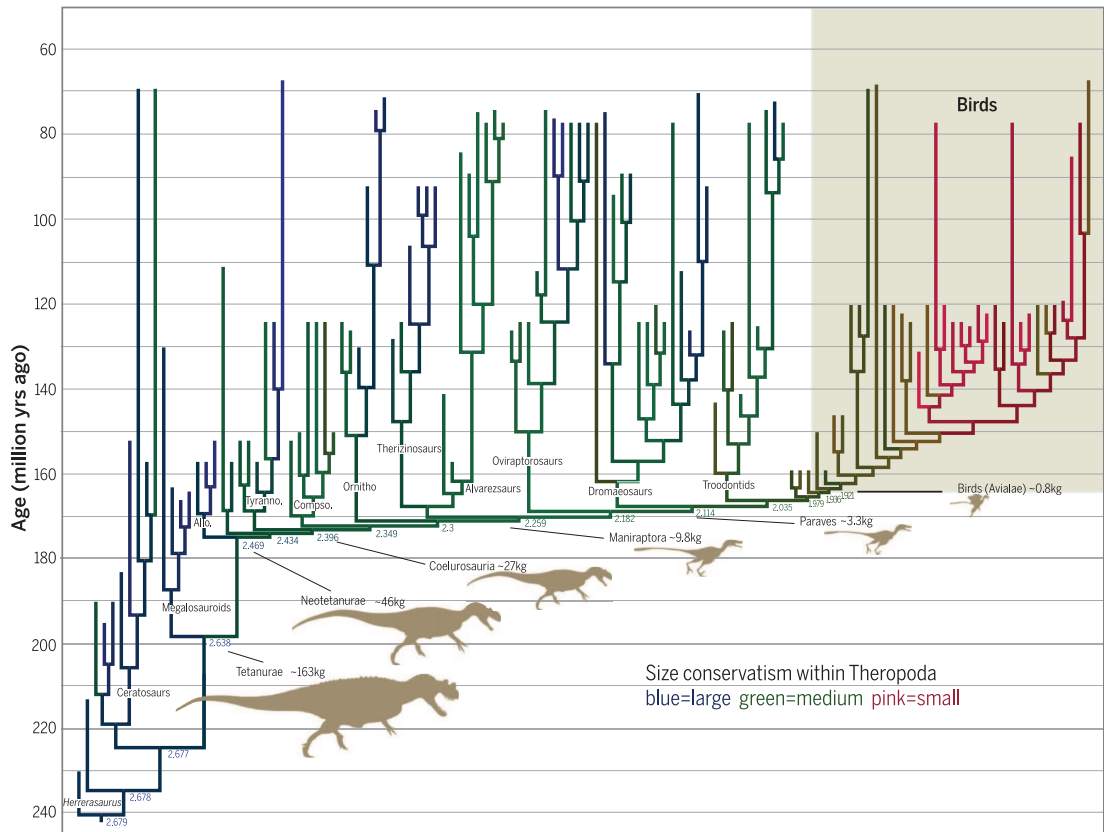
The absence of inter-microchromosomal rearrangements both between avian species and along the dinosaur lineage supports the hypothesis that, at an interchromosomal level, microchromosomes represent blocks of conserved interchromosomal synteny, (suggesting that these are essentially large HSBs). When testing the hypothesis therefore that there are enrichments for genes that have an evolutionary advantage to staying together as a single microchromosome, only an enrichment for immune genes on chromosome 16 (consistent with the presence of the MHC) was found. These results suggest that while microchromosomes represent highly conserved blocks of interchromosomal synteny there is limited evidence to support the hypothesis that this is due to a clustering of associated genes on the same chromosome. Species that exhibit a high degree of interchromosomal rearrangement (mammals, reptiles and amphibians) all tend to have large, repeat-rich genomes that, as described in section 1.5, appear to correlate with a higher rate of rearrangement. The results presented in this thesis suggest that some avian lineages (such as the parrots) also undergo a similar degree of chromosomal change but without the correspondingly large, repeat rich genome. Instead, comparisons of the zebra finch and the budgerigar suggest that the high chromosomal mutation rates seen in both lineages may in fact be changes that have occurred

in response to the exploitation of evolutionary niches, which ultimately end in fixed interchromosomal rearrangements. In the majority of other bird species however, it appears that such fixation is prevented, resulting in maintenance of an overall stable avian karyotype.

Perhaps more interesting biologically however was the fact that GO analysis of the dinosaur EBRs revealed significant enrichment for genes associated with sequence variants related to body size (human height given the data set used chicken-human orthologues). In fact, despite modelling a number of different analyses, including variation in the size of EBR flanking regions and differences in background gene list, enrichment for size (height) was always significant and consistent throughout all of the analyses. Sankoff (2009) stated that EBRs are where the 'action' in genome evolution takes place and therefore it could be expected that EBRs are enriched for GO terms where there is an evolutionary advantage for rearrangement of the genes associating them with phenotypic features that are under rapid change, particularly those that are important for adaptation (Larkin et al. 2009). In this example, if genes for height were present in EBRs that were involved in dinosaur evolution then we would expect this group of animals to have undergone rapid change in body size during the period of study.

It is widely acknowledged from paleontological records that theropod dinosaurs grew rapidly in the early phase of their radiation. Fossil evidence indicates that around 230 mya the mean body mass within this group was around 10kg (Benson et al. 2014). In the following 40 million years, body size increased to up to 160kg, prior to a pattern of prolonged decrease in body size between 200 and 160 mya. A recent study by Lee et al (2014) found a pattern of sustained miniaturization and accelerated skeletal innovation in the bird stem lineage as illustrated in Figure 6-14. Using femur length as a means of inferring body mass in dinosaur fossils, the authors concluded that the effect was particularly pronounced within the stem lineage that led to birds and suggested that the development of feathers along the bird stem may have facilitated the reduction in body size by providing an efficient means of body insulation (Lee et al. 2014). Similar patterns were found by Benson and colleagues (2014) who concluded that this reduction in body size ultimately to around 100g is likely to have contributed to the ecological radiation of Mesozoic birds and to their survival (along with small squamates and mammals) through the K-Pg extinction event while non-avian dinosaurs with higher body mass perished (Benson et al. 2014). The rapid evolutionary rates identified in the avian stem lineage in both of these studies (Benson et al. 2014; Lee et al. 2014) may have contributed to the morphological and ecological diversity seen among these species. The results found in this chapter for enrichments in genes

related to body size change therefore correspond to these paleontology results; in addition, the rapid rates of evolutionary change identified by these authors may also correspond to the presence of the 'avian-style' genome found here.



**Figure 6-14: Illustration of body size change among the theropod lineage from their emergence around 235mya to the divergence of modern birds (Lee et al. 2014).**

### 6.5.7 Conclusions

Results presented here suggest that a combined bioinformatics and molecular cytogenetic approach is a powerful means of ancestral genome reconstruction where both methods complement each other in order to extend the resolution for which ancestral karyotype reconstruction can be performed. Achieving a level of detail not possible with each method in isolation, this approach provides proof of principle from which further studies of genome evolution and comparative genomics can follow as well as providing a glimpse into the past to gain a picture of avian and dinosaurian evolution.

In summary, it appears that around 275 mya, the classic 'avian' karyotype was fundamentally in place and remained largely intact in this lineage until the turtle divergence 260 mya. At this point a reduction in interspersed elements and corresponding genome size reduction along with a

series of chromosomal fusions resulted in the avian genome structure similar to that of modern birds. This karyotypic structure is therefore likely to have been present in most dinosaurs, perhaps due to, or a contributory factor to, the evolution of lighter bones and ultimately the requirements for flight. The gene enrichments associated with body size change found in EBRs - areas where greater recombination takes place, also correspond to the reduction in body size seen among the dinosaurs suggesting that chromosome rearrangement may have been a driver for body size change.

## 7 General Discussion

This thesis has made a significant contribution towards the understanding of cytogenetically associated reproductive issues and reproductive isolation, both from an individual organism perspective and from the point of view of evolution and speciation. From a technical standpoint, it demonstrates that tools developed to study one aspect (e.g. translocations in individual animals) are applicable for the study of the other (i.e. chromosomal changes in gross genomic evolution). From a biological standpoint, this thesis also provides examples that chromosomal changes that can lead to infertility can also play a part in speciation. Finally, the work provides mechanistic insight into the biological reasons why chromosomal rearrangements occur in both extant and extinct species.

As mentioned in the general introduction, molecular cytogenetics is predominantly used a) to diagnose genetic disease, b) to facilitate genomic mapping and c) to study genome evolution. This thesis contains aspects of each of these three elements. It provides insight into the formation of and subsequent evolution, of a phylogenetic class (birds) with a signature karyotype - unusually driven by intra- rather than interchromosomal change. Furthermore, study of the evolution of long dead (but nonetheless fascinating) animals, the dinosaurs, was made possible through the examination of recently assembled genomes of extant species.

Specifically, the thesis was successful in the pursuit of its stated aims as follows:

1. Sub-telomeric sequences from pig, cattle and chicken genome assemblies were isolated, BAC probes selected and labelled and multiprobe devices developed for all three species. Two of these (pig and chicken) became commercial products (with cattle in development); the pig device being taken forward as a routine screening tool that proved effective in screening for both overt and cryptic translocations (specific aims 1 and 2).
2. The chicken device developed in specific aim 2 was tested on a range of species proving to be an effective tool for screening of evolutionary chromosome rearrangements among *Galliformes*. Further development of this tool through the incorporation of probes containing evolutionary conserved regions (identified by colleagues at the Royal Veterinary College London) allowed for analysis of interchromosomal changes of



phylogenetically distant birds. These worked successfully on all birds attempted, confirming the hypothesis that microchromosomes remain largely conserved in most species, either as discrete entities (most birds) or fused 'intact' to form larger chromosomes in the falcon and parrot species under investigation.

3. The technology developed in specific aim 2 (evolutionarily conserved probes) was successfully used to upgrade the scaffold based genome assemblies of pigeon and peregrine falcon to chromosome level, similar to that achieved by more traditional (Sanger) approaches. Tools were made available to do the same on a range of other avian species and proof of principle established for application to any fragmented animal genome assembly.
4. Bioinformatic tools in conjunction with resources developed in specific aims 1, 2 and 3 were used to re-create the overall likely genome structure (karyotype) of both saurian and avian ancestors. The gross evolutionary changes that occurred along the dinosaur (Saurischian, Therapoda, Maniraptora, Paraves, Avialae) lineage and several avian (*Galliformes*, *Anseriformes*, *Psittaciformes*, *Ratite*) lineages were retraced. Gene ontology analysis of multi-species homologous synteny blocks (msHSBs) and evolutionary breakpoint regions (EBRs) of chromosomes revealed clade specific enrichment for genes involved in chromosome rearrangement and formation of the signature fragmented karyotype of birds and dinosaurs. Most notably, EBR enrichment for genes related to change in body size was identified, consistent with the rapid change in dinosaur morphology (size) during that period.

### **7.1.1 Tools for Studying Chromosome Rearrangements in Mammals and Birds (Chapters 3-4)**

The tools developed specifically to screen for translocations in pigs (chapter 3, specific aim 1) and evolutionary changes in birds (chapter 4, specific aim 2) have now been fully commercialised by collaborators at Cytocell and are available to purchase directly from the company. The porcine device is routinely used in our lab and other labs to screen for chromosomal rearrangements in pigs. Our screening service alone has also expanded from one UK based breeding company to a further four large European breeding centres, in part due to the success of the product which is considered to be the 'gold-standard' method of translocation screening

and is advertised by the breeding companies to their clients as a measure of their quality control. An increasing interest in the effect of translocations on fertility in the cattle breeding industry suggests that there will be a similar high level of demand for the equivalent cattle device once this has been completed. Marketing materials developed for the porcine device are illustrated in Figure-7-1.

**Cytocell**  
an OGT company  
multiprobe

**OGT**

## Chromoprobe Multiprobe® Porcine

The Cytocell Chromoprobe Multiprobe Porcine is a useful tool in the pig breeding industry to identify translocations that have implications in fertility. The system allows the user to perform multiple FISH experiments on a single slide and to detect chromosomal aberrations (aneuploidy, chromosome breakage and rearrangements) present in the pig karyotype. The Chromoprobe Multiprobe Porcine utilizes directly-labelled locus-specific subtelomeric BAC probes, allowing the visualisation of any chromosomal aberrations under the fluorescence microscope without expert knowledge of the pig karyotype.

**The Chromoprobe Multiprobe Porcine delivers:**

- **Optimised and simple FISH protocol** – A simple FISH hybridisation without pepsin, RNase treatment or paraformaldehyde fixation and wash procedure, allows simultaneous analysis of the entire genome on one slide
- **Easy to use** – Template slides with marked areas corresponding to squares of the Multiprobe device allow cell cultures to optimally contact the probes during hybridisation for high-quality results
- **Sample flexibility** – The Chromoprobe Multiprobe Porcine is developed to work with metaphase and interphase cultures prepared in standard 3:1 methanol, acetic acid fixative
- **Complete system** – The Chromoprobe Multiprobe Porcine includes all ancillary reagents required (i.e. DAPI and hybridisation mix) for straightforward processing

1:q	2:q	3:q	4:q	5:q	6:q	7:q	8:q
9:q	10:q	11:q	12:q	13:q	14:q	15:q	16:q
17:q	18:q	X:q					

Optimised and simple protocol

Easy to use

Sample flexibility

Complete system

*"The Chromoprobe Multiprobe Porcine is in routine use in our laboratory and has allowed us to identify a cryptic translocation that is only detectable using this technology. This finding delivered significant financial implications to a large pig breeding program. Standard karyotyping tests failed to detect the rearrangement at the original testing regimen."*  
Professor Darren Griffin, University of Kent, Canterbury, UK

**Cytocell ZooFISH myProbes**

- The Chromoprobe Multiprobe Porcine can be customised according to requirements
- Individual subtelomeres or any other specific combinations of probes can be developed as required – both as devices or in traditional liquid form
- Porcine paints 1–18 X:Y are also available as either a Multiprobe device or individually labelled (in red and green) or multicolour (red green and blue) liquid probes

**Oxford Gene Technology**  
Biogenetics Science Park, Begbroke Hill, Woodstock Road, Begbroke, Oxfordshire, OX5 1PF, UK  
T: +44(0)1865 958828 (US: 914-467-5285) E: contact@ogt.com W: www.ogt.com

**OGT**


Figure-7-1:Marketing material created by Cytocell for the Porcine multiprobe device.

As a result of the efforts outlined in chapter 3, it is hoped that a cattle-based device will appear in the Cytocell catalogue in the next year.

The chicken multiprobe device has also generated interest within the avian genomics community and has been successfully adapted by Cytocell to incorporate additional probes in order to map viral integration sites within the chicken genome, thereby illustrating a large degree of flexibility for future bespoke projects. It has also now formed the basis of two teaching workshops held at Sharjah University in the United Arab Emirates. Marketing materials developed for the chicken device are illustrated in Figure 7-2.




**Cytocell**  
an OGT company  
multiprobe



ogt

---

## Chromoprobe Multiprobe® Chicken





Optimised and simple protocol

Easy to use

Sample flexibility

Complete system



## Chromoprobe Multiprobe Chicken

The chicken genome consists of a large number of microchromosomes; therefore, analysis of the chicken genome by classical cytogenetics is extremely complex. The Cytocell Chromoprobe Multiprobe Chicken is designed to identify complex chromosome rearrangements present in the chicken karyotype\* (1-28, Z and W). The system utilises directly-labelled whole chromosome painting (WCP) probes and locus-specific subtelomere BAC probes that allow the visualisation of chromosomal rearrangements under the fluorescent microscope without expert knowledge of the chicken karyotype.

\* The chicken karyotype consists of 38 autosomal chromosomes plus Z and W sex chromosomes, for which 39 pairs of autosomal chromosomes and Z and W have been sequenced. Cytocell's Chicken Multiprobe system contains probes for chromosomes that have sequencing data available.

**The Chromoprobe Multiprobe Chicken delivers:**

- **Optimised and simple protocol** — A simple FISH hybridisation without pre- or RNase treatment or paraformaldehyde fixation and wash procedures, allows simultaneous analysis of chicken chromosomes 1-28, Z and W on one slide
- **Easy to use** — Template slides with marked areas corresponding to squares of the Multiprobe device, allow cell cultures to optimally contact the probes during hybridisation for high quality results
- **Sample flexibility** — The Chromoprobe Multiprobe Chicken is developed to work with metaphase and interphase cultures prepared in standard 3:1 methanol, acetic acid fixative. It has been shown to work with metaphase cultures of other avian species (Galliformes) therefore suggesting the device may be useful in comparative genomic studies<sup>1,2</sup>
- **Complete system** — The Chromoprobe Multiprobe Chicken includes all ancillary reagents (i.e DAPI and hybridisation mix) for straightforward processing

10:pq	11:pq	12:pq	13:pq	14:pq	15:pq	16:pq	17:pq
1,4,3	18:pq	2,5,8	19:pq	11,7,9	20:pq	Z,W	21:pq
22:pq	23:pq	24:pq	25:pq	26:pq	27:pq	28:pq	

Figure 1: The Chromoprobe Multiprobe Chicken device layout. Each square of the device carries either chicken chromosome paints for three different chromosomes (macro) labelled in red, green and aqua (Texas Red®, FITC and aqua species, respectively), or subtelomeric BAC probes labelled in red and green fluorochromes (Texas Red, FITC). The probes are reversibly dried on to the first 23 squares of the device using a proprietary process. FISH signals are visible with a DAPI/FITC/Texas Red® triple filter or individually through specific single fluorescence filters.

"We utilised a customised Chromoprobe Multiprobe Chicken to map the chromosomal locations of viral integration sites. The customisation process was simple and allowed us to tailor the product to suit our precise requirements."  
Professor Venugopal Nar, Pirbright Institute, UK

**Cytocell ZooFISH myProbes**

- The Chromoprobe Multiprobe Chicken can be customised according to requirements.

**References**

1. Lefebvre, P.E. et al (2014) Novel tools for characterizing inter and intra chromosomal rearrangements in avian microchromosomes. Chromosome Res.
2. Romanos, M.N. et al (2014) Reconstruction of gross avian genome structure: reorganisation and evolution suggests that the chicken lineage most closely resembles the dinosaur avian ancestor. BMC Genomics.

**Oxford Gene Technology**  
Biological Science Park, Begbroke Hill, Woodstock Road, Begbroke, Oxfordshire, OX8 1PF, UK  
T: +44(0)1865 856958 (US: 914-467-5285) E: contact@ogt.com W: www.ogt.com

The information on this page, for use in advertising or other promotional material, is available in the US.  
This document and its contents are © Oxford Gene Technology © 2016. All rights reserved. OGT is a trademark of Oxford Gene Technology. P-Limita, Cytosol, myProbes, Multiprobe and Texas Red are trademarks of Cytocell Ltd.



1600187 01/16

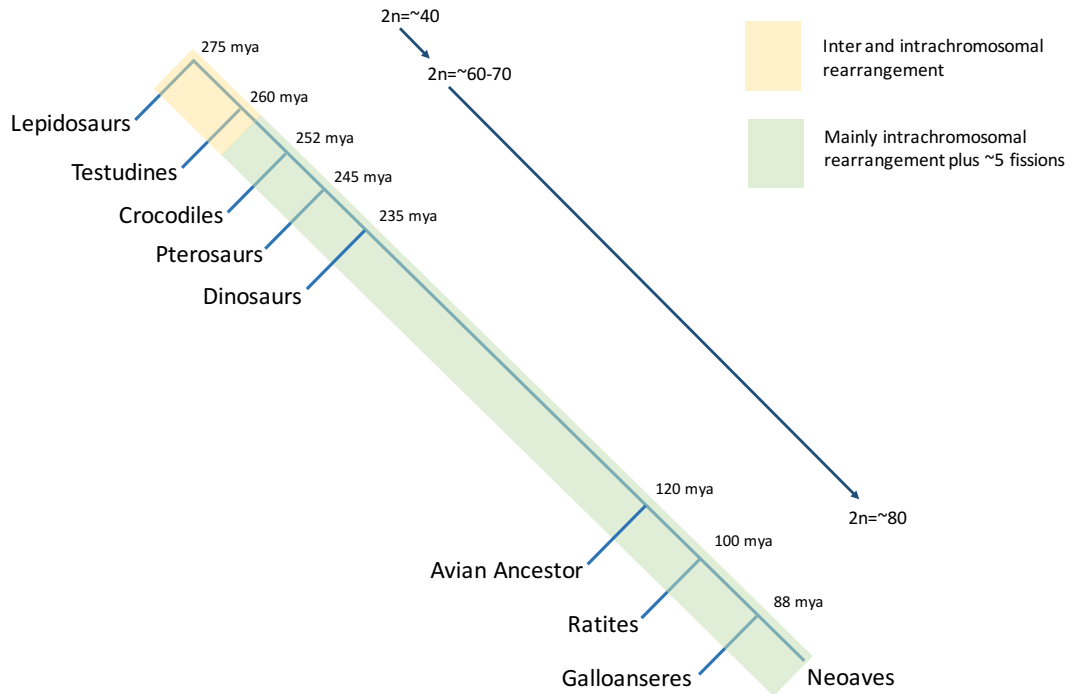
**Figure 7-2: Marketing material created by Cytocell for the Chicken multiprobe device.**

Future adaptation of the chicken device to incorporate the high success BACs described in chapter 4 (specific aim 2) will ultimately provide a valuable, cost effective tool for the avian comparative genomics community as well as provide proof of principle that the method can be applied to other genomes of interest beyond those of avian species (see section 6.4.3). In particular, these probes have been useful for addressing questions associated with avian (and dinosaur) evolution as outlined in chapters 4-6 and discussed further below.

### 7.1.2 Formation of the Signature Avian Karyotype (Chapters 4-6)

The signature avian karyotype alluded to many times in this thesis contains both macro and microchromosomes. It would be wrong however to suggest that the presence of both macro- and microchromosomes alone were a unique feature of avian genome organization. Indeed, microchromosomes are typical of most amniotes; mammals and crocodylia are exceptions. As described in section 1.3.3, many reptiles, such as snakes, turtles and lizards (but, surprisingly not crocodylians (Cohen 1970), the only extant examples of non-avian archosaurs) all possess microchromosomes. The greatest number and smallest size of microchromosomes are however typically found among birds. Burt (2002) hypothesized that microchromosomes were present in

the common dinosaur ancestor that gave rise to birds and that a series of fissions in the avian lineage resulted in the basic pattern of  $2n=80$  (~30 pairs of microchromosomes) with this structure becoming fixed before the Paleognathae-Neognathae divergence 100mya. In this study however, the evidence suggests that such a fixation may have been much earlier at around 260mya, with the basic pattern established in dinosaurs and pterosaurs and remaining relatively stable thereafter as illustrated in Figure 7-3.



**Figure 7-3: Changes to diploid number along archosauromorpha lineage to birds.**

In this study (chapter 6, specific aim 4 and published in Romanov et al. 2014) evidence is provided that leads to suggested possible mechanisms why, with relatively rare exceptions, avian genomes remain evolutionarily stable interchromosomally. Absence of interchromosomal rearrangement either suggests an evolutionary advantage to retaining this signature avian configuration or else little opportunity for change. Evidence of considerable intrachromosomal change in pigeons (chapter 5 of this thesis) and *Passeriformes* species (chapter 6, Romanov et al. 2014; Skinner and Griffin 2012) suggest however that intrachromosomal change proceeds largely un-hindered and can accelerate in line with rapid speciation events. Indeed, the near absence of interchromosomal rearrangement is no barrier to diversity and a direct correlation has been reported between the rates of speciation and intrachromosomal rearrangement (King 1995). There may even be an evolutionary advantage to maintaining a karyotypic structure formed of many compact, gene rich microchromosomes (Romanov et al. 2014).

In birds, it is a reasonable assumption that the characteristic stable gross karyotypic structure has a reduced opportunity for chromosome rearrangement as there are low numbers of recombination hotspots, fewer repeat structures such as transposable elements, and fewer endogenous retroviruses. All of these genomic features have been previously demonstrated to provide substrates for interchromosomal rearrangement and all are sparser in avian, compared to other genomes (Ellegren 2013). In previous studies it has been argued that the signature avian karyotype evolved in response to shrinking of the genome in birds as a result of the metabolic demands of flight (Gregory 2009, Andrews et al. 2009). This thesis however indicates that the basic karyotype structure was in place long before avian genome size reduction. Average genome size in non dinosaur and, non avian saurians (lepidosaurs, turtles and crocodiles) is around 3Gb (Janes et al. 2010) and is significantly smaller in saurischian (1.78pg) in comparison to ornithischian dinosaurs (2.49pg) (Organ et al. 2007) as illustrated in Figure 7-4.

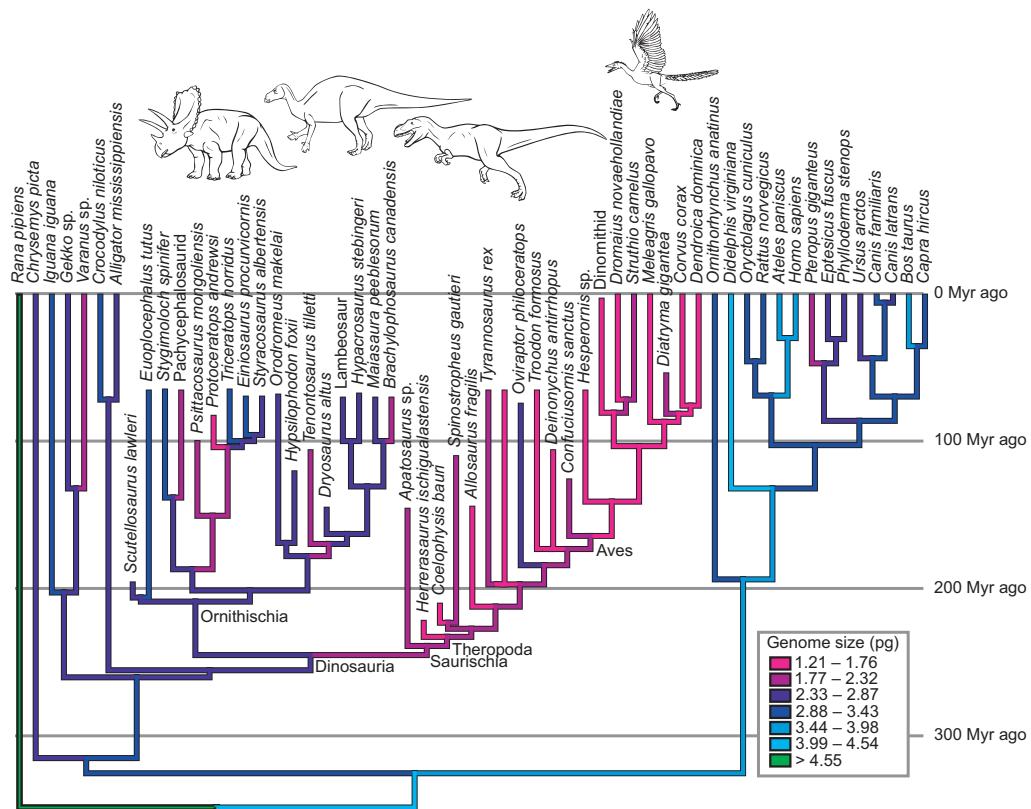


Figure 7-4: Haploid genome sizes illustrating the significantly smaller genome sizes along the theropod lineage originating early in the evolution of archosaurs (Organ et al. 2007).

Although flight evolution may be a factor in genome size reduction therefore (pterosaurs are reported to have smaller genomes than other Avemetatarsalians (Organ and Shedlock 2009) and bats have smaller genomes than other mammals (Hughes and Hughes 1995)), other factors

are clearly in play as flight only evolved in theropods approximately 150mya (Garner et al. 1999). It is possible therefore that evolution of the karyotype was a driver of genome size reduction rather than the other way around.

Explanations as to why some species such as falcons and parrots exhibit such a high degree of interchromosomal rearrangement, particularly microchromosomal fusion remain topics of debate. There are other groups (such as kingfishers) that have an unusually high ( $2n=130+$ ) number of chromosomes. Indeed, both higher and lower than usual deviations from the typical ( $2n\sim 80$ ) organization can occur in the same group e.g. the Adélie penguin ( $2n=96$ ) and a lower than average number in the emperor penguin ( $2n=72$ ). This suggests that similar mechanisms can cause both a rapid reduction and a rapid increase in chromosome number. The short time period over which these changes occur in the penguins and the rearranged karyotypes of the *Falconiformes* (but not the sister group *Strigiformes*) and the *Psittaciformes* (but not the sister group the *Passeriformes*) suggest that these changes can happen quickly in evolutionary terms. Vertebrates with large, repeat-rich genomes (such as mammals and amphibians) frequently demonstrate rapid intra- and interchromosomal rearrangements (Eichler & Sankoff 2003). The results presented here suggest that birds too can undergo similar changes in certain groups although there is little evidence that these highly rearranged avian genomes are particularly large, or more repeat rich than other avian genomes. Zebra finch and budgerigar comparisons indicate a similarly high rate of chromosome mutation in both groups but these features appear to be a result of fixed interchromosomal rearrangements that have arisen due to exploitation of evolutionary niches, while in others avian species such fixation is prevented resulting in maintenance of the classic avian style karyotype. Why some become fixed, and others do not, is a relatively understudied field, although clues may lie in the study of GO terms present in EBRs. Farré and colleagues found a correlation between EBRs and specific avian adaptive features in individual species, including forebrain development in the budgerigar (one of the species investigated in this thesis), consistent with this species being not only a vocal-learner but having distinctive neuronal connections compared to other vocal-learners (Farré et al. 2016). As more genomes become available with better assemblies these analyses may well point to adaptive phenotypic features of individual orders and families.

The work presented in this thesis to assemble the genomes of the peregrine falcon and the pigeon (chapter 5) will have the benefit of increasing the number of chromosomally assembled avian genomes from 5 to 7, therefore adding substantially to analyses such as these, both in

terms of data points but also in terms of coverage of the phylogenetic tree. With this in mind, even chromosomally assembled genomes are by no means perfect. Results in chapter 3 of this thesis highlight how even the best assembled genomes must be considered a work in progress. Until genome assembly methods catch up in terms of quality with the sheer number of sequencing outputs, the fragmented nature of scaffold assembled genomes will continue to be an important consideration when performing this type of analysis. The ancestral reconstruction results presented here and elsewhere are however a significant step forwards considering the current technology available, providing valuable insight into the genomic structure of long-dead ancestral species.

Regardless of the exceptions to the rule, the fact remains that the signature karyotype described here is nonetheless a successful one. This was already apparent in birds (Griffin et al. 2007) however, as a result of this thesis it seems that this statement can be extended to non-avian dinosaurs and (possibly) pterosaurs as well. As described, the greater possible combination of gametes from having more chromosomes and the increase in overall recombination rate per chromosome (despite the lack of recombination hot spots) ultimately has the effect of generating variation, the driver of natural selection. Burt (2002) suggested that a higher recombination rate has also contributed to the unique genomic features seen in microchromosomes such as high GC-content, low repeats and high gene-density which subsequently led to the maintenance of the typical avian karyotype. The potential for greater variation may therefore explain why we observed such diversity among the dinosaurs, why they were the dominant land vertebrate for hundreds of millions of years and why, even then after the majority were wiped out by the K-Pg event 66mya, they were able to diversify in a second great adaptive radiation into a fantastically diverse class of organisms - neornithine birds (Jetz et al. 2012).

### **7.1.3 Future Work: Mammals and Beyond**

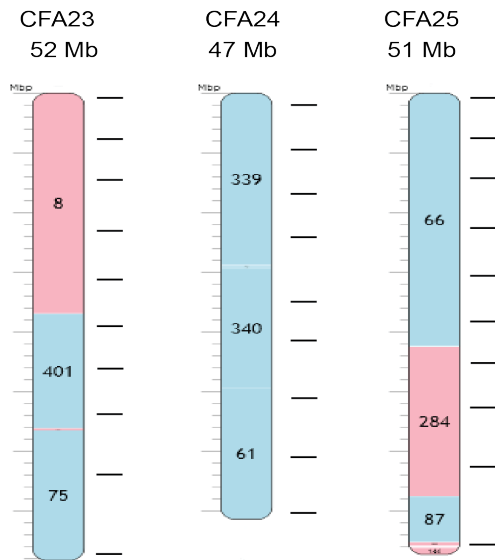
The set of cross-species avian BACs developed in this thesis can be, theoretically, applied to all avian species both as a means of addressing individual comparative genomic questions (chapters 4 and 6) but also as a method to upgrade the NGS scaffold based genomes of sequenced avian genomes (chapter 5). In fact, RACA predictions have already been analysed by the Larkin lab for the ostrich, the budgerigar and the saker falcon. Work is currently underway in the lab to use this set of BACs to map the PCF outputs produced by RACA for each of these species with early results showing equivalent levels of success as illustrated for the pigeon and the peregrine falcon

genomes mapped in chapter 5. Preliminary data testing these BACS on chromosomes of the reptile species reported in chapter 6 also indicate that further fine mapping with the full set of avian BACs will be possible without needing to develop a set of reptile specific BACs.

The most obvious phylogenetic class on which the future efforts should be focused is the mammals. To date of the >5,000 extant species only 20 have genomes assembled to the level of nearly one long scaffold per chromosome. These represent a tiny proportion of the total, with a bias towards primates (Rogers & Gibbs 2014), rodents (Fang et al. 2014) and artiodactyls (Larkin et al. 2012). Several hundred genomes are already being assembled to scaffold level by individual projects or consortia such as Genome10K (Koepfli et al. 2015). Mammals are the most studied class of organisms in the scientific literature, largely due to their use as biomedical models (e.g. mouse, rat, rabbit, pig), companion animals (e.g. cat and dog) and food production (e.g. cattle, pig, sheep) as well as a degree of self-interest in *Homo sapiens*. Many mammalian species are also on the IUCN Red list meaning that tools for the study of ecology and conservation are essential.

Next generation sequencing (NGS) in mammals has already had a significant impact on personalised medicine (predominantly in humans), agriculture (pig, cattle, sheep etc.) and basic studies (mainly mouse). As has been highlighted in this thesis (chapter 5) these studies rely on integration of the NGS data with an initial chromosome-level reference assembly built with traditional mapping technologies and longer-read capillary (Sanger) sequencing. Most of the current *de novo* NGS mammalian genome efforts, like their avian counterparts, do not have these chromosome level assemblies for the reasons already given (see section 1.1.4). Even for those mammals of sufficient interest to warrant the additional funding to take advantage of newer technologies (Optical mapping, Dovetail, BioNano, PACBio – see section 1.1.6) that generate longer sequencing reads, these are nonetheless still large scaffolds. As previously described (section 1.1.6) extending across fragile regions, centromeres or large heterochromatin regions remains a problem. Perhaps the best example is the most recent assembly of the dog genome (using Dovetail technology). These are impressively sized super-scaffolds but they are, still sub-chromosomal in size (see Figure 7-5) and would benefit from a cytogenetic mapping approach in order to complete the assembly.





**Figure 7-5: Dovetail dog scaffolds aligned to the reference assembly of dog (CFA – *Canis familiaris*) chromosomes 23-25. Numbers inside chromosomes represent scaffold IDs with black lines to the right of each chromosome indicating putative positions of BACs selected every 5 Mb suggesting that the majority of Dovetail scaffolds could be ordered and oriented along the chromosomes. Image was provided by Harris Lewin (University of California, Davis).**

The lack of chromosome-level assemblies for the vast majority of sequenced mammalian genomes (a number that is increasing all the time) limits their use for a range of essential aspects of evolutionary and applied genetics, especially genotype-phenotype correlations, gene assisted selection and breeding, and finding QTNs of economic importance. Indeed, chromosome level assemblies will soon be required for agricultural mammals that are used in developing countries (e.g. camels, yaks, buffalo) to foster more efficient food production, therefore contributing to global food security. The role of EBRs as ‘hotspots’ of chromosome evolution is relatively well established (Larkin et al. 2009; this thesis), however, difficulties in detecting HSBs from both a small number of large-range assemblies or a large number of fragmented assemblies still limits the scope of this type of analysis. An inexpensive method to upgrade the current mammalian genomes assemblies produced with the contemporary sequencing techniques (to chromosome level assemblies) is therefore critical to test hypotheses about the role of these ‘building blocks’ of genomes in evolution. Of particular interest are the studies of the exceptional cases when genomes have very few (like the Indian muntjac) or a very large number of chromosomes (like the viscacha rat) (Contreras et al. 1990).

To build a universal mammalian universal BAC set might be a greater challenge than in birds as mammalian genomes are approximately three times larger (therefore requiring three times as

many BACs to achieve the same level of mapping resolution). Additionally, mammalian genomes have a much larger proportion of repetitive sequence than birds (Hughes and Piontkivska 2005; Schmid et al. 2015), therefore potentially impeding efforts to select evolutionary conserved sequences. On the other hand, mammalian genome projects are better funded, assembly quality is usually higher, and the availability of large numbers of high quality BAC clones (specifically human, pig and cattle – see chapter 3) is much better. Considering that this current avian BAC set appears to be showing high success rates on lizard and turtle chromosomes it seems that universal BAC sets for all vertebrate and ultimately all animal groups could be on the horizon. Coupled with interactive comparative cytogenomic browsers such as Evolution Highway (thanks to the Larkin lab), the results described in this thesis illustrate tools that can be used for studying the biological significance of chromosome level assemblies for thousands of genomes.

#### **7.1.4 Personal Perspectives**

Whilst I am very pleased with the work presented in this thesis, another important aspect to this PhD is the incredible journey that these last four years have been. There are so many highlights that it is difficult to know where to begin. As a keen traveller I have relished the opportunity to travel internationally, whether to receive training (San Diego Zoo was a particular highlight!), or to provide training (in the United Arab Emirates) or to attend international conferences (5 of which I have been fortunate enough to give oral presentations at). As well as enjoying these experiences on a personal level, meeting so many varied and interesting people during each of these trips has emphasised the international nature of scientific research and reinforced the importance of collaboration across borders and across disciplines (as well as being a lot of fun!). Development work with Ensembl, to load the 'dinosaur' genome generated here into a web format shows great promise for outreach activities and has been a fascinating insight into a different area of genome research. Working with industry collaborators has proven to be very productive both on the research side, and for the companies involved, with at least two products now commercially available and a commercial screening service established. In addition, I have been a co-author on 7 manuscripts, 3 of which are published, 2 have been submitted and 2 are in preparation – all of which I feel very proud to be part of. Finally, the avian work presented here has inspired a grant application for funding to repeat the project but with a mammalian focus, which I hope will be successful soon. All in all, the last four years has been a very exciting, very productive time during which I have met some fantastic people and made many amazing memories which will stay with me forever.

## 8 References:

- Alekseyev, M.A. & Pevzner, P.A., 2009. Breakpoint graphs and ancestral genome reconstructions. *Genome research*, 19(5), pp.943–57.
- Alföldi, J. et al., 2011. The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature*, 477(7366), pp.587–91.
- Alkan, C., Sajjadian, S. & Eichler, E.E., 2010. Limitations of next-generation genome sequence assembly. *Nature Methods*, 8(1), pp.61–65.
- Andersson, L. & Georges, M., 2004. Domestic-animal genomics: deciphering the genetics of complex traits. *Nature Reviews Genetics*, 5(3), pp.202–212.
- Andrews, C.B., Mackenzie, S.A. & Gregory, T.R., 2009. Genome size and wing parameters in passerine birds. *Proceedings. Biological sciences / The Royal Society*, 276(1654), pp.55–61.
- Arnheim, N., Calabrese, P. & Nordborg, M., 2003. Hot and cold spots of recombination in the human genome: the reason we should find them and how this can be achieved. *American journal of human genetics*, 73(1), pp.5–16.
- Auton, A. et al., 2015. A global reference for human genetic variation. *Nature*, 526(7571), pp.68–74.
- Avdeyev, P. et al., 2016. Reconstruction of Ancestral Genomes in Presence of Gene Gain and Loss. *Journal of computational biology : a journal of computational molecular cell biology*.
- Avise, J.C. & Robinson, T.J., 2008. Hemiplasy: a new term in the lexicon of phylogenetics. *Systematic biology*, 57(3), pp.503–7.
- Ayala, F.J. & Coluzzi, M., 2005. Chromosome speciation: humans, *Drosophila*, and mosquitoes. *Proceedings of the National Academy of Sciences of the United States of America*, 102 Suppl (suppl\_1), pp.6535–42.
- Backström, N. et al., 2010. The recombination landscape of the zebra finch *Taeniopygia guttata* genome. *Genome research*, 20(4), pp.485–95.
- Badenhorst, D. et al., 2015. Physical Mapping and Refinement of the Painted Turtle Genome (*Chrysemys picta*) Inform Amniote Genome Evolution and Challenge Turtle-Bird Chromosomal Conservation. *Genome biology and evolution*, 7(7), pp.2038–50.
- Bailey, J.A. et al., 2004. Hotspots of mammalian chromosomal evolution. *Genome biology*, 5(4), p.R23.
- Baker, M., 2012. *De novo* genome assembly: what every biologist should know. *Nature Methods*, 9(4), pp.333–337.
- Balakrishnan, C.N. et al., 2010. Gene duplication and fragmentation in the zebra finch major histocompatibility complex. *BMC biology*, 8(1), p.29.

Beçak, W., et al., 1964. Close karyological kinship between the reptilian suborder Serpentes and the class Aves. *Chromosoma*, 15(5), pp.606-617.

Bejjani, B.A. & Shaffer, L.G., 2006. Application of array-based comparative genomic hybridisation to clinical diagnostics. *The Journal of molecular diagnostics : JMD*, 8(5), pp.528–33.

Bellott, D.W. et al., 2010. Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. *Nature*, 466(7306), pp.612–616.

Benson, R.B.J. et al., 2014. Rates of Dinosaur Body Mass Evolution Indicate 170 Million Years of Sustained Ecological Innovation on the Avian Stem Lineage H. Morlon, ed. *PLoS Biology*, 12(5), p.e1001853.

Benton, M.J. & Twitchett, R.J., 2003. How to kill (almost) all life: the end-Permian extinction event. *Trends in Ecology & Evolution*, 18(7), pp.358–365.

Benton, M.J. et al., 2013. Exceptional vertebrate biotas from the Triassic of China, and the expansion of marine ecosystems after the Permo-Triassic mass extinction. *Earth-Science Reviews*, 125, pp.199–243.

Benton, M.J. et al., 2015. Constraints on the timescale of animal evolutionary history. *Palaeontologia Electronica*, 18(1), pp.1–106.

Benton, M.J., Forth, J. & Langer, M.C., 2014. Models for the rise of the dinosaurs. *Current biology : CB*, 24(2), pp.R87–95.

Berlin, K. et al., 2014. Assembling Large Genomes with Single-Molecule Sequencing and Locality Sensitive Hashing, *Cold Spring Harbor Labs Journals*.

Bernardi, G., 2000. Isochores and the evolutionary genomics of vertebrates. *Gene*, 241(1), pp.3–17.

BirdLife International, 2014. We have lost over 150 bird species since 1500. Presented as part of the BirdLife State of the world's birds website.

Bittel, D.C. et al., 2009. Refining the 22q11.2 deletion breakpoints in DiGeorge syndrome by aCGH. *Cytogenetic and genome research*, 124(2), pp.113–20.

Böhne, A. et al., 2008. Transposable elements as drivers of genomic and biological diversity in vertebrates. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, 16(1), pp.203–15.

Borgaonkar, D.S., 1975. *Chromosomal Variation in Man*.

Borisov, A.B. et al., 2003. Rapid response of cardiac obscurin gene cluster to aortic stenosis: differential activation of Rho-GEF and MLCK and involvement in hypertrophic growth. *Biochemical and Biophysical Research Communications*, 310(3), pp.910–918.

Bonnet, C. et al., 2010. Microdeletion at chromosome 4q21 defines a new emerging syndrome with marked growth restriction, mental retardation and absent or severely delayed speech. *Journal of medical genetics*, 47(6), pp.377–84.

Bourque, G. & Pevzner, P.A., 2002. Genome-Scale Evolution: Reconstructing Gene Orders in the Ancestral Species. *Genome Res.*, 12(1), pp.26–36.

Branco, M.R. & Pombo, A., 2006. Intermingling of Chromosome Territories in Interphase Suggests Role in Translocations and Transcription-Dependent Associations. *PLoS Biology*, 4(5), p.e138.

Brenner, S. et al., 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnology*, 18(6), pp.630–634.

Brown, J.D. & O'Neill, R.J., 2010. Chromosomes, conflict, and epigenetics: chromosomal speciation revisited. *Annual review of genomics and human genetics*, 11, pp.291–316.

Brown, W.R.A. et al., 2003. The chicken as a model for large-scale analysis of vertebrate gene function. *Nature reviews. Genetics*, 4(2), pp.87–98.

Brusatte, S.L. et al., 2008. The first 50Myr of dinosaur evolution: macroevolutionary pattern and morphological disparity. *Biology letters*, 4(6), pp.733–6.

Brusatte, S.L. et al., 2011. Footprints pull origin and diversification of dinosaur stem lineage deep into Early Triassic. *Proceedings. Biological sciences / The Royal Society*, 278(1708), pp.1107–13.

Burkardt, D.D. et al., 2011. Distinctive phenotype in 9 patients with deletion of chromosome 1q24-q25. *American journal of medical genetics. Part A*, 155A(6), pp.1336–51.

Burt, D.W. et al., 1999. The dynamics of chromosome evolution in birds and mammals. *Nature*, 402(6760), pp.411–3.

Burt, D.W., 2002. Origin and evolution of avian microchromosomes. *Cytogenetic and genome research*, 96(1-4), pp.97–112.

Cáceres, M., Puig, M. & Ruiz, A., 2001. Molecular characterization of two natural hotspots in the *Drosophila buzzatii* genome induced by transposon insertions. *Genome research*, 11(8), pp.1353–64.

Cano et al., (1992). Isolation and partial characterisation of DNA from the bee *Proplebeia dominicana* (Apidae : Hymenoptera) in 25-40 million year old amber. *Medical science research*, 20(7), pp.249–251.

Carter, N.P. et al., 1992. Reverse chromosome painting: a method for the rapid analysis of aberrant chromosomes in clinical cytogenetics. *Journal of Medical Genetics*, 29(5), pp.299–307.

Chandley, A.C., 1991. On the parental origin of *de novo* mutation in man. *Journal of Medical Genetics*, 28(4), pp.217–223.

- Chen, N. et al., 2012. Identification of avian W-linked contigs by short-read sequencing. *BMC genomics*, 13(1), p.183.
- Cheung, V.G. et al., 2001. Integration of cytogenetic landmarks into the draft sequence of the human genome. *Nature*, 409(6822), pp.953–8.
- Chiappe, L.M. and Dyke, G.J., 2006. The early evolutionary history of birds. *Journal-Paleontological Society Of Korea*, 22(1), p.133
- Chiari, Y. et al., 2012. Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). *BMC biology*, 10(1), p.65.
- Choudhuri, J. V et al., 2004. GenAlyzer: interactive visualisation of sequence similarities between entire genomes. *Bioinformatics (Oxford, England)*, 20(12), pp.1964–5.
- Chowdhary, B.P. et al., 1998. Emerging Patterns of Comparative Genome Organization in Some Mammalian Species as Revealed by Zoo-FISH. *Genome Res.*, 8(6), pp.577–589.
- Christidis, L., 1990. *Animal cytogenetics 4: Chordata 3 B: Aves*. Gebrüder Borntraeger, Berlin, Germany.
- Clarke, J.A. et al., 2005. Definitive fossil evidence for the extant avian radiation in the Cretaceous. *Nature*, 433(7023), pp.305–8.
- Clayton, D.F., Balakrishnan, C.N. & London, S.E., 2009. Integrating genomes, brain and behavior in the study of songbirds. *Current biology : CB*, 19(18), pp.R865–73.
- Cohen, M.M. & Gans, C., 1970. The chromosomes of the order Crocodylia. *Cytogenetic and Genome Research*, 9(2), pp.81–105.
- Commins, J., Toft, C. & Fares, M.A., 2009. Computational biology methods and their application to the comparative genomics of endocellular symbiotic bacteria of insects. *Biological procedures online*, 11(1), pp.52–78.
- Conrad, D.F. et al., 2010. Origins and functional impact of copy number variation in the human genome. *Nature*, 464(7289), pp.704–12.
- Contreras, L.C., Torres-Mura, J.C. & Spotorno, A.E., 1990. The largest known chromosome number for a mammal, in a South American desert rodent. *Experientia*, 46(5), pp.506–508.
- Copeland, N. et al., 1993. A genetic linkage map of the mouse: current applications and future prospects. *Science*, 262(5130), pp.57–66.
- Cortez, D. et al., 2014. Origins and functional evolution of Y chromosomes across mammals. *Nature*, 508(7497), pp.488–93.
- Costantini, M. et al., 2016. The Anolis Lizard Genome: An Amniote Genome without Isochores? *Genome biology and evolution*, 8(4), pp.1048–55.

Coyne, J.A., Aulard, S. & Berry, A., 1991. Lack of underdominance in a naturally occurring pericentric inversion in *Drosophila melanogaster* and its implications for chromosome evolution. *Genetics*, 129(3), pp.791–802.

Crawford, N.G. et al., 2012. More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. *Biology letters*, 8(5), pp.783–6.

Cribiu, E.P. et al., 2001. International System for Chromosome Nomenclature of Domestic Bovids (ISCNDB 2000). *Cytogenetics and cell genetics*, 92(3-4), pp.283–99.

Crick, H.Q.P. & Ratcliffe, D.A., 1995. The Peregrine *Falco peregrinus* breeding population of the United Kingdom in 1991. *Bird Study*, 42(1), pp.1–19.

Cunningham, F. et al., 2015. Ensembl 2015. *Nucleic acids research*, 43, pp.D662–9.

Dalloul, R.A. et al., 2010. Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. *PLoS biology*, 8(9), p.e1000475.  
Darwin, C., 1859. *The Origin of Species*.

Darwin, C., 1868. *The variation of animals and plants under domestication*. O. Judd, ed.,

Dauwerse, J.G. et al., 1992. Multiple colors by fluorescence in situ hybridisation using ratio-labelled DNA probes create a molecular karyotype. *Human Molecular Genetics*, 1(8), pp.593–598.

Dawson, A.J. et al., 2002. Cryptic chromosome rearrangements detected by subtelomere assay in patients with mental retardation and dysmorphic features. *Clinical genetics*, 62(6), pp.488–494.

De Lorenzi, L. et al., 2012. Reciprocal translocations in cattle: frequency estimation. *Journal of animal breeding and genetics*, 129(5), pp.409–16.

de Oliveira, E.H.C. et al., 2005. Chromosome reshuffling in birds of prey: the karyotype of the world's largest eagle (Harpy eagle, *Harpia harpyja*) compared to that of the chicken (*Gallus gallus*). *Chromosoma*, 114(5), pp.338–43.

de Oliveira, E.H.C. et al., 2010. Reciprocal chromosome painting between white hawk (*Leucopternis albicollis*) and chicken reveals extensive fusions and fissions during karyotype evolution of Accipitridae (Aves, *Falconiformes*). *Chromosome research*, 18(3), pp.349–55.

de Oliveira, E.H.C. et al., 2013. Chromosome painting in three species of *Buteoninae*: a cytogenetic signature reinforces the monophyly of South American species. *PloS one*, 8(7), p.e70071.

Deakin, J.E. & Ezaz, T., 2014. Tracing the evolution of amniote chromosomes. *Chromosoma*, 123(3), pp.201–216.

DeBry, R.W. & Seldin, M.F., 1996. Human/mouse homology relationships. *Genomics*, 33(3), pp.337–51.

Delneri, D. et al., 2003. Engineering evolution to study speciation in yeasts. *Nature*, 422(6927), pp.68–72.

Dennis Jr, Glynn et al., 2003. DAVID: Database for Annotation, Visualisation, and Integrated Discovery. *Genome Biology*, 4(9), pp.1–11.

Derjushcheva, S. et al., 2004. High chromosome conservation detected by comparative chromosome painting in chicken, pigeon and passerine birds. *Chromosome research*, 12(7), pp.715–23.

DeSalle, R. et al., 1992. DNA sequences from a fossil termite in Oligo-Miocene amber and their phylogenetic implications. *Science*, 257(5078), pp.1933–1936.

Drewitt, E.J. & Dixon, N., 2008. Diet and prey selection of urban-dwelling Peregrine Falcons in southwest England. *British Birds*, (101(2)), p.p.58.

Driscoll, C.A., Macdonald, D.W. & O'Brien, S.J., 2009. From wild animals to domestic pets, an evolutionary view of domestication. *Proceedings of the National Academy of Sciences*, 106(Supplement\_1), pp.9971–9978.

Drmanac, R. et al., 2010. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science (New York, N.Y.)*, 327(5961), pp.78–81.

Ducos, A. et al., 2007. Chromosomal control of pig populations in France: 2002–2006 survey. *Genetics Selection Evolution*, 39(5), pp.583–597.

Ducos, A. et al., 2008. Cytogenetic screening of livestock populations in Europe: an overview. *Cytogenetic and genome research*, 120(1-2), pp.26–41.

Dumas, L. et al., 2007. Gene copy number variation spanning 60 million years of human and primate evolution. *Genome research*, 17(9), pp.1266–77.

Eiben, B. et al., 1990. Cytogenetic analysis of 750 spontaneous abortions with the direct-preparation method of chorionic villi and its implications for studying genetic causes of pregnancy wastage. *American journal of human genetics*, 47(4), pp.656–63.

Eichler, E.E. & Sankoff, D., 2003. Structural dynamics of eukaryotic chromosome evolution. *Science (New York, N.Y.)*, 301(5634), pp.793–7.

Eid, J. et al., 2009. Real-time DNA sequencing from single polymerase molecules. *Science (New York, N.Y.)*, 323(5910), pp.133–8.

Ellegren, H. et al., 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*, 491(7426), pp.756–60.

Ellegren, H., 2013. The evolutionary genomics of birds. *Annual Review of Ecology, Evolution, and Systematics*, 44(1), pp.239–259.

Elsik, C.G. et al., 2009. The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution. *Science*, 324(5926), pp.522–528.



- Estrada, K. et al., 2009. A genome-wide association study of northwestern Europeans involves the C-type natriuretic peptide signaling pathway in the etiology of human height variation. *Human molecular genetics*, 18(18), pp.3516–24.
- Evangelidou, P. et al., 2013. Implementation of high resolution whole genome array CGH in the prenatal clinical setting: advantages, challenges, and review of the literature. *BioMed research international*, 2013, p.346762.
- Ezcurra, M.D., Scheyer, T.M. & Butler, R.J., 2014. The origin and early evolution of Sauria: reassessing the permian Saurian fossil record and the timing of the crocodile-lizard divergence. *PloS one*, 9(2), p.e89165.
- Fang, X. et al., 2014. Genome-wide adaptive complexes to underground stresses in blind mole rats *Spalax*. *Nature Communications*, 5.
- Farmer, C.G. & Sanders, K., 2010. Unidirectional airflow in the lungs of alligators. *Science (New York, N.Y.)*, 327(5963), pp.338–40.
- Farré, M. et al., 2011. Assessing the role of tandem repeats in shaping the genomic architecture of great apes. *PloS one*, 6(11), p.e27239.
- Farré, M. et al., 2016. Novel insights into chromosome evolution in birds, archosaurs, and reptiles. *Genome Biology and Evolution*, p.evw166.
- Ferguson-Lees, J. & Christie, D.A., 2001. *Raptors of the world*, Houghton Mifflin Harcourt.
- Ferguson-Smith, M.A. & Trifonov, V., 2007. Mammalian karyotype evolution. *Nature reviews. Genetics*, 8(12), pp.950–62.
- Ferguson-Smith, M.A., 1997. Genetic analysis by chromosome sorting and painting: phylogenetic and diagnostic applications. *European journal of human genetics : EJHG*, 5(5), pp.253–65.
- Feuk, L. et al., 2005. Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies. *PLoS genetics*, 1(4), p.e56.
- Feuk, L., Carson, A.R. & Scherer, S.W., 2006. Structural variation in the human genome. *Nature reviews. Genetics*, 7(2), pp.85–97.
- Fillon, V. et al., 2007. FISH mapping of 57 BAC clones reveals strong conservation of synteny between *Galliformes* and *Anseriformes*. *Animal genetics*, 38(3), pp.303–7.
- Flint, J. et al., 1995. The detection of subtelomeric chromosomal rearrangements in idiopathic mental retardation. *Nature genetics*, 9(2), pp.132–40.
- Frankl-Vilches, C. et al., 2015. Using the canary genome to decipher the evolution of hormone-sensitive gene regulation in seasonal singing birds. *Genome biology*, 16(1), p.19.
- Frey, N. & Olson, E.N., 2002. Calsarcin-3, a Novel Skeletal Muscle-specific Member of the Calsarcin Family, Interacts with Multiple Z-disc Proteins. *Journal of Biological Chemistry*, 277(16), pp.13998–14004.

Friedman, J.M. et al., 2006. Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation. *American journal of human genetics*, 79(3), pp.500–13.

Froenicke, L. et al., 2006. Are molecular cytogenetics and bioinformatics suggesting diverging models of ancestral mammalian genomes? *Genome research*, 16(3), pp.306–10.

Froenicke, L., 2005. Origins of primate chromosomes - as delineated by Zoo-FISH and alignments of human and mouse draft genome sequences. *Cytogenetic and genome research*, 108(1-3), pp.122–38.

Gadea, J., 2005. Sperm factors related to in vitro and in vivo porcine fertility. *Theriogenology*, 63(2), pp.431–444.

Gadea, J., Sellés, E. & Marco, M.A., 2004. The predictive value of porcine seminal parameters on fertility outcome under commercial conditions. *Reproduction in domestic animals*, 39(5), pp.303–308.

Gadelha Pereira Fontenele, E. et al., 2015. Association Study of GWAS-Derived Loci with Height in Brazilian Children: Importance of MAP3K3, MMP24 and IGF1R Polymorphisms for Height Variation. *Hormone Research in Paediatrics*, 84(4), pp.248–253.

Ganapathy, G. et al., 2014. High-coverage sequencing and annotated assemblies of the budgerigar genome. *GigaScience*, 3(1), pp.1–9.

Garrick, D. and Ruvinsky, A. eds., 2014. *The genetics of cattle*. CABI.

Gatesy, S., 2002. Locomotor evolution on the line to modern birds. In *Mesozoic Birds: Above the Heads of Dinosaurs*. University of California Press, pp. 432–447.

Giannuzzi, G. et al., 2013. Hominoid fission of chromosome 14/15 and the role of segmental duplications. *Genome research*, 23(11), pp.1763–73.

Gibbs, R.A. et al., 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*, 428(6982), pp.493–521.

Gill, F. & D.D., 2016. *IOC World Bird List (v 6.1)*.

Giuffra, E. et al., 2000. The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics*, 154(4), pp.1785–91.

Gordon, J.L., Byrne, K.P. & Wolfe, K.H., 2009. Additions, losses, and rearrangements on the evolutionary route from a reconstructed ancestor to the modern *Saccharomyces cerevisiae* genome. *PLoS genetics*, 5(5), p.e1000485.

Graphodatsky, A., Ferguson-Smith, M.A. & Stanyon, R., 2012. A short introduction to cytogenetic studies in mammals with reference to the present volume. *Cytogenetic and Genome Research*, 137(2-4), pp.83–96.

Graves, J.A.M., 2014. Avian sex, sex chromosomes, and dosage compensation in the age of genomics. *Chromosome Research*, 22(1), pp.45–57.

Green, E.D., 2001. Strategies for the systematic sequencing of complex genomes. *Nature Reviews Genetics*, 2(8), pp.573–583.

Gregory, T.R., 2009. A Bird's-Eye View Of The C-Value Enigma: Genome Size, Cell Size, And Metabolic Rate In The Class Aves. *Evolution*, 56(1), pp.121-130.

Griffin, D.K. & Finch, K.A., 2005. The genetic and cytogenetic basis of male infertility. *Human fertility*, 8(1), pp.19–26.

Griffin, D.K. et al., 1999. Micro- and macrochromosome paints generated by flow cytometry and microdissection: tools for mapping the chicken genome. *Cytogenetics and cell genetics*, 87(3-4), pp.278–81.

Griffin, D.K. et al., 2007. The evolution of the avian genome as revealed by comparative molecular cytogenetics. *Cytogenetic and genome research*, 117(1-4), pp.64–77.

Griffin, D.K. et al., 2008. Whole genome comparative studies between chicken and turkey and their implications for avian genome evolution. *BMC genomics*, 9(1), p.168.

Groenen, M.A.M. et al., 2009. A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. *Genome research*, 19(3), pp.510–9.

Groenen, M.A.M. et al., 2012. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*, 491(7424), pp.393–398.

Gudbjartsson, D.F. et al., 2008. Many sequence variants affecting diversity of adult human height. *Nature Genetics*, 40(5), pp.609–615.

Gustavsson, I., 1979. Distribution and effects of the 1/29 Robertsonian translocation in cattle. *Journal of dairy science*, 62(5), pp.825–35.

Gustavsson, I., 1990. Chromosomes of the pig. *Advances in veterinary science and comparative medicine*, 34, pp.73–107.

Guttenbach, M. et al., 2003. Comparative chromosome painting of chicken autosomal paints 1-9 in nine different bird species. *Cytogenetic and genome research*, 103(1-2), pp.173–84.

Habermann, F.A. et al., 2001. Arrangements of macro- and microchromosomes in chicken cells. *Chromosome Research*, 9(7), pp.569–584.

Hansmann, T. et al., 2009. Cross-species chromosome painting corroborates microchromosome fusion during karyotype evolution of birds. *Cytogenetic and genome research*, 126(3), pp.281–304.

Hanson, D. et al., 2014. Identifying biological pathways that underlie primordial short stature using network analysis. *Journal of molecular endocrinology*, 52(3), pp.333–44.

Harada, N. et al., 2002. A 4q21-q22 deletion in a girl with severe growth retardation. *Clinical genetics*, 61(3), pp.226–8.

Harry, D.E. et al., 2003. A first-generation map of the turkey genome. *Genome*.

Hassold, T. & Hunt, P., 2001. To err (meiotically) is human: the genesis of human aneuploidy. *Nature reviews. Genetics*, 2(4), pp.280–91.

Haussler, D. et al., 2009. Genome 10K: A Proposal to Obtain Whole-Genome Sequence for 10,000 Vertebrate Species. *The Journal of heredity*, 100(6), pp.659–74.

Hedges, D.J. et al., 2011. Comparison of Three Targeted Enrichment Strategies on the SOLiD Sequencing Platform. *Genetics*, ed. PLoS ONE, 6(4), p.e18595.

Hedges, S.B. et al., 2015. Tree of life reveals clock-like speciation and diversification. *Molecular biology and evolution*, 32(4), pp.835–45.

Hiroi, N. et al., 2013. Copy number variation at 22q11.2: from rare variants to common mechanisms of developmental neuropsychiatric disorders. *Molecular psychiatry*, 18(11), pp.1153–65.

Hirschhorn, J.N. & Lettre, G., 2009. Progress in genome-wide association studies of human height. *Hormone research*, 71 Suppl 2, pp.5–13.

Hoffmann, A.A. & Rieseberg, L.H., 2008. Revisiting the Impact of Inversions in Evolution: From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation? *Annual review of ecology, evolution, and systematics*, 39, pp.21–42.

Hollstein, R. et al., 2015. HACE1 deficiency causes an autosomal recessive neurodevelopmental syndrome. *Journal of medical genetics*, 52(12), pp.797–803.

Holmes, D. & Ottinger, M., 2003. Birds as long-lived animal models for the study of aging. *Experimental Gerontology*, 38(11-12), pp.1365–1375.

Horsley, S.W. et al., 1998. Del (18p) shown to be a cryptic translocation using a multiprobe FISH assay for subtelomeric chromosome rearrangements. *Journal of medical genetics*, 35(9), pp.722–726.

Hu, F. et al., 2014. Probabilistic Reconstruction of Ancestral Gene Orders with Insertions and Deletions. *IEEE/ACM transactions on computational biology and bioinformatics / IEEE, ACM*, 11(4), pp.667–72.

Hu, X. et al., 2009. Advanced technologies for genomic analysis in farm animals and its application for QTL mapping. *Genetica*, 136(2), pp.371–86.

Huang, Y. et al., 2013. The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. *Nature genetics*, 45(7), pp.776–783.

Hughes, A.L. & Hughes, M.K., 1995. Small genomes for better flyers. *Nature*, 377(6548), pp.391–391.

Hughes, A.L. & Friedman, R., 2008. Genome size reduction in the chicken has involved massive loss of ancestral protein-coding genes. *Molecular biology and evolution*, 25(12), pp.2681–8.

Hulse-Post, D.J. et al., 2005. Role of domestic ducks in the propagation and biological evolution of highly pathogenic H5N1 influenza viruses in Asia. *Proceedings of the National Academy of Sciences*, 102(30), pp.10682–10687.

International Chicken Genome Sequencing Consortium (ICGSC), 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, 432(7018), pp.695–716.

Ijdo, J.W. et al., 1991. Origin of human chromosome 2: an ancestral telomere-telomere fusion. *Proceedings of the National Academy of Sciences*, 88(20), pp.9051–9055.

International Human Genome Sequencing, 2004. Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011), pp.931–45.

Ioannou, D. et al., 2011. Multicolour interphase cytogenetics: 24 chromosome probes, 6 colours, 4 layers. *Molecular and cellular probes*, 25(5-6), pp.199–205.

Ioannou, D. et al., 2012. Twenty-four chromosome FISH in human IVF embryos reveals patterns of post-zygotic chromosome segregation and nuclear organisation. *Chromosome research*, 20(4), pp.447–60.

Islam, F. et al., 2014. Comparison of the Chromosome Structures between the Chicken and Three Anserid Species, the Domestic Duck (*Anas platyrhynchos*), Muscovy Duck (*Cairina moschata*), and Chinese Goose (*Anser cygnoides*), and the Delineation of their Karyotype Evolution. *The Journal of Poultry Science*, 51(1), pp.1–13.

Itoh, Y. & Arnold, A.P., 2005. Chromosomal polymorphism and comparative painting analysis in the zebra finch. *Chromosome Research*, 13(1), pp.47–56.

Jacobs, P.A. & Strong, J.A., 1959. A Case of Human Intersexuality Having a Possible XXY Sex-Determining Mechanism. *Nature*, 183(4657), pp.302–303.

Jacobs, P.A., 1992. The chromosome complement of human gametes. *Oxford reviews of reproductive biology*, 14, pp.47–72.

Janes, D.E. et al., 2010. Genome evolution in Reptilia, the sister group of mammals. *Annual review of genomics and human genetics*, 11, pp.239–264.

Jarvis, E.D. et al., 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science (New York, N.Y.)*, 346(6215), pp.1320–31.

Jarvis, E.D., 2016. Perspectives from the Avian Phylogenomics Project: Questions that Can Be Answered with Sequencing All Genomes of a Vertebrate Class. *Annual Review of Animal Biosciences*, 4(1), pp.45–59.

Jensen-Seaman, M.I. et al., 2004. Comparative recombination rates in the rat, mouse, and human genomes. *Genome research*, 14(4), pp.528–38.

Jetz, W. et al., 2012. The global diversity of birds in space and time. *Nature*, 491(7424), pp.444–448.

Karim, L. et al., 2011. Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. *Nature genetics*, 43(5), pp.405–13.

Kaiser, O. et al., 2003. Whole genome shotgun sequencing guided by bioinformatics pipelines—an optimized approach for an established technique. *Journal of Biotechnology*, 106(2-3), pp.121–133.

Kallioniemi, A. et al., 1992. Comparative genomic hybridisation for molecular cytogenetic analysis of solid tumors. *Science*, 258(5083), pp.818–821.

Kasai, F., O'Brien, P.C.M. & Ferguson-Smith, M.A., 2012. Reassessment of genome size in turtle and crocodile based on chromosome measurement by flow karyotyping: close similarity to chicken. *Biology letters*, 8(4), pp.631–635.

Kaufman, J. et al., 1999. The chicken B locus is a minimal essential major histocompatibility complex. *Nature*, 401(6756), pp.923–5.

Kawai, A. et al., 2009. The ZW sex chromosomes of *Gekko hokouensis* (Gekkonidae, Squamata) represent highly conserved homology with those of avian species. *Chromosoma*, 118(1), pp.43–51.

Kawakami, T. et al., 2014. A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. *Molecular Ecology*, 23(16), pp.4035–4058.

Kawashima; Mayer; Farinelli. WO Patent 91/44152, 1998.

Kidwell, M.G., 2001. Transposon-Induced Hotspots for Genomic Instability. *Genome Research*, 11(8), pp.1321–1322.

Kim, J. et al., 2013. Reference-assisted chromosome assembly. *Proceedings of the National Academy of Sciences of the United States of America*, 110(5), pp.1785–90.

Kim, J.-J. et al., 2010. Identification of 15 loci influencing height in a Korean population. *Journal of human genetics*, 55(1), pp.27–31.

King, M., 1995. *Species evolution: the role of chromosome change*. Cambridge University Press.

Kinsella, R.J. et al., 2011. Ensembl BioMart: a hub for data retrieval across taxonomic space. *Database : the journal of biological databases and curation*, 2011(0)

Kirkpatrick, M., 2010. How and why chromosome inversions evolve. *PLoS biology*, 8(9), p.e1000501.

Komlósi, K. et al., 2015. Phenotypic variability in a Hungarian patient with the 4q21 microdeletion syndrome. *Molecular Cytogenetics*, 8(1), p.16.

Kossiva, L. et al., 2010. Too short stature, too many stigmata. *BMJ case reports*, 2010.

- Knight, S.J. et al., 1996. Development and clinical application of an innovative fluorescence in situ hybridisation technique which detects submicroscopic rearrangements involving telomeres. *European journal of human genetics : EJHG*, 5(1), pp.1–8.
- Knight, S.J.L. & Flint, J., 2000. Perfect endings: a review of subtelomeric probes and their use in clinical diagnosis. *Journal of medical genetics*, 37(6), pp.401–409.
- Koepfli, K.P., Paten, B. and O'Brien, S.J., 2015. The genome 10K project: a way forward. *Annu. Rev. Anim. Biosci.*, 3(1), pp.57–111.
- Koonin, E. V., 2005. Orthologs, paralogs, and evolutionary genomics. *Annual review of genetics*, 39, pp.309–38.
- Koren, S. et al., 2012. Hybrid error correction and *de novo* assembly of single-molecule sequencing reads. *Nature biotechnology*, 30(7), pp.693–700.
- Kumar, S. & Hedges, S.B., 1998. A molecular timescale for vertebrate evolution. *Nature*, 392(6679), pp.917–20.
- Lander, E.S. et al., 2001. Initial sequencing and analysis of the human genome. *Nature*, 409(6822), pp.860–921.
- Larkin, D.M. et al., 2003. A Cattle–Human Comparative Map Built with Cattle BAC-Ends and Human Genome Sequence. *Genome Research*, 13(8), pp.1966–1972.
- Larkin, D.M. et al., 2006. Comparative mapping of mink chromosome 8p: in situ hybridisation of seven cattle BAC clones. *Animal genetics*, 37(4), pp.429–30.
- Larkin, D.M. et al., 2009. Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. *Genome research*, 19(5), pp.770–777.
- Larkin, D.M. et al., 2012. Whole-genome resequencing of two elite sires for the detection of haplotypes under selection in dairy cattle. *Proceedings of the National Academy of Sciences of the United States of America*, 109(20), pp.7693–8.
- Larkin, D.M. et al., 2014. Cytogenetics and chromosome maps. In D. J. Garrick & A. Ruvinsky, eds. *The Genetics of Cattle*. Wallingford: CABI, pp. 103–129.
- Lee, H. et al., 2016. Third-generation sequencing and the future of genomics, *Cold Spring Harbor Labs Journals*.
- Lee, M.S.Y. et al., 2014. Sustained miniaturization and anatomical innovation in the dinosaurian ancestors of birds. *Science (New York, N.Y.)*, 345(6196), pp.562–566.
- Lejeune, J. et al., 1963. Three cases of partial deletion of the short arm of a 5 chromosome. *Comptes rendus hebdomadaires des séances de l'Académie des sciences*, 257, pp.3098–102.
- Lejeune, J., Gautier, M. & Turpin, R., 1959. Study of somatic chromosomes from 9 mongoloid children. *Comptes rendus hebdomadaires des séances de l'Académie des sciences*, 248(11), pp.1721–2.

Lewin, H.A. et al., 2009. Every genome sequence needs a good map. *Genome research*, 19(11), pp.1925–1928.

Lindblad-Toh, K. et al., 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, 438(7069), pp.803–19.

Lithgow, P.E., O'Connor R.E. et al., 2014. Novel tools for characterising inter and intra chromosomal rearrangements in avian microchromosomes. *Chromosome Research*, 22(1), pp.85–97.

Liu, L. et al., 2012. Comparison of Next-Generation Sequencing Systems. *Journal of Biomedicine and Biotechnology*, 2012, pp.1–11.

Loman, N.J., Quick, J. & Simpson, J.T., 2015. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nature Methods*, 12(8), pp.733–735.

Longrich, N.R., Tokaryk, T. & Field, D.J., 2011. Mass extinction of birds at the Cretaceous-Paleogene (K-Pg) boundary. *Proceedings of the National Academy of Sciences of the United States of America*, 108(37), pp.15253–7.

Louis, A., Muffato, M. & Roest Crolius, H., 2013. Genomicus: five genome browsers for comparative genomics in eukaryota. *Nucleic acids research*, 41(Database issue), pp.D700–5.

Lovell, P. V et al., 2014. Conserved syntenic clusters of protein coding genes are missing in birds. *Genome biology*, 15(12), p.565.

Lyon, M.F., 2003. Transmission ratio distortion in mice. *Annual review of genetics*, 37, pp.393–408.

Lyson, T.R. et al., 2010. Transitional fossils and the origin of turtles. *Biology letters*, 6(6), pp.830–3.

Ma, J. et al., 2006. Reconstructing contiguous regions of an ancestral genome. *Genome research*, 16(12), pp.1557–65.

Ma, J., 2010. A probabilistic framework for inferring ancestral genomic orders. In 2010 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, pp. 179–184.

Ma, J., 2011. Reconstructing the history of large-scale genomic changes: biological questions and computational challenges. *Journal of computational biology*, 18(7), pp.879–93.

Mak, A.C.Y. et al., 2016. Genome-Wide Structural Variation Detection by Genome Mapping on Nanochannel Arrays. *Genetics*, 202(1), pp.351–62.

Marais, G., 2003. Biased gene conversion: implications for genome and sex evolution. *Trends in genetics : TIG*, 19(6), pp.330–8.

Mardis, E.R., 2008. Next-generation DNA sequencing methods. *Annual review of genomics and human genetics*, 9, pp.387–402.



Mardis, E.R., 2013. Next-generation sequencing platforms. *Annual review of analytical chemistry* (Palo Alto, Calif.), 6, pp.287–303.

Margulies, M. et al., 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057), p.376.

Martin, C.L. & Warburton, D., 2015. Detection of Chromosomal Aberrations in Clinical Practice: From Karyotype to Genome Sequence. *Annual review of genomics and human genetics*, 16, pp.309–26.

Martinez, R.N. et al., 2011. A basal dinosaur from the dawn of the dinosaur era in southwestern Pangaea. *Science* (New York, N.Y.), 331(6014), pp.206–10.

Masabanda, J.S. et al., 2004. Molecular cytogenetic definition of the chicken genome: the first complete avian karyotype. *Genetics*, 166(3), pp.1367–1373.

Matsubara, K. et al., 2006. Evidence for different origin of sex chromosomes in snakes, birds, and mammals and step-wise differentiation of snake sex chromosomes. *Proceedings of the National Academy of Sciences of the United States of America*, 103(48), pp.18190–5.

Matsubara, K. et al., 2012. Intra-genomic GC heterogeneity in sauropsids: evolutionary insights from cDNA mapping and GC(3) profiling in snake. *BMC genomics*, 13(1), p.604.

Matsuda, Y. et al., 2005. Highly conserved linkage homology between birds and turtles: bird and turtle chromosomes are precise counterparts of each other. *Chromosome Research*, 13(6), pp.601–615.

Mayr, G. et al., 2007. The tenth skeletal specimen of *Archaeopteryx*. *Zoological Journal of the Linnean Society*, 149(1), pp.97–116.

McLachlan, R.I. & O'Bryan, M.K., 2010. Clinical Review: State of the art for genetic testing of infertile men. *The Journal of clinical endocrinology and metabolism*, 95(3), pp.1013–24.

McPherson, J.D. et al., 2001. A physical map of the human genome. *Nature*, 409(6822), pp.934–41.

McQueen, H.A., Siriaco, G. & Bird, A.P., 1998. Chicken Microchromosomes Are Hyperacetylated, Early Replicating, and Gene Rich. *Genome Res.*, 8(6), pp.621–630.

McVean, G.A. et al., 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491(7422), pp.56–65.

Meador, S. et al., 2010. Genome assembly quality: assessment and improvement using the neutral indel model. *Genome research*, 20(5), pp.675–84.

Meyer, H. von., 1861. *Archaeopteryx lithographica* (Vogel-Feder) und *Pterodactylus* von Solnhofen. *Neues Jahrbuch für Mineralogie, Geognosie, Geologie und Petrefakten-Kunde*, pp.678–679.

Miller, J.R., Koren, S. & Sutton, G., 2010. Assembly algorithms for next-generation sequencing data. *Genomics*, 95(6), pp.315–27.

Mitelman, F., 2005. Cancer cytogenetics update 2005. *Atlas Genet Cytogenet Oncol Haematol*.

Morisson, M. et al., 2007. The chicken RH map: current state of progress and microchromosome mapping. *Cytogenetic and genome research*, 117(1-4), pp.14–21.

Muffato, M. & Roest Crolius, H., 2008. Paleogenomics in vertebrates, or the recovery of lost genomes from the mist of time. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 30(2), pp.122–34.

Muffato, M. et al., 2010. Genomicus: a database and a browser to study gene synteny in modern and ancestral genomes. *Bioinformatics (Oxford, England)*, 26(8), pp.1119–21.

Munné, S., 2006. Chromosome abnormalities and their relationship to morphology and development of human embryos. *Reproductive BioMedicine Online*, 12(2), pp.234–253.

Murphy, W.J. et al., 2005. Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science (New York, N.Y.)*, 309(5734), pp.613–7.

Nadeau, J.H. & Taylor, B.A., 1984. Lengths of chromosomal segments conserved since divergence of man and mouse. *Proceedings of the National Academy of Sciences*, 81(3), pp.814–818.

Nakatani, Y. et al., 2007. Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. *Genome research*, 17(9), pp.1254–1265.

Nanda, I. et al., 1999. 300 million years of conserved synteny between chicken Z and human chromosome 9. *Nature genetics*, 21(3), pp.258–9.

Nanda, I. et al., 2006. Extensive gross genomic rearrangements between chicken and Old World vultures (Falconiformes: Accipitridae). *Cytogenetic and genome research*, 112(3-4), pp.286–95.

Nanda, I. et al., 2007. Chromosome repatterning in three representative parrots (*Psittaciformes*) inferred from comparative chromosome painting. *Cytogenetic and genome research*, 117(1-4), pp.43–53.

Nanda, I. et al., 2011. Synteny conservation of chicken macrochromosomes 1-10 in different avian lineages revealed by cross-species chromosome painting. *Cytogenetic and genome research*, 132(3), pp.165–81.

Navarro, A. & Barton, N., 2003a. Chromosomal speciation and molecular divergence-accelerated evolution in rearranged chromosomes. *Science (New York, N.Y.)*, 300(5617), pp.321–4.

Navarro, A. & Barton, N., 2003b. Accumulating Postzygotic Isolation Genes In Parapatry: A New Twist On Chromosomal Speciation. *Evolution*, 57(3), pp.447–459.

Nesbitt, S.J. et al., 2013. The oldest dinosaur? A Middle Triassic dinosauriform from Tanzania. *Biology letters*, 9(1), p.20120949.

Nesje, M. et al., 2000. Genetic relationships in the peregrine falcon (*Falco peregrinus*) analysed by microsatellite DNA markers. *Molecular Ecology*, 9(1), pp.53–60.

Nicholson, D.B. et al., 2015. Climate-mediated diversification of turtles in the Cretaceous. *Nature Communications*, 6, p.7848.

Nie, W. et al., 2009. Avian comparative genomics: reciprocal chromosome painting between domestic chicken (*Gallus gallus*) and the stone curlew (*Burhinus oedicephalus*, *Charadriiformes*)-an atypical species with low diploid number. *Chromosome research*, 17(1), pp.99–113.

Nie, W. et al., 2015. Multidirectional chromosome painting substantiates the occurrence of extensive genomic reshuffling within *Accipitriformes*. *BMC Evolutionary Biology*, 15(1), p.205.

Nishida-Umehara, C. et al., 2007. The molecular basis of chromosome orthologies and sex chromosomal differentiation in palaeognathous birds. *Chromosome research*, 15(6), pp.721–34.

Nishida, C. et al., 2008. Characterization of chromosome structures of Falconinae (Falconidae, *Falconiformes*, *Aves*) by chromosome painting and delineation of chromosome rearrangements during their differentiation. *Chromosome research*, 16(1), pp.171–81.

Nishida, C. et al., 2013. Karyotype reorganization with conserved genomic compartmentalization in dot-shaped microchromosomes in the Japanese mountain hawk-eagle (*Nisaetus nipalensis orientalis*, *Accipitridae*). *Cytogenetic and genome research*, 141(4), pp.284–94.

Nishida, C. et al., 2014. Dynamic chromosome reorganization in the osprey (*Pandion haliaetus*, *Pandionidae*, *Falconiformes*): relationship between chromosome size and the chromosomal distribution of centromeric repetitive DNA sequences. *Cytogenetic and genome research*, 142(3), pp.179–189.

Noor, M.A. et al., 2001. Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences of the United States of America*, 98(21), pp.12084–8.

Noor, M.A.F. & Bennett, S.M., 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, 103(6), pp.439–44.

Nowell, P.C. & Hungerford, D.A., 1960. Chromosome studies on normal and leukemic human leukocytes. *Journal of the National Cancer Institute*, 25, pp.85–109.

O'Connor, C., 2008. Karyotyping for Chromosomal abnormalities. *Nature Education*, p.1 (1): 27.

O'Connor, P.M. & Claessens, L.P.A.M., 2005. Basic avian pulmonary design and flow-through ventilation in non-avian theropod dinosaurs. *Nature*, 436(7048), pp.253–6.

O'Meally, D. et al., 2012. Are some chromosomes particularly good at sex? Insights from amniotes. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, 20(1), pp.7–19.

Ohno, S., 1970. *Evolution by Gene Duplication*, Berlin, Heidelberg: Springer Berlin Heidelberg.

Oleksyk, T.K. et al., 2012. A locally funded Puerto Rican parrot (*Amazona vittata*) genome sequencing project increases avian data and advances young researcher education. *GigaScience*, 1(1), p.14.

Organ, C.L. et al., 2007. Origin of avian genome size and structure in non-avian dinosaurs. *Nature*, 446(7132), pp.180–4.

Organ, C.L., Moreno, R.G. & Edwards, S. V, 2008. Three tiers of genome evolution in reptiles. *Integrative and Comparative Biology*, 48(4), pp.494–504.

Organ, C.L. & Shedlock, A.M., 2009. Palaeogenomics of pterosaurs and the evolution of small genome size in flying vertebrates. *Biology letters*, 5(1), pp.47–50.

Paternoster, L. et al., 2011. Adult height variants affect birth length and growth rate in children. *Human molecular genetics*, 20(20), pp.4069–75.

Paudel, Y. et al., 2013. Evolutionary dynamics of copy number variation in pig genomes in the context of adaptation and domestication. *BMC Genomics*, 14(1), p.449.

Pendleton, M. et al., 2015. Assembly and diploid architecture of an individual human genome via single-molecule technologies. *Nature Methods*, 12(8), pp.780–786.

Peng, Q., Pevzner, P.A. & Tesler, G., 2006. The fragile breakage versus random breakage models of chromosome evolution. *PLoS computational biology*, 2(2), p.e14.

Penney, D. et al., 2013. Absence of ancient DNA in sub-fossil insect inclusions preserved in “Anthropocene” Colombian copal. *PLoS one*, 8(9), p.e73150.

Perry, G.H. et al., 2006. Hotspots for copy number variation in chimpanzees and humans. *Proceedings of the National Academy of Sciences*, 103(21), pp.8006–8011.

Perry, G.H. et al., 2007. Diet and the evolution of human amylase gene copy number variation. *Nature Genetics*, 39(10), pp.1256–1260.

Pevzner, P. & Tesler, G., 2003. Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 100(13), pp.7672–7.

Pichugin, A.M. et al., 2001. Estimation of the Minimal Size of Chicken *Gallus gallus domesticus* Microchromosomes via Pulsed-Field Electrophoresis. *Russian Journal of Genetics*, 37(5), pp.535–538.

Pielberg, G. et al., 2002. Unexpectedly high allelic diversity at the KIT locus causing dominant white color in the domestic pig. *Genetics*, 160(1), pp.305–11.

Pinkel, D. et al., 1986. Cytogenetic Analysis by In Situ Hybridisation with Fluorescently Labeled Nucleic Acid Probes. *Cold Spring Harbor Symposia on Quantitative Biology*, 51(0), pp.151–157.

Pinton, A. et al., 2000. Chromosomal abnormalities in hypoproliferic boars. *Hereditas*, 132(1), pp.55–62.

Pokorná, M. et al., 2011. Strong conservation of the bird Z chromosome in reptilian genomes is revealed by comparative painting despite 275 million years divergence. *Chromosoma*, 120(5), pp.455–68.

Pokorná, M. et al., 2012. Conservation of chromosomes syntenic with avian autosomes in squamate reptiles revealed by comparative chromosome painting. *Chromosoma*, 121(4), pp.409–18.

Pop, M., 2009. Genome assembly reborn: recent computational challenges. *Briefings in bioinformatics*, 10(4), pp.354–66.

Postlethwait, J.H., 2000. Zebrafish Comparative Genomics and the Origins of Vertebrate Chromosomes. *Genome Research*, 10(12), pp.1890–1902.

Price, T.D., 2002. Domesticated birds as a model for the genetics of speciation by sexual selection. In *Springer Netherlands*, pp. 311–327.

Putnam, N.H. et al., 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science (New York, N.Y.)*, 317(5834), pp.86–94.

Putnam, N.H. et al., 2016. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome research*, p.gr.193474.115–.

Quick, J. et al., 2016. Real-time, portable genome sequencing for Ebola surveillance. *Nature*, 530(7589), pp.228–232.

Randhawa, I.A.S. et al., 2015. Composite Selection Signals for Complex Traits Exemplified Through Bovine Stature Using Multibreed Cohorts of European and African *Bos taurus*. *G3 (Bethesda, Md.)*, 5(7), pp.1391–401.

Rands, C.M. et al., 2013. Insights into the evolution of Darwin's finches from comparative analysis of the *Geospiza magnirostris* genome sequence. *BMC genomics*, 14(1), p.95.

Rao, M. et al., 2012. A duck RH panel and its potential for assisting NGS genome assembly. *BMC genomics*, 13(1), p.513.

Ratcliffe, D.A., 1993. *The Peregrine Falcon* 2nd edn., London: Poyser.

Raudsepp, T. et al., 2002. Cytogenetic analysis of California condor (*Gymnogyps californianus*) chromosomes: comparison with chicken (*Gallus gallus*) macrochromosomes. *Cytogenetic and genome research*, 98(1), pp.54–60.

Raudsepp, T. et al., 2004. Exceptional conservation of horse-human gene order on X chromosome revealed by high-resolution radiation hybrid mapping. *Proceedings of the National Academy of Sciences*, 101(8), pp.2386–2391.

Rauhut, O.W.M. et al., 2012. A new rhynchocephalian from the late jurassic of Germany with a dentition that is unique amongst tetrapods. *PloS one*, 7(10), p.e46839.

Ravnan, J.B. et al., 2006. Subtelomere FISH analysis of 11 688 cases: an evaluation of the frequency and pattern of subtelomere rearrangements in individuals with developmental disabilities. *Journal of medical genetics*, 43(6), pp.478–489.

Redon, R. et al., 2006. Global variation in copy number in the human genome. *Nature*, 444(7118), pp.444–454.

Reed, K.M. et al., 2005. A comparative genetic map of the turkey genome. *Cytogenetic and genome research*, 111(2), pp.118–27.

Reed, K.M., Chaves, L.D. & Mendoza, K.M., 2007. An integrated and comparative genetic map of the turkey genome. *Cytogenetic and genome research*, 119(1-2), pp.113–26.

Rens, W. et al., 2006. Cross-species chromosome painting. *Nature protocols*, 1(2), pp.783–90.

Rhoads, A. & Au, K.F., 2015. PacBio Sequencing and Its Applications. *Genomics, Proteomics & Bioinformatics*, 13(5), pp.278–289.

Rickman, L. et al., 2005. Prenatal diagnosis by array-CGH. *European journal of medical genetics*, 48(3), pp.232–40.

Rieseberg, L.H., 2001. Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution*, 16(7), pp.351–358.

Roberts, R.J. et al., 2013. The advantages of SMRT sequencing. *Genome Biology* 2013 14:7, 14(7), pp.133–138.

Robinson, T.J., Ruiz-Herrera, A. & Avise, J.C., 2008. Hemiplasy and homoplasy in the karyotypic phylogenies of mammals. *Proceedings of the National Academy of Sciences of the United States of America*, 105(38), pp.14477–81.

Rodrigues, B.S. et al., 2014. Chromosomal studies on *Coscoroba coscoroba* (Aves: *Anseriformes*) reinforce the Coscoroba-Cereopsis clade. *Biological Journal of the Linnean Society*, 111(2), pp.274–279.

Rodríguez-Gil, J.E. & Estrada, E., 2013. Boar Reproduction: Fundamentals and New Biotechnological Trends. In S. Bonet et al., eds. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 589–607.

Rodríguez, A. et al., 2010. Reproductive consequences of a reciprocal chromosomal translocation in two Duroc boars used to provide semen for artificial insemination. *Theriogenology*, 74(1), pp.67–74.

Rogers, J. & Gibbs, R.A., 2014. Comparative primate genomics: emerging patterns of genome content and dynamics. *Nature Reviews Genetics*, 15(5), pp.347–359.

Romanov, M.N. et al., 2014. Reconstruction of gross avian genome structure, organization and evolution suggests that the chicken lineage most closely resembles the dinosaur avian ancestor. *BMC genomics*, 15(1), p.1060.

Rommens, J.M. et al., 1989. Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science (New York, N.Y.)*, 245(4922), pp.1059–65.

Ronaghi, M. et al., 1996. Real-Time DNA Sequencing Using Detection of Pyrophosphate Release. *Analytical Biochemistry*, 242(1), pp.84–89.

Ronquist, F. & Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19(12), pp.1572–1574.

Ross, M.T. et al., 2005. The DNA sequence of the human X chromosome. *Nature*, 434(7031), pp.325–37.

Rothberg, J.M. et al., 2011. An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, 475(7356), pp.348–352.

Rothschild, M.F. & Ruvinsky, A., 2011. *The genetics of the pig*, CABI.

Rowley, J.D., 1973. A New Consistent Chromosomal Abnormality in Chronic Myelogenous Leukaemia identified by Quinacrine Fluorescence and Giemsa Staining. *Nature*, 243(5405), pp.290–293.

Ruiz-Herrera, A., Farré, M. & Robinson, T.J., 2011. Molecular cytogenetic and genomic insights into chromosomal evolution. *Heredity*, 108(1), pp.28–36.

Saccone, S. et al., 1992. The highest gene concentrations in the human genome are in telomeric bands of metaphase chromosomes. *Proceedings of the National Academy of Sciences*, 89(11), pp.4913–4917.

Saccone, S. et al., 1993. Correlations between isochores and chromosomal bands in the human genome. *Proceedings of the National Academy of Sciences of the United States of America*, 90(24), pp.11929–33.

Sakai, H. et al., 2015. The power of single molecule real-time sequencing technology in the *de novo* assembly of a eukaryotic genome. *Scientific reports*, 5, p.16780.

Salzberg, S.L. & Yorke, J.A., 2005. Beware of mis-assembled genomes. *Bioinformatics (Oxford, England)*, 21(24), pp.4320–1.

Sanger, F., Nicklen, S. & Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), pp.5463–7.

Sankoff, D., 1999. Genome rearrangement with gene families. *Bioinformatics*, 15(11), pp.909–917.

Sankoff, D., 2009. The where and wherefore of evolutionary breakpoints. *Journal of biology*, 8(7), p.66.

Santana, J.A.M., Gardner, L.I. & Neu, R.L., 1977. The X Isochromosome-X Syndrome [46,X,i(Xq)]: Report of Three Cases with Review of the Phenotype. *Clinical Pediatrics*, 16(11), pp.1021–1026.

Schinzel, A., 2001. Catalogue of Unbalanced Chromosome Aberrations in Man, Walter de Gruyter.

Schmid, M. et al., 2005. Second report on chicken genes and chromosomes 2005. *Cytogenetic and genome research*, 109(4), pp.415–79.

Schmid, M. et al., 2015. Third Report on Chicken Genes and Chromosomes 2015. *Cytogenetic and genome research*, 145(2), pp.78–179.

Schrock, E. et al., 1996. Multicolor Spectral Karyotyping of Human Chromosomes. *Science*, 273(5274), pp.494–497.

Schubert, I. & Lysak, M.A., 2011. Interpretation of karyotype evolution should consider chromosome structural constraints. *Trends in Genetics*, 27(6), pp.207–216.

Schulte, P. et al., 2010. The Chicxulub asteroid impact and mass extinction at the Cretaceous-Paleogene boundary. *Science (New York, N.Y.)*, 327(5970), pp.1214–8.

Scriven, P.N., 2001. Robertsonian translocations--reproductive risks and indications for preimplantation genetic diagnosis. *Human Reproduction*, 16(11), pp.2267–2273.

Sémon, M. & Duret, L., 2006. Evolutionary origin and maintenance of coexpressed gene clusters in mammals. *Molecular biology and evolution*, 23(9), pp.1715–23.

Shaffer, H.B. et al., 2013. The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome biology*, 14(3), p.R28.

Shaffer, L. & Tommerup, N., 2005. *ISCN 2005: An International System for Human Cytogenetic Nomenclature (2005) : Recommendations of the International Standing Committee on Human Cytogenetic Nomenclature*, Karger Medical and Scientific Publishers.

Shah, K. et al., 2003. The genetic basis of infertility. *Reproduction*, 126(1), pp.13–25.

Shapiro, M.D. et al., 2013. Genomic diversity and evolution of the head crest in the rock pigeon. *Science (New York, N.Y.)*, 339(6123), pp.1063–7.

Shedlock, A.M. & Edwards, S.V., 2009. Amniotes. In *The Timetree of Life*. OUP Oxford, pp. 375–379.

Shendure, J. et al., 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science (New York, N.Y.)*, 309(5741), pp.1728–32.

Shendure, J. & Ji, H., 2008. Next-generation DNA sequencing. *Nature Biotechnology*, 26(10), pp.1135–1145.

Shetty, S., Griffin, D.K. & Graves, J.A.M., 1999. Comparative Painting Reveals Strong Chromosome Homology Over 80 Million Years of Bird Evolution. *Chromosome Research*, 7(4), pp.289–295.

Shibusawa, M. et al., 2002. Chromosome rearrangements between chicken and guinea fowl defined by comparative chromosome painting and FISH mapping of DNA clones. *Cytogenetic and genome research*, 98(2-3), pp.225–30.



- Shibusawa, M. et al., 2004. Karyotypic evolution in the *Galliformes*: an examination of the process of karyotypic evolution by comparison of the molecular cytogenetic findings with the molecular phylogeny. *Cytogenetic and genome research*, 106(1), pp.111–9.
- Sigurdardottir, S. et al., 1999. Clinical, cytogenetic, and fluorescence in situ hybridisation findings in two cases of “complete ring” syndrome. *American journal of medical genetics*, 87(5), pp.384–90.
- Simmonds, N.W., 1976. Evolution of crop plants.
- Singer, G.A.C. et al., 2005. Clusters of co-expressed genes in mammalian genomes are conserved by natural selection. *Molecular biology and evolution*, 22(3), pp.767–75.
- Skaletsky, H. et al., 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, 423(6942), pp.825–37.
- Skinner, B.M. & Griffin, D.K., 2011. Intrachromosomal rearrangements in avian genome evolution: evidence for regions prone to breakpoints. *Heredity*, 108(1), pp.37–41.
- Skinner, B.M. et al., 2009. Comparative genomics in chicken and Pekin duck using FISH mapping and microarray analysis. *BMC genomics*, 10(1), p.357.
- Skinner, B.M. et al., 2014. Global patterns of apparent copy number variation in birds revealed by cross-species comparative genomic hybridisation. *Chromosome Research*, 22(1), pp.59–70.
- Smith et al., 2000. Integration of the genetic and physical maps of the chicken macrochromosomes. *Animal genetics*, 31(1), pp.20–27.
- Smith, C.A. et al., 2009. The avian Z-linked gene DMRT1 is required for male sex determination in the chicken. *Nature*, 461(7261), pp.267–71.
- Smith, J., Bruley, C.K., et al., 2000. Differences in gene density on chicken macrochromosomes and microchromosomes. *Animal Genetics*, 31(2), pp.96–103.
- Smith, L.M. et al., 1986. Fluorescence detection in automated DNA sequence analysis. *Nature*, 321(6071), pp.674–9.
- Smith, T.P.L. et al., 2016. Approaches Taken, Progress Made, and Enhanced Utility of Long Read-based Goat, Swine, Cattle and Sheep Reference Genomes. In *Plant and Animal Genome*.
- Speicher, M.R. & Carter, N.P., 2005. The new cytogenetics: blurring the boundaries with molecular biology. *Nature reviews. Genetics*, 6(10), pp.782–92.
- Speicher, M.R., Gwyn Ballard, S. & Ward, D.C., 1996. Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nature genetics*, 12(4), pp.368–75.
- Stankiewicz, P. & Lupski, J.R., 2002. Molecular-evolutionary mechanisms for genomic disorders. *Current Opinion in Genetics & Development*, 12(3), pp.312–319.
- Steel, M. & Penny, D., 2000. Parsimony, likelihood, and the role of models in molecular phylogenetics. *Molecular biology and evolution*, 17(6), pp.839–50.

Stergiakouli, E. et al., 2014. Genome-wide association study of height-adjusted BMI in childhood identifies functional variant in ADCY3. *Obesity (Silver Spring, Md.)*, 22(10), pp.2252–9.

Stern, C., 1999. Chromosome translocations in couples with in-vitro fertilization implantation failure. *Human Reproduction*, 14(8), pp.2097–2101.

Strasburg, J.L. et al., 2009. Genomic patterns of adaptive divergence between chromosomally differentiated sunflower species. *Molecular biology and evolution*, 26(6), pp.1341–55.

Stringham, S.A. et al., 2012. Divergence, Convergence, and the Ancestry of Feral Populations in the Domestic Rock Pigeon,

Su, N. et al., 2014. Role of FGF/FGFR signaling in skeletal development and homeostasis: learning from mouse models. *Bone Research*, 2, p.14003.

Sun, S.C. et al., 2014. A *de novo* chromosomal abnormality in Cri du Chat syndrome. *Indian journal of pediatrics*, 81(7), pp.722–5.

Tagliarini, M.M. et al., 2011. Maintenance of syntenic groups between Cathartidae and *Gallus gallus* indicates symplesiomorphic karyotypes in new world vultures. *Genetics and molecular biology*, 34(1), pp.80–3.

Takenouchi, T. et al., 2012. 12q14 microdeletion syndrome and short stature with or without relative macrocephaly. *American journal of medical genetics. Part A*, 158A(10), pp.2542–4.

Tassano, E. et al., 2013. Genotype-Phenotype Correlation of 2q37 Deletions Including NPPC Gene Associated with Skeletal Malformations B. He, ed. *PLoS ONE*, 8(6), p.e66048.

Tatton-Brown, K. et al., 2014. Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. *Nature genetics*, 46(4), pp.385–8.

Taylor, T.H. et al., 2014. The origin, mechanisms, incidence and clinical consequences of chromosomal mosaicism in humans. *Human reproduction update*, 20(4), pp.571–81.

Teague, B. et al., 2010. High-resolution human genome structure by single-molecule analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 107(24), pp.10848–53.

Telenius, H. et al., 1992. Cytogenetic analysis by chromosome painting using dop-pcr amplified flow-sorted chromosomes. *Genes, Chromosomes and Cancer*, 4(3), pp.257–263.

Tempest, H.G. & Simpson, J.L., 2010. Role of preimplantation genetic diagnosis (PGD) in current infertility practice. *Int J Infertil Fetal Med*, 1, pp.1–10.

The BPEX Yearbook 2013-14, A. and H.D.B., 2014.

The Chimpanzee Sequencing and Analysis Consortium, 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, 437(7055), pp.69–87.

Thomas, S. et al., 2014. A homozygous PDE6D mutation in Joubert syndrome impairs targeting of farnesylated INPP5E protein to the primary cilium. *Human mutation*, 35(1), pp.137–46.

Todd, N.B., 1970. Karyotypic fissioning and Canid phylogeny. *Journal of Theoretical Biology*, 26(3), pp.445–480.

Torres, E.M., Williams, B.R. & Amon, A., 2008. Aneuploidy: cells losing their balance. *Genetics*, 179(2), pp.737–46.

Trask, B.J., 2002. Human cytogenetics: 46 chromosomes, 46 years and counting. *Nature reviews. Genetics*, 3(10), pp.769–78.

Tsien, R. Y.; Ross, P.; Fahnestock, M.; Johnston, A. J. WO Patent 91/06678, 1991.

Tucker, V., 1998. Gliding flight: speed and acceleration of ideal falcons during diving and pull out. *The Journal of experimental biology*, 201(Pt 3), pp.403–14.

United States Department of Agriculture, F.A.S., 2015. *Livestock and Poultry: World Markets and Trade*.

Uno, Y. et al., 2012. Inference of the protokaryotypes of amniotes and tetrapods and the evolutionary processes of microchromosomes from comparative gene mapping. *PloS one*, 7(12), p.e53027.

Utsunomiya, Y.T. et al., 2013. Genome-wide association study for birth weight in Nellore cattle points to previously described orthologous genes affecting human and bovine height. *BMC Genetics*, 14(1), p.52.

van Duyvenvoorde, H.A. et al., 2014. Copy number variants in patients with short stature. *European journal of human genetics : EJHG*, 22(5), pp.602–9.

Varricchio, D.J. et al., 2008. Avian paternal care had dinosaur origin. *Science (New York, N.Y.)* 322(5909), pp.1826–8.

Venter, J.C. et al., 2001. The sequence of the human genome. *Science (New York, N.Y.)*, 291(5507), pp.1304–51.

Ventura, M. et al., 2003. Neocentromeres in 15q24-26 map to duplicons which flanked an ancestral centromere in 15q25. *Genome research*, 13(9), pp.2059–68.

Völker, M. et al., 2010. Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution. *Genome research*, 20(4), pp.503–511.

Wallis, J.W. et al., 2004. A physical map of the chicken genome. *Nature*, 432(7018), pp.761–764.

Warr, A. et al., 2015. Identification of Low-Confidence Regions in the Pig Reference Genome (Sscrofa10.2). *Frontiers in genetics*, 6, p.338.

Warren, W.C. et al., 2010. The genome of a songbird. *Nature*, 464(7289), pp.757–762.

Waters, P.D., Wallis, M.C. & Marshall Graves, J.A., 2007. Mammalian sex--Origin and evolution of the Y chromosome and SRY. *Seminars in cell & developmental biology*, 18(3), pp.389–400.

Waterston, R.H. et al., 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420(6915), pp.520–62.

Weir, J.T. & Schluter, D., 2007. The latitudinal gradient in recent speciation and extinction rates of birds and mammals. *Science (New York, N.Y.)*, 315(5818), pp.1574–6.

Weishampel, D.O., 2004. *The Dinosauria: Second Edition*, University of California Press.

Weiss, M.M. et al., 1999. Comparative genomic hybridisation. *Molecular pathology : MP*, 52(5), pp.243–51.

White, M.J.D., 1969. Chromosomal Rearrangements and Speciation in Animals. *Annual Review of Genetics*, 3(1), pp.75–98.

Wienberg, J. & Stanyon, R., 1997. Comparative painting of mammalian chromosomes. *Current Opinion in Genetics & Development*, 7(6), pp.784–791.

Wienberg, J., 2004. The evolution of eutherian chromosomes. *Current opinion in genetics & development*, 14(6), pp.657–66.

Wienberg, J., L. Froenicke, 2000 and Stanyon R. Insights into Mammalian Genome Organization and Evolution by Molecular Cytogenetics. In *Comparative Genomics*. Springer US, pp. 207–244.

Widén, E. et al., 2010. Distinct variants at LIN28B influence growth in height from birth to adulthood. *American journal of human genetics*, 86(5), pp.773–82.

Witmer, L.M., 2002. The debate on avian ancestry: phylogeny, function, and fossils. *Mesozoic birds: above the heads of dinosaurs*, pp.3-30.

Wurster, D.H. & Benirschke, K., 1970. Indian Momtjac, *Muntiacus muntjak*: A Deer with a Low Diploid Chromosome Number. *Science*, 168(3937), pp.1364–1366.

Yang, F. et al., 1999. A complete comparative chromosome map for the dog, red fox, and human and its integration with canine genetic maps. *Genomics*, 62(2), pp.189–202.

Yunis, J. & Prakash, O., 1982. The origin of man: a chromosomal pictorial legacy. *Science*, 215(4539), pp.1525–1530.

Zarrei, M. et al., 2015. A copy number variation map of the human genome. *Nature Reviews Genetics*, 16(3), pp.172–183.

Zhan, X. et al., 2013. Peregrine and saker falcon genome sequences provide insights into evolution of a predatory lifestyle. *Nature genetics*, 45(5), pp.563–566.

Zhang, G. et al., 2012. The Genome of Darwin's Finch (*Geospiza fortis*). *GigaScience*, 1, p.13.

Zhang, G. et al., 2014a. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science (New York, N.Y.)*, 346(6215), pp.1311–20.

Zhang, G. et al., 2014b. Comparative genomic data of the Avian Phylogenomics Project. *GigaScience*, 3(1), p.26.

Zhang, G. et al., 2015. Genomics: Bird sequencing project takes off. *Nature*, 522(7554), p.34.

Zhang, J. et al., 2015 Improving the ostrich genome assembly using optical mapping data. *GigaScience*, 4, p.24.

Zhang, J., Wang, X. & Podlaha, O., 2004. Testing the chromosomal speciation hypothesis for humans and chimpanzees. *Genome research*, 14(5), pp.845–51.

Zhao, J. et al., 2010. The role of height-associated loci identified in genome wide association studies in the determination of pediatric stature. *BMC Medical Genetics*, 11(1), p.96.

Zheng, Q. & Wang, X.-J., 2008. GOEAST: a web-based software toolkit for Gene Ontology enrichment analysis. *Nucleic acids research*, 36(Web Server issue), pp.W358–63.

Zhou, Q. et al., 2014. Complex evolutionary trajectories of sex chromosomes across bird taxa. *Science (New York, N.Y.)*, 346(6215), p.1246338.

## 9 Appendix

**Table S 1: List of BACs and their placement in the pigeon (CLI – *Columba livia*) genome.**

CLI Chrom	BAC Name	CLI Start Position	CLI End Position	GGA Homolog
1	CH261-89C18	396,281	567,747	1
1	TGMCBA-206D5	5,887,762	6,035,883	1
1	CH261-89G23	19,959,217	20,196,167	1
1	CH261-119K2	29,209,698	29,446,014	1
1	CH261-25P18	62,666,825	62,909,469	1
1	CH261-36B5	73,883,062	74,090,665	1
1	CH261-29N14	100,209,284	100,405,963	1
1	CH261-18J16	110,416,860	110,675,329	1
1	TGMCBA-204O4	114,507,041	114,681,760	1
1	CH261-118M1	128,109,054	128,338,364	1
1	TGMCBA-146O14	136,320,261	136,484,766	1
1	CH261-9B17	137,796,422	138,001,016	1
1	CH261-168O17	144,561,115	144,807,137	1
1	CH261-83O13	149,854,217	150,093,874	1
1	CH261-107E2	159,438,366	159,667,425	1
1	CH261-58K12	170,413,895	170,616,548	1
1	CH261-184E5	176,515,781	176,749,790	1
1	CH261-98G4	193,759,470	193,942,338	1
2	CH261-192C19	173,423	349,921	2
2	CH261-169N6	5,647,981	5,843,562	2
2	CH261-172N3	13,675,655	13,823,925	2
2	CH261-177K1	23,905,092	24,148,670	2
2	CH261-186J5	31,688,248	31,916,077	2
2	CH261-40G6	33,678,681	33,866,544	2
2	CH261-50C15	36,314,195	36,549,331	2
2	TGMCBA-340P4	42,461,146	42,616,308	2
2	CH260-1M4	63,744,914	63,946,548	2
2	TGMCBA-78C11	81,476,525	81,602,766	2
2	CH261-169E4	86,072,127	86,301,386	2
2	CH261-1J20	95,371,258	95,606,649	2
2	CH261-44D16	101,725,243	101,930,318	2
2	CH261-44H14	110,741,710	110,947,426	2
3	CH261-115J5	1,540,042	1,722,678	3
3	CH261-18I9	3,676,313	3,794,600	3
3	TGMCBA-295P5	9,545,415	9,664,441	3
3	CH261-130M12	15,313,919	15,479,372	3
3	CH261-160I6	19,952,734	20,180,816	3
3	CH261-97P20	34,895,064	35,131,528	3
3	CH261-17B14	60,023,468	60,247,486	3
3	TGMCBA-250J17	71,111,183	71,282,517	3
3	CH261-169K18	90,003,534	90,144,106	3
3	TGMCBA-64D9	98,414,246	98,535,331	3
3	CH261-120H23	108,368,313	108,549,014	3
4	TGMCBA-104H6	3,490,482	3,701,988	4
4	CH261-93H1	9,640,680	9,856,611	4
4	CH261-18C6	22,788,584	23,041,430	4
4	CH261-85H10	31,613,037	31,858,379	4
4	CH261-89P6	40,544,253	40,685,944	4
4	TGMCBA-216A16	51,589,566	51,743,409	4

Table S 1 continued: List of BACs and their placement in the pigeon (CLI – *Columba livia*) genome.

CLI Chrom	BAC Name	CLI Start Position	CLI End Position	GGA Homolog
4A	TGMCBA-200G5	490,480	635,804	4p
4A	CH261-111A15	1,360,305	1,582,612	4p
4A	TGMCBA-220A5	5,648,252	5,796,609	4p
4A	TGMCBA-280M7	9,036,323	9,177,873	4p
4A	TGMCBA-330J11	12,812,354	12,944,510	4p
4A	CH261-83E1	13,392,588	13,564,675	4p
4A	CH261-71L6	13,450,684	13,638,480	4p
5	TGMCBA-145C6	1,710,548	1,859,041	5
5	CH261-49B22	5,927,861	6,121,261	5
5	CH261-122F8	10,171,952	10,342,381	5
5	CH261-78F13	24,604,583	24,791,602	5
5	CH261-161B22	46,225,996	46,384,725	5
5	TGMCBA-24C1	46,604,285	46,755,808	5
6	CH261-56K7	4,013,225	4,174,879	7
6	TGMCBA-34L13	8,331,372	8,435,629	7
6	CH261-112D24	8,397,152	8,564,543	7
6	TGMCBA-224O13	13,534,346	13,575,713	7
6	CH261-180H18	20,168,959	20,402,201	7
6	CH261-186K14	25,773,892	25,950,076	7
7	CH261-94G14	3,400,195	3,584,009	6
7	TGMCBA-382J4	7,624,890	7,762,444	6
7	CH261-67H5	17,053,102	17,302,575	6
7	CH261-49F3	17,589,591	17,702,311	6
8	CH261-34H16	4,376,843	4,571,727	8
8	CH261-69H1	13,068,255	13,268,303	8
8	CH261-107D8	13,630,315	13,855,214	8
8	TGMCBA-208D17	19,420,087	19,627,402	8
8	CH261-96D24	20,458,664	20,651,954	8
8	TGMCBA-252A4	25,168,874	25,296,956	8
9	CH261-187M16	7,035,104	7,223,504	9
9	CH261-95N3	14,005,701	14,199,366	9
9	CH261-183N19	21,568,422	21,743,024	9
10	CH261-71G18	10,392,190	10,610,577	10
10	CH261-115G24	15,122,158	15,334,323	10
11	CH261-154H1	8,213,836	8,429,006	11
11	TGMCBA-192A10	12,753,590	12,871,235	11
11	CH261-121N21	13,587,290	13,829,922	11
11	CH261-138H13	15,068,803	15,249,380	11
12	TGMCBA-305E19	422,066	600,021	12
12	CH261-4M5	5,426,227	5,618,549	12
12	TGMCBA-342P15	6,330,361	6,471,420	12
12	CH261-95H20	6,933,184	7,154,139	12
12	CH261-90N18	10,856,889	11,015,439	12
12	CH261-60P3	11,734,413	11,876,407	12
13	TGMCBA-266O5	5,276,001	5,446,961	13
13	TGMCBA-321B13	10,471,364	10,621,097	13
13	CH261-59M8	11,909,002	12,109,283	13
13	CH261-115I12	11,973,595	12,167,816	13
14	CH261-69D20	5,467,792	5,655,151	14
14	TGMCBA-205N19	12,402,979	12,547,940	14
14	CH261-122H14	12,614,373	12,812,489	14
14	CH261-49P24	14,213,620	14,377,578	14

**Table S 1 continued: List of BACs and their placement in the pigeon (CLI – *Columba livia*) genome.**

CLI Chrom	BAC Name	CLI Start Position	CLI End Position	GGA Homolog
15	CH261-90P23	238,561	423,949	15
15	CH261-48M1	1,225,404	1,409,289	15
15	CH261-40D6	11,465,150	11,658,292	15
15	TGMCBA-266G23	12,272,126	12,426,987	15
17	TGMCBA-185B22	312,641	414,140	17
17	TGMCBA-67H23	1,411,836	1,532,268	17
17	TGMCBA-375I5	3,223,829	3,353,117	17
17	CH261-113A7	7,979,387	8,126,436	17
17	TGMCBA-197G19	8,378,667	8,559,970	17
17	CH261-42P16	8,694,779	8,866,801	17
20	TGMCBA-250E3	730,827	879,685	20
20	TGMCBA-341F20	1,945,983	2,073,663	20
20	TGMCBA-225I12	4,255,164	4,412,511	20
18	CH261-72B18	4,545,824	4,718,147	18
18	CH261-118D24	6,007,607	6,178,052	18
19	TGMCBA-84A3	2,440,374	2,564,988	19
19	CH261-50H12	2,936,830	3,088,242	19
19	TGMCBA-356O18	6,637,725	6,785,311	19
19	TGMCBA-307H9	7,221,622	7,338,338	19
19	CH261-10F1	7,752,151	7,892,680	19
24	CH261-103F4	1,101,545	1,253,496	24
24	TGMCBA-111K1	1,818,895	1,959,659	24
24	CH261-65O4	2,805,683	2,957,241	24
24	TGMCBA-82A15	5,129,435	5,389,869	24
21	CH261-83I20	2,018,298	2,212,146	21
21	CH261-122K8	3,544,023	3,725,136	21
23	CH261-191G17	43,769	245,643	23
23	CH261-105P1	3,723,223	3,907,962	23
23	CH261-49G9	4,046,994	4,268,615	23
23	CH261-90K11	4,819,211	4,981,476	23
27	TGMCBA-23C5	1,256,046	1,414,598	27
27	CH261-66M16	3,044,636	3,220,773	27
27	CH261-28L10	3,422,500	3,638,617	27
27	TGMCBA-324P4	4,742,254	4,839,372	27
22	CH261-40J9	2,864,743	3,043,053	22
22	CH261-18G17	3,282,539	3,497,361	22
22	TGMCBA-113N13	3,291,901	3,451,771	22
26	CH261-170L23	1,786,864	1,981,437	26
26	TGMCBA-297G21	2,350,596	2,459,583	26
26	TGMCBA-332G15	2,610,752	2,686,442	26
26	TGMCBA-97D20	2,866,387	2,951,800	26
28	CH261-64A15	317,591	487,671	28
25	TGMCBA-65M1	322,016	549,575	25
25	CH261-59C21	552,119	696,008	25
25	CH261-127K7	594,734	707,995	25
Z	CH261-129A16	7,630,958	7,856,832	Z
Z	TGMCBA-200J22	12,806,452	12,986,657	Z
Z	TGMCBA-270I9	18,018,657	18,175,405	Z
Z	CH261-133M4	36,150,190	36,342,409	Z



Table S 2: List of BACs and their assigned position in the peregrine falcon (FPE - *Falco peregrinus*).

FPE Chrom	BAC Name	FPE Start Position	FPE End Position	GGA Homolog
1	CH261-179F2	1,769,902	1,939,111	6
1	TGMCBA-382J4	12,951,026	13,088,580	6
1	CH261-94G14	14,968,854	15,152,668	6
1	CH261-49F3	29,207,285	29,320,005	6
1	CH261-165L8	29,411,174	29,606,901	6
1	CH261-67H5	29,607,021	29,856,494	6
1	TGMCBA-375I5	43,808,073	43,937,361	17
1	TGMCBA-67H23	45,628,922	45,749,354	17
1	TGMCBA-185B22	46,747,050	46,848,549	17
1	CH261-113A7	50,886,316	51,033,365	17
1	TGMCBA-197G19	51,285,596	51,466,899	17
1	CH261-42P16	51,601,708	51,773,730	17
1	CH261-69M11	51,870,473	52,039,486	17
1	CH261-78F13	54,125,925	54,312,944	5
1	CH261-2I23	75,843,902	76,011,796	5
1	TGMCBA-24C1	99,258,695	99,410,218	5
1	CH261-161B22	99,629,778	99,788,507	5
1	CH261-115G24	111,013,651	111,225,816	10
1	CH261-71G18	115,737,397	115,955,784	10
1	CH261-118E15	116,642,169	116,826,602	10
1	TGMCBA-310P11	118,627,420	118,800,758	10
1	TGMCBA-48L18	124,613,053	124,746,909	10
2	CH261-18C6	3,828,519	4,081,365	4
2	CH261-93H1	10,226,917	10,442,848	4
2	CH261-89P6	30,373,516	30,515,207	4
2	CH261-185L11	37,310,737	37,576,536	4
2	TGMCBA-104H6	41,155,714	41,367,220	4
2	CH261-85H10	51,935,828	52,181,170	4
2	TGMCBA-216A16	64,295,283	64,449,126	4
2	TGMCBA-266G23	76,025,372	76,180,233	15
2	CH261-40D6	76,794,067	76,987,209	15
2	CH261-90P23	81,620,770	81,806,158	15
2	CH261-48M1	82,607,613	82,791,498	15
2	TGMCBA-231D20	82,729,435	82,933,618	15
2	CH261-10F1	86,035,749	86,176,278	19
2	TGMCBA-307H9	87,812,819	87,929,535	19
2	TGMCBA-356O18	88,365,846	88,513,432	19
2	CH261-50H12	92,062,915	92,214,327	19
2	TGMCBA-84A3	92,586,169	92,710,783	19
2	CH261-118D24	101,569,217	101,739,662	18
2	CH261-67N15	104,750,634	104,934,652	18
2	TGMCBA-263I20	105,080,854	105,226,158	18
2	CH261-72B18	105,825,027	105,997,350	18
2	CH261-137B21	107,818,800	107,970,563	18

**Table S 2 continued: List of BACs and their assigned position in the peregrine falcon (FPE - *Falco peregrinus*).**

FPE Chrom	BAC Name	FPE Start Position	FPE End Position	GGA Homolog
3	CH261-17J16	4,388,682	4,569,835	2
3	CH261-44H14	19,351,414	19,557,130	2
3	CH261-44D16	32,284,640	32,489,715	2
3	CH261-1J20	38,608,309	38,843,700	2
3	CH261-169E4	47,913,572	48,142,831	2
3	TGMCBA-78C11	52,612,192	52,738,433	2
3	TGMCBA-340P4	81,395,866	81,551,028	2
3	CH261-10A18	84,877,903	85,078,588	21
3	CH261-83I20	86,889,512	87,083,360	21
3	CH261-122K8	89,299,995	89,481,108	21
3	TGMCBA-134A8	91,212,071	91,368,030	21
3	TGMCBA-272G9	93,173,632	93,373,543	23
3	CH261-49G9	93,208,304	93,429,925	23
3	TGMCBA-48O8	93,738,766	93,864,280	23
3	CH261-90K11	93,980,521	94,142,786	23
3	CH261-191G17	95,510,529	95,711,325	23
3	TGMCBA-173N15	96,820,938	96,956,342	23
4	TGMCBA-146O14	2,652,611	2,817,116	1
4	CH261-118M1	9,832,849	10,062,159	1
4	CH261-29N14	31,068,771	31,265,450	1
4	TGMCBA-204O4	45,366,528	45,541,247	1
4	CH261-9B17	51,496,123	51,700,717	1
4	CH261-168O17	60,605,379	60,851,401	1
4	CH261-83O13	65,898,481	66,138,138	1
4	CH261-107E2	75,482,630	75,711,689	1
4	CH261-58K12	86,458,159	86,660,812	1
4	CH261-184E5	92,560,045	92,794,054	1
4	CH261-98G4	109,803,734	109,986,602	1
5	CH261-123O22	591,762	769,808	2
5	CH261-50C15	4,266,071	4,501,207	2
5	CH261-40G6	14,443,305	14,631,168	2
5	CH261-186J5	23,070,569	23,298,398	2
5	CH261-177K1	28,445,531	28,689,109	2
5	CH261-172N3	38,770,276	38,918,546	2
5	CH261-169N6	46,750,639	46,946,220	2
5	CH261-192C19	52,636,360	52,812,858	2
5	CH261-64A15	53,490,343	53,660,423	28
5	CH261-186C5	53,495,773	53,631,229	28
5	TGMCBA-37M13	54,432,484	54,687,626	28
5	TGMCBA-231J13	55,185,524	55,260,375	28
5	CH261-101C8	58,059,907	58,254,681	28
5	CH261-72A10	58,102,614	58,312,770	28
5	TGMCBA-205N19	63,302,288	63,447,249	14
5	CH261-122H14	63,513,682	63,711,798	14
5	CH261-49P24	65,112,929	65,276,887	14
5	CH261-69D20	73,293,730	73,481,089	14
5	CH261-4M5	82,319,299	82,511,621	12
5	TGMCBA-342P15	83,223,433	83,364,492	12
5	CH261-95H20	83,826,256	84,047,211	12
5	CH261-60P3	88,627,485	88,769,479	12
5	TGMCBA-305E19	98,539,798	98,717,753	12

**Table S 2 continued: List of BACs and their assigned position in the peregrine falcon (FPE - *Falco peregrinus*).**

FPE Chrom	BAC Name	FPE Start Position	FPE End Position	GGA Homolog
6	CH261-89G23	10,446,498	10,683,448	1
6	CH261-119K2	19,696,979	19,933,295	1
6	CH261-120J2	24,096,835	24,330,448	1
6	TGMCBA-206D5	41,056,811	41,204,932	1
6	CH261-25P18	46,942,920	47,185,564	1
6	CH261-36B5	55,034,817	55,242,420	1
7	TGMCBA-295P5	1,289,154	1,408,180	3
7	CH261-97P20	9,963,097	10,199,561	3
7	CH261-120H23	14,528,321	14,709,022	3
7	CH261-17B14	25,651,084	25,875,102	3
7	TGMCBA-64D9	54,631,818	54,752,903	3
7	CH261-169K18	63,023,043	63,163,615	3
7	TGMCBA-250J17	81,884,632	82,055,966	3
8	TGMCBA-224O13	1,132,822	1,279,373	7
8	TGMCBA-34L13	4,189,270	4,293,527	7
8	CH261-112D24	4,255,050	4,422,441	7
8	CH261-56K7	10,218,588	10,380,242	7
8	CH261-180H18	21,999,669	22,232,911	7
8	CH261-186K14	27,604,602	27,780,786	7
8	TGMCBA-356H21	29,616,579	29,737,495	7
8	CH261-38E18	40,553,271	40,744,381	7
8	TGMCBA-266O5	47,410,005	47,580,965	13
8	CH261-115I12	56,156,247	56,350,468	13
8	CH261-59M8	56,214,780	56,415,061	13
8	TGMCBA-321B13	57,702,966	57,852,699	13
9	CH261-122F8	7,933,610	8,104,039	5
9	CH261-49B22	12,292,562	12,485,962	5
9	TGMCBA-145C6	15,594,186	15,742,679	5
9	TGMCBA-341F20	20,743,260	20,870,940	20
9	TGMCBA-225I12	23,052,441	23,209,788	20
9	TGMCBA-250E3	35,510,107	35,658,965	20
10	TGMCBA-346F6	880,293	1,052,657	8
10	CH261-96D24	5,534,165	5,727,455	8
10	TGMCBA-208D17	6,558,717	6,766,032	8
10	TGMCBA-252A4	15,226,860	15,354,942	8
10	CH261-34H16	19,447,695	19,642,579	8
11	CH261-115J5	2,052,095	2,234,731	3
11	CH261-18I9	4,859,089	5,081,522	3
11	CH261-160I6	18,681,358	18,909,440	3
11	CH261-130M12	23,382,802	23,548,255	3
12	CH261-183N19	2,407,945	2,582,547	9
12	TGMCBA-150E19	6,231,531	6,390,404	9
12	CH261-95N3	12,584,968	12,778,633	9
12	CH261-187M16	15,855,972	16,044,372	9
12	TGMCBA-321L6	18,422,783	18,593,906	9
12	TGMCBA-217A3	23,665,643	23,931,449	9
12	CH261-68O18	24,071,040	24,262,587	9

**Table S 2 continued: List of BACs and their assigned position in the peregrine falcon (FPE - *Falco peregrinus*).**

FPE Chrom	BAC Name	FPE Start Position	FPE End Position	GGA Homolog
13	TGMCBA-220A5	984,589	1,132,946	4
13	CH261-111A15	5,198,586	5,420,893	4
13	TGMCBA-330J11	13,848,705	13,980,861	4
13	CH261-83E1	14,428,939	14,601,026	4
13	CH261-71L6	14,487,035	14,674,831	4
13	TGMCBA-280M7	18,550,043	18,691,593	4
13	CH261-183B15	23,358,556	23,557,443	4
14	CH261-154H1	6,359,240	6,574,410	11
14	CH261-138H13	12,944,748	13,125,325	11
14	CH261-121N21	14,364,206	14,606,838	11
14	TGMCBA-192A10	21,608,634	21,726,279	11
15	CH261-154H17	688,846	886,426	24
15	TGMCBA-82A15	1,426,889	1,687,323	24
15	CH261-65O4	2,235,455	2,387,013	24
15	TGMCBA-111K1	5,836,178	5,976,942	24
15	CH261-103F4	6,542,341	6,694,292	24
16	TGMCBA-332G15	1,025,611	1,101,301	26
16	TGMCBA-297G21	1,252,470	1,361,457	26
16	CH261-186M13	1,316,933	1,493,593	26
16	TGMCBA-97D20	2,508,432	2,593,845	26
16	CH261-170L23	4,735,265	4,929,838	26
17	CH261-18G17	1,055,179	1,270,001	22
17	TGMCBA-113N13	1,100,769	1,260,639	22
17	CH261-40J9	1,509,487	1,687,797	22
17	TGMCBA-151I22	4,174,379	4,345,795	22
18	TGMCBA-324P4	1,010,187	1,107,305	27
18	TGMCBA-23C5	1,859,124	2,017,676	27
18	CH261-66M16	2,792,864	2,969,001	27
18	CH261-28L10	3,170,728	3,386,845	27
18	CH261-100E5	3,386,940	3,551,666	27
Z	CH261-129A16	16,110,876	16,336,750	Z
Z	TGMCBA-200J22	21,286,370	21,466,575	Z
Z	CH261-133M4	24,519,647	24,711,866	Z
Z	TGMCBA-270I9	35,885,879	36,042,627	Z

Chapter 6 Supplementary Figures:

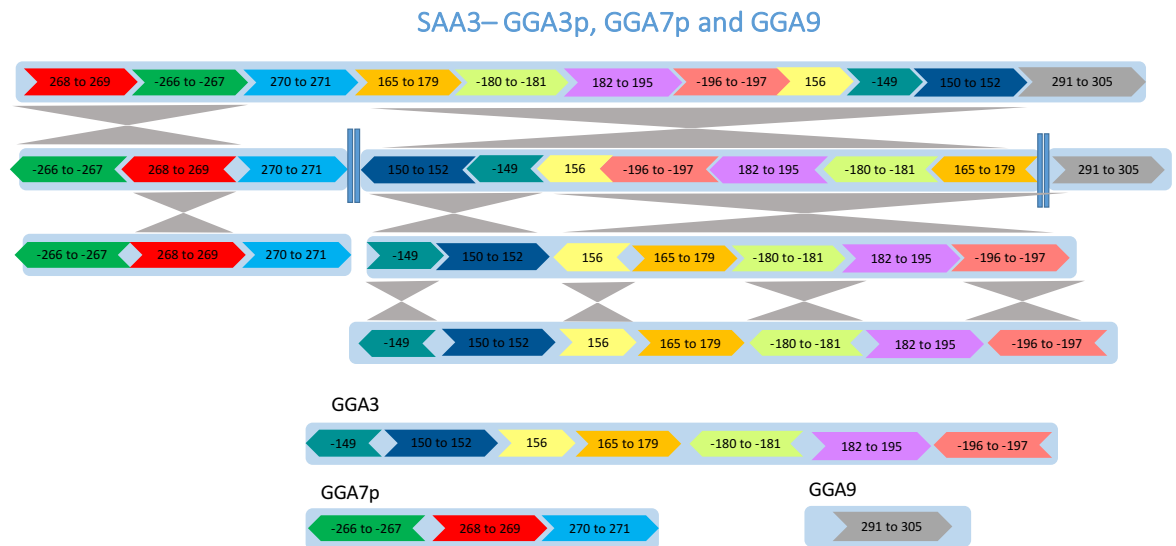


Figure S 1: Inter and intrachromosomal rearrangements identified that occurred from SAA3 to GGA3, GGA7 and GGA9 (the grey triangles represent inversions and the blue bars represent fissions).

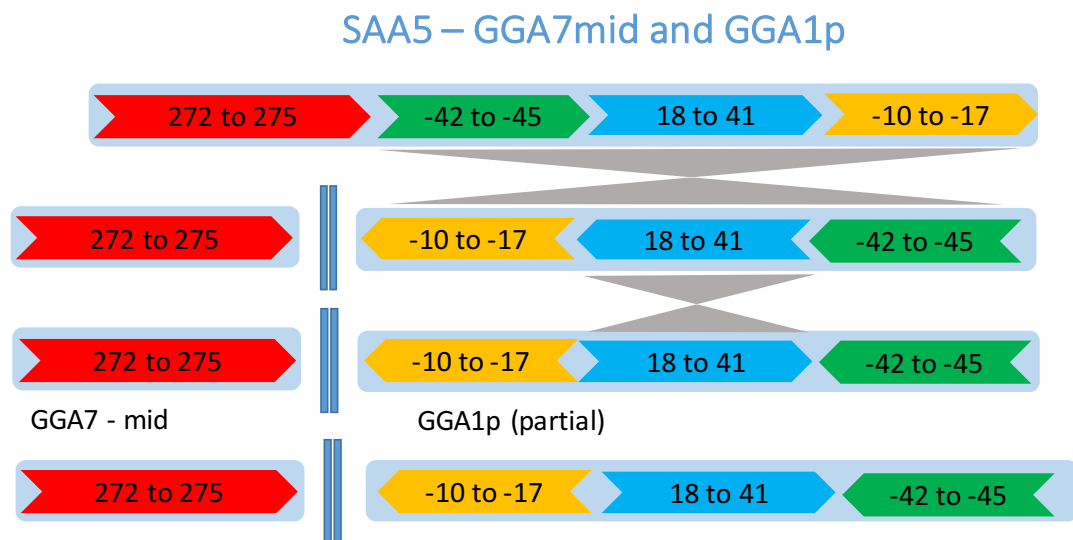


Figure S 2: Inter and intrachromosomal rearrangements identified that occurred from SAA5 to GGA7 and GGA1 (the grey triangles represent inversions and the blue bars represent fissions).

### SAA8 – GGA6 and GGA13

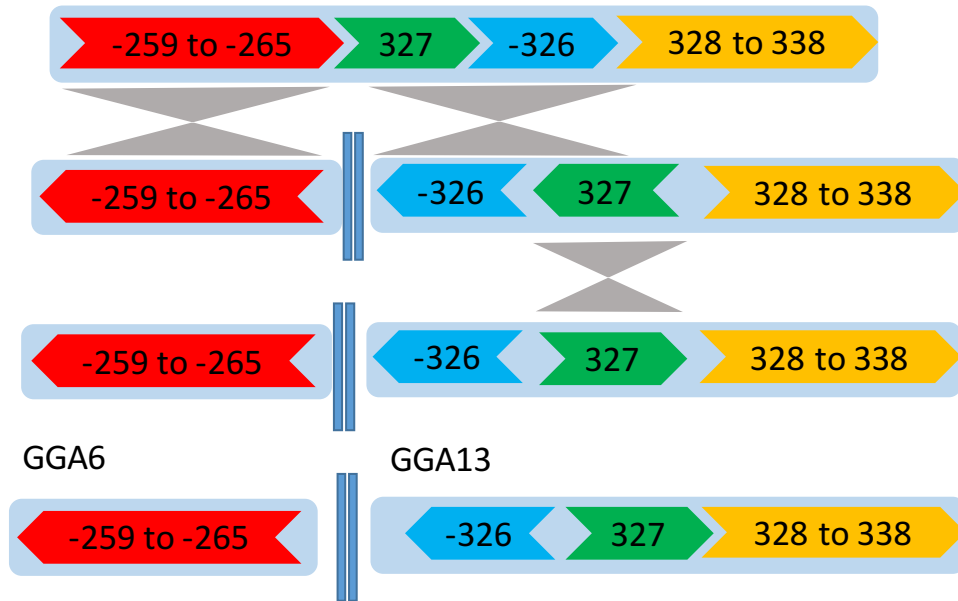


Figure S 3: Inter and intrachromosomal rearrangements identified that occurred from SAA8 to GGA6 and GGA13 (the grey triangles represent inversions and the blue bars represent fissions).

### SAA9 – GGA8 and GGA27

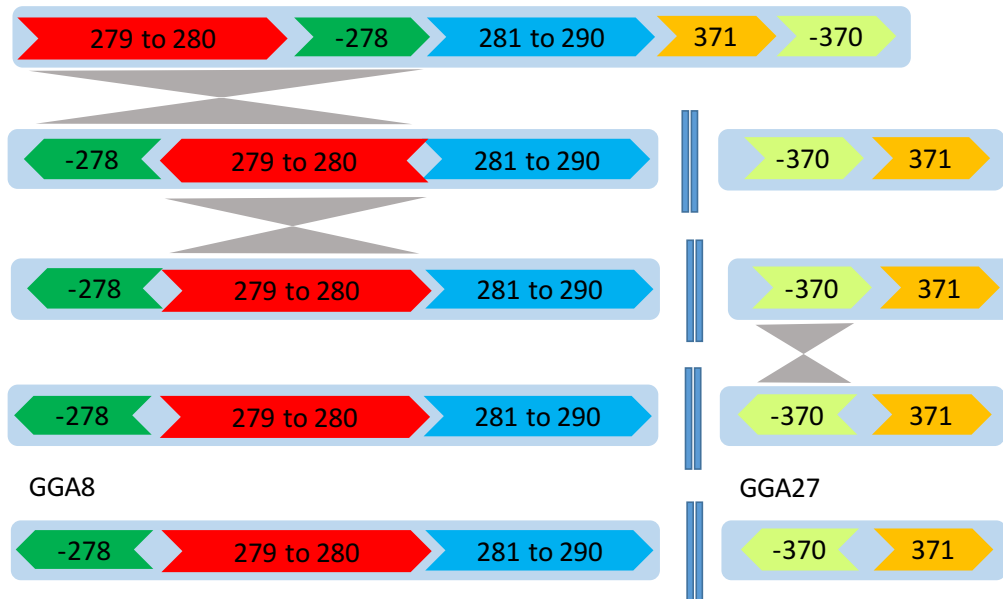


Figure S 4: Inter and intrachromosomal rearrangements identified that occurred from SAA9 to GGA8 and GGA27 (the grey triangles represent inversions and the blue bars represent fissions)..

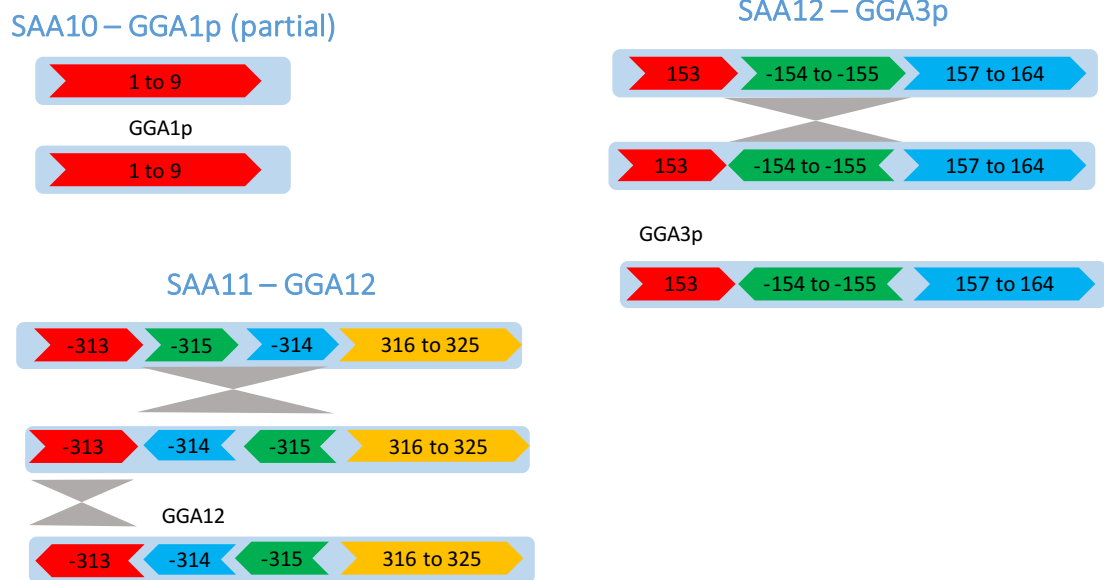


Figure S 5: Intrachromosomal rearrangements identified that occurred from SAA10 to GGA1; SAA11 to GGA12 and SAA12 to GGA3 (the grey triangles represent inversions)..

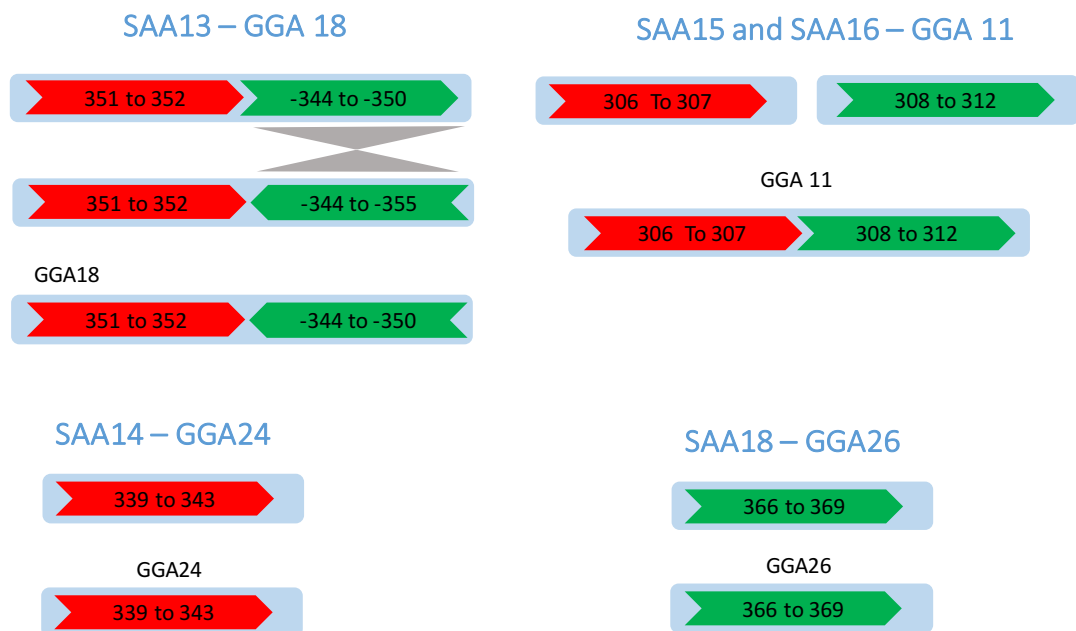
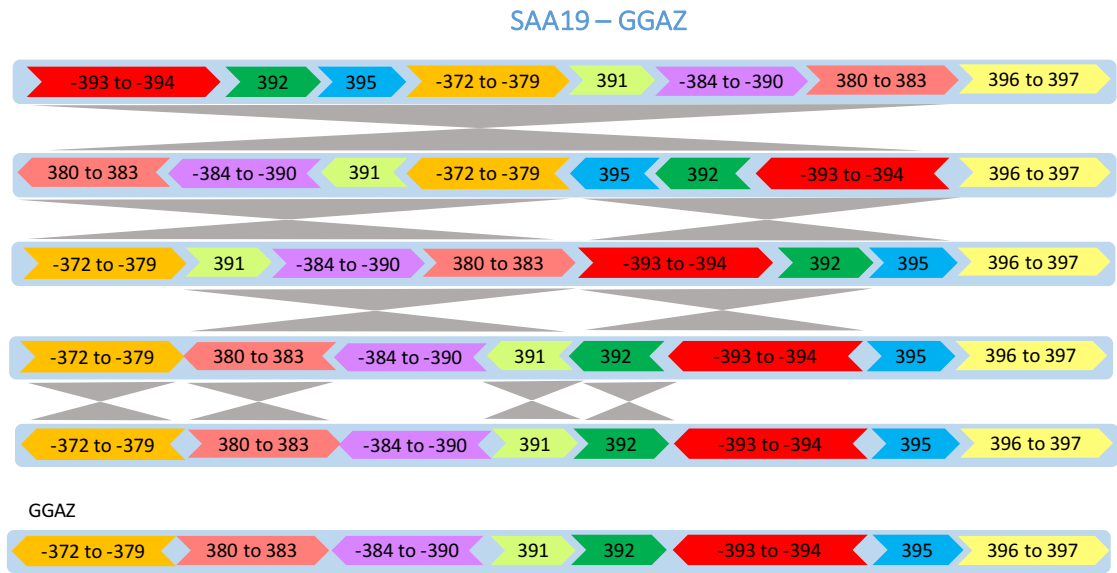


Figure S 6: Intrachromosomal rearrangements identified that occurred from SAA13 to GGA18; SAA15 and SAA16 that fused to form GGA11; SAA14 to GGA24 and SAA18 to GGA26 (the grey triangles represent inversions).



**Figure S 7: Intrachromosomal rearrangements identified that occurred from SAA19 to GGAZ (the grey triangles represent inversions).**



## 10 Publications arising from this work

### Research Manuscripts:

- Lithgow PE\*, **O'Connor R.E.\***, et al. *'Novel tools for characterising inter and intra chromosomal rearrangements in avian microchromosomes'*. Chromosome research. 2014;22(1):85-97. \*Joint first authors
- Romanov MN, Farré M, Lithgow PE, Fowler KE, Skinner BM, **O'Connor R.E.**, et al. *'Reconstruction of gross avian genome structure, organization and evolution suggests that the chicken lineage most closely resembles the dinosaur avian ancestor'*. BMC genomics. 2014;15(1):1060.
- Schmid M, Smith J, Burt DW, Aken BL, Antin PB, Archibald AL,...**O'Connor R.E.**... et al. *'Third Report on Chicken Genes and Chromosomes 2015'*. Cytogenetic and Genome Research. 2015;145(2):78-179.
- Damas J.\* and **O'Connor, R.E.\*** et al. (Genome Research - submitted). *'Upgrading short read animal genome assemblies to chromosome level using comparative genomics and a universal probe set'*. \*Joint first authors
- **O'Connor R.E.** et al. (Animal Genetics - submitted). *'Isolation of subtelomeric sequences of porcine chromosomes reveals errors in the pig genome assembly'*.

### Published Abstracts:

- **O'Connor R.E.** et al. 2014. Identification of Chromosomal Translocations in Pigs using FISH with Subtelomeric Probes and the development of a novel screening tool for their application "21st International Colloquium on Animal Cytogenetics and Gene Mapping." Chromosome Res 22 (2014): 393-437.
- Griffin, D.K., **O'Connor R.E.** et al. 2014. Avian cytogenetics goes functional. "21st International Colloquium on Animal Cytogenetics and Gene Mapping." Chromosome Res 22 (2014): 393-437.
- Sadraie, M., Fowler K.E., **O'Connor R.E.**, et al., 2015. Evaluation of aneuploidy of autosome chromosomes in boar sperm samples. 13th Annual Preimplantation Genetic Diagnosis International Society (PGDIS) Meeting Abstracts." Chromosome Res 22 (2014): 573-643.
- Griffin, D.K., **O'Connor R.E.** et al., 2015 Avian Chromonomics goes functional. Chromosome Research. pp. 413–414
- **O'Connor, R.E.** et al., 2015. Reconstruction of the putative Saurian karyotype and the hypothetical chromosome rearrangements that occurred along the Dinosaur lineage. Chromosome Research, 23(2), p.343.
- Sadraie, M., Fowler K.E., **O'Connor R.E.**, et al., 2015. Evaluation of aneuploidy of autosome chromosomes in boar sperm. In Chromosome Research. pp. 384–385.
- Lithgow, P.E., **O'Connor R.E.** et al., 2015. Inter and intra chromosomal rearrangements in avian microchromosomes. Chromosome Research. pp. 413–414.
- Romanov, M.N.,... **O'Connor R.E.** et al., 2015. Avian ancestral karyotype reconstruction and differential rates of inter-and intrachromosomal change in different lineages. Chromosome Research, 23(2), p.414.
- **O'Connor R.E.** et al., 2016. Upgrading Molecular Cytogenetics to Study Reproduction and Reproductive Isolation in Mammals, Birds, and Dinosaurs. Cytogenetic and Genome Research, 148(2-3), pp.83–155.

# Novel tools for characterising inter and intra chromosomal rearrangements in avian microchromosomes

Pamela E Lithgow · Rebecca O'Connor · Deborah Smith · Gothami Fonseka ·  
Abdullah Al Mutery · Claudia Rathje · Richard Frodsham · Patricia O'Brien ·  
Fumio Kasai · Malcolm A. Ferguson-Smith · Benjamin M. Skinner ·  
Darren K. Griffin

Published online: 3 April 2014  
© Springer Science+Business Media Dordrecht 2014

**Abstract** Avian genome organisation is characterised, in part, by a set of microchromosomes that are unusually small in size and unusually large in number. Although containing about a quarter of the genome, they contain around half the genes and three quarters of the total chromosome number. Nonetheless, they continue to

believe analysis by cytogenetic means. Chromosomal rearrangements play a key role in genome evolution, fertility and genetic disease and thus tools for analysis of the microchromosomes are essential to analyse such phenomena in birds. Here, we report the development of chicken microchromosomal paint pools, generation of pairs of specific microchromosome BAC clones in chicken, and computational tools for in silico comparison of the genomes of microchromosomes. We demonstrate the use of these molecular and computational tools across species, suggesting their use to generate a clear picture of microchromosomal rearrangements between avian species. With increasing numbers of avian genome sequences that are emerging, tools such as these will find great utility in assembling genomes de novo and for asking fundamental questions about genome evolution from a chromosomal perspective.

Responsible Editors: Darren K. Griffin and Beth A. Sullivan.

Pamela E Lithgow, Rebecca O'Connor, Deborah Smith and Gothami Fonseka are joint first authors.  
Benjamin M Skinner and Darren K Griffin are joint last authors

**Electronic supplementary material** The online version of this article (doi:10.1007/s10577-014-9412-1) contains supplementary material, which is available to authorized users.

P. E. Lithgow · R. O'Connor · D. Smith · G. Fonseka ·  
A. Al Mutery · C. Rathje · B. M. Skinner · D. K. Griffin (✉)  
School of Biosciences, University of Kent,  
Canterbury CT2 7AF, UK  
e-mail: d.k.griffin@kent.ac.uk

A. Al Mutery  
Department of Applied Science, University of Sharjah,  
Sharjah, United Arab Emirates

R. Frodsham  
Cytocell Ltd, Newmarket Road, Cambridge, UK

P. O'Brien · F. Kasai · M. A. Ferguson-Smith  
School of Clinical and Veterinary Medicine,  
University of Cambridge, Madingley Road,  
Cambridge CB3 0ES, UK

B. M. Skinner  
Department of Pathology, University of Cambridge,  
Tennis Court Road, Cambridge CB2 1QP, UK

**Keywords** Evolution · Avian · Chromosomes · Birds · Breakpoint

## Abbreviations

AC	<i>Anser cygnoides</i> , goose
APL	<i>Anas platyrhynchos</i> , duck
BAC	Bacterial artificial chromosome
CUN	<i>Chlamydotis undulate</i> , Houbara Bustard
DOP-PCR	Degenerate oligonucleotide-primed polymerase chain reaction
EBR	Evolutionary breakpoint region
FISH	Fluorescent in situ hybridization
FRU	<i>Falco rusticolus</i> , Gyr falcon
GGA	<i>Gallus gallus</i> , chicken

HSB	Homologous synteny blocks
MUN	<i>Melopsittacus undulates</i> , budgerigar
NGS	Next-generation sequencing
PI	Propidium iodide
TGU	<i>Taeniopygia guttata</i> , zebra finch

## Introduction

Avian genome organisation is relatively unique in nature and characterised by a small overall size (The International Chicken Genome Sequencing Consortium 2004; McQueen et al. 1998; Smith et al. 2000; Habermann et al. 2001) and a highly distinctive karyotype (Christidis 1990; Masabanda et al. 2004; Griffin et al. 2007). The most striking feature is a series of microchromosomes that are both unusually small in size and unusually large in number. This represents a particular challenge to the chromosome biologist when trying to classify each chromosome and any differences that might exist between individuals, strains or species. Being GC-rich and gene-dense, microchromosomes account for 23 % of the genome, 48 % of the genes (McQueen et al. 1998; Smith et al. 2000; Habermann et al. 2001; Burt 2002) and three quarters of the total chromosome number but belie analysis by classical cytogenetic means. Microchromosome number and structure also appear to be highly conserved (Griffin et al. 2008; Skinner et al. 2009; Völker et al. 2010) with large-scale interchromosomal rearrangements apparent in only around one third of species. Intrachromosomally, analysis by molecular cytogenetics is impeded by the small size of each chromosome.

Masabanda et al. (2004) reported the generation of individual microchromosomal paints generated by microdissection and amplification of single microchromosomes from chicken (*Gallus gallus*). This remains, to date, the only complete definition of any avian karyotype and, despite working well within chicken, the chromosome paints were of limited utility cross-species, working successfully only on turkey (Griffin et al. 2008 and our unpublished data). Moreover, attempts to re-amplify and clone these paints have met with little or no success, and the paints have degraded over time (unpublished data). With the benefit of hindsight, this has been particularly disappointing since there are still virtually no physical or bioinformatic genomic resources pertaining to the very smallest of the microchromosomes (29–38, termed the “D group”).

The generation of chicken chromosome paints and bacterial artificial chromosome (BACs; for chromosomes 1–28 plus Z and W) has allowed both detailed definition of the chicken karyotype (Griffin et al. 1999; Habermann et al. 2001; Masabanda et al. 2004) and extensive comparative genomics (e.g. Shetty et al. 1999; Raudsepp et al. 2002; Shibusawa et al. 2002; Itoh and Arnold 2005; Griffin et al. 2007; Nanda et al. 2007, 2011; Nishida et al. 2008, 2013; Kasai et al. 2012a). Chromosome paints for the macrochromosomes have worked successfully in numerous avian species (including emu, parrots and falcons) and even in some reptiles (including a turtle, the red ear slider and the Nile crocodile) (Kasai et al. 2012a). Conversely, the use of microchromosomal paints has been restricted to just a few species (Griffin et al. 1999; Shetty et al. 1999; Nishida et al. 2008, 2013; Hansmann et al. 2009; Nie et al. 2009; Nanda et al. 2011), due in part to interpretation often being complicated by the presence of more than one microchromosome recognized by each paint. Cross-species BAC mapping has met with even less success. For instance, chicken BACs used on turkey (*Meleagris gallopavo*) appear to work with approximately 70 % success (Griffin et al. 2008), which reduces to less than 40 % on Pekin duck (*Anas platyrhynchos*) chromosomes (Skinner et al. 2009) and little success at all beyond the Galloanserae (unpublished results). Studies on zebra finch (*Taeniopygia guttata*) have taken chicken BACs applied to chicken chromosomes, then isolated BACs containing homologous sequences in zebra finch to create comparative maps (Völker et al. 2010). This general approach however relies on the existence of a well-defined BAC library and whole genome assembly of the species of interest (as it is for chicken).

Analysis of intrachromosomal rearrangements in the microchromosomes is more limited still. Indeed, we are not aware of any detailed cross-species analyses of microchromosomes between birds other than that of Rao et al. who assembled radiation hybrid maps in duck and thereby created a duck genome assembly from which comparative genomics to chicken could be performed (Rao et al. 2012). We have previously reported the use of GenAlyser on chicken, turkey and zebra finch macrochromosomes, describing intrachromosomal rearrangements that were also confirmed by FISH (Völker et al. 2010; Skinner and Griffin 2012). To date, however, such analyses have not been extended to the microchromosomes.

Analysis of the microchromosomes and thus investigations on the inter- and intrachromosomal relationships between individuals, species and strains is therefore impeded by a lack of adequate tools. The purpose of this study was to make a significant advance in the development and use of such tools from both in silico and lab-based perspectives. We produced a panel of microchromosomal paints and BACs that work reliably across species. In addition, we present new data on the comparative genomics of microchromosomes in three species for which fully assembled genomes exist.

## Materials and methods

### Cell culture and chromosome preparation

Chromosome preparations were made from cultured fibroblasts derived from 5- to 7-day-old embryos from fertilized eggs (chicken, duck and budgerigar), or muscle biopsy fibroblasts (Gyr Falcon and Houbara) following collagenase treatment. Lymphocyte cultures were established for zebra finch following separation of white blood cells over a Histopaque-1077 (Sigma) gradient; cells were cultured for 72 h in RPMI with 10 % chicken serum, pen/strep, L-glutamine and concanavalin A (100 µg/ml). Chromosome preparation followed standard protocols (Griffin et al. 1999; Ahlroth 2000). Mitostatic treatment with colcemid at a final concentration of 0.1 µg/ml for 1 h at 37 °C, hypotonic treatment with 75 mM KCl for 15 min at 37 °C and fixation with 3:1 methanol:acetic acid.

### Karyotyping and analysis

Metaphases were stained with a combination of DAPI (1.5 µg/ml) and propidium iodide (PI, 0.6 µg/ml) in Vectashield antifade medium (Vector labs USA). Image capture involved an Olympus BX61 epifluorescence microscope with cooled CCD camera and SmartCapture (Digital Scientific) capture system; SmartType (Digital Scientific UK) was used for karyotyping purposes.

### Selection and preparation of chicken BAC clones

BAC clones (ranging in size between 150,000 and 210,000 kb) were selected near the ends of each microchromosome from the CHORI-261 Chicken BAC library. BAC clones were generated by mini prep (Qiagen) then directly labelled by nick translation with FITC or

Cy3.5. Alternatively BAC clones were generated by mini prep (Qiagen) then directly labelled by patent-protected in-house technologies (Cytocell Ltd) with FITC or Texas Red. Labelled BACs were transferred to an Octochrome device (eight chromosomes per slide) marketed by Cytocell limited ([www.cytocell.com](http://www.cytocell.com)) using air-drying.

### Generation of chromosome paints

Microchromosomes were flow-sorted into nine pools (eight pools used in this study) by fluorescence-activated chromosome sorting at the Cambridge Resource Centre for Comparative Genomics (Cambridge Veterinary School, University of Cambridge, UK) as previously described (Griffin et al. 1999; Masabanda et al. 2004; Kasai et al. 2012b). Microchromosome paints were amplified by DOP-PCR (Telenius et al. 1992) and labelled by a secondary round of DOP-PCR incorporating FITC or Texas red.

### FISH, dual FISH, microscopy, capture

Fluorescent in situ hybridization (FISH) slides with metaphase preparations were aged for 1 h at room temperature. BACs and/or chromosome paints directly labelled with FITC, TR or CY3.5 were applied to the metaphase preparations and sealed under a cover slip before simultaneous denaturation of probe and target DNA for 2 min on a 75 °C hotplate. Hybridization was carried out in a humidified chamber for 24–72 h at 37 °C. Following post-hybridization washes (2 min in 0.4× SSC at 72 °C; 30 s in 2× SSC/0.05 % Tween 20 at RT), slides were counterstained using Vectashield antifade medium with DAPI (Vector Labs, USA). Dual-colour FISH was achieved by simultaneous hybridization of FITC and TR or Cy3.5 labelled probes. Slides were analysed on an Olympus BX-61 epifluorescence microscope equipped with a cooled CCD camera and appropriate filters. Images were captured using SmartCapture 3 (Digital Scientific UK).

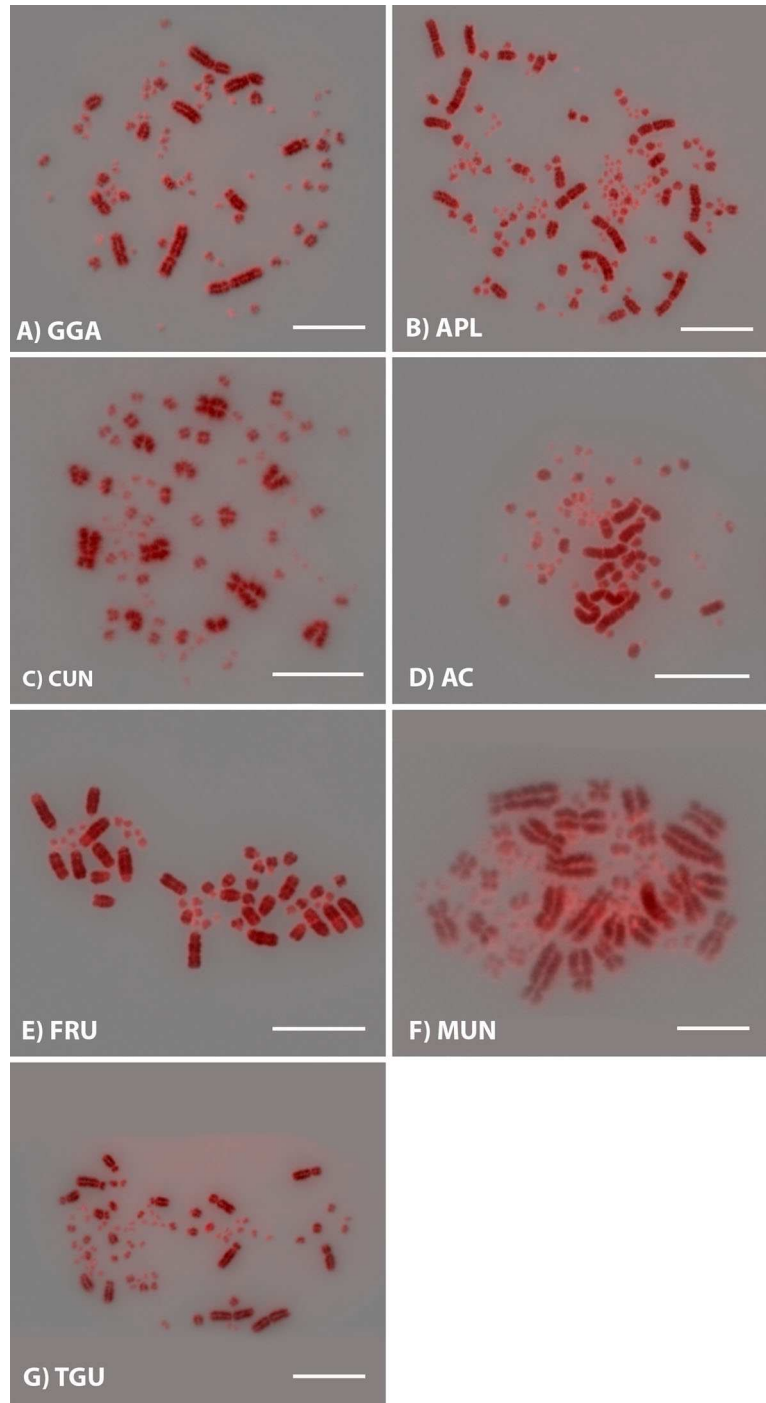
### GenAlyser analysis of chicken, turkey and zebra finch

To visualise large-scale intrachromosomal rearrangements, we aligned whole-chromosome sequences of chicken microchromosomes 11–28 and their turkey and zebra finch orthologs using the program GenAlyzer (Choudhuri et al. 2004) with default settings. The chicken–zebra finch alignments were already available from our previous study (Völker et al. 2010).

Subsequently, to aid visualization, the GenAlyzer output matches (of 100+ base pairs) were combined into contiguous blocks using a custom script. This script combined direct or inverted matches where there was a consecutive run of at least five matches. If a distance

of 40 kb occurred with no matches, a new block was called. Blocks of at least 250 kb were plotted, to remove spurious matches caused by repetitive content and to focus on the larger rearrangements. The chromosomes were manually segmented based on these charts, and the

**Fig. 1** DAPI (pseudo coloured black/white) and propidium iodide (red) stained metaphase chromosomes for **a** chicken (GGA), **b** duck (APL), **c** houbara (CUN), **d** goose (AC), **e** gyrfalcon (FRU), **f** budgerigar (MUN) and **g** zebra finch (TGU). Scale bar 10  $\mu$ m



segments numbered and ordered relative to turkey. The Multiple Genomes Rearrangement tool on the GRIMM web server (<http://grimm.ucsd.edu/MGR/>) (Bourque and Pevzner 2002) was used to calculate optimal rearrangement pathways between each species, and to reconstruct a likely potential chicken–turkey ancestor, in the manner of Mlynarski et al. (2010). The series of possible rearrangements from the chicken–turkey ancestor to each species was considered, and for each rearrangement, the segment ends flanking the breakpoints were noted. Within each lineage, the number of times a segment end was involved in a rearrangement was counted.

## Results

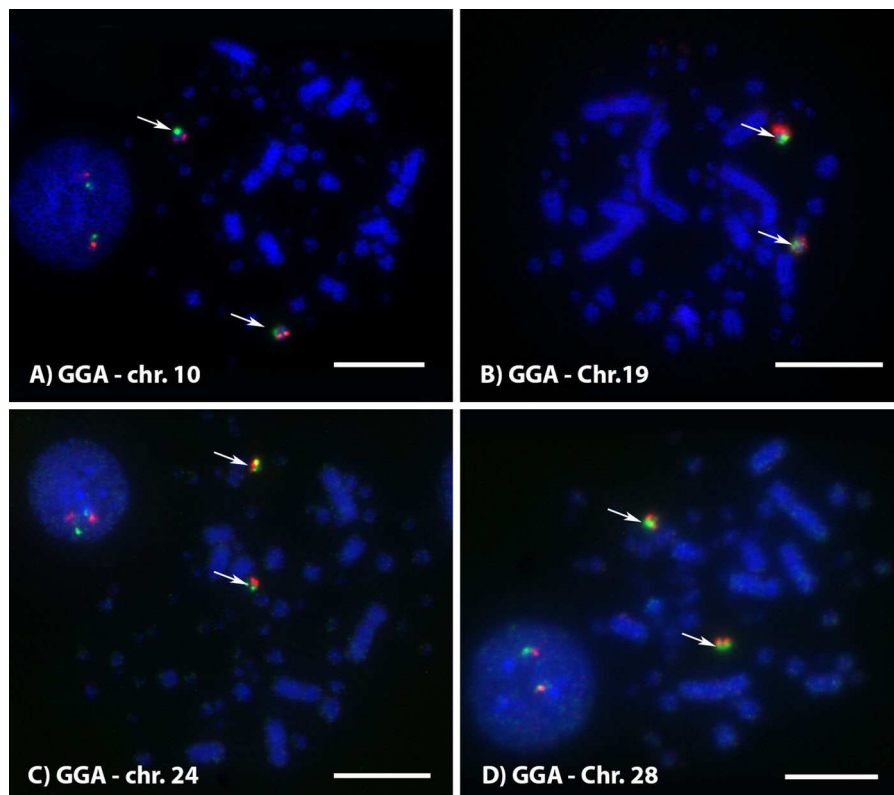
### Karyotyping

Chromosome preparation and DAPI and PI staining of metaphase chromosomes was successful in a number of

avian species (chicken—GGA, duck—APL, houbara—CUN, goose—AC, gyrfalcon—FRU, budgerigar—MUN and zebra finch—TGU) (Fig. 1). We took the decision to visualise PI as “red on black” and DAPI as “black on white” then merged the images. These seemed to give the most distinct pictures (presumably due to the preferential recognition of the AT-rich and GC-rich regions in the two dyes) and limited banding information is even apparent in some preparations (Fig. 1a, b, e).

### Microchromosome FISH

To allow detailed evaluation of chicken microchromosomes, BACs located at the subtelomeric regions of the p and q arms of chromosomes 10–28 were selected (supplementary Table 1). Co-hybridisation of p and q arm BACs to the same microchromosomes on chicken metaphase preparations allowed confirmation of the location of the BAC clones (Fig. 2). Bright,



**Fig. 2** Example of dual FISH results for BACs to microchromosome arms p (FITC) and q (Cy3.5) to confirm correct mapping **a** microchromosome 10 p arm (CH261-94C12) and q arm (CH261-118E15), **c** microchromosome 19 p arm (CH261-167A1) and q arm (CH261-189E4), **c** microchromosome 24 p arm

(CH261-90H16) and q arm (CH261-154H17) and **d** microchromosome 28 p arm (CH261-16I3) and q arm (CH261-179 N23) on chicken (GGA) chromosomes. *Arrows* highlight co-localisation of the BACs. *Scale bar* 10 μm



specific BAC signals on all chromosomes 11–28 were apparent and the Octochrome device (Cytocell) allowed fast, efficient hybridization on metaphases of chicken and other species.

Nine pools of microchromosomal paints, mostly representing more than one chromosome, were successfully generated using material produced from flow-sorted chicken microchromosomes Kasia et al (2012b) (eight were used in this study). Following fluorescent labelling, the microchromosome pools were applied across a number of species. When hybridised onto chicken metaphase chromosomes, the microchromosome pools (R1–R8) contained between one and five primary signals with between two and eight additional dim signals (Table 1, Fig. 3). Thereafter, assignment of the microchromosome paints to specific chromosomes was achieved by dual FISH with BACs to the q arm of each microchromosome (Fig. 3). Table 1 summarises the assignment of the microchromosome pools (R1–R8) to specific

microchromosomes. Dual FISH confirmed that R1 and R2 or R5 hybridise to the same microchromosomes but with differing intensities, R2 and R5 hybridised to the same microchromosomes with very similar intensity by visual inspection (supplementary Fig. 1).

Cross-species FISH was performed using the microchromosome paints on zebra finch, gyrfalcon, budgerigar and houbara metaphase chromosomes (Fig. 4, supplementary Fig. 2, Table 2). FISH also revealed chromosomes from gyrfalcon and budgerigar where the microchromosome paint (R1, R4 or R6) labelled the end of macrochromosomes indicating evolutionary fusion events (Fig. 5). Additionally, paint R2 decorated a single entire chromosomes in budgerigar suggesting a fusion of the homologues of chicken chromosomes 10 and 12.

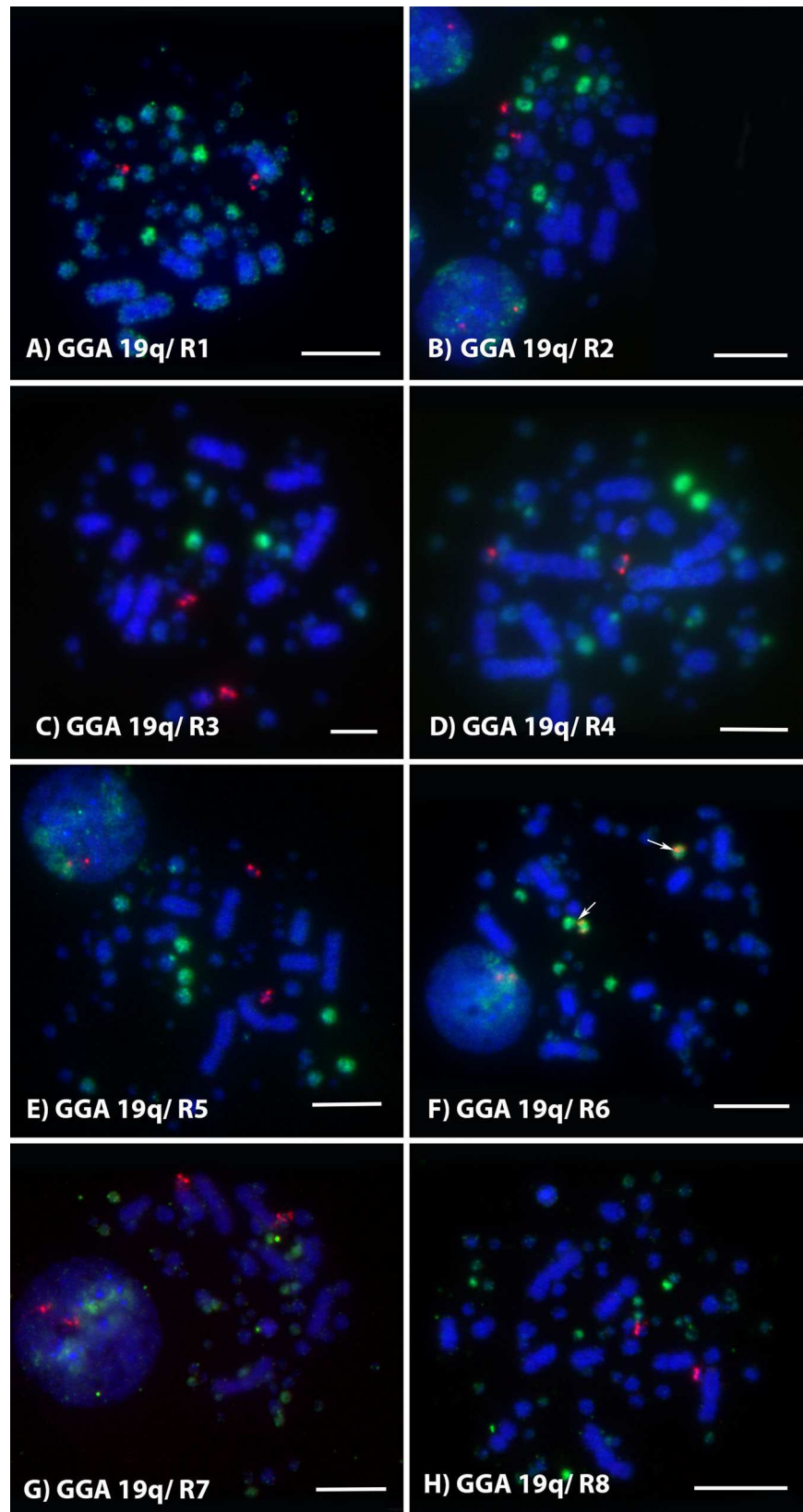
To extend the evaluation of microchromosomes across species, FISH was also performed using the microchromosome BACs on zebra finch, gyrfalcon,

**Table 1** Assignment of microchromosomes to microchromosome pools (R1–R8)

BAC/Paint	R1	R2	R3	R4	R5	R6	R7	R8
10	+	+++	–	–	+++	–	–	–
11	+++	+	–	–	++	–	–	–
12	+	+++	–	–	+++	–	–	–
13	–	–	+++	++	–	–	–	–
14	–	–	–	+++	–	–	–	–
15	–	–	–	+	–	–	–	–
16	–	–	–	–	–	–	–	–
17	–	–	–	–	–	+++	–	–
18	–	–	–	–	–	+++	–	–
19	–	–	–	–	–	+++	–	–
20	–	–	–	+	–	–	–	–
21	–	–	–	–	–	–	–	–
22	–	–	–	–	–	–	–	–
23	–	–	–	–	–	–	–	–
24	–	–	–	–	–	–	–	–
25	–	–	–	–	–	–	+	–
26	–	–	–	–	–	–	–	–
27	–	–	–	–	–	–	+	–
28	–	–	–	–	–	–	–	+
Number of primary signals	1	2	1	2	2	3	5	3
Number of secondary signals	2–3	2–3	4	3–4	3–4	2–8	3	2–3

Co-localisation was examined by FISH of BACS to the q arm (list of clones' supplementary table) of each microchromosome with microchromosome paint pools. The number of “+” indicates the strength of signal observed for the microchromosome paint to which the BAC probe is localised

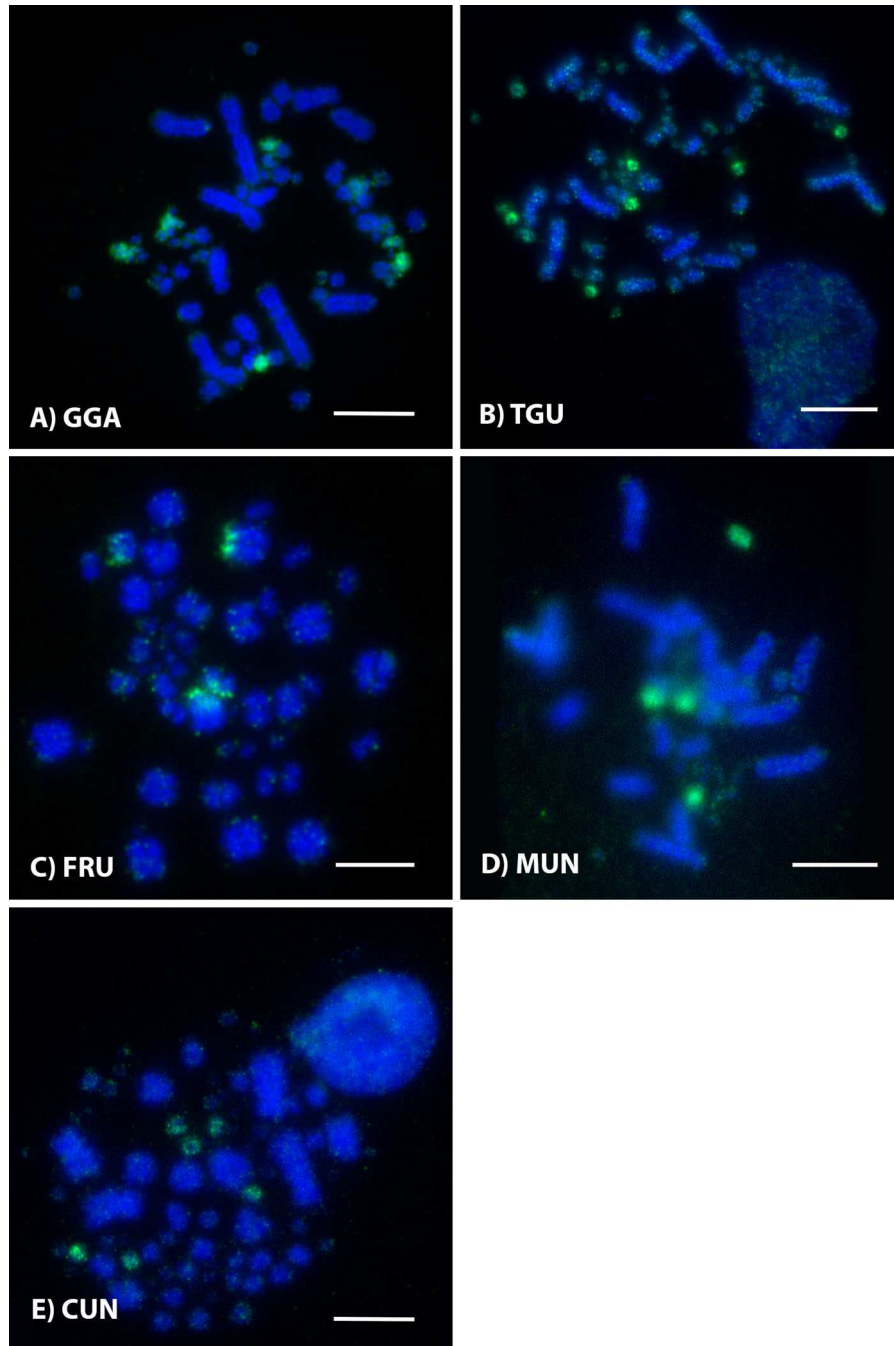
**Fig. 3** Example of dual FISH results. BAC for the q arm of microchromosome 19 (CH261-189E4) and microchromosome paint pool **a** R1, **b** R2, **c** R3, **d** R4, **e** R5, **f** R6, **g** R8, and **h** R9 on chicken (GGA) chromosomes. *Arrows* highlight co-localisation of the BAC for microchromosome 19 with microchromosome paint pool R6. *Scale bar* 10  $\mu$ m





budgerigar and houbara metaphase chromosomes. Hybridisation was observed for some of the microchromosomes from each species (supplementary Table 2). For instance, hybridisation was observed

for BACs against microchromosome 23 (p and q arms) when hybridised to zebra finch, gyrfalcon, budgerigar and houbara metaphase chromosomes (Fig. 6).



**Fig. 4** Example of ZooFISH results. Microchromosome paint pool R6 on **a** chicken (GGA), **b** zebra finch (TGU), **c** gyrfalcon (FRU), **d** budgerigar (MUN) and **e** houbara (CUN) metaphase chromosomes. Scale bar 10  $\mu$ m

**Table 2** Number of pairs of signals observed for microchromosome pools (R1–R6) in a number of avian species (chicken, zebra finch, gyrfalcon, budgerigar, houbara)

	Paint ID					
	R1	R2	R3	R4	R5	R6
Chicken	1	2	1	2	2	3
Zebra finch	1	2	1	2	2	3
Gyrfalcon	1	2	1	2 (1 fused to macro)	2	2 (1 fused to macro)
Budgerigar	1 (fused to macro)	2	1	2 (1 fused to macro)	2	2
Houbara	1	2	1	2	2	3

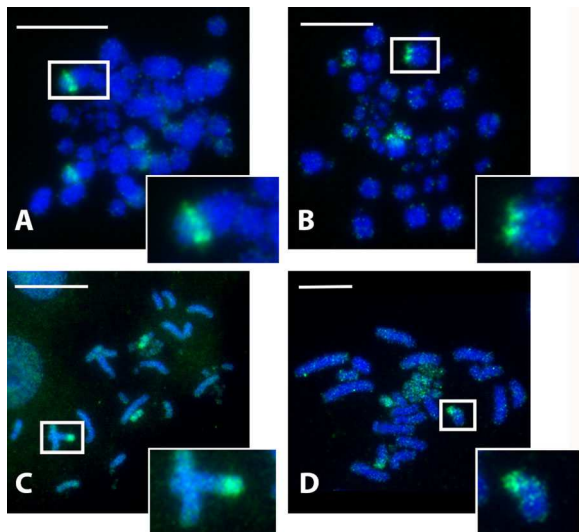
ZooFISH was performed with microchromosome paint pools (R1–R6). The number of pairs of strong signals observed is recorded (for chicken this is the “primary” signal)

Comparative analysis of chicken, turkey, zebra finch microchromosomes using GenAlyzer and our own in-house algorithm

The chicken–zebra finch–turkey alignments for the macrochromosomes (1–10+Z) were already available from our previous studies (Völker et al. 2010; Skinner and Griffin 2012), however further analysis allowed construction of comparative maps in the microchromosomes. A bioinformatic analysis of all available microchromosomes (GGA11–28) for chicken, turkey and zebra finch successfully produced comprehensive comparative maps using a previously developed

in-house algorithm (Skinner and Griffin 2012). An example is given in Fig. 7 for orthologues of chicken chromosome 18. In all analyses, no evidence of intermicrochromosomal rearrangements was seen in any of the three species, however numerous intramicrochromosomal rearrangements were apparent. In total, 38, 56 and 58 intrachromosomal rearrangements (all inversions) were detected in the microchromosomes (GGA11–28) of chicken, turkey and zebra finch, respectively. From direct-sequence analysis of chicken, zebra finch and turkey genomes, a total of 141 homologous synteny blocks (HSBs) were identified ranging in size from 3.65 to 13,000 kb, with mean and median sizes of 1,087 and 326 kb, respectively.

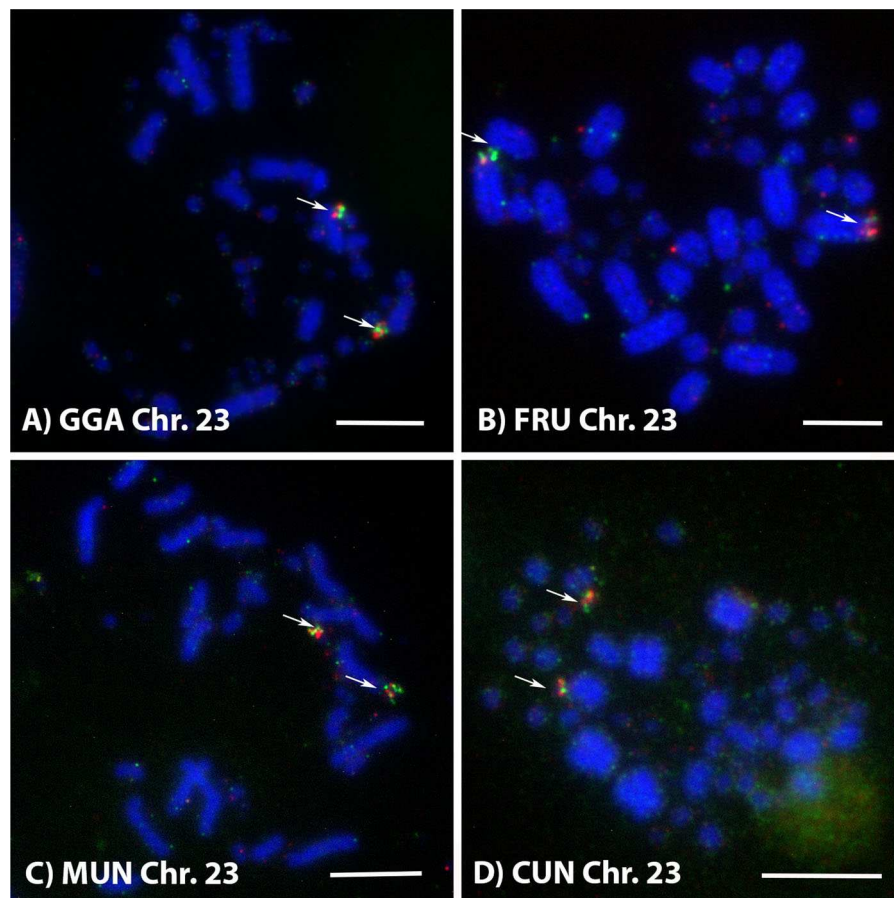
In chicken chromosomes 11–28, and their turkey and zebra finch orthologs, 282 segment ends were identified, of which 243 were involved in rearrangements. The most parsimonious predicted pathways from the chicken–turkey ancestor suggested that 87 breakpoint regions (35.8 %) recurred in different lineages, whereas 133 breakpoint regions (54.7 %) recurred in either the same or different lineages.



**Fig. 5** Example of ZooFISH results highlighting fusion of microchromosomes to macrochromosomes. **a** Microchromosome paint pool R4 on gyrfalcon, **b** microchromosome paint pool R6 on gyrfalcon, **c** microchromosome paint pool R1 on budgerigar, **d** microchromosome paint pool R4 on budgerigar. Scale bar 10  $\mu$ m

## Discussion

In this study, we demonstrate that the combination of in silico and lab-based experimental approaches provides a powerful approach for the classification of avian microchromosomes and for performing comparative genomics. In birds, comparative genomics at the chromosomal scale has been limited mostly to the macrochromosomes (Habermann et al. 2001; Masabanda et al. 2004; Griffin et al. 2007, 2008; Fillon et al. 2007; Skinner et al. 2009; Völker et al.



**Fig. 6** Example of cross-species dual FISH results for chicken BACs to microchromosome 23 arms p (FITC, CH261-191G17) and q (Cy3.5, CH261-9 K11) on **a** chicken (GGA), **b** falcon

(FRU), **c** budgerigar (MUN) and **d** Houbara (CUN). Arrows highlight co-localisation of the BACs. Scale bar 10  $\mu$ m

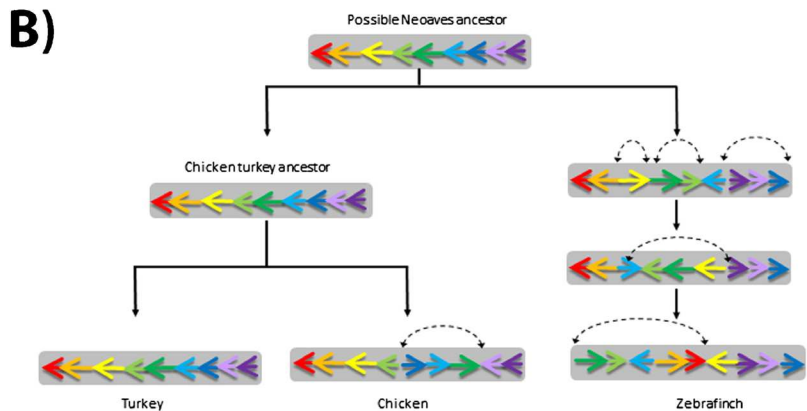
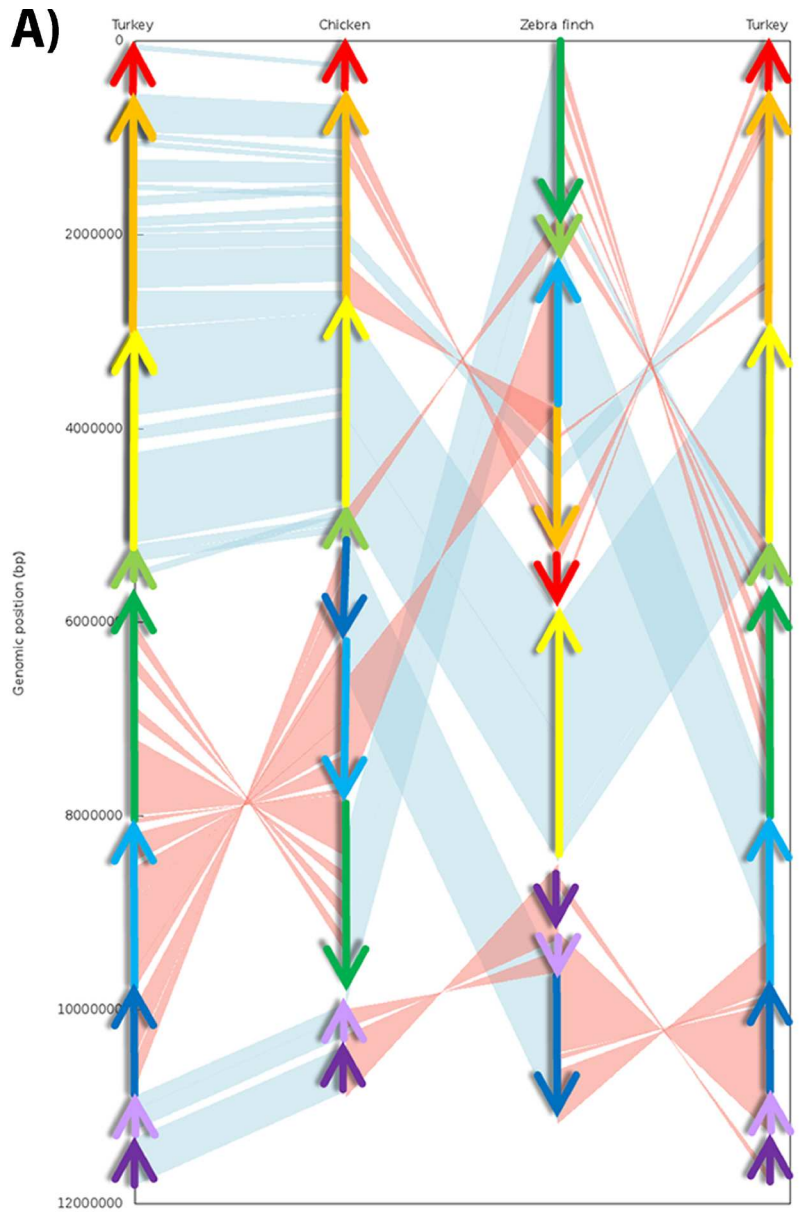
2010); the tools and resources developed here will provide a means of extending these studies to the microchromosomes. We demonstrated that interchromosomal studies using significantly improved FISH tools work efficiently cross-species. Furthermore, intrachromosomal studies, with rearrangements in chicken, turkey and zebra finch were demonstrated in an even more detailed level.

Current studies of avian cytogenetics and genomics suggest that avian genomes remain remarkably conserved in terms of chromosome number and that interchromosomal changes are relatively rare. When they occur, they tend to recur with several examples of homoplasmy (Griffin et al 2007; Skinner and Griffin 2012). Cross-species FISH is particularly useful in species where chromosome rearrangement is commonplace, e.g. for characterising interchromosomal rearrangements such as those seen in the Falconinae and Psittaciformes (Nanda et al. 2007;

Nishida et al. 2008). Previous work has highlighted macrochromosomal rearrangements and regions of macrochromosomes without assignment, indicating the fusion of microchromosomes (Nanda et al. 2007; Nishida et al. 2008). This study reliably confirms for gyrfalcon and budgerigar that microchromosomes are indeed fused to macrochromosomes. Following the assignment of the microchromosomes to the pools (R1–R8) using the BAC clones, it is possible to infer that in the budgerigar the orthologues of chicken microchromosomes 11 and 14 are fused to macrochromosomes and in the gyrfalcon the chicken orthologue of microchromosome 14 and one of the microchromosomes 16, 17 or 18 are fused to macrochromosomes. Further investigation is required to confirm the identity of the macrochromosomes that are involved.

Despite the interchromosomal stability of the microchromosomes, we can detect evidence for many

**Fig. 7** Comparative analysis of chicken chromosome 18, turkey chromosome 20, and zebra finch chromosome 18 using output from our in-house algorithm based on Synteny Tracker analysis. **a** Coloured arrows show the division of the chromosomes into segments oriented with respect to turkey. **b** A predicted series of inversions from Neoaves common ancestor leading to chicken chromosome 18, turkey chromosome 20, and zebra finch chromosome 18. Inverted segments are indicated with dotted arrows





intrachromosomal rearrangements. This agrees with our previous work on the macrochromosomes of the same species (Völker et al. 2010; Skinner and Griffin 2012). Here though, we suggest that about half of the breakpoints recurred, with maybe one third of the breakpoints recurring in different lineages—more extensive than observed amongst the macrochromosomes (36 and 10 %, respectively). Potentially, this can be explained by the higher gene density and lower repetitive content on the microchromosomes versus the macrochromosomes: fewer substrates for non-allelic recombination, and fewer places within the chromosome where breakpoints will not disrupt genes or regulatory elements.

We thus provide tools and proof of principle studied for the generation of accurate cytogenetic maps in avian microchromosomes, an essential prerequisite for further analysis (Lewin et al. 2009). Molecular cytogenetic tools are as relevant as they have ever been given that next-generation sequencing (NGS) technologies, can often lose mapping information when attempting to generate de novo assemblies. For the foreseeable future, sequence assemblies will require further information from physical mapping data, and the probe sets such as those generated here will help link sequence scaffolds to whole chromosomes. Indeed, because there are numerous avian genomes recently sequenced by NGS, we can apply the tools developed here to study them more closely at a chromosomal level. This will help construct complete and reference-able genome assemblies for many of the ongoing avian vertebrate genome projects. Without the link back to physical mapping data, newly sequenced genomes will remain simply catalogues of genes (at best, collections of scaffolds) with limited reference to the overall genomic structure and organisation.

Chromosomal mapping information is essential to help define the functional role of whole chromosomes, homologous synteny blocks (HSBs) and evolutionary breakpoint regions (EBRs) in evolution, speciation and phenotype. We know that chromosomal changes lead to, or perpetuate, speciation (Rieseberg 2001; Noor et al. 2001; Delneri et al. 2003), however the use of cytogenetic genome maps generated by a combination of in silico and lab-based tools will allow us to address the underlying molecular reasons for specific chromosomal rearrangements. In other words, it remains unclear if certain chromosomal changes associated with speciation arise because there is an adaptive value to a specific

chromosomal configuration or karyotype. We and others have reported that chromosome breakpoints can be re-used (Sankoff 1999; Stankiewicz and Lupski 2002; Skinner and Griffin 2012). Indeed Larkin et al. (2009) suggested different evolutionary pathways for HSBs, and EBRs suggested that natural selection can preserve blocks of genes (HSBs), thus maintaining evolutionary advantageous processes through genome organisation (Larkin et al. 2009). Moreover EBRs may be used and re-used to help generate phenotypic variation and novel combinations of genes that help promote adaptation.

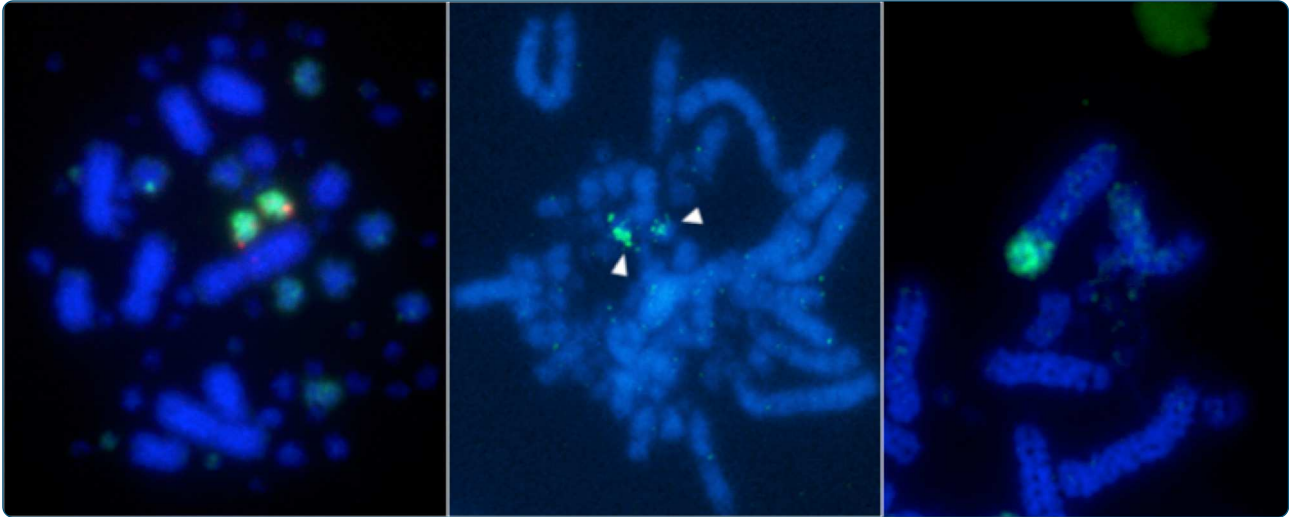
There are fundamental biological questions relating to the significance of EBRs and HSBs, the potential specialisation of gene biological function in different genomic compartments, and the roles of repetitive and transposable elements in genome organisation and evolution. As we develop additional tools to study complete chromosomes (especially the microchromosomes) at the sequence level, such questions will become answerable in birds.

**Acknowledgments** We thank Cherry Valley Farms, the Central Veterinary Research Laboratory in Dubai, and the Falcon Hospital Dubai for providing feathers and blood samples from which the DNA samples were extracted. This study was supported by grant BB/K008161 from the BBSRC.

## References

- Ahlroth MK, Kola EH, Kulomaa MS et al (2000) Characterization and chromosomal localization of the chicken avidin gene family. *Anim Genet* 31:367–375
- Bourque G, Pevzner PA (2002) Genome-scale evolution: reconstructing gene orders in the ancestral species. *Genome Res* 12:26–36
- Burt DW (2002) Origin and evolution of avian microchromosomes. *Cytogenet Genome Res* 96:97–112
- Christidis L (1990) *Animal cytogenetics: vol 4. Chordata. Aves, vol 3, Part 2.* 116
- Choudhuri JV, Schleiermacher C, Kurtz S, Giegerich R (2004) GenAlyzer: interactive visualization of sequence similarities between entire genomes. *Bioinformatics* 20:1964–5
- Delneri D, Colson I, Grammenoudi S et al (2003) Engineering evolution to study speciation in yeasts. *Nature* 422:68–72
- Fillon V, Vignoles M, Crooijmans RPMA et al (2007) FISH mapping of 57 BAC clones reveals strong conservation of synteny between Galliformes and Anseriformes. *Anim Genet* 38:303–307
- Griffin DK, Haberman F, Masabanda J et al (1999) Micro- and macrochromosome paints generated by flow cytometry and microdissection: tools for mapping the chicken genome. *Cytogenet Cell Genet* 87:278–281
- Griffin DK, Robertson LBW, Tempest HG, Skinner BM (2007) The evolution of the avian genome as revealed by comparative molecular cytogenetics. *Cytogenet Genome Res* 117:64–77
- Griffin DK, Robertson LB, Tempest HG et al (2008) Whole genome comparative studies between chicken and turkey

- and their implications for avian genome evolution. *BMC Genomics* 9:168
- Habermann FA, Cremer M, Walter J et al (2001) Arrangements of macro- and microchromosomes in chicken cells. *Chromosome Res* 9:569–584
- Hansmann T, Nanda I, Volobouev V et al (2009) Cross-species chromosome painting corroborates microchromosome fusion during karyotype evolution of birds. *Cytogenet Genome Res* 126:281–304
- Itoh Y, Arnold AP (2005) Chromosomal polymorphism and comparative painting analysis in the zebra finch. *Chromosome Res* 13:47–56
- Kasai F, O'Brien PCM, Martin S, Ferguson-Smith MA (2012a) Extensive homology of chicken macrochromosomes in the karyotypes of *Trachemys scripta elegans* and *Crocodylus niloticus* revealed by chromosome painting despite long divergence times. *Cytogenet Genome Res* 136:303–307
- Kasai F, O'Brien PCM, Ferguson-Smith MA (2012b) Reassessment of genome size in turtle and crocodile based on chromosome measurement by flow karyotyping: close similarity to chicken. *Biol Lett* 8:631–635
- Larkin DM, Pape G, Donthu R et al (2009) Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. *Genome Res* 19:770–777
- Lewin HA, Larkin DM, Pontius J, O'Brien SJ (2009) Every genome sequence needs a good map. *Genome Res* 19:1925–1928
- Masabanda JS, Burt DW, O'Brien PCM et al (2004) Molecular cytogenetic definition of the chicken genome: the first complete avian karyotype. *Genetics* 166:1367–1373
- McQueen HA, Siriaco G, Bird AP (1998) Chicken microchromosomes are hyperacetylated, early replicating, and gene rich. *Genome Res* 8:621–630
- Mlynarski EE, Obergfell CJ, O'Neill MJ, O'Neill RJ (2010) Divergent patterns of breakpoint reuse in Muroid rodents. *Mamm Genome* 21:77–87
- Nanda I, Karl E, Griffin DK et al (2007) Chromosome repatterning in three representative parrots (Psittaciformes) inferred from comparative chromosome painting. *Cytogenet Genome Res* 117:43–53
- Nanda I, Benisch P, Fetting D et al (2011) Synteny conservation of chicken macrochromosomes 1–10 in different avian lineages revealed by cross-species chromosome painting. *Cytogenet Genome Res* 132:165–181
- Nie W, O'Brien PCM, Ng BL et al (2009) Avian comparative genomics: reciprocal chromosome painting between domestic chicken (*Gallus gallus*) and the stone curlew (*Burhinus oedicnemus*, Charadriiformes)—an atypical species with low diploid number. *Chromosome Res* 17:99–113
- Nishida C, Ishijima J, Kosaka A et al (2008) Characterization of chromosome structures of Falconinae (Falconidae, Falconiformes, Aves) by chromosome painting and delineation of chromosome rearrangements during their differentiation. *Chromosome Res* 16:171–181
- Nishida C, Ishijima J, Ishishita S et al (2013) Karyotype reorganization with conserved genomic compartmentalization in dot-shaped microchromosomes in the Japanese mountain Hawk-Eagle (*Nisaetus nipalensis orientalis*, Accipitridae). *Cytogenet Genome Res* 141:284–294
- Noor MA, Grams KL, Bertucci LA, Reiland J (2001) Chromosomal inversions and the reproductive isolation of species. *Proc Natl Acad Sci U S A* 98:12084–12088
- Rao M, Morisson M, Faraut T et al (2012) A duck RH panel and its potential for assisting NGS genome assembly. *BMC Genomics* 13:513
- Raudsepp T, Houck ML, O'Brien PC et al (2002) Cytogenetic analysis of California condor (*Gymnogyps californianus*) chromosomes: comparison with chicken (*Gallus gallus*) macrochromosomes. *Cytogenet Genome Res* 98:54–60
- Rieseberg LH (2001) Chromosomal rearrangements and speciation. *Trends Ecol Evol* 16:351–358
- Sankoff D (1999) Genome rearrangement with gene families. *Bioinformatics* 15:909–917
- Shetty S, Griffin DK, Graves JAM (1999) Comparative painting reveals strong chromosome homology over 80 million years of bird evolution. *Chromosome Res* 7:289–295
- Shibusawa M, Nishida-Umehara C, Masabanda J et al (2002) Chromosome rearrangements between chicken and guinea fowl defined by comparative chromosome painting and FISH mapping of DNA clones. *Cytogenet Genome Res* 98:225–230
- Skinner B, Griffin D (2012) Intrachromosomal rearrangements in avian genome evolution: evidence for regions prone to breakpoints. *Heredity (Edinb)* 108:37–41
- Skinner BM, Robertson LBW, Tempest HG et al (2009) Comparative genomics in chicken and Pekin duck using FISH mapping and microarray analysis. *BMC Genomics* 10:357
- Smith J, Bruley CK, Paton IR et al (2000) Differences in gene density on chicken macrochromosomes and microchromosomes. *Anim Genet* 31:96–103
- Stankiewicz P, Lupski JR (2002) Genome architecture, rearrangements and genomic disorders. *Trends Genet* 18:74–82
- Telenius H, Carter NP, Bebb CE et al (1992) Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* 13:718–725
- The International Chicken Genome Sequencing Consortium (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432:695–716
- Völker M, Backström N, Skinner BM et al (2010) Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution. *Genome Res* 20:503–511



# Reconstruction of gross avian genome structure, organization and evolution suggests that the chicken lineage most closely resembles the dinosaur avian ancestor

Romanov *et al.*

RESEARCH ARTICLE

Open Access

# Reconstruction of gross avian genome structure, organization and evolution suggests that the chicken lineage most closely resembles the dinosaur avian ancestor

Michael N Romanov<sup>1†</sup>, Marta Farré<sup>2†</sup>, Pamela E Lithgow<sup>1</sup>, Katie E Fowler<sup>1,10</sup>, Benjamin M Skinner<sup>3</sup>, Rebecca O'Connor<sup>1</sup>, Gothami Fonseka<sup>1</sup>, Niclas Backström<sup>4</sup>, Yoichi Matsuda<sup>5</sup>, Chizuko Nishida<sup>6</sup>, Peter Houde<sup>7</sup>, Erich D Jarvis<sup>8</sup>, Hans Ellegren<sup>4</sup>, David W Burt<sup>9</sup>, Denis M Larkin<sup>2\*†</sup> and Darren K Griffin<sup>1\*†</sup>

## Abstract

**Background:** The availability of multiple avian genome sequence assemblies greatly improves our ability to define overall genome organization and reconstruct evolutionary changes. In birds, this has previously been impeded by a near intractable karyotype and relied almost exclusively on comparative molecular cytogenetics of only the largest chromosomes. Here, novel whole genome sequence information from 21 avian genome sequences (most newly assembled) made available on an interactive browser (Evolution Highway) was analyzed.

**Results:** Focusing on the six best-assembled genomes allowed us to assemble a putative karyotype of the dinosaur ancestor for each chromosome. Reconstructing evolutionary events that led to each species' genome organization, we determined that the fastest rate of change occurred in the zebra finch and budgerigar, consistent with rapid speciation events in the Passeriformes and Psittaciformes. Intra- and interchromosomal changes were explained most parsimoniously by a series of inversions and translocations respectively, with breakpoint reuse being commonplace. Analyzing chicken and zebra finch, we found little evidence to support the hypothesis of an association of evolutionary breakpoint regions with recombination hotspots but some evidence to support the hypothesis that microchromosomes largely represent conserved blocks of synteny in the majority of the 21 species analyzed. All but one species showed the expected number of microchromosomal rearrangements predicted by the haploid chromosome count. Ostrich, however, appeared to retain an overall karyotype structure of  $2n = 80$  despite undergoing a large number (26) of hitherto un-described interchromosomal changes.

**Conclusions:** Results suggest that mechanisms exist to preserve a static overall avian karyotype/genomic structure, including the microchromosomes, with widespread interchromosomal change occurring rarely (e.g., in ostrich and budgerigar lineages). Of the species analyzed, the chicken lineage appeared to have undergone the fewest changes compared to the dinosaur ancestor.

**Keywords:** Ancestral karyotype, Avian genome, Chromosome evolution, Dinosaur

\* Correspondence: [dlarkin@rvc.ac.uk](mailto:dlarkin@rvc.ac.uk); [d.k.griffin@kent.ac.uk](mailto:d.k.griffin@kent.ac.uk)

†Equal contributors

<sup>2</sup>Department of Comparative Biomedical Sciences, Royal Veterinary College, University of London, London NW1 0TU, UK

<sup>1</sup>School of Biosciences, University of Kent, Canterbury CT2 7NJ, UK

Full list of author information is available at the end of the article



## Background

The mechanisms of genome evolution are most often considered from the perspective of individual genes or gene families; there is nonetheless increasing evidence supporting the functional role and significance of events at a chromosomal (cytogenetic) level [1]. To date, bird genomes remain relatively understudied from an overall genome organization perspective; however, the recent availability of multiple avian genome sequence assemblies [2] allows us to consider the role of chromosomal change in the evolution of Aves from their dinosaur ancestors. Chromosome rearrangements between species can cause or reinforce reproductive isolation through reduced fitness of hybrid offspring due to a compromised ability to synapse and segregate chromosomes at meiosis [3,4]. Moreover, reduced interspecific recombination in rearranged regions is thought to promote the accumulation of incompatibility loci in such regions [5-7]. The purpose of this study was to gain further insight into the mechanism of bird evolution through the multiple comparative analyses of chromosomal segments and breakpoints.

Unraveling the mechanisms and relevance of bird karyotype evolution has hitherto been impeded by a karyotype that is difficult to define because of indistinct banding on the macrochromosomes and a preponderance of cytogenetically indistinguishable microchromosomes. Indeed, to date, only a single avian karyotype (chicken) has been fully defined using a combination of BAC/cosmid clones and chromosome paints generated by flow cytometry and microdissection [8]. Moreover, karyotypes are broadly similar in overall pattern from species to species. For instance, at a cytogenetic level, two thirds of bird species have a chromosome number of around  $2n = 80$  with similar numbers of macro- and microchromosomes suggesting little interchromosomal changes between species [9]. Molecular insights into interchromosomal differences between species (and the evolutionary events that have led to them) have focused mostly on the largest macrochromosomes. These studies applied chicken chromosome paints [10] to the chromosomes of numerous other species (reviewed in [11]) in zoo-FISH experiments. Such investigations have provided much insight into intermacrochromosomal rearrangements between birds with the underlying message that the ancestral pattern has remained largely unaltered in the majority of species. Rare exceptions include significant chromosome rearrangement in Psittaciformes (parrots etc.), Falconiformes (falcons) and Sphenisciformes (penguins) [11]. There are also individual changes associated with representative orders, e.g., fission of chromosome 1 in Passeriformes (songbirds) and of chromosome 2 in certain Galliformes (land fowl) (reviewed in [11]). Studies of interchromosomal changes involving the microchromosomes are much more limited as the flow cytometry methods used to generate the chromosome

paints [10] do not have the resolution to isolate individual microchromosomes.

Using chicken BAC clones, studies provide a low-resolution appraisal of intrachromosomal rearrangements between chicken and other species [12-14] (turkey, duck, zebra finch, respectively). This approach, however, is limited in its ability to identify the molecular coordinates of evolutionary breakpoints. The availability of whole assembled genomes [15-17] allows comparative genomics at a much more detailed level of resolution than can be achieved by cross-species FISH. Burt *et al.* [18] were the first to use bioinformatics to define inter-species analysis of whole avian chromosomes at a genomic level (chicken-human). The publication of the chicken genome sequence [15] provided more detailed information, establishing conserved synteny between chicken and human whole genome assemblies. In the ten years since, only conserved synteny comparisons have been made between the chromosomes of two [14,19], or at most three [20,21] avian species.

The use of whole genome assemblies to study cytogenetic phenomena has raised interest in the study of comparative cytogenetics from the perspective of evolutionary breakpoint regions (EBRs) and homologous synteny blocks (HSBs). To date, the majority of such studies have focused on mammals [22], however, analysis of other groups, such as birds, is essential in order to establish whether mammalian systems are representative of, or an exception to, general patterns observed in other animal groups. Larkin *et al.* [22] found that, in mammals, EBRs can lie in gene-dense regions. In the human genome EBRs also lie in regions with more zinc finger protein genes, more genes whose function is associated with environmental stimulus response, as well as more segmental duplications, CNVs, SNPs and retrotransposed genes. Such "EBR genes" appear to be related to lineage-specific biology and adaptive features [22-24]. EBRs are also frequently reused, i.e. there are regions of the genome that are prone to chromosomal breakage leading to translocations, inversions and fissions [25,26]. Comparison of sequence assemblies in chicken, zebra finch and turkey suggests that breakpoint reuse is higher in birds than in mammals [20,21]. The data in birds also suggests a key role for recombination-based mechanisms in the generation of chromosome rearrangements in that EBR location is consistent with elevated levels of genetic recombination at these loci [14]. This is consistent with the notion that, if recombination drives chromosomal rearrangements and assuming an evolutionarily conserved recombination landscape [27-29], EBRs might be enriched in genomic regions with elevated recombination rates. Not all species show an association of chromosomal breakage and elevated recombination however, e.g., insects [30,31] and mammals. Indeed, in mammals Larkin *et al.* [22] suggested that the

highest levels of recombination are located between the EBRs rather than in association with them.

HSBs have been defined in all animal species thus far examined for conserved chromosomal synteny [32]. Larkin *et al.* [22] argue that the continued presence of HSBs in all species may indicate a selective advantage to the retention of gene combinations in close proximity. Supporting evidence is found in the fact that multispecies HSBs (msHSBs) involving nine mammals plus chicken, unlike EBRs, are enriched in gene ontology (GO) terms for organismal development, central nervous system, and brain function in the human genome. Others argue that the idea of close proximity and any resulting correlation in expression patterns (if present) are not necessarily adaptive or required (e.g., [33,34]). Given that around three quarters of avian chromosomes are small, cytogenetically indistinguishable microchromosomes, and that overall karyotype structure appears broadly similar between at least two thirds of bird species, a high degree of conserved chromosomal synteny is inferred [9]. This raises the hypothesis that avian karyotypes are evolutionarily static; however, for this to be tested, we would first need to establish that inter-microchromosomal rearrangements are rare or absent in most birds. If true, we would subsequently hypothesize that, like HSBs in mammals, individual whole microchromosomes are enriched for functional GO terms (regardless of any intrachromosomal rearrangements between them).

A detailed account of the chromosomal differences and changes that have occurred during the evolution of avian chromosomes is an essential prerequisite for any further insights into functional and/or mechanistic relevance. The combination of comparative analysis by bioinformatics and chromosome painting has the potential to do this, provided the appropriate tools are developed and used. The purpose of this study was thus to examine multiple avian genomes recently sequenced [2,35], reconstruct the common ancestral karyotype and thence the evolutionary events that led to extant karyotypes. Furthermore, we tested the hypothesis that EBRs occurring in two lineages (chicken and zebra finch) are associated with elevated levels of genetic recombination and assessed the degree to which EBRs are reused in avian evolution. Finally, we tested the hypothesis that whole microchromosomes essentially constitute interchromosomal HSBs (i.e. that rearrangements between them are rare or absent) and that each microchromosome consists of functionally enriched GO terms.

## Results

### Genomic data and visualization of HSBs and EBRs

Results from this study were derived from HSB and EBR data from a total of 21 avian genomes and one outgroup reptile species loaded to an interactive, publicly available

chromosome browser Evolution Highway [36]. This now allows for multispecies cytogenetic comparison in birds [37]. For six bird species (chicken, turkey, Pekin duck, zebra finch and budgerigar) and one lizard outgroup (Carolina anole - *Anolis carolinensis*), a combination of large scaffold size (manifested by  $N50 > 10$  Mb) and supporting molecular cytogenetic data (cross-species chromosome painting) allowed us to make chromosomal or near chromosomal comparison, orientation of HSBs and reconstruction of ancestral chromosome rearrangements. Evolution Highway screenshots for avian species and lizard outgroup compared to chicken chromosomes 5 and 11 are illustrated in Figure 1 (these chromosomes chosen throughout as they give the clearest representative examples in both FISH and bioinformatics analyses).

### FISH analysis

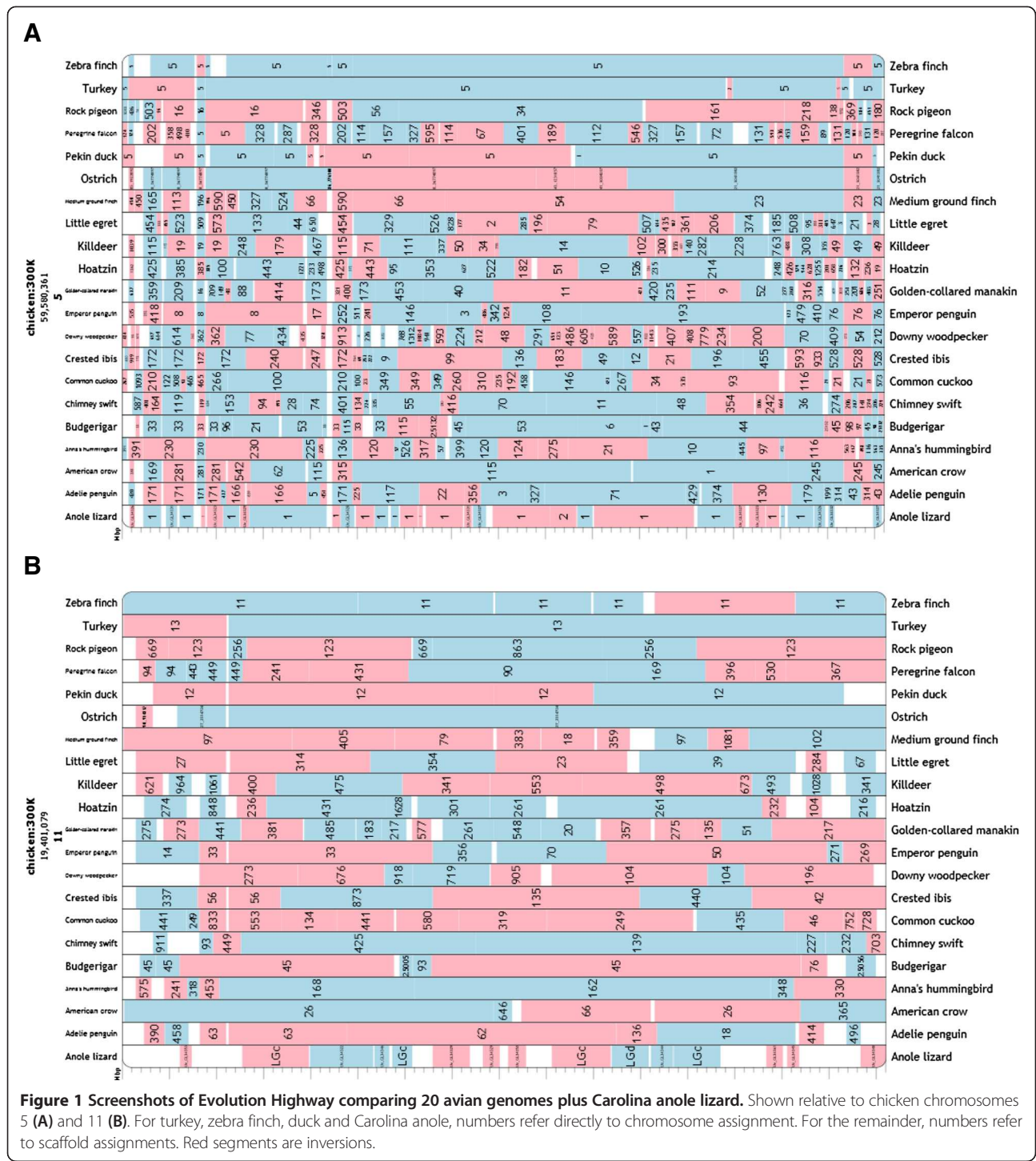
Reconstructions of scaffold-based assemblies also relied, in part, on previously published zoo-FISH (BAC and chromosome painting) data for the macro- and microchromosomes of chicken, turkey, duck and zebra finch [12-14] as well as newly generated data in this study as follows: we used seven new chicken microchromosomal paints A-G [21], verifying their assignments with chicken BACs (see Additional file 1) by dual color FISH and painting them onto ostrich and budgerigar metaphases.

For chicken, turkey, duck and zebra finch, zoo-FISH has been previously described [12-14]. For ostrich, no further differences between this species and chicken microchromosomes were found (Table 1 and Figure 2). For budgerigar, analysis reveals a more complex pattern incorporating several of the microchromosomes, namely six hitherto undescribed fusions (Table 1 and Figure 2).

### Reconstruction of ancestral karyotypes and chromosomal changes

A combination of FISH and bioinformatic analyses allowed reconstruction of ancestral chromosomes 1-5 for all birds, and chromosomes 6-28 + Z for Neognathae (see Methods). As a frame of reference, we used the new phylogenetic tree of another recent study [35]. Figure 3A indicates the comparative genomics of ancestral chromosome 5 and its orthologs, and 3B the changes that occurred in the orthologs of chicken chromosome 11. Although the outgroup did not have sufficient coverage to generate an "all-avian" ancestral chromosome directly for chromosome 11, the avian ancestral rearrangement is inferred from the identical patterns present in ostrich and chicken.

Overall, analysis suggests that, of the six species, the chicken lineage underwent the least number of intrachromosomal rearrangements (i.e. chicken was most similar to the common avian ancestor, probably a bipedal feathered dinosaur). Of the 46 rearrangements observed in the turkey lineage since the divergence from chicken 30 MYA



(million years ago), 19 were on chromosome 1 (we believe that this may be a slight overestimate due to assembly errors in the turkey genome). The analysis also suggests that ostrich lineage underwent 44 intrachromosomal changes on chromosomes 1–5 since the divergence from the common avian ancestor (approximately 100 MYA), and the duck 28 changes since the galliform-anseriform divergence

(~65 MYA). A faster rate of change was seen in the zebra finch and the budgerigar lineages, 41 in the former and 39 in the latter, occurring since the passeriform-psittaciform divergence (~54 MYA, Figure 4A). For the orthologs of chromosomes 6–28 + Z, in the absence of meaningful data from the lizard outgroup (i.e. there was minimal comparative data available), our analysis focused on the Neognathae



**Table 1 Comparative mapping of chicken chromosome paints A–G, and their ostrich and budgerigar orthologs**

Chromosome paint ID	Chicken chromosome(s)	Ostrich orthologs (all microchromosomes)	Budgerigar orthologs
A	11	1 pair	Fusion as part of chromosome 5
B	10 and 12	2 pairs	2 pairs of microchromosomes (no apparent fissions/fusions at this resolution)
C	13	1 pair	1 pair of microchromosomes (no apparent fissions/fusions at this resolution)
D	13 and 14	1 pair	1 microchromosome pair +1 arm of chromosome 8 = fission and fusion at this resolution
E	10 and 12	2 pairs	1 pair = fusion
F	16, 17 and 18	3 pairs	2 pairs = fusion
G	~5 pairs smaller than 18	No result	3 pairs = 2 fusions (although some signals are weak so may be failure of hybridization)

**Note:**

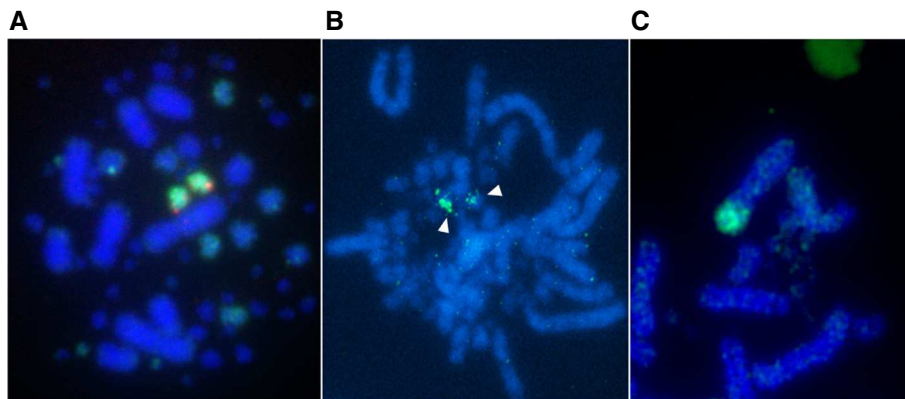
Bioinformatic approaches detected further rearrangements that are beyond the resolution of zoo-FISH. BACs that confirmed these assignments are given in Additional file 1: Table S1.

alone (using ostrich as an outgroup, Figure 4B). Again the chicken lineage appeared to have the least number of changes compared to the ancestor and the greatest rate of change was seen in the zebra finch since the passeriform-psittaciform divergence 54 MYA (68 for zebra finch and 79 for budgerigar). For all chromosomes, the intrachromosomal events are most parsimoniously explained by a series of inversions, and the interchromosomal rearrangements by a series of translocations. We next tested the robustness of our analysis in a series of additional MGRA simulations and iterations, excluding one species at a time from the set of six species (see Methods). We were interested to know if this would affect the general chicken-like pattern of the reconstructed avian ancestor. Results showed that, although the number of reconstructed contiguous ancestral regions (CARs) tended to decrease slightly if more fragmented (scaffold-based) genome assemblies (i.e. those of budgerigar and ostrich) were excluded, near identical order of msHSBs were observed within each CAR regardless of excluding one species. The number of changes and their timescales (hence rates of change) are presented in Figure 4A (for all

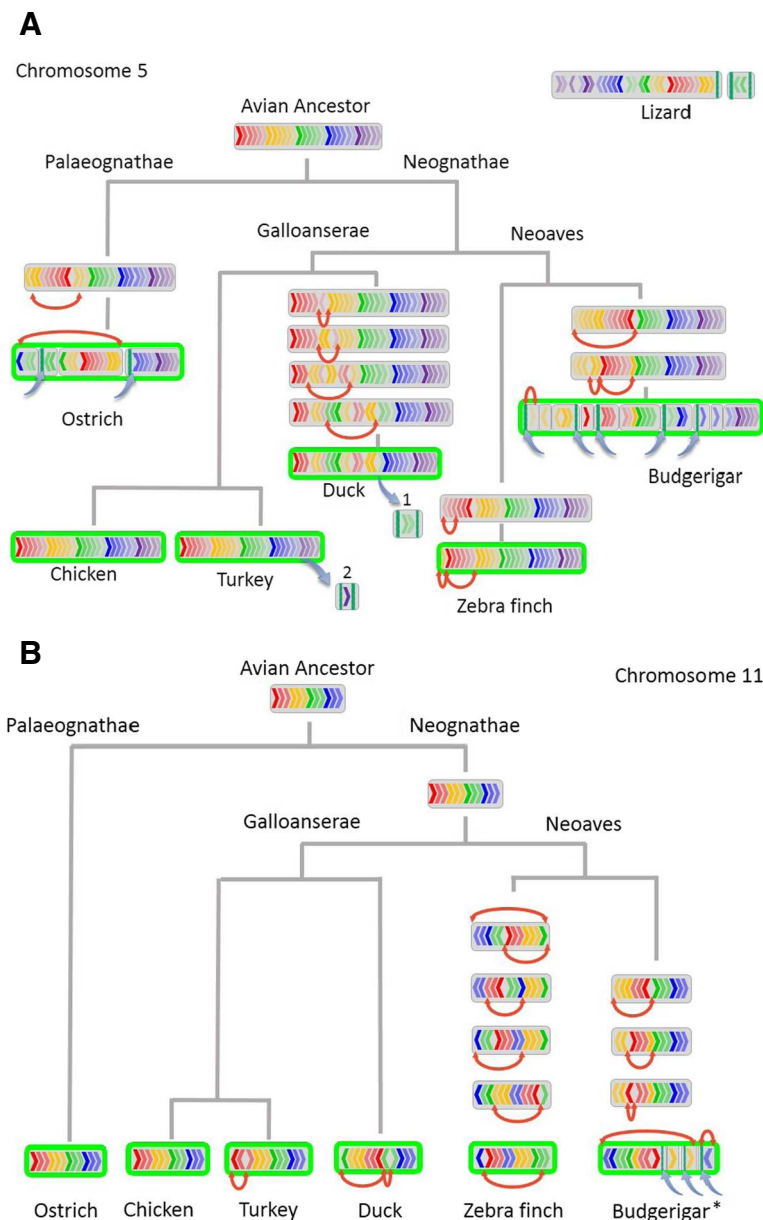
avian chromosomes 1–5) and 4B for the Neognathae (chromosomes 6–28 + Z).

A combination of FISH and bioinformatic data revealed a total of 26 interchromosomal and 44 intrachromosomal changes that have occurred in the ostrich lineage since divergence of the common avian ancestor ~100 MYA (Table 2 and Figure 4A). Most changes that occurred in the duck, chicken and turkey lineages appear to have done so since the galliform-anseriform divergence ~65 MYA. Notably, most of the changes seen in budgerigar and zebra finch lineages each appear to be different from one another, thereby suggesting that nearly all changes have occurred in the ~54 million years since the Passeriformes and the Psittaciformes diverged (Figure 4 and Table 2).

Closer analysis of the breakpoints to address the question of breakpoint reuse (see Background) identified, in chicken chromosomes 1–5 (and their turkey, duck, zebra finch, budgerigar and ostrich orthologs), 620 segment ends, of which 421 were involved in rearrangements. The most parsimonious predicted pathways from the common avian ancestor suggested that 100 breakpoint



**Figure 2 Chromosome painting experiment using chromosome paint A. (A)** On chicken chromosomes; dual FISH with a chromosome 11 BAC (red) confirms that this chromosome paint (green) maps to chromosome 11. **(B)** Painting one chromosome pair in ostrich; and **(C)** painting the terminal q arm of chromosome 5 in budgerigar.



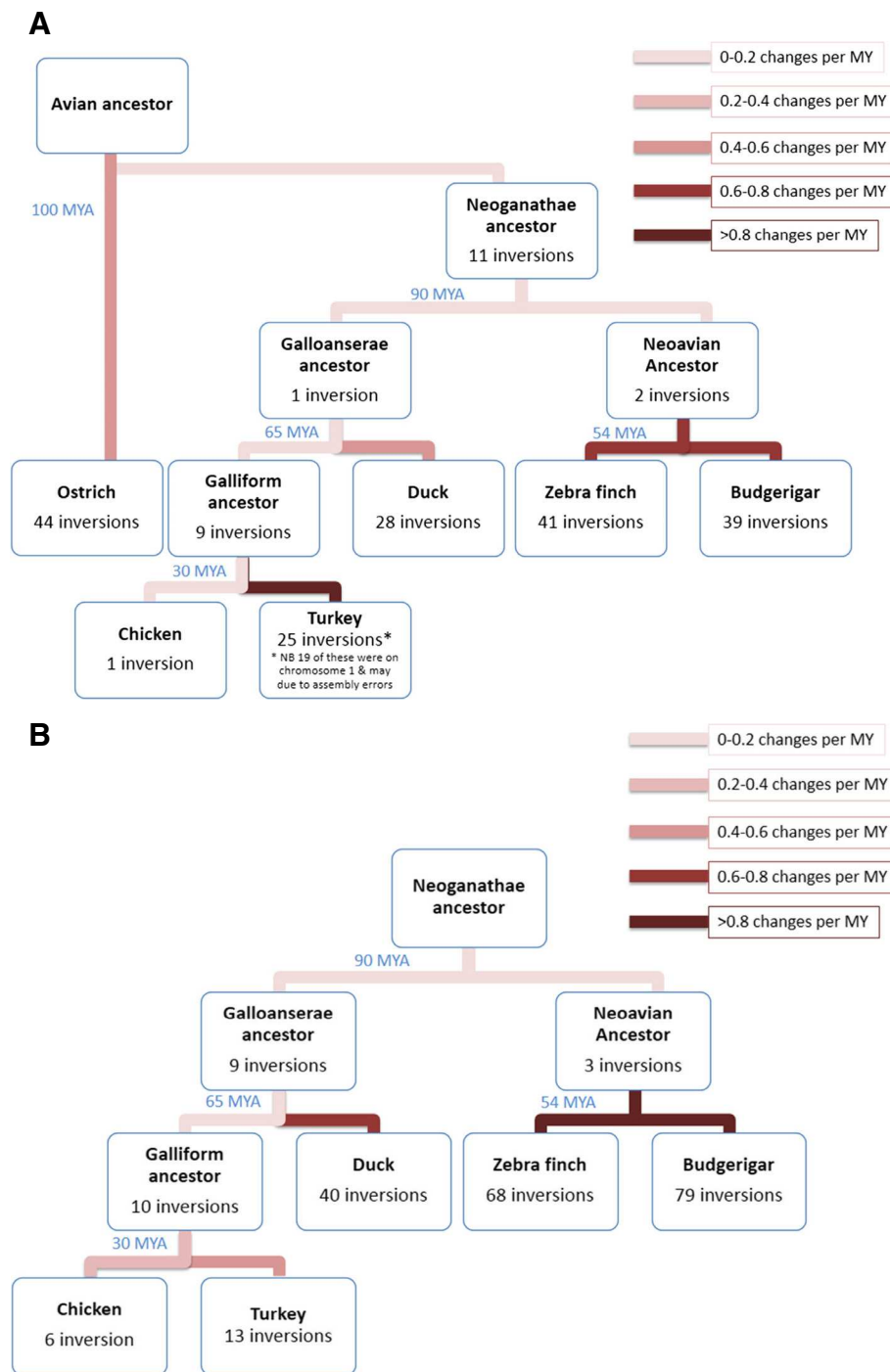
**Figure 3 Ancestral arrangement of chromosomes in six species and the rearrangements led to the extant pattern.** Exemplified for chicken chromosomes 5 (A; Carolina anole lizard arrangement also indicated) and 11 (B). Rainbow patterned arrows within the chromosomes represent the HSBs, red curved arrows indicate chromosome inversions, blue arrows indicate chromosome translocations, green outline indicates the chromosome painting results. As the arrangement for ostrich and Neognathae ancestors were the same, the avian ancestor could be derived (unlike for other chromosomes smaller than 5). \*In budgerigar, FISH indicates fusion to a larger chromosome.

regions (23.8%) recurred in different lineages, whereas 214 breakpoint regions (50.8%) recurred in either the same or different lineages. In chicken chromosomes 4p, 6–28 and Z, and their turkey, duck, zebra finch and budgerigar orthologs, 560 segment ends were identified, of which 428 were involved in rearrangements. The most parsimonious predicted pathways from the common avian ancestor suggested that 109 breakpoint regions (25.5%) recurred in different lineages, whereas 210 breakpoint

regions (49.1%) recurred in either the same or different lineages.

#### EBRs and recombination in chicken and zebra finch

As also mentioned in the Background section, we tested the hypothesis that the presence of EBRs was related to the regional recombination rate. Given the quality of the genetic maps and the data available in this study, this could be achieved for the chicken and zebra finch only.



**Figure 4 Total number of chromosomal inversions in six extant species as they diverged from the ancestor.** The inversions most parsimoniously explain the patterns seen in these species. **(A)** For chromosomes 1–5, sufficient coverage of the lizard outgroup allowed conclusions to be drawn from an avian ancestor. **(B)** For chromosomes 6–28+Z, ostrich was used as an outgroup due to the lack of coverage in the lizard. Greatest rates of change were seen in zebra finch and budgerigar. The phylogenetic tree is based on [35].

In chicken the analysis revealed no association between presence of EBR and the regional recombination rate. The 1 Mb non-overlapping windows containing EBRs ( $n = 35$ ) had an average recombination rate of  $2.80 (\pm 3.00, SD)$  cM/

Mb while windows without EBRs ( $n = 963$ ) had an average recombination rate of  $2.90 (\pm 3.00)$  cM/Mb (Wilcoxon's test,  $W = 13492$ ,  $P = 0.42$ ; randomization test, empirical difference in mean between classes =  $-0.11$ ,  $P = 0.28$ ; Figure 5).

**Table 2 Total numbers of inter- and intrachromosomal rearrangements since divergence from avian ancestor 100 MYA**

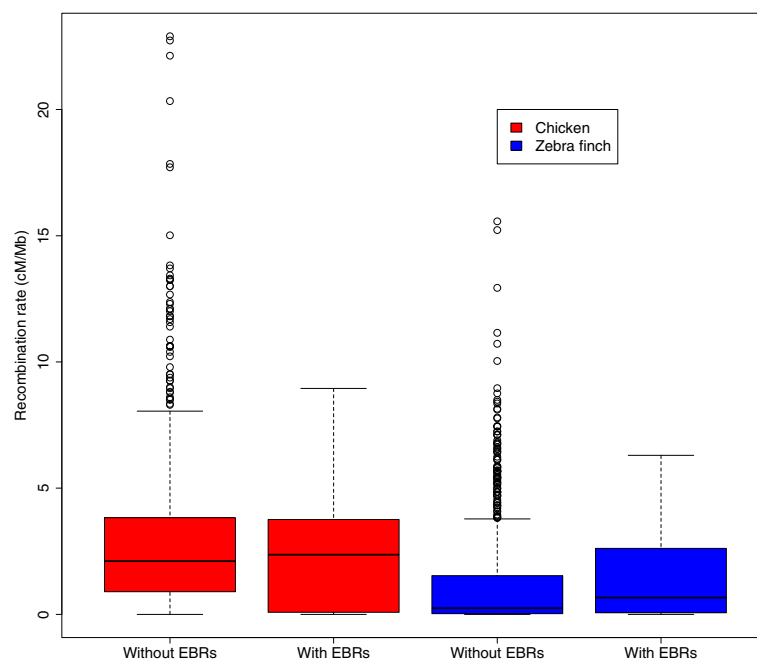
Species	Ostrich	Chicken	Turkey	Duck	Zebra finch	Budgerigar
No. of interchromosomal changes (as determined by FISH) from avian ancestor	0	1	1	0	2	8
No. of interchromosomal changes (determined using bioinformatics) from avian ancestor	26	1	5	1	2	40
No. of intrachromosomal changes from avian ancestor in chromosomes 1–5 (excluding 4p)	44	22	46	40	54	52
No. of intrachromosomal changes from Neognathae ancestor in chromosomes 6–28 + 4p + Z	Not applicable	25	32	49	71	82

In zebra finch, 1 Mb non-overlapping windows with EBRs ( $n = 31$ ) had a slightly higher recombination rate than windows without ( $n = 952$ ; 1.60 vs. 1.29 cM/Mb), although this was not statistically significant (Wilcoxon's test,  $P = 0.1$ ; randomization test, empirical difference in mean between classes = 0.31,  $P = 0.1$ ; Figure 5).

**Interchromosomal changes in multiple species and GO of microchromosomes**

For chicken, turkey, zebra finch and duck, intermacrochromosomal changes have been previously described, i.e. chromosome 4 fusion for chicken, chromosome 2 fission for turkey, chromosome 1 fission for zebra finch, and no changes in duck [12-14] in these four species. In the current analyses, however, results suggested that there were at least 26 interchromosomal differences between chicken and ostrich, and 40 between chicken

and budgerigar for all chromosomes (Table 2), with the changes in the budgerigar lineage occurring since the passeriform-psittaciform divergence (~54 MYA). Considering microchromosomes alone and using data pertaining to numbers of interchromosomal rearrangements for the remaining 15 species [37], results suggested that microchromosomal rearrangement was rare, except where the species of interest had been previously known to have an unusually large or small number of chromosomes (Table 3). In other words, as illustrated in Figure 6, there was a statistically significant correlation ( $R^2 = 0.3$ ;  $P = 0.03$ ) between number of interchromosomal rearrangements and published deviation from a haploid chromosome number of 40. The exception to this "rule" was the ostrich ( $2n = 80$ ), with 26 interchromosomal differences, 11 involving the microchromosomes, results suggesting significant rearrangement while maintaining the



**Figure 5 Rates of recombination and their association with EBRs for chicken (red) and zebra finch (blue).** In chicken, recombination rates are near identical in windows with and without EBRs (2.90 and 2.80, respectively). In zebra finch recombination rates are slightly higher in windows with EBRs (1.60 and 1.29, respectively) but the difference does not reach statistical significance ( $P = 0.1$  for both tests used).

**Table 3 Total number of interchromosomal rearrangements involving microchromosomes in 21 avian species compared to chicken**

Species	Total number of interchromosomal changes involving macro- and microchromosomes	Interchromosomal changes between microchromosomes only	Haploid chromosome number (difference from $n = 40$ )
Adélie penguin ( <i>Pygoscelis adeliae</i> )	6	0	48 (8)
American crow ( <i>Corvus brachyrhynchos</i> )	1	0	40 (0)
Common cuckoo ( <i>Cuculus canorus</i> )	1	0	?
Pekin duck ( <i>Anas platyrhynchos</i> )	0	0	40 (0)
Little egret ( <i>Egretta garzetta</i> )	4	1	33 (7)
Emperor penguin ( <i>Aptenodytes forsteri</i> )	5	0	36 (5)
Peregrine falcon ( <i>Falco peregrinus</i> )	6	4	25 (15)
Zebra finch ( <i>Taeniopygia guttata</i> )	0	0	40 (0)
Hoatzin ( <i>Ophisthocomus hoazin</i> )	3	0	?
Anna's hummingbird ( <i>Calypte anna</i> )	0	0	37 (3)
Crested ibis ( <i>Nipponia nippon</i> )	6	0	34 (6)
Killdeer ( <i>Charadrius vociferous</i> )	1	0	38 (2)
Golden collared manakin ( <i>Manacus vitellinus</i> )	0	0	?
Medium ground finch ( <i>Geospiza fortis</i> )	0	0	?
Ostrich ( <i>Struthio camelus</i> )	11	0	40 (0)
Budgerigar ( <i>Melopsittacus undulatus</i> )	11	2	29 (11)
Rock dove ( <i>Columba livia</i> )	1	0	40 (0)
Chimney swift ( <i>Chaetura pelagica</i> )	1	0	?
Turkey ( <i>Meleagris gallopavo</i> )	2	0	40 (0)
Downy woodpecker ( <i>Picoides pubescens</i> )	4	1	?
Chicken ( <i>Gallus gallus</i> )	0	0	39 (1)

As detected by bioinformatic approaches [37] and compared to the published haploid number of chromosomes in each species [9]. For counts of all interchromosomal rearrangements in the bird genomes see [37].

overall karyotypic structure. Indeed, if ostrich is excluded from the analysis outlined in Table 3 and Figure 6, the statistical significance of the association increases markedly ( $R^2 = 0.7$ ,  $P = 0.0002$ ).

Once we had established (above) that rearrangements were rare in the microchromosomes, then this led to the hypothesis that each microchromosome contained functionally enriched GO categories (see Background). We found evidence to support this hypothesis only for chromosome 16 (enriched for immune function) when  $P < 0.05$  and a false discovery rate (FDR) threshold of 0.05 were applied. Nonetheless several chromosomes had a significant  $P$  value but did not pass the FDR threshold: for chromosome 11 enrichment categories were apparent for drug/cafeine metabolism as well as hemophilic cell adhesion; for chromosome 12 genes for nucleotide binding were clustered together; for chromosome 13 there were enrichment categories for GTPase regulator activity; phosphatase activity in chromosome 15; chromosome 17 for glycosylation and glycoprotein related processes; chromosome 18 for cytoskeletal and motor protein related

genes; and chromosome 20 for genes involved in apoptosis and cell death.

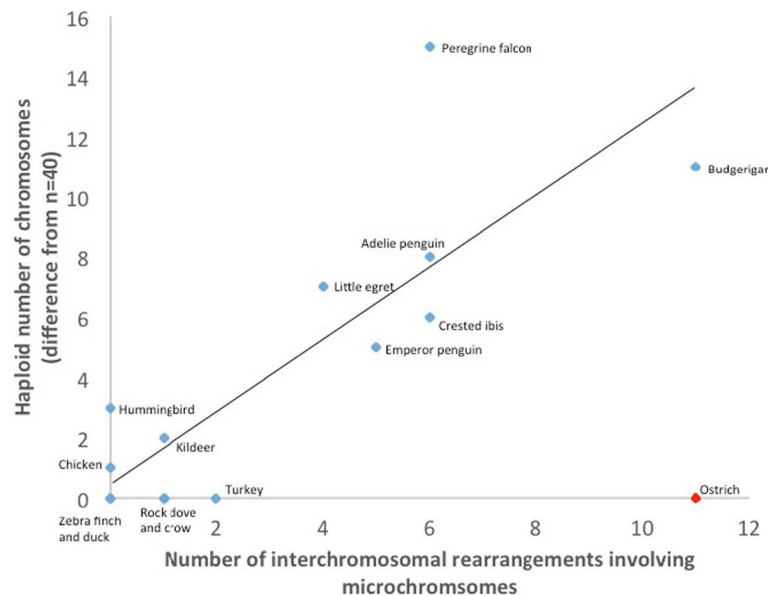
We thus find evidence to support our hypothesis that microchromosomes represent highly conserved blocks of interchromosomal synteny but find limited evidence to support the hypothesis that one possible explanation for this is a clustering of genes of associated function on the same chromosome.

## Discussion

The results presented here signify the most comprehensive appraisal of avian comparative cytogenetics to date. They provide a more detailed reconstruction of avian genome evolution than could be achieved by zoo-FISH analysis alone and demonstrate proof of principle from which further studies of genome evolution and comparative genomics can ensue.

We used a highly interactive avian genome dataset from the Evolution Highway comparative chromosome browser [37,38] that, as has already been demonstrated in mammals, can be applied to compare the chromosome organization





**Figure 6 Number of interchromosomal rearrangements involving microchromosomes.** Plotted against deviation from  $n = 40$  for each species in which chromosome number is published (Table 3). Analysis suggests that haploid chromosome number effectively is a reflection of number of microchromosomal rearrangement, except in ostrich (red dot). Best-fit line is drawn excluding ostrich outlier ( $R^2 = 0.7$ ,  $P = 0.0002$  if ostrich is excluded;  $R^2 = 0.3$ ,  $P = 0.03$  if ostrich is included).

of individual or multiple species. The ultimate aim for this browser is that, in chromosomes for all avian species uploaded, HSBs will be displayed with reference to the chromosome number, as is currently the case for turkey, zebra finch and duck, or to specific scaffolds for other birds. In future, this will be achieved by a number of strategies: (a) by improved scaffold sizes, e.g., using optical mapping such as has been achieved to some degree in ostrich and budgerigar in this study; (b) by linkage to radiation hybrid (RH) maps such as was achieved for duck in this study (see also [19]); (c) by association with known linkage and other physical maps (e.g., [39,40]); (d) by use of novel algorithms to order and orient scaffolds into longer chromosomal fragments or whole chromosomes using comparative genome information and pair-end reads (reference-assisted chromosome assembly; [41]); (e) by systematic FISH mapping to chromosomes of orthologous clones derived from the individual scaffolds. We are currently concentrating our efforts on the development of FISH probes that will identify not only on which chromosomes the scaffolds lie in the species of interest, but also the order in which they appear on the chromosome. With current technology, however, even the best-assembled genomes (e.g., assisted with optical mapping) require a degree of intervention by molecular cytogenetics in order to generate a complete picture of overall genome organization. Given the efforts that have been made to sequence the genomes of the birds recently by current technologies [2], it is questionable how many of them will be re-sequenced using newer technologies that generate large

scaffolds. A note of caution is relevant here: no genome assembly is “perfect” - the results reported here and elsewhere represent the state of the art in terms of what can be reasonably gleaned with the current technology available. Our future studies will focus on the systematic molecular characterization by zoo-FISH of as many scaffolds and EBRs as time and resources allow.

Earlier cytogenetic data suggested that, for the majority of bird species, karyotypic patterns are broadly similar to one another [9,11,14,20]. This purportedly extends to ratite birds [42-44]; however, further analysis presented in this study challenges this notion. That is, we identified 26 interchromosomal rearrangements in ostrich compared to the ancestor. Moreover, the question of whether the conserved interchromosomal synteny seen in the macrochromosomes applies to the microchromosomes has hitherto been beyond the resolution of contemporary methodology. This study is the first to classify inter-microchromosomal rearrangements in any species; we provide evidence that interchromosomal rearrangements are nonetheless rare, except in cases (around 1/3 of species) where we already knew that karyotypes were highly rearranged [9]. Ostrich is the exception and it will be interesting to note whether this applies to other ratite birds.

Microchromosomes are not a uniquely avian feature. They are also found in some primitive amphibians (Cryptobranchidae and Hynobiidae have 14–19 pairs [45,46]), most (but not all) reptiles (snakes have around 20 pairs [47]), but paradoxically not Crocodylia [48] – the closest phylogenetic lineage to birds. Indeed microchromosomes

are typical of most amniotes (mammals and crocodilians being exceptions); however, the greatest number and smallest size of microchromosomes are typically found among birds. Burt [49] in a “fission-fusion” hypothesis suggested that most microchromosomes were already present in the common dinosaur ancestor that gave rise to birds (which probably had already evolved a small genome size and karyotype of around  $2n = 60$  including 20 pairs of microchromosomes) but that chromosome fission created the remainder, presumably including the smallest ones. In the current study, the similar number of chromosomes amongst most species but relatively large number of rearrangements between ostrich and all the other birds studied suggest that a basic pattern of  $2n = 80$  (~30 pairs of microchromosomes) became fixed before the Palaeognathae-Neognathae divergence 100 MYA but that interchromosomal rearrangement was still relatively common in birds at the time. Another alternative is that ratite birds underwent further adaptive changes that may be associated with the very different phenotypes present in this clade alone. The paucity of inter-microchromosomal rearrangements between most Neognathae (if the evidence presented here is representative, this would presumably include the 2/3 of Neognathae species where  $2n = \sim 80$ ) supports our hypothesis that the microchromosomes represent blocks of conserved synteny at an interchromosomal level. An absence of interchromosomal rearrangement could either suggest an evolutionary advantage to retaining this particular configuration or a lack of opportunity for chromosome rearrangement. The latter might be explained by few recombination hotspots, transposable elements or endogenous retroviruses, all of which have been associated with chromosomal change. Both inter- and intrachromosomal change can arise via these mechanisms, and thus the rapid amount of intrachromosomal but not interchromosomal change in our representative passeriform species, the zebra finch, suggest that there may be an evolutionary advantage to keeping microchromosomes numerous, gene dense, compact and evolutionarily static. Stasis in evolution can, however, arise via alternative interpretations; it may be that the mutational mechanisms underlying chromosomal changes are different in birds or that lack of adaptive value, rather than purifying selection, slows down the rate of chromosomal changes. At the time of writing no sequences have yet been associated with the very smallest of the avian microchromosomes (29–38) and this is an issue that will require rectifying in future avian genome projects using more sophisticated technologies.

The rate of chromosomal change in any eukaryotic organism, and the speciation that ultimately arises from it, is dependent on two factors: the rate of mutation and the rate of fixation [18]. The mutation rate of chromosomes is, in turn, related to the frequency of homologous sites

[49]. Repeat structures in general, and transposable elements in particular, provide substrates for chromosomal rearrangement. In a genome that is constrained by size (perhaps, as has been suggested, because of the energy requirements associated with flight [50,51]), the opportunity for mutation is reduced and only fission (or intrachromosomal rearrangement such as inversion) can occur. This would explain first why the avian genome is the most fragmented of any vertebrate genome (i.e. birds have the most chromosomes) and second why there have been few interchromosomal rearrangements in most species. There are also possible advantages of multiple chromosomes in a karyotype in terms of generating variation, the driver of natural selection. That is, more chromosomes lead to more combinations of gametes as well as an increase in recombination rate as there has to be at least one obligatory chiasma per chromosome. The absence of positive selection for much change in chromosome number is a possible explanation of why there was little fixation of any interchromosomal changes among birds although inbreeding and genetic drift may play a role [18,49,52,53]. Burt [49] suggested that a higher recombination rate is another constraint that has resulted in the properties we most associate with microchromosomes (e.g., high GC-content, low repeats, high gene-density) and led to the maintenance of the typical avian karyotype with both macro- and microchromosomes and few rearrangements between them.

A constraint of overall karyotype structure does not preclude intrachromosomal rearrangements. Indeed there is a correlation between the rates of speciation and intrachromosomal rearrangement [4]. In the current study, the rapid rate of intrachromosomal rearrangement in the zebra finch would argue for a relationship between intrachromosomal rearrangement and speciation in birds given the Passeriformes represent over half of all species. Such mechanisms could be mediated through an increase in localized repeat content. Hotspots of recombination have previously been reported to also play a role [14] and in this study we tested the hypothesis further utilizing “zebra finch only” and “chicken only” breakpoints comparing them to previously reported genetic maps of each species [37,54,55]. In chicken, recombination rates were near identical in regions with breakpoints compared to those without. In zebra finch, the difference in rates between regions containing EBRs and regions without EBRs, although similar in magnitude to that previously reported [14], failed to reach statistical significance (at  $P < 0.05$ ). This therefore casts doubt on our original findings, thereby either suggesting that our hypothesis should be rejected or that the numbers in the study were not sufficiently large to reach statistical significance. A further alternative explanation is that the available recombination maps have too low marker density (typically Mb scale) to

pick up local recombination rate variation at a sufficiently detailed scale (Kb scale) to detect associations with EBRs. Study of a greater number of species in this manner using high-density linkage maps or population based recombination rate estimates may resolve the paradoxical difference between [14] and the current study.

Some avian species undergo a radical departure from the typical ( $2n = \sim 80$ ) avian genome organization. The presence of an unusually high chromosome number in the Adélie penguin ( $2n = 96$ ) and a lower than average number in the emperor penguin ( $2n = 72$ ) (but both associated with high degrees of inter-microchromosomal rearrangement) suggest that similar mechanisms can act to either reduce or increase chromosome number rapidly. Evidence from the penguins and the rearranged karyotypes of the Falconiformes and the Psittaciformes suggest that these changes can happen in a relatively short time. Mammals, reptiles and amphibians with larger, repeat-rich genomes have the potential to undergo rapid intra- and interchromosomal rearrangements and the results presented here suggest that birds too can undergo similar changes in certain groups. We are not, however, aware of any evidence to suggest that highly rearranged avian genomes are especially large, or significantly more repeat-rich than other avian genomes. Comparisons of the zebra finch and the budgerigar suggest that mutation rates of chromosomes may well be similarly high in both groups but that they are features associated with exploiting evolutionary niches in certain groups that serve to fix interchromosomal rearrangements, while in others such fixation is prevented and the overall avian karyotype maintained. Such processes are, to date, undiscovered but possible clues might lie in the study of GO terms present in EBRs. In an associated study, a correlation between EBRs and specific avian adaptive features in individual species has been demonstrated. This included forebrain development in budgerigar, one of the six species focused upon in this study and consistent with this species being not only vocal-learner but having distinctive neuronal connections compared to other vocal-learners [37]. As more genomes become available with better assemblies, these analyses may well point to adaptive phenotypic features of individual orders and families.

Finally, we observed that it appears to be the chicken that seems to have undergone the fewest chromosomal changes compared to the ancestor. There are interesting parallels between this study and another study [56] examining sex chromosome evolution. While our data demonstrates that autosomes have been reorganized least in chicken chromosomes 1–5 in comparison to the common avian ancestor, Zhou *et al.* [56] conclude that the ancestral sex chromosome organization is observed closer to that of the Palaeognathae (ostrich and emu). Zhou *et al.* [56] show less degradation of the sex chromosomes and a closer synteny to the lizard. As, in this study, we only examined the

Z chromosome in the Neognathae (for the reasons given), further studies will be required to establish whether sex chromosomes and autosomes preserve their ancestry differently in the different lineages. The question also arises of whether chicken and related species, having undergone the fewest chromosomal changes, have undergone the fewest adaptive changes compared to the avian ancestor. Most authors agree that the dinosaur ancestors of birds were bipedal and terrestrial, relatively small (small size being an immediate pre-adaptation to flight) and had limited flying ability, not unlike Galliformes [57]. On the other hand, the earliest known Ornithurae along the presumed direct line to modern birds were either fully aquatic or amphibious (e.g., *Gansus* [58]) and details of their anatomy, including webbed feet, have been likened to ducks [59,60]. The oldest relatively certain fossil representative of Neornithes (modern birds) is aquatic, and identified as a Galloanseres (e.g., *Vegavis* [61]). However, the fossil record may be difficult to interpret due to geographic and depositional sampling biases, limited understanding of functional anatomy, and the uncertainty that avian ancestors were ecologically and behaviorally typical of the larger groups to which they belonged. As an independent record of the actual substance of inheritance of living birds, genomic characteristics such as chromosomal arrangement complement a fossil record that may imperfectly represent actual neornithine forebears. Thus, chromosomal rearrangements may provide information on the ecological adaptations of avian ancestors that the fossil record may never be able to establish unambiguously [62].

## Conclusions

In summary, this study represents the most comprehensive appraisal of changes in overall avian genome structure hitherto reported. We provide further insight on previously reported roles of genetic recombination in chromosome rearrangement and on the functional significance of karyotype stability in the avian genome. Here, we establish that the chicken lineage contains the fewest number of chromosomal changes compared to the dinosaur ancestor relative to the other five species studied. At this stage it would be unwise automatically to infer that this means that the chicken has the fewest number of adaptive changes also. This will nonetheless be the topic of future study.

## Methods

### Presentation of multiple avian genome assemblies

In order to present and visualize comparative cytogenetics and identify HSBs and EBRs in multiple avian species, an interactive, comparative chromosome browser Evolution Highway was used [38]. All blocks of synteny were identified and displayed relative to chromosomes of the reference chicken genome (ICGSC *Gallus\_gallus*-4.0/galGal4).

Evolution Highway was used to display the sequence coordinates of all syntenic fragments (SF) and HSBs in each genome [37]). We made use of the set of HSBs and SFs that contained rearrangements that are  $\geq 300$  Kb in the reference genome. This set, together with two other separate sets that visualize HSBs and SFs that are larger than 100 Kb and 500 Kb in the reference genome, is publicly available from the Evolution Highway website [36] (Figure 1) and are further described in [37].

For the purposes of this study, 21 avian genomes plus one outgroup species were utilized to address the questions set out in the Background section and made up of the following: of these 21, 17 were recently sequenced and presented [2] including common cuckoo, peregrine falcon, American crow, little egret, crested ibis, domestic pigeon, hoatzin, golden-collared manakin, medium ground finch, downy woodpecker, Adélie penguin, emperor penguin, Anna's hummingbird, chimney swift, killdeer, budgerigar and ostrich. Conserved blocks of synteny are presented as scaffolds (scaffold 1 being the largest and the rest numbered accordingly to size) in relation to chicken chromosomes. Chromosome-level assembly and analysis of conserved synteny had been previously reported for the largest (macro-) chromosomes of chicken, turkey and zebra finch [14,20,21]. Thus, the turkey (TGC Turkey\_2.01/melGal1) and zebra finch (WUGSC 3.2.4/taeGut1) genomes were presented in Evolution Highway with reference to published chromosome number (e.g., chromosome 11 in chicken corresponds to chromosome 12 in duck and 13 in turkey; see Figure 1). Chromosome-level assembly of the Pekin duck genome was constructed from available genome scaffolds [63] using an original RH mapping approach through hybrid sequencing (Faraut *et al.*, personal communication). Pekin duck was added and presented with reference to published chromosome number. The Carolina anole was the only reptile outgroup genome available with reference to whole chromosomes and therefore this was chosen for this study as the outgroup for reconstruction of the ancestral chromosomes (see the sub-section *Establishment of ancestral avian karyotypes*).

Of the 17 newly sequenced species, two (ostrich and budgerigar) were selected for studies involving reconstruction of the ancestral chromosomes. These species, thanks to optical mapping, had the largest N50 ( $>10$  Mb) and were also the species on which we performed zoo-FISH studies due to the availability of material for chromosome preparation. These and the remaining 15 species were used for defining EBRs to compare with recombination rate and for establishing interchromosomal conserved synteny among the microchromosomes [37].

#### **Karyotype and zoo-FISH analysis**

For chromosome analysis, rapidly dividing embryonic fibroblasts or white blood cells were arrested in metaphase

using colchicine (Sigma), swollen using 75 mM KCl and fixed to glass slides using 3:1 methanol : acetic acid mix. Metaphases were stained with a combination of DAPI and propidium iodide in VECTASHIELD® antifade medium (Vector Laboratories). Image capture involved an Olympus BX61 epifluorescence microscope with cooled CCD camera; SmartCapture system and SmartType software (Digital Scientific UK) were used for capturing and karyotyping purposes, respectively. Microchromosome paints described elsewhere [21] were generated by flow cytometry, then amplified and directly labeled with FITC using DOP-PCR. BAC clone DNAs were used to verify chromosome paint alignment and were extracted by miniprep (QIAprep Spin Miniprep Kit, QIAGEN), then directly labeled by nick translation with FITC or Cy3.5.

For FISH, metaphases were probed with chicken chromosome paints and BACs generated above. Briefly, probes were dissolved in a formamide buffer and applied, under a coverslip, and then sealed using rubber cement. Simultaneous denaturation of probe and genomic DNA on a 75°C hotplate preceded hybridization at 37°C (overnight for same species FISH, three days for zoo-FISH). Post-hybridization washes (2 minutes in  $0.4 \times$  SSC at 73°C; 30 seconds in  $2 \times$  SSC/0.5% Tween 20 at room temperature) were followed by chromosome counterstaining using VECTASHIELD® antifade medium with DAPI and viewed as above using epifluorescence and SmartCapture (Digital Scientific UK).

#### **Establishment of ancestral avian karyotypes**

In total six avian species (chicken, turkey, duck, zebra finch, ostrich and budgerigar) plus one lizard outgroup species (Carolina anole) were chosen for reconstruction of the ancestral karyotypes (for the reasons given in the sub-section *Presentation of multiple avian genome assemblies*). A combination of bioinformatics, zoo-FISH and karyotyping allowed us to make reconstructions of the order and orientation of scaffolds and thence the ancestral chromosomes. To reconstruct a putative avian ancestor as inferred from orthology maps the Multiple Genomes Rearrangements and Ancestors (MGRA) tool on the Algorithmic Biology Lab web server at St. Petersburg Academic University of the Russian Academy of Sciences [64,65] was used as follows: using Evolution Highway, pairwise alignments for turkey, duck, zebra finch, budgerigar and ostrich were visualized relative to the chicken whole genome sequence as a reference at the 300 Kb resolution. The orthology map of the Carolina anole, also visualized by Evolution Highway, was used as an input for the MGRA program and included in the analysis as an outgroup. Orthologous regions observed in all the species compared were defined as mHSBs and served as MGRA inputs for individual genomes. The hypothetical ancestral genome was determined using the phylogenetic tree information for this set of six species [35].



For chromosomes 1–5, 80% of the avian genomes were also represented by orthologous sequences in the Carolina anole outgroup. In this case we could therefore reconstruct the ancestral chromosomes for all birds. For chromosomes 6–28 and Z, we used ostrich as the outgroup (thus only drawing conclusions about the Neognathae), as only ~9% of the genome had orthologous sequences represented in the lizard outgroup. Where the ostrich and Neognathae ancestor had the same arrangement of HSBs, we could infer the avian ancestor (as with chromosome 11, Figure 3).

In order to test the robustness of our analysis in a series of additional MGRA simulations and iterations, we established if exclusion of one species at a time from the set of six species would affect the overall pattern of the reconstructed avian ancestor genome organization.

#### Reconstruction of evolutionary events guided by MGRA

The positions of CARs and HSBs or SFs within each species genome were noted, allowing correlation with our previously published FISH based physical mapping data in chicken turkey, duck and zebra finch [12–14] and that derived by cross-species chromosome painting in former publications [66,67] and in the current study. These data were previously acquired by cross-species FISH of chicken BACs and chromosome paints onto turkey, duck, ostrich and budgerigar chromosomes, and same-species FISH of orthologous zebra finch BACs onto zebra finch chromosomes.

The available karyotypic, FISH and bioinformatic data were combined to generate the “best-fit” model for chromosomal evolution in the six avian species of interest, i.e. the one with the minimum number of rearrangements. The MGRA tool was used on the whole genome datasets to reconstruct the evolutionary events that, most parsimoniously, led to the arrangement seen in the extant species. For the most part, the changes suggested by MGRA were accepted as the most parsimonious involving the minimum inversions for intrachromosomal rearrangements and fissions/fusions for interchromosomal rearrangements (the process of defining the inversions is illustrated in Figure 3; see also [20]). In cases where apparent interchromosomal rearrangements (such as translocations) had occurred, the MGRA solution was cross-referenced with the reconstructions on a chromosome-by-chromosome basis using the Multiple Genome Rearrangements (MGR) tool [68,69] and with zoo-FISH data. In cases of disagreement on the pattern of rearrangements, three independent observers with extensive cytogenetic expertise manually checked and decided the rearrangement pattern. When a whole, otherwise independent, block (scaffold or chromosome) was classed as inverted, this was counted in the analysis as a true inversion if a different orientation was recovered for two or more species (example shown in Figure 3b for chromosome 11 in zebra finch).

#### Identification of EBRs and breakpoint reuse

We used the EBRs defined in [37] that involved a single reference chromosome (intrachromosomal EBRs) and more than one reference chromosome (interchromosomal EBRs) in target species' chromosomes or scaffolds [70]. Interchromosomal EBRs delineated interchromosomal rearrangements, which were then compared with published chromosome number [9], or more specifically deviation from  $n = 40$ ; correlation coefficient  $R^2$  was calculated using Microsoft Excel. In order to determine breakpoint reuse, the series of possible rearrangements from the common avian ancestor (with lizard as the outgroup, chromosomes 1–5) or Neognathae ancestor (with ostrich as the outgroup, chromosomes 4p, Z and 6–28) to each species was considered, and for each rearrangement, the segment ends flanking the breakpoints were noted. Within each lineage, the number of times a segment end was involved in a rearrangement was counted and reuse classified if it occurred more than once in any lineage or between lineages.

#### Recombination rate analyses

We used the chicken- and finch-specific EBRs defined in [37] to compare with chicken-specific recombination rates and zebra finch-specific EBRs with zebra-finch recombination rates. This differed from our previous approach [14] in which we examined all EBRs between three species compared to the zebra finch genetic map. Zebra finch-specific EBRs coordinates initially identified in chicken chromosomes were translated into zebra finch chromosome coordinates (WUGSC 3.2.4/taeGut1) using the correspondence between coordinates of finch HSB boundaries in the chicken and finch chromosome assemblies [37]. In this way all chicken-specific and zebra finch-specific EBRs identified at 300 Kb resolution were compared directly with genetic maps in chicken and zebra finch, respectively.

We obtained sex-averaged recombination rate estimates for 1 Mb non-overlapping windows by comparing genetic and physical positions of SNPs distributed along the chicken and zebra finch genomes (data from [54,55]). To assess if the recombination rate differed between regions with and without chromosomal breakpoints, we partitioned the recombination data into two classes, one with windows containing at least one breakpoint and one with windows without breakpoints, using the zebra finch and chicken breakpoint data [37]. We applied a non-parametric test (Wilcoxon's rank sum test with continuity correction as implemented in R [71]) to assess the level of significance for the difference in recombination rates between classes. Since the sample size differed considerably between classes (i.e. windows not containing EBRs vastly exceeded those that contained EBRs) we also applied a randomization test in R [71]. We randomly sampled the same number of windows as those containing EBRs in each respective taxon ( $n = 31$  for zebra finch,  $n = 35$  for

chicken) from the entire sample 10,000 times. Lastly, we calculated the average recombination rate in the random sample of windows for each iteration to obtain an expected distribution.

### GO analysis of microchromosomes

In order to ask whether individual microchromosomes were enriched for specific GO categories, whole gene sets for each microchromosome were collated and loaded both into DAVID [72,73] and GOEAST [74,75]. Specifically, Ensembl gene ID data and gene name for each microchromosome were extracted from the BioMart Ensembl Genes 75 Database [76,77], using galGal4 as the dataset. In order to eliminate any “significant” results arising through the presence of multiple copies of genes in the same family being present on the same chromosome, gene families were reduced to a single representative member. Downloaded gene IDs and gene names were then copied into a spreadsheet for further analysis using DAVID and GOEAST. Gene IDs for each microchromosome were uploaded into DAVID Bioinformatics Resources 6.7, using Ensembl Gene ID as the list identifier and subsequently analyzed using the Functional Annotation Clustering tool. Cluster data from each microchromosome gene list output was downloaded into Microsoft Excel and filtered using an enrichment score of 1.3 and above and a *P* value less than 0.05 to edit the list for clusters considered to be significant. BioMart (Ensembl) derived gene names for each microchromosome were also uploaded into GOEAST using *Gallus gallus* as the reference. Batch-gene analysis was performed by GOEAST, and enriched GO term outputs with a *P* value less than 0.05 were considered to be significant. The GO results obtained from GOEAST were downloaded into Microsoft Excel and presented with graphic files created directly from GOEAST for each microchromosome where results were available. Finally, in order to correct for multiple sampling error, an FDR threshold of 0.05 was used.

### Additional file

**Additional file 1: Table S1.** BAC clones used to confirm chromosome paint assignments.

### Abbreviations

BAC: Bacterial artificial chromosome; CAR: Contiguous ancestral region; cM: Centimorgan; CNV: Copy number variation; EBR: Evolutionary breakpoint region; FDR: False discovery rate; FISH: Fluorescent in situ hybridization; GC: Guanine-cytosine; GO: Gene ontology; HSB: Homologous synteny block; Kb: Kilobase; Mb: Megabase; msHSB: Multispecies homologous synteny block; MGR: Multiple Genome Rearrangements; MGRA: Multiple Genomes Rearrangements and Ancestors; MY: Million years; MYA: Million years ago; SD: Standard deviation; SF: Syntenic fragment; SNP: Single nucleotide polymorphism.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

MNR led the project on a day-to-day basis including the reconstruction of chromosome rearrangements; this was also performed by PEL, ROC, BMS and KEF. The FISH experimentation and analysis was performed by PEL, ROC, KEF (chicken, budgerigar) and by YM and CN (ostrich). The Evolution Highway database was created and all the avian species uploaded by MF and DML. GO analysis was performed by ROC, MNR, MF and PH, while analysis of genetic recombination was performed by NB and HE. DKG, DML and DWB conceived the project, which was largely supervised by DKG. DKG wrote the first draft of the manuscript with significant sections written by DML, HE, PE and DWB. All authors had input on subsequent drafts and DKG coordinated the submission process. All authors read and approved the final manuscript.

### Authors' information

Michael N Romanov and Marta Farré, joint first authors.  
Denis M Larkin and Darren K Griffin, joint last and corresponding authors.

### Acknowledgements

The authors would like to thank Alain Vignal and Thomas Faraut of INRA Toulouse (France) for access to the duck chromosome assembly data. This research was funded in part by PL-Grid Infrastructure (DML), Biotechnology and Biological Sciences Research Council BB/K008161 (DML, DKG), BB/K008226/1 (DML), BB/J010170/1 (DML, MF) and a knowledge transfer partnership award (DKG and Cytocell Ltd). The authors are grateful to Malcolm Ferguson-Smith's lab (Cambridge, UK) for producing the flow-sorted chicken microchromosome paints. We also thank Cytocell Ltd (Cambridge, UK) for technical support in FISH technologies.

### Author details

<sup>1</sup>School of Biosciences, University of Kent, Canterbury CT2 7NJ, UK.  
<sup>2</sup>Department of Comparative Biomedical Sciences, Royal Veterinary College, University of London, London NW1 0TU, UK. <sup>3</sup>Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK.  
<sup>4</sup>Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden. <sup>5</sup>Laboratory of Animal Genetics, Department of Applied Molecular Biosciences, Graduate School of Bioagricultural Sciences, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8601, Japan. <sup>6</sup>Department of Natural History Sciences, Faculty of Science, Hokkaido University, Kita 10, Nishi 8, Kita-ku, Sapporo, Hokkaido 060-0810, Japan. <sup>7</sup>Department of Biology, New Mexico State University, Las Cruces, NM 88003, USA. <sup>8</sup>Department of Neurobiology, Duke University Medical Center, Box 3209, Durham, NC 27710, USA. <sup>9</sup>Department of Genomics and Genetics, The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh EH25 9PS, UK.  
<sup>10</sup>Current address: School of Human and Life Sciences, Canterbury Christ Church University, Canterbury, Kent CT1 1QU, UK.

Received: 11 November 2014 Accepted: 27 November 2014

Published: 11 December 2014

### References

1. Lewin HA, Larkin DM, Pontius J, O'Brien SJ: **Every genome sequence needs a good map.** *Genome Res* 2009, **19**:1925–1928.
2. Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW, Ödeen A, Cui J, Zhou Q, Xu L, Pan H, Wang Z, Jin L, Zhang P, Hu H, Yang W, Hu J, Xiao J, Yang Z, Liu Y, Xie Q, Yu H, Lian J, Wen P, Zhang F, Li H, et al: **Comparative genomics reveals insights into avian genome evolution and adaptation.** *Science* 2014, **346**:1311–1320.
3. White BJ, Crandall C, Raveche ES, Hjio JH: **Laboratory mice carrying three pairs of Robertsonian translocations: establishment of a strain and analysis of meiotic segregation.** *Cytogenet Cell Genet* 1978, **21**:113–138.
4. King M: *Species evolution. The Role of Chromosome Change.* Cambridge: Cambridge University Press; 1993.
5. Delneri D, Colson I, Grammenoudi S, Roberts IN, Louis EJ, Oliver SG: **Engineering evolution to study speciation in yeasts.** *Nature* 2003, **422**:68–72.
6. Noor MA, Grams KL, Bertucci LA, Reiland J: **Chromosomal inversions and the reproductive isolation of species.** *Proc Natl Acad Sci U S A* 2001, **98**:12084–12088.

7. Rieseberg LH: **Chromosomal rearrangements and speciation.** *Trends Ecol Evol* 2001, **16**:351–358.
8. Masabanda JS, Burt DW, O'Brien PC, Vignal A, Fillon V, Walsh PS, Cox H, Tempest HG, Smith J, Habermann F, Schmid M, Matsuda Y, Ferguson-Smith MA, Crooijmans RP, Groenen MA, Griffin DK: **Molecular cytogenetic definition of the chicken genome: the first complete avian karyotype.** *Genetics* 2004, **166**:1367–1373.
9. Christidis L: **Aves.** In *Animal Cytogenetics. Volume 4: Chordata 3 B*. Edited by John B, Kayano H, Levan A. Berlin: Gebrüder Borntraeger; 1990.
10. Griffin DK, Haberman F, Masabanda J, O'Brien P, Bagga M, Sazanov A, Smith J, Burt DW, Ferguson-Smith M, Wienberg J: **Micro- and macrochromosome paints generated by flow cytometry and microdissection: tools for mapping the chicken genome.** *Cytogenet Cell Genet* 1999, **87**:278–281.
11. Griffin DK, Robertson LB, Tempest HG, Skinner BM: **The evolution of the avian genome as revealed by comparative molecular cytogenetics.** *Cytogenet Genome Res* 2007, **117**:64–77.
12. Griffin DK, Robertson LB, Tempest HG, Vignal A, Fillon V, Crooijmans RP, Groenen MA, Deryushcheva S, Gaginskaya E, Carré W, Waddington D, Talbot R, Völker M, Masabanda JS, Burt DW: **Whole genome comparative studies between chicken and turkey and their implications for avian genome evolution.** *BMC Genomics* 2008, **9**:168.
13. Skinner BM, Völker M, Ellis M, Griffin DK: **An appraisal of nuclear organisation in interphase embryonic fibroblasts of chicken, turkey and duck.** *Cytogenet Genome Res* 2009, **126**:156–164.
14. Völker M, Backström N, Skinner BM, Langley EJ, Bunzey SK, Ellegren H, Griffin DK: **Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution.** *Genome Res* 2010, **20**:503–511.
15. International Chicken Genome Sequencing Consortium: **Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution.** *Nature* 2004, **432**:695–716.
16. Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Künstner A, Searle S, White S, Vilella AJ, Fairley S, Heger A, Kong L, Ponting CP, Jarvis ED, Mello CV, Minx P, Lovell P, Velho TA, Ferris M, Balakrishnan CN, Sinha S, Blatti C, London SE, Li Y, Lin YC, George J, Sweedler J, Southey B, Gunaratne P, Watson M, et al: **The genome of a songbird.** *Nature* 2010, **464**:757–762.
17. Dalloul RA, Long JA, Zimin AV, Aslam L, Beal K, Blomberg LA, Bouffard P, Burt DW, Crasta O, Crooijmans RP, Cooper K, Coulombe RA, De S, Delany ME, Dodgson JB, Dong JJ, Evans C, Frederickson KM, Flicek P, Florea L, Folkerts O, Groenen MA, Harkins TT, Herrero J, Hoffmann S, Megens HJ, Jiang A, de Jong P, Kaiser P, Kim H, et al: **Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis.** *PLoS Biol* 2010, **8**:e1000475.
18. Burt DW, Bruley C, Dunn IC, Jones CT, Ramage A, Law AS, Morrice DR, Paton IR, Smith J, Windsor D, Sazanov A, Fries R, Waddington D: **The dynamics of chromosome evolution in birds and mammals.** *Nature* 1999, **402**:411–413.
19. Rao M, Morisson M, Faraut T, Bardes S, Fève K, Labarthe E, Fillon V, Huang Y, Li N, Vignal A: **A duck RH panel and its potential for assisting NGS genome assembly.** *BMC Genomics* 2012, **13**:513.
20. Skinner BM, Griffin DK: **Intrachromosomal rearrangements in avian genome evolution: evidence for regions prone to breakpoints.** *Heredity (Edinb)* 2012, **108**:37–41.
21. Lithgow PE, O'Connor R, Smith D, Fonseka G, Al Mutery A, Rathje C, Frodsham R, O'Brien P, Ferguson-Smith MA, Skinner BM, Griffin DK: **Novel tools for characterising inter and intra chromosomal rearrangements in avian microchromosomes.** *Chromosome Res* 2014, **22**:85–97.
22. Larkin DM, Pape G, Donthu R, Auvil L, Welge M, Lewin HA: **Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories.** *Genome Res* 2009, **19**:770–777.
23. Bovine Genome Sequencing and Analysis Consortium, Elisk CG, Tellam RL, Worley KC, Gibbs RA, Muzny DM, Weinstock GM, Adelson DL, Eichler EE, Elmitki L, Guigó R, Hamernik DL, Kappes SM, Lewin HA, Lynn DJ, Nicholas FW, Raymond A, Rijnkels M, Skow LC, Zdobnov EM, Schook L, Womack J, Alioto T, Antonarakis SE, Astashyn A, Chapple CE, Chen HC, Chrast J, Câmara F, Ermolaeva O, et al: **The genome sequence of taurine cattle: a window to ruminant biology and evolution.** *Science* 2009, **324**:522–528.
24. Groenen MA, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, Rogel-Gaillard C, Park C, Milan D, Megens HJ, Li S, Larkin DM, Kim H, Frantz LA, Caccamo M, Ahn H, Aken BL, Anselmo A, Anthon C, Auvil L, Badaoui B, Beattie CW, Bendixen C, Berman D, Blecha F, Blomberg J, Bolund L, Bosse M, Botti S, Bujie Z, et al: **Analyses of pig genomes provide insight into porcine demography and evolution.** *Nature* 2012, **491**:393–398.
25. Sankoff D: **Genome rearrangement with gene families.** *Bioinformatics* 1999, **15**:909–917.
26. Stankiewicz P, Lupski JR: **Molecular-evolutionary mechanisms for genomic disorders.** *Curr Opin Genet Dev* 2002, **12**:312–319.
27. Dumont BL, Payseur BA: **Evolution of the genomic recombination rate in murid rodents.** *Genetics* 2011, **187**:643–657.
28. Garcia-Cruz R, Pacheco S, Brieno MA, Steinberg ER, Mudry MD, Ruiz-Herrera A, Garcia-Caldes M: **A comparative study of the recombination pattern in three species of Platyrrhini monkeys (primates).** *Chromosoma* 2011, **120**:521–530.
29. Segura J, Ferretti L, Ramos-Onsins S, Capilla L, Farré M, Reis F, Oliver-Bonet M, Fernández-Bellón H, Garcia F, Garcia-Caldés M, Robinson TJ, Ruiz-Herrera A: **Evolution of recombination in eutherian mammals: insights into mechanisms that affect recombination rates and crossover interference.** *Proc Biol Sci* 2013, **280**:20131945.
30. Ranz JM, Maurin D, Chan YS, von Grotthuss M, Hillier LW, Roote J, Ashburner M, Bergman CM: **Principles of genome evolution in the *Drosophila melanogaster* species group.** *PLoS Biol* 2007, **5**:e152.
31. Stevison LS, Hoehn KB, Noor MA: **Effects of inversions on within- and between-species recombination and divergence.** *Genome Biol Evol* 2011, **3**:830–841.
32. Eichler EE, Sankoff D: **Structural dynamics of eukaryotic chromosome evolution.** *Science* 2003, **301**:793–797.
33. Singer GA, Lloyd AT, Huminiecki LB, Wolfe KH: **Clusters of co-expressed genes in mammalian genomes are conserved by natural selection.** *Mol Biol Evol* 2005, **22**:767–775.
34. Sémon M, Duret L: **Evolutionary origin and maintenance of coexpressed gene clusters in mammals.** *Mol Biol Evol* 2006, **23**:1715–1723.
35. Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SYW, Faircloth BC, Nabholz B, Howard JT, Suh A, Weber CC, da Fonseca RR, Li J, Zhang F, Li H, Zhou L, Narula N, Liu L, Ganapathy G, Boussau B, Bayzid MS, Zavidovych V, Subramanian S, Gabaldón T, Capella-Gutiérrez S, Huerta-Cepas J, Rekepalli B, Munch K, Schierup M, et al: **Whole-genome analyses resolve the early branches in the tree of life of modern birds.** *Science* 2014, **346**:1320–1331.
36. **Evolution Highway.** [http://evolutionhighway.ncsa.uiuc.edu]
37. Farré M, Narayan J, Slavov G, Auvil L, Li C, Jarvis ED, Burt DW, Griffin DK, Larkin DM: **Chromosome dynamics is associated with ancestral and lineage-specific phenotypes in birds, archosaurs, and other reptiles.** *Proc Natl Acad Sci U S A*.
38. Murphy WJ, Larkin DM, Everts van der Wind A, Bourque G, Tesler G, Auvil L, Beever JE, Chowdhary BP, Galibert F, Gatzke L, Hitte C, Meyers SN, Milan D, Ostrander EA, Pape G, Parker HG, Raudsepp T, Rogatcheva MB, Schook LB, Skow LC, Welge M, Womack JE, O'Brien SJ, Pevzner PA, Lewin HA: **Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps.** *Science* 2005, **309**:613–617.
39. Wallis JW, Aerts J, Groenen MA, Crooijmans RP, Layman D, Graves TA, Scheer DE, Kremitzki C, Fedele MJ, Mudd NK, Cardenas M, Higginbotham J, Carter J, McGrane R, Gaige T, Mead K, Walker J, Albracht D, Davito J, Yang SP, Leong S, Chinwalla A, Sekhon M, Wylie K, Dodgson J, Romanov MN, Cheng H, de Jong PJ, Osoegawa K, Nefedov M, et al: **A physical map of the chicken genome.** *Nature* 2004, **432**:761–764.
40. Romanov MN, Dodgson JB: **Cross-species overgo hybridization and comparative physical mapping within avian genomes.** *Animal Genet* 2006, **37**:397–399.
41. Kim J, Larkin DM, Cai Q, Cai Q, Zhang Y, Ge R-L, Auvil L, Capitanu B, Zhang G, Lewin HA, Ma J: **Reference-assisted chromosome assembly.** *Proc Natl Acad Sci U S A* 2013, **110**:1785–1790.
42. Shetty S, Griffin DK, Graves JA: **Comparative painting reveals strong chromosome homology over 80 million years of bird evolution.** *Chromosome Res* 1999, **7**:289–295.
43. Guttenbach M, Nanda I, Feichtinger W, Masabanda JS, Griffin DK, Schmid M: **Comparative chromosome painting of chicken autosomal paints 1–9 in nine different bird species.** *Cytogenet Genome Res* 2003, **103**:173–184.
44. Nishida-Umehara C, Tsuda Y, Ishijima J, Ando J, Fujiwara A, Matsuda Y, Griffin DK: **The molecular basis of chromosome orthologies and sex chromosomal differentiation in palaeognathous birds.** *Chromosome Res* 2007, **15**:721–734.
45. Morescalchi A: **Phylogenetic aspects of karyological evidence.** In *Major Patterns in Vertebrate Evolution*. Edited by Hecht MK, Goody PC, Hecht BM. New York, London: Plenum Press; 1977:149–167.

46. Morescalchi A, Odierna G, Olmo E: **Karyology of the primitive salamanders, family Hynobiidae.** *Experientia* 1979, **35**:1434–1436.
47. Mengden GA, Stock AD: **Chromosomal evolution in Serpentes: a comparison of G and C chromosome banding patterns of some colubrid and boid genera.** *Chromosoma* 1980, **79**:53–64.
48. King M, Honeycutt R, Contreras N: **Chromosomal repatterning in crocodiles: C, G and N-banding and the in situ hybridization of 18S and 26S rRNA cistrons.** *Genetica* 1986, **70**:191–201.
49. Burt DW: **Origin and evolution of avian microchromosomes.** *Cytogenet Genome Res* 2002, **96**:97–112.
50. Gregory TR: **A bird's-eye view of the C-value enigma: genome size, cell size, and metabolic rate in the class Aves.** *Evolution* 2002, **56**:121–130.
51. Andrews CB, Mackenzie SA, Gregory TR: **Genome size and wing parameters in passerine birds.** *Proc Biol Sci* 2009, **276**:55–61.
52. Bush GL, Case SM, Wilson AC, Patton JL: **Rapid speciation and chromosomal evolution in mammals.** *Proc Natl Acad Sci U S A* 1977, **74**:3942–3946.
53. Fontdevila A, Ruiz A, Ocaña J, Alonso G: **Evolutionary history of *Drosophila buzzatii*. II. How much has chromosomal polymorphism changed in colonization?** *Evolution* 1982, **36**:843–851.
54. Groenen MA, Wahlberg P, Foglio M, Cheng HH, Megens HJ, Crooijmans RP, Besnier F, Lathrop M, Muir WM, Wong GK, Gut I, Andersson L: **A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate.** *Genome Res* 2009, **19**:510–519.
55. Backström N, Forstmeier W, Schielzeth H, Mellenius H, Nam K, Bolund E, Webster MT, Öst T, Schneider M, Kempnaers B, Ellegren H: **The recombination landscape of the zebra finch *Taeniopygia guttata* genome.** *Genome Res* 2010, **20**:485–495.
56. Zhou Q, Zhang J, Bachtrog D, An N, Huang Q, Jarvis ED, Gilbert MT, Zhang G: **Complex evolutionary trajectories of sex chromosomes across bird taxa.** *Science* 2014, **346**:1246338.
57. Witmer LM: **The debate on avian ancestry: phylogeny, function, and fossils.** In *Mesozoic Birds: Above the Heads of Dinosaurs*. Edited by Chiappe LM, Witmer LM. Berkeley: University of California Press; 2002:3–30.
58. Chiappe LM, Dyke GJ: **The early evolutionary history of birds.** *J Paleont Soc Korea* 2006, **22**:133–151.
59. You HL, Lamanna MC, Harris JD, Chiappe LM, O'Connor J, Ji SA, Lü JC, Yuan CX, Li DQ, Zhang X, Lacovara KJ, Dodson P, Ji Q: **A nearly modern amphibious bird from the Early Cretaceous of northwestern China.** *Science* 2006, **312**:1640–1643.
60. Nudds RL, Atterholt J, Wang X, You HL, Dyke GJ: **Locomotor abilities and habitat of the Cretaceous bird *Gansus yumenensis* inferred from limb length proportions.** *J Evol Biol* 2013, **26**:150–154.
61. Clarke JA, Tambussi CP, Noriega JI, Erickson GM, Ketchum RA: **Definitive fossil evidence for the extant avian radiation in the Cretaceous.** *Nature* 2005, **433**:305–308.
62. Gatesy SM: **Locomotor evolution on the line to modern birds.** In *Mesozoic Birds: Above the Heads of Dinosaurs*. Edited by Chiappe LM, Witmer LM. Berkeley: University of California Press; 2002:432–447.
63. Huang Y, Li Y, Burt DW, Chen H, Zhang Y, Qian W, Kim H, Gan S, Zhao Y, Li J, Yi K, Feng H, Zhu P, Li B, Liu Q, Fairley S, Magor KE, Du Z, Hu X, Goodman L, Tafer H, Vignal A, Lee T, Kim KW, Sheng Z, An Y, Searle S, Herrero J, Groenen MA, Crooijmans RP, et al: **The duck genome and transcriptome provide insight into an avian influenza virus reservoir species.** *Nat Genet* 2013, **45**:776–783.
64. **MGRA (Multiple Genome Rearrangements and Ancestors) web server, beta version.** [http://mgra.bioinf.spbau.ru/]
65. Alekseyev MA, Pevzner PA: **Breakpoint graphs and ancestral genome reconstructions.** *Genome Res* 2009, **19**:943–957.
66. Nanda I, Karl E, Volobouev V, Griffin DK, Scharl M, Schmid M: **Extensive gross genomic rearrangements between chicken and Old World vultures (Falconiformes: Accipitridae).** *Cytogenet Genome Res* 2006, **112**:286–295.
67. Nishida C, Ishijima J, Kosaka A, Tanabe H, Habermann FA, Griffin DK, Matsuda Y: **Characterization of chromosome structures of Falconinae (Falconidae, Falconiformes, Aves) by chromosome painting and delineation of chromosome rearrangements during their differentiation.** *Chromosome Res* 2008, **16**:171–181.
68. **MGR: Multiple Genome Rearrangements.** [http://grimm.ucsd.edu/MGR/]
69. Bourque G, Pevzner PA: **Genome-Scale Evolution: Reconstructing Gene Orders in the Ancestral Species.** *Genome Res* 2002, **12**:26–36.
70. Ma J, Zhang L, Suh BB, Raney BJ, Burhans RC, Kent JW, Blanchette M, Haussler D, Miller W: **Reconstructing contiguous regions of an ancestral genome.** *Genome Res* 2006, **16**:1557–1565.
71. **The R Project for Statistical Computing.** [http://www.r-project.org/]
72. **DAVID Functional Annotation Bioinformatics Microarray Analysis.** [http://david.abcc.ncifcrf.gov/]
73. Huang DW, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources.** *Nature Protoc* 2009, **4**:44–57.
74. **Gene Ontology Enrichment Analysis Software Toolkit (GOEAST).** [http://omicslab.genetics.ac.cn/GOEAST/]
75. Zheng Q, Wang XJ: **GOEAST: a web-based software toolkit for Gene Ontology enrichment analysis.** *Nucleic Acids Res* 2008, **36**(Web Server issue):W358–W363.
76. **BioMart – MartView.** [http://www.biomart.org/biomart/martview]
77. Kasprzyk A: **BioMart: driving a paradigm change in biological data management.** *Database (Oxford)* 2011, **2011**:bar049.

doi:10.1186/1471-2164-15-1060

**Cite this article as:** Romanov et al.: Reconstruction of gross avian genome structure, organization and evolution suggests that the chicken lineage most closely resembles the dinosaur avian ancestor. *BMC Genomics* 2014 **15**:1060.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit





# Third Report on Chicken Genes and Chromosomes 2015

Prepared by

Michael Schmid<sup>1</sup> Jacqueline Smith<sup>3</sup> David W. Burt<sup>3</sup> Bronwen L. Aken<sup>5</sup> Parker B. Antin<sup>26</sup> Alan L. Archibald<sup>3</sup>  
Chris Ashwell<sup>29</sup> Perry J. Blackshear<sup>30</sup> Clarissa Boschiero<sup>48</sup> C. Titus Brown<sup>31,32</sup> Shane C. Burgess<sup>27</sup>  
Hans H. Cheng<sup>33</sup> William Chow<sup>6</sup> Derrick J. Coble<sup>40</sup> Amanda Cooksey<sup>28</sup> Richard P.M.A. Crooijmans<sup>17</sup>  
Joana Damas<sup>8</sup> Richard V.N. Davis<sup>34</sup> Dirk-Jan de Koning<sup>18</sup> Mary E. Delany<sup>36</sup> Thomas Derrien<sup>20</sup> Takele T. Desta<sup>10</sup>  
Ian C. Dunn<sup>3</sup> Matthew Dunn<sup>6</sup> Hans Ellegren<sup>19</sup> Lél Eöry<sup>3</sup> Ionas Erb<sup>24</sup> Marta Farré<sup>8</sup> Mario Fasold<sup>2</sup>  
Damarius Fleming<sup>40</sup> Paul Flicek<sup>5</sup> Katie E. Fowler<sup>11</sup> Laure Frésard<sup>22</sup> David P. Froman<sup>38</sup> Valerie Garceau<sup>3</sup>  
Paul P. Gardner<sup>51,52</sup> Almas A. Gheyas<sup>3</sup> Darren K. Griffin<sup>12</sup> Martien A.M. Groenen<sup>17</sup> Thomas Haaf<sup>1</sup>  
Olivier Hanotte<sup>10</sup> Alan Hart<sup>3,13</sup> Julien Häsler<sup>7</sup> S. Blair Hedges<sup>39</sup> Jana Hertel<sup>2</sup> Kerstin Howe<sup>6</sup> Allen Hubbard<sup>35</sup>  
David A. Hume<sup>3</sup> Pete Kaiser<sup>4</sup> Darek Kedra<sup>24</sup> Stephen J. Kemp<sup>14</sup> Christophe Klopp<sup>23</sup> Kalmia E. Kniel<sup>35</sup>  
Richard Kuo<sup>3</sup> Sandrine Lagarrigue<sup>21</sup> Susan J. Lamont<sup>40</sup> Denis M. Larkin<sup>8</sup> Raman A. Lawal<sup>10</sup>  
Sarah M. Markland<sup>35</sup> Fiona McCarthy<sup>28</sup> Heather A. McCormack<sup>3</sup> Marla C. McPherson<sup>36</sup> Akira Motegi<sup>53</sup>  
Stefan A. Muljo<sup>41</sup> Andrea Münsterberg<sup>15</sup> Rishi Nag<sup>5</sup> Indrajit Nanda<sup>1</sup> Michael Neuberger<sup>7</sup> Anne Nitsche<sup>2</sup>  
Cedric Notredame<sup>24</sup> Harry Noyes<sup>14</sup> Rebecca O'Connor<sup>12</sup> Elizabeth A. O'Hare<sup>43</sup> Andrew J. Oler<sup>42</sup>  
Sheila C. Ommeh<sup>49</sup> Helio Pais<sup>16</sup> Michael Persia<sup>44</sup> Frédérique Pitel<sup>22</sup> Likit Preeyanon<sup>31</sup> Pablo Prieto Barja<sup>24</sup>  
Elizabeth M. Pritchett<sup>35</sup> Douglas D. Rhoads<sup>45</sup> Charmaine M. Robinson<sup>37</sup> Michael N. Romanov<sup>12</sup>  
Max Rothschild<sup>40</sup> Pierre-François Roux<sup>21</sup> Carl J. Schmidt<sup>35</sup> Alisa-Sophia Schneider<sup>1</sup> Matt Schwartz<sup>46</sup>  
Steve M. Searle<sup>6</sup> Michael A. Skinner<sup>9</sup> Craig A. Smith<sup>50</sup> Peter F. Stadler<sup>2</sup> Tammy E. Steeves<sup>51</sup> Claus Steinlein<sup>1</sup>  
Liang Sun<sup>35</sup> Minoru Takata<sup>54</sup> Igor Ulitsky<sup>25</sup> Qing Wang<sup>35</sup> Ying Wang<sup>36</sup> Wesley C. Warren<sup>27</sup>  
Jonathan M.D. Wood<sup>6</sup> David Wragg<sup>22</sup> Huaijun Zhou<sup>36</sup>

<sup>1</sup>Department of Human Genetics, University of Würzburg, Würzburg, and <sup>2</sup>Bioinformatics Group, Department of Computer Science and Interdisciplinary Centre for Bioinformatics, University of Leipzig, Leipzig, Germany; <sup>3</sup>Division of Genetics and Genomics, The Roslin Institute and R(D)SVS, and <sup>4</sup>Division of Infection and Immunity, The Roslin Institute, University of Edinburgh, Edinburgh, <sup>5</sup>European Bioinformatics Institute, EMBL, and <sup>6</sup>Wellcome Trust Sanger Institute, Hinxton, <sup>7</sup>Medical Research Council Laboratory of Molecular Biology, Cambridge, <sup>8</sup>Department of Comparative Biomedical Sciences, Royal Veterinary College, University of London, and <sup>9</sup>Section of Virology, Department of Medicine, Imperial College London, London, <sup>10</sup>School of Life Sciences, University of Nottingham, Nottingham,

Supported by Biotechnology and Biological Sciences Research Council (BBSRC); USDA Agriculture and Food Research Initiative Competitive Grant; NIH Intramural Research Program of the National Institute of Allergy and Infectious Diseases; European Union FP-7 project QUANTOMICS; Deutsche Forschungsgemeinschaft (DFG); Swiss National Science Foundation; Lady Tata Memorial Trust; French 'Agence Nationale de la Recherche' EpiBird; Intramural Research Program of the NIEHS, NIH; Department for International Development (DFID).

Correspondence to:  
Michael Schmid  
Department of Human Genetics, University of Würzburg  
Biozentrum, Am Hubland  
DE-97074 Würzburg (Germany)  
E-Mail m.schmid@biozentrum.uni-wuerzburg.de

Jacqueline Smith  
Division of Genetics and Genomics  
The Roslin Institute, University of Edinburgh  
Edinburgh EH25 9RG (UK)  
E-Mail jacqueline.smith@roslin.ed.ac.uk

David W. Burt  
Division of Genetics and Genomics  
The Roslin Institute, University of Edinburgh  
Edinburgh EH25 9RG (UK)  
E-Mail dave.burt@roslin.ed.ac.uk

genomes for the traces of tooth-related genes recovered multiple pseudogene fossils of enamel and dentin genes with multiple frame-shift or exon-deletion mutations. The majority of these mutations differ in diverse avian genomes, suggesting that they occurred independently in evolution. However, all birds analyzed by the Avian Genome Consortium shared the same deletions in 4 enamel genes (*ENAM*, *AMELX*, *AMTN*, and *MMP20*) and 1 dentin-related gene (*DSPP*) suggesting that the common ancestor of all birds likely had no mineralized teeth [Meredith et al., 2014].

Another amazing feature of birds is their most advanced vertebrate visual system. They exhibit the ability to distinguish colors over a wider range of wavelengths than mammals. Unlike mammals, birds likely have retained the ancestral tetrapod set of cones [Zhang G et al., 2014b]. For the majority of vertebrate visual opsin genes, birds had a higher number of copies compared to mammals. The number of opsin gene classes (4) found in most birds suggests that birds are most likely tetrachromatic [Zhang G et al., 2014b], with the exception of penguins who had only 3 classes of opsin genes [Li C et al., 2014] suggesting a 3-chromatic vision, consistent with earlier observations in aquatic mammals who also lost 1 or 2 of cone pigments [Newman and Robinson, 2005].

In conclusion, the comparative analysis of 48 avian genomes proved to be a powerful tool to reveal multiple signatures of genome adaptations related to avian ability to fly. The evolutionary stability of avian karyotypes is likely related to the reduction of transposable and other repetitive sequences in avian genomes. Avian-specific segmental deletions of gene paralogs together with shorter genes and intergenic regions made gene regulation fast and energy-efficient. The skeleton modifications that resulted in a smaller number of light-weight bones were accompanied by accelerated evolution of genes involved in ossification. Avian genomes tend to show relatively small variation in regulatory gene sequences compared to mammals reflecting the high degree of adaptation and specialization of bird genomes, probably inherited from their dinosaur ancestor.

### Avian Cytogenetics Goes Functional

(Prepared by D.K. Griffin, M.N. Romanov, R. O'Connor, K.E. Fowler, and D.M. Larkin)

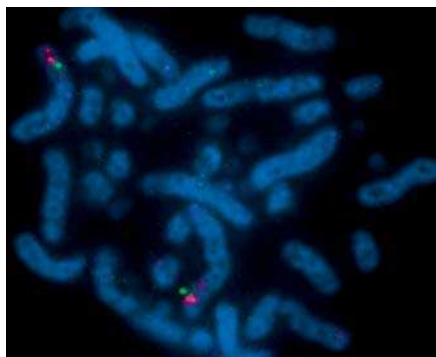
It is now over 10 years since the first avian genome [International Chicken Genome Sequencing Consortium, 2004] and the first complete avian karyotype [Ma-

sabanda et al., 2004] were both published; however, until 2014, avian cytogenetics has focused heavily on descriptive studies [e.g. Griffin et al., 2007, 2008; Skinner et al., 2009; Völker et al., 2010] with less attention to its functional relevance. Last year, however, saw 2 landmark efforts in the chromosomal studies of birds: a special issue of *Chromosome Research* in April and the announcement of recently completed sequences of multiple new avian genomes in *Science* and the *BMC* journals (taking the total number sequenced to over 50) in December. Studying the chromosomes of birds is, perhaps for the first time, telling us more about avian biology, function and evolution than it ever has.

### What Do We Know So Far? Karyotypic Stability

The near-unique nature of the avian karyotype has remained a consistently reported feature of bird biology since the first chromosome preparations were made. Although many animal groups have microchromosomes, the small size and abundant number of chromosomes in avian species set birds apart genomically from other vertebrate groups. To the best of our knowledge, there are over 1,000 published avian karyotypes, most comprehensively summarized by Christidis [1990], with several hundred added since this review. All of these karyotypes are partial however, with usually only 5–10 pairs of chromosomes easily distinguished, and the rest homogeneously classified. Moreover, the vast majority of karyotypes hardly differ from each another, with rare exceptions including the stone curlew (*Burhinus oedicephalus*;  $2n = 40$ ), the beach thick knee (*Esacus magnirostris*;  $2n = 40$ ), several hornbills ( $2n = 42$ ), kingfishers and hoopoes (*Upupa epops*;  $2n > 120$ ) at each end of the numerical spectrum [Christidis, 1990]. Indeed, even since the advent of zoolish, the identification of an interchromosomal rearrangement in a bird is a relatively uncommon event [Griffin et al., 2007].

Central to our understanding of avian biology and evolution is establishing the reasons *why* avian karyotypes are evidently so stable. Clues to such an enquiry might lie in those rare exceptions to the rule. For instance, the Falconiformes (falcons, etc.) and Psittaciformes (parrots, etc.) have noticeably undergone numerous evolutionary changes. Moreover, it is noteworthy that when interchromosomal change occurs, it tends to recur. The best example of this is a fusion of the ancestral chromosomes 4 and 10; an event that appears to have occurred independently throughout evolution in chicken (*Gallus gallus*), greylag goose (*Anser anser*), collared dove (*Streptopelia decaocto*) and probably other species also [Griffin et al., 2007]. In this review, we



**Fig. 4.** FISH of 2 BACs for chicken microchromosome 19 (green, p arm; red, q arm) to peregrine falcon (*Falco peregrinus*) chromosomes. A fusion to a falcon macrochromosome is apparent [our unpubl. data].

examine some of the latest tools and preliminary solutions that are being used to understand the underlying mechanisms that lead to chromosome rearrangements in birds (and in eukaryotes in general).

If we accept that interchromosomal change occurs only rarely in birds, then it is reasonable to assume that this happens usually only when there is an adaptive value to doing so. In most species, phenotypic diversity is usually associated with wholesale changes in karyotype structure. Aves as a phylogenetic class underwent a series of rapid speciation events beginning ~65 million years ago (Mya) and ending ~50 Mya. Chromosomal change is usually a cause or consequence of speciation (i.e. a species barrier), but until recently, the microchromosomes that constitute the majority of the avian karyotype, have not been amenable to study. The latest studies, however, have paved the way for a flurry of research activity that not only describes the avian karyotype in more detail, but might also provide functional clues as to its nature.

#### *New Molecular Cytogenetic Tools*

Lithgow et al. [2014] produced a set of chromosome paints and bacterial artificial chromosomes (BACs) that will start the process of characterizing the microchromosomes and their changes over evolutionary time. They reported the development of chicken microchromosomal paint pools and generation of pairs of specific microchromosome BAC clones with some success in zoo-FISH experiments. For instance, they detected a fusion of the ancestral chicken chromosome 23 orthologue to a macrochromosome in gyrfalcon (*Falco rusticolus*). A FISH image of BACs hybridized to peregrine falcon (*Falco peregrinus*) chromosomes [unpubl. data] is shown in figure 4.

McPherson et al. [2014] examined the Japanese quail (*Coturnix japonica*). Comparing chicken and turkey BAC clones on mitotic and meiotic chromosomes, they demonstrated that high-resolution FISH is practicable. Ishishita et al. [2014] also assessed the distribution of centromeric repetitive sequences on both micro- and macrochromosomes. It is therefore now possible to achieve full, high-resolution characterization of all avian chromosomes in all species studied, including the elusive chromosome 16 and the D-group (smallest) chromosomes. There are several current strategies to fill the gaps; one of these is by the use of PacBio, a novel single-molecule real-time sequencing platform, targeting the sequence of smaller chromosomes using sorted chromosome preps, and assembling contigs into scaffolds and super-scaffolds from optical maps [Ganapathy et al., 2014].

#### *What Have Sequence Assemblies Taught Us?*

The progress of genome assembly in birds has been slow in comparison to other animal groups such as mammals. Following chicken [International Chicken Genome Sequencing Consortium, 2004], it took a further 6 years until the second and third avian genome sequences were published, namely those of the zebra finch (*Taeniopygia guttata*, a model for neurological function, especially learned vocalization) [Warren et al., 2010] and turkey (*Meleagris gallopavo*) [Dalloul et al., 2010]. More recently, the Pekin duck (*Anas platyrhynchos*) [Huang et al., 2013] was added along with 2 falcon species (*Falco peregrinus* and *F. cherrug*) [Zhan et al., 2013] and many others (table 4). The availability of these assembled genomes provided the opportunity for comparative genomics at a chromosomal level. In 2010, we made the first comparison of 2 species using genome assembly information from the macrochromosomes [Völker et al., 2010]. A similar comparison more recently was made in chicken compared to duck [Rao et al., 2012], and then a 3-way comparison (allowing studies of the direction of change) in chicken, turkey and zebra finch [Skinner and Griffin, 2012; Lithgow et al., 2014]. The principal features of chromosomal change in birds are homologous synteny blocks (HSBs), which are demarked by evolutionary breakpoint regions (EBRs). While analyzing these features, some general patterns have started to emerge. The first is that, although interchromosomal change is rare, intrachromosomal changes are commonplace. Breakpoint reuse is also commonplace, significantly more so than in mammals, and there is some evidence of an association between chromosomal breakage and non-allelic homologous recombination (NAHR) [Völker et al., 2010].

Zhang G et al. [2014b] used a whole-genome shotgun strategy to generate new whole-genome sequences from 45 bird species representing many of the major clades and at least 1 representative from over 90% of all avian orders. Around 20 species had a high (50× or greater) coverage and these were the subjects of further cytogenetic studies. These included the common ostrich (*Struthio camelus*) and the budgerigar (*Melopsittacus undulatus*), which were further assembled using data from optical mapping experiments [Ganapathy et al., 2014]. This had the effect of significantly increasing the assembly's N50 scaffold sizes to around 15 Mb, and these were subsequently used with those already assembled by chromosome (chicken, turkey, zebra finch, and duck). Romanov et al. [2014] made use of novel whole-genome sequence information from 21 avian genome sequences available on an interactive browser (Evolution Highway). By focusing on the 6 best-assembled genomes (chicken, turkey, duck, zebra finch, ostrich, and budgerigar), a putative karyotype of the avian ancestor (probably a bipedal feathered dinosaur) was assembled for each chromosome. The evolutionary events were reconstructed that led to each of the 6 species' genome organization. Intra- and interchromosomal changes appear best explained most parsimoniously by a series of inversions and translocations with common breakpoint reuse. Microchromosomes represent conserved blocks of synteny in most of the 21 species, and a series of interchromosomal changes in the ostrich were also described that would not have been predicted by karyotype analysis alone. These results suggest that mechanisms exist to preserve a static overall avian karyotype/genomic structure, including the microchromosomes, with rare interchromosomal change (e.g. in ostrich and budgerigar lineages); this is discussed in depth in the next section. Of the species examined, it seemed that chicken had the least number of chromosomal rearrangements compared to the dinosaur ancestor. From Evolution Highway it is also possible to assess rates of chromosomal evolution in birds. Zhang G et al. [2014b] suggest that birds have a lower chromosomal rearrangement rate than mammals but nonetheless can undergo 'bursts' of rearrangement, e.g. during the evolution of vocal learning. This finding corroborates those of Romanov et al. [2014] that identified the zebra finch and budgerigar as the 2 species with the most chromosomal rearrangements from the avian ancestor.

If we accept that chicken and its galliform relatives underwent the least number of chromosomal changes whilst diverging from the ancestral bird, we also must consider

whether they also have undergone the fewest phenotypic changes. In other words: is the dinosaur avian ancestor more like a land fowl than any other bird? The most ancient near-certain fossil representative of modern birds (Neornithes) was almost certainly aquatic (for example, *Vegavis*, a genus of birds from the Late Cretaceous epoch) and has been identified as a Galloanseres. Indeed, the earliest known bird-like creatures in the fossil record (e.g. the Ornithurae *Gansus*) were either fully aquatic or at least amphibious, and it has been suggested that, due to the fact that they had webbed feet (as well as other traits), they were more like ducks [Romanov et al., 2014]. On the other hand, most authors agree that the dinosaur ancestors of birds were terrestrial, feathered, bipedal, relatively small and with limited flying ability – not unlike a chicken. At best we can determine therefore, the ancestral birds were most likely more phenotypically associated with the Galloanseres, and the confusion of whether they were more akin to water- or land fowl may be due to interpretations based on depositional sampling biases, limited understanding of functional anatomy, and whether the individuals that have been discovered are actually fully representative of the groups to which they belonged. Chromosomal evidence provides an independent record of the functional material of inheritance in living birds and, as such, can complement a fossil record that is always likely to be incomplete.

Of all species studied so far, it seems clear that the rearrangement of chromosomes is non-random [Pevzner and Tesler, 2003; Larkin et al., 2009]. The reasons for this non-random nature warrant deeper investigation. According to mammalian evidence, evolutionarily conserved HSBs appear to evolve in different ways from the dynamic and ever-changing EBRs; whether this is true of birds remains to be seen. In mammals, chromosomal breakpoints are correlated with sequences of segmentally duplicated or repetitive DNA [Bovine Genome Sequencing and Analysis Consortium et al., 2009; Larkin et al., 2009; Groenen et al., 2012; Ruiz-Herrera et al., 2012], and species-specific EBRs are correlated with regions enriched for transposable elements (TEs) [Bovine Genome Sequencing and Analysis Consortium et al., 2009; Groenen et al., 2012]. In mammals, EBRs and HSBs largely contain genes with notably different functional ontologies, e.g. organismal development in HSBs [Larkin et al., 2009] and lineage-specific biology and adaptive features in EBRs [Bovine Genome Sequencing and Analysis Consortium et al., 2009; Larkin et al., 2009; Groenen et al., 2012]. It has been suggested therefore that chromosome rearrangements and the respective gene ontologies con-



tained within HSBs and EBRs help to explain lineage-specific phenotypes in mammals. Mammalian and avian genomes are very different however (not least because of the interchromosomal stability of avian genomes), and thus the question remains about whether the patterns that have been observed in mammals will apply to birds also. Birds have less repetitive DNA through the elimination of repetitive sequences [International Chicken Genome Sequencing Consortium, 2004; Shedlock, 2006; Zhang G et al., 2014b] so that the avian genome is constrained by size, primarily because of gene loss as well as lineage-specific erosion of repetitive elements and large segmental deletions. In addition to their karyotypic stability, bird genomes also have a very high degree of evolutionary stasis at nucleotide sequence and gene synteny levels. Nonetheless, one of the key findings was the detection of non-neutral evolutionary changes in functional genes as well as non-coding regions. Many of these changes coincide with adaptations to different lifestyles and niches and display homoplasy [Zhang G et al., 2014b].

The non-random nature of chromosome rearrangement in birds, the reasons for the apparent interchromosomal (but not intrachromosomal) stability of avian karyotypes (see next section), the role of TEs and NAHR, the relationship to phenotype, the question of whether spatial organization of ancestral gene networks is maintained in bird and other reptile lineages, and the question of whether lineage-specific EBRs alter gene order in networks that had adaptive value, all require further investigation. Harnessing the data from over 50 avian genomes (undoubtedly with many more on the way) and employing tools such as Evolution Highway will give us unprecedented insight into avian chromosome evolution and its relationship to avian biology.

#### *Why Is the Avian Karyotype Structure Conserved Inter- but Not Intrachromosomally?*

Burt's 'fission-fusion' hypothesis suggested that most avian microchromosomes became fixed in the common dinosaur ancestor with a karyotype of  $2n \approx 60$  including 20 microchromosome pairs [Burt, 2002]. The remainder, including the smallest, probably was created by further fission. Romanov et al. [2014] suggested that a basic pattern of  $2n = 80$  (~30 microchromosome pairs) was fixed before the Palaeognathae-Neognathae divergence 100 Mya. The subsequent paucity of intermicrochromosomal rearrangements between most Neognathae indicates an evolutionary advantage either to retaining this pattern or a lack of opportunity for change. For instance, an explanation for such evolutionary stasis might be that the un-

derlying mutational mechanisms of chromosomal changes are fundamentally different in birds compared to other amniotes through a lack of adaptive value, rather than purifying selection, slowing down the rate of change. Much of this could be explained, in part, by a paucity of copy number variants (CNVs; including segmental duplications), recombination hotspots, TEs and/or endogenous retroviruses; however, this would not explain why interchromosomal change is rare but intrachromosomal change is common, particularly in groups that have undergone rapid speciation such as Passeriformes.

The rate of chromosome rearrangement (and subsequent speciation) depends on: (1) the mutation rate and (2) the fixation rate [Burt et al., 1999]. The first of these is related to the frequency of homologous sites [Burt, 2002]. Repeat structures in general (e.g. CNVs), and TEs in particular, provide substrates for chromosomal rearrangement. In a genome constrained by size, the opportunity for mutation is reduced and only fission (or intrachromosomal change, e.g. inversion) can occur. This provides an explanation why (1) avian genomes are more fragmented than any other vertebrate (birds have the most chromosomes) and (2) why there have been fewer interchromosomal rearrangements. There might also be advantages to retaining multiple chromosomes in a karyotype through the generation of variation, the driver of natural selection. That is, a karyotype with more chromosomes leads to a greater number of genetic variants that the gametes produce and an increase in recombination rate due to the fact that there needs to be at least 1 obligatory chiasma per chromosome. Burt [2002] proposed that a higher recombination rate has also led to the features that we most associate with microchromosomes (high GC content, low repeats, high gene density, etc.) and resulted in the formation and fixation of the archetypal avian karyotype with both macro- and microchromosomes and little interchromosomal rearrangement. Such a constraint, however, does not preclude rearrangement within the individual chromosomes. Romanov et al. [2014] and King [1995] argue that an increase in intrachromosomal rearrangement correlates with bursts of speciation in birds, perhaps mediated by an increase in localized repeat content.

Some birds nonetheless have a significantly different karyotype from the standard  $2n \approx 80$ . This can occur within one closely related group, e.g. Adélie penguin (*Pygoscelis adeliae*;  $2n = 96$ ) and the emperor penguin (*Aptenodytes forsteri*;  $2n = 72$ ) (but both associated with high degrees of intermicrochromosomal rearrangement), thereby suggesting that similar mechanisms can both re-

duce or increase chromosome number in relatively short time frames. Comparisons of chromosomal change in the zebra finch and the budgerigar suggest that rearrangement rates are similarly high in both groups to which they belong (Passeriformes and Psittaciformes, respectively) but that the latter is capable of fixing interchromosomal rearrangements, while the former is not. The mechanisms underpinning these differences are, as yet, unknown, but studies of the gene ontology terms of species-specific EBRs might provide clues. As more avian genomes with better assemblies are analyzed, this may indicate adaptive phenotypic features associated with specific gene ontologies typical of individual orders, families or genera.

### *The Sex Chromosomes*

Worthy of especial consideration is the conserved sex chromosome ZW system that is present in all birds apart from the Palaeognathae. Their independent origin from the XY system does not escape the fact that similar mechanisms appear to have run in parallel, for instance genes on the Z chromosome (like the mammalian X) have undergone selection for male-advantage functions. Like the Y chromosome, the W is small (albeit medium-sized by avian standards), heterochromatic and gene poor. Graves [2014] suggests that the W chromosome is at a more advanced stage of differentiation than the Y chromosome as it has accumulated more LINEs and lost more genes during its evolution. Pokorná et al. [2014] considered multiple sex chromosomes and meiotic drive in a range of amniotes. This study noted that the single ZW system in birds contrasts with that of other reptile and amniote groups; they raised a very exciting hypothesis that this contrast may possibly be related to the differential involvement of sex-specific sex chromosomes in female meiosis (females being the heterogametic sex). Early in the assembly of the chicken genome, the quality of the build of both the Z and W sex chromosomes was very poor, and limited studies existed on sex determination. Since this, the Z chromosome was painstakingly assembled and sequenced BAC by BAC [Bellott et al., 2010], and is now one of the best-assembled chromosomes in the chicken genome. The same is now expected for the W sex chromosome, which currently is very poorly assembled [Chen et al., 2012]. Zhou Q et al. [2014] conclude that the ancestral sex chromosome organization is closer to that of the Palaeognathae (ostrich and emu) and demonstrated that there is less degradation of the sex chromosomes and a closer synteny with non-avian reptile species.

### *Copy Number Variation*

Redon et al. [2006] first highlighted the impact of CNV in the human genome. This seminal study heralded a new era in cytogenetics and has subsequently been applied to many other species and groups including birds. Skinner et al. [2014] provided a global overview of apparent cross-species CNVs in birds using cross-species array-CGH. Griffin and Burt [2014] pointed out issues of definition in that 'copy number variation', strictly speaking, refers to polymorphisms *within a species*. The question arises therefore whether results of cross-species array-CGH represent genuine variation in copies of orthologous genes between species. Skinner et al. [2014] stated that 'difference in gene copy number between species is a question of gene duplication, segmental duplications etc. and may be driven by expansion and contraction of paralogs within different gene families.' Nonetheless, this paper provided a broad appraisal of apparent cross-species CNVs in 16 avian species. Microchromosomes appear to have more apparent CNVs than macrochromosomes. Indeed, in species with microchromosomal fusions such as Falconiformes, the fused 'former microchromosomes' still retained their ancestral features such as a higher degree of cross-species CNVs. Skinner et al. [2014] reported that ~50% of the apparent cross-species CNVs overlap with known chicken-specific CNVs. In terms of gene ontology, there appears to be a general enrichment in immune response and antigen presentation genes as well as 5 CNV regions perfectly correlated with the unique loss of sexual dichromatism. More specifically, there were also suggestions of CNVs involved in diet in turkey (proteolytic digestion/degradation of trypsin inhibitors), and correlation of the unique migratory behaviour of common quail among fowl through the following genes: *OBSCN* associated with hypertrophy of myofibrils, and *MAPK8IP3* implicated in respiratory gaseous exchange [Skinner et al., 2014]. There were also suggestions of an association with muscle activity in falcons through the gain of *MYOZ3*, preferentially expressed in fast-twitch myofibers and skeletal muscle and an association between immune function in the common quail (*Coturnix coturnix*) and silver pheasant (*Lophura nycthemera*) (*LEAP2* and *ITCH*) as well as homeotic genes in common pheasant and California quail (*SCML2* and *DLX5*). Finally, Skinner et al. [2014] identified cross-species CNVs associated with brain development and neuronal function in turkey (e.g. loss of *CTXN1*), common quail (gain of *LRFN5*) and duck (e.g. *DLGAP2*).

## Conclusions

The most recent advances in avian cytogenetics have culminated in great promise not only for the study of bird karyotypes, but also for providing insight into the mechanisms of chromosome evolution in general. New avenues for investigation include gene regulation; for instance, it will become necessary to map accurately the physical location of polyadenylation and transcription start sites, important reference points that define promoters and post-transcriptional regulation. It will also become possible to sequence full-length transcripts, to allow accurate identification of alternate splicing events and their controlling elements. The ENCODE (Encyclopedia of DNA Elements) project has helped to define functional elements of the human genome, including those aforementioned as well as other chromatin signals, e.g. active chromatin, enhancers, insulators, methylation domains, etc. An effort of agENCODE is underway to include agriculturally important birds such as chicken, turkey, duck, quail, and perhaps ostrich. The study of cytogenetics will be essential here in helping to define higher-order structures in nuclear organization that show regulatory interactions within and between chromosomes. Finally, reconstruction of evolutionary events allows us to study genome organization and function not only in extant but, by extrapolation, in extinct species also. Reconstruction of avian-reptilian ancestral karyotypes will allow us to define chromosomal rearrangements in long-dead species that have captured the public imagination. Here be dragons!

### Hypermethylated Chromosome Regions in Chicken and Other Birds

(Prepared by M. Schmid, C. Steinlein, A.-S. Schneider, I. Nanda, and T. Haaf)

The advent of specific antibodies against the different nucleosides and nucleotides has promoted direct cytogenetic analyses of the various DNA classes along eukaryote chromosomes. These antibodies were first produced by the group of Bernard F. Erlanger some decades ago [Erlanger and Beiser, 1964; Garro et al., 1968; Sawicki et al., 1971; Erlanger et al., 1972]. They were produced by immunizing rabbits to bovine serum albumin (BSA) conjugated to one of the DNA bases. The antibodies are reactive with the BSA conjugate used to induce them and also with single-stranded DNA [Erlanger and Beiser, 1964]. They are highly specific for the base and show little or no cross-reaction with the other bases. Over the years, a se-

ries of such polyclonal antibodies were produced, with specificities for a number of nucleosides, nucleotides and dinucleotides [Dev et al., 1972; Erlanger et al., 1972; Miller, 1973]. In the early 1990s, the first monoclonal antibodies against 5-methylcytosine (5-MeC) and other modified nucleosides were produced [Reynaud et al., 1992] and subsequently used for chromosome staining [Barbin et al., 1994; Miniou et al., 1994; Montpellier et al., 1994; Bernardino et al., 1996].

Of special interest were antisera specific for 5-MeC which were initially applied by the group of Orlando J. Miller to the chromosomes of several mammalian species, including human, chimpanzee, gorilla, cattle, mouse, and kangaroo rat [Miller et al., 1974; Schreck et al., 1974, 1977; Schnedl et al., 1975, 1976]. Using an immunofluorescence technique and anti-5-MeC antibodies, they showed that methylated DNA can be detected in fixed metaphase chromosomes after they have been UV-irradiated to generate regions of single-stranded DNA. In these species, the methylated regions corresponded to the locations of repetitive DNA, i.e. to the heterochromatic regions of all or a subset of the chromosomes in the karyotypes. Subsequently, this technique was applied to chromosomes of further mammalian species [Vasilikaki-Baker and Nishioka, 1983; Bernardino et al., 2000] and to human chromosomes [Barbin et al., 1994; Montpellier et al., 1994; Bernardino et al., 1996; Kokalj-Vokac et al., 1998], even including cases of inherited chromosome aberrations [Breg et al., 1974] and leukemia cell lines [Bensaada et al., 1998].

With one exception [Grützner et al., 2001], no immunofluorescence studies on the distribution of hypermethylated regions in bird chromosomes have been published. The present report is a brief summary of the results obtained for avian chromosomes in an ongoing project on the hypermethylation patterns in vertebrate chromosomes [Schmid et al., in preparation].

Mitotic chromosomes of 13 species from 7 orders, belonging to both modern (Neognathae) and primitive (Palaeognathae) birds (table 13), were prepared from embryonic or skin fibroblast cell cultures following standard techniques (colcemid treatment, exposure to hypotonic solution, fixation with methanol:acetic acid). Hypermethylated DNA was detected by indirect immunofluorescence using a monoclonal antibody against 5-MeC. In double-stranded DNA, the methyl groups are hidden in the phosphodiester backbone of the double helix and not accessible to the antibody. The anti-5-MeC antibody recognizes and binds to its target only if the DNA is in the single-stranded configuration. Therefore, the slides with

1 **Upgrading short read animal genome assemblies to chromosome level using**  
2 **comparative genomics and a universal probe set**

3

4 Joana Damas<sup>1\*</sup>, Rebecca O'Connor<sup>2\*</sup>, Marta Farré<sup>1</sup>, Vasileios Panagiotis E. Lenis<sup>3</sup>, Henry  
5 Martell<sup>2</sup>, Anjali Mandawala<sup>2,4</sup>, Katie Fowler<sup>4</sup>, Sunitha Joseph<sup>2</sup>, Martin T. Swain<sup>3</sup>, Darren K.  
6 Griffin<sup>2\*\*</sup>, Denis M. Larkin<sup>1\*\*</sup>

7

8 <sup>1</sup>Department of Comparative Biomedical Sciences, Royal Veterinary College, Royal College  
9 Street, University of London, London, NW1 0TU, UK

10 <sup>2</sup>School of Biosciences, University of Kent, Canterbury, CT2 7NY, UK

11 <sup>3</sup>Institute of Biological, Environmental and Rural Sciences, Aberystwyth University,  
12 Aberystwyth, SY23 3DA, UK

13 <sup>4</sup>School of Human and Life Sciences, Canterbury Christ Church University, Canterbury, CT1  
14 1QU, UK

15

16

17 \*Joint first authors

18 \*\*Joint last authors

19

20 <sup>§</sup>Corresponding author: Denis M. Larkin, Department of Comparative Biomedical Sciences,  
21 Royal Veterinary College, Royal College Street, NW1 0TU, London, UK. Phone number:  
22 +44-207-121-1906. E-mail: dlarkin@rvc.ac.uk

23

24 Running title: Building chromosome-level assemblies for animals

25

26 Keywords: *de novo* assembly, avian chromosome assembly, evolutionary breakpoint  
27 regions, peregrine falcon, pigeon, chicken



1 **ABSTRACT**

2 Most recent initiatives to sequence and assemble new species' genomes *de-novo* fail to  
3 achieve the ultimate endpoint to produce a series of contigs, each representing one whole  
4 chromosome. Even the best-assembled genomes (using contemporary technologies) consist  
5 of sub-chromosomal sized scaffolds. To circumvent this problem, we developed a novel  
6 approach that combines computational algorithms to merge scaffolds into chromosomal  
7 fragments, scaffold verification by PCR and physical mapping to chromosomes. Multi-  
8 genome-alignment-guided probe selection led to the development of a set of universal avian  
9 BAC clones that permit rapid anchoring of multiple scaffold loci to chromosomes on all avian  
10 genomes. As proof of principle we assembled genomes of the pigeon (*Columbia livia*) and  
11 peregrine falcon (*Falco peregrinus*) to chromosome level comparable, in continuity, to avian  
12 reference genomes. Both species are of interest for breeding, cultural, food and/or  
13 environmental reasons. Pigeon has a typical avian karyotype ( $2n=80$ ) while falcon ( $2n=50$ ) is  
14 highly rearranged compared to the avian ancestor. Using chromosome breakpoint data, we  
15 established that avian interchromosomal breakpoints appear in the regions of low density of  
16 conserved non-coding elements (CNEs) and that the chromosomal fission sites are further  
17 limited to long CNE "deserts". This corresponds with fission being the rarest type of  
18 rearrangement in avian genome evolution. High-throughput multiple hybridization and rapid  
19 capture strategies using the current BAC set provide the basis for assembling numerous avian  
20 (and possibly other reptilian) species while the overall strategy for scaffold assembly and  
21 mapping provides the basis for an approach that could be applied to any animal genome.

## 1 INTRODUCTION

2 The ability to sequence complex animal genomes quickly and cheaply has initiated numerous  
3 genome projects beyond those of agricultural/medical importance (e.g., Hu et al. 2009;  
4 Groenen et al. 2012) and inspired ambitious undertakings to sequence thousands of species  
5 (Zhang et al. 2014a; Koepfli et al. 2015). *De novo* genome assembly efforts ultimately aim to  
6 create a series of contigs, each representing a single chromosome, from p- to q- terminus  
7 (“chromosome level” assembly). Assembling genomes using next generation sequencing  
8 (NGS) technology however typically relies on integration of the NGS data with a pre-existing  
9 chromosome-level reference assembly built with previous sequencing/mapping technologies  
10 (Larkin et al. 2012). Indeed, use of short read NGS data rarely produces assemblies at a  
11 similar level of integrity as those provided by traditional methodologies because of: a) an  
12 inability of NGS to generate long error-free contigs or scaffolds to cover chromosomes  
13 completely; and b) a paucity of inexpensive mapping technologies to upgrade NGS genomes  
14 to chromosome level. Even for projects with sufficient read-depths and long insert libraries,  
15 software algorithms at best, produce sub-chromosomal sized “scaffolds” requiring physical  
16 mapping to assemble chromosomes. Newer technologies such as optical mapping (Teague  
17 et al. 2010) including BioNano (Mak et al. 2016), Dovetail (Putnam et al. 2016), and PacBio  
18 long read sequencing (Rhoads and Au 2015) provide a long-term solution to this problem. To  
19 date, however, such approaches suffer from multiple limitations: for instance, BioNano contigs  
20 do not extend across multiple DNA nick site regions, centromeres or large heterochromatin  
21 blocks while PacBio sequencing requires hundreds of micrograms of high molecular weight  
22 DNA which is often not easy to obtain.

23

24 Bioinformatic approaches, e.g., the Reference-Assisted Chromosome Assembly (RACA; Kim  
25 et al. 2013), were developed to approximate near chromosome-sized fragments for a *de novo*  
26 assembled NGS genome. Their use, however, requires a genome from the same phylogenetic  
27 order of the target species being assembled to chromosomes (Kim et al. 2013), sequencing  
28 of long-insert libraries and, at best, produces sub-chromosome sized predicted chromosome

1 fragments (PCFs) that require further verification and subsequent chromosome assembly.  
2 RACA applied to the Tibetan antelope and blind mole rat genomes significantly improved  
3 continuities of these assemblies but they still contain more than one large PCF for most  
4 chromosomes (Kim et al. 2013; Fang et al. 2014).

5

6 A dearth of chromosome-level assemblies for nearly all newly sequenced genomes limits their  
7 use for critical aspects of evolutionary and applied genomics. Chromosome-level assemblies  
8 are essential for species that are regularly bred (e.g., for food or conservation) because a  
9 known order of DNA markers facilitates establishment of phenotype-to-genotype associations  
10 for gene-assisted selection and breeding (Andersson and Georges 2004). While such  
11 assemblies are established for popular livestock species, they are not available for those  
12 species widely used in developing countries (e.g., camels, yaks, buffalo, ostrich, quail) or  
13 species bred for conservation reasons (e.g., falcons). Chromosome-level information is  
14 essential for addressing basic biological questions pertaining to overall genome (karyotype)  
15 evolution and speciation (Lewin et al. 2009). Karyotype differences between species arise  
16 from DNA aberrations in germ cells that were fixed throughout evolution. These are associated  
17 with repetitive sequences used for non-allelic homologous recombination (NAHR) in  
18 evolutionary breakpoint regions (EBRs) where ancestral chromosomes break and/or combine  
19 in descendant species genomes (Murphy et al. 2005). An alternative theory however, suggests  
20 that proximity of DNA regions in chromatin is the main driver of rearrangements and repetitive  
21 sequences play a minor role (Branco and Pombo 2006). Regardless of the mechanism,  
22 comparisons of multiple animal genomes show that, between EBRs, are evolutionary stable  
23 homologous synteny blocks (HSBs). Our studies in mammals (Larkin et al. 2009) and birds  
24 (Farré et al. 2016) suggest that at least the largest HSBs are maintained non-randomly and  
25 are highly enriched for conserved non-coding elements (CNEs) many of which are gene  
26 regulatory sequences and miRNA (Zhang et al. 2014b). We recently hypothesized that a  
27 higher fraction of elements under negative selection involved in gene regulation and  
28 chromosome structure in avian genomes (~7%) (Zhang et al. 2014b) compared to mammals

1 (~4%) (Lindblad-Toh et al. 2011) could contribute to some avian-specific phenotypes and the  
2 evolutionary stability of most avian karyotypes (Farré et al. 2016). Whilst a high density of  
3 CNEs in avian multi-species (ms)HSBs supports this hypothesis (Farré et al. 2016) a more  
4 definitive answer might be obtained by examining the fate of CNEs in the “interchromosomal  
5 EBRs” (flanking interchromosomal rearrangements) of an avian genome with a highly  
6 rearranged karyotype.

7

8 In this study we focused on two avian genomes. The first, the peregrine falcon (*Falco*  
9 *peregrinus*) has an atypical karyotype ( $2n=50$ ) (Nishida et al. 2008). Its ability to fly at speeds  
10  $>300$  km/h and its enhanced visual acuity make it the fastest predator on Earth (Tucker et al.  
11 1998). A prolonged period of extinction risk due to persecution around the World War II and  
12 secondary poisoning from organochlorine pesticides (e.g., DDT) in the 1950s-60s (Ferguson-  
13 Lees and Christie 2005) led to its placement on the CITES list of endangered species. The  
14 second avian genome that was focused on here, the pigeon (*Columba livia*) has a typical avian  
15 karyotype ( $2n=80$ ) similar to those of reference avian genomes: chicken, turkey and zebra  
16 finch. Pigeon is one of the earliest examples of domestication in birds (Driscoll et al. 2009)  
17 contemporarily used as food and in sporting circles (Price 2002). Pigeon breeds can vary  
18 significantly in appearance with color, pattern, head crest, body shape, feathers, tails,  
19 vocalization and flight display variations (Price 2002) inspiring considerable interest in  
20 identifying the genetic basis for these variations (Stringham et al. 2012; Shapiro et al. 2013).  
21 For the above reasons, both species genomes were sequenced (Shapiro et al. 2013; Zhan et  
22 al. 2013), however their assemblies are highly fragmented and chromosome-level assemblies  
23 are thus essential.

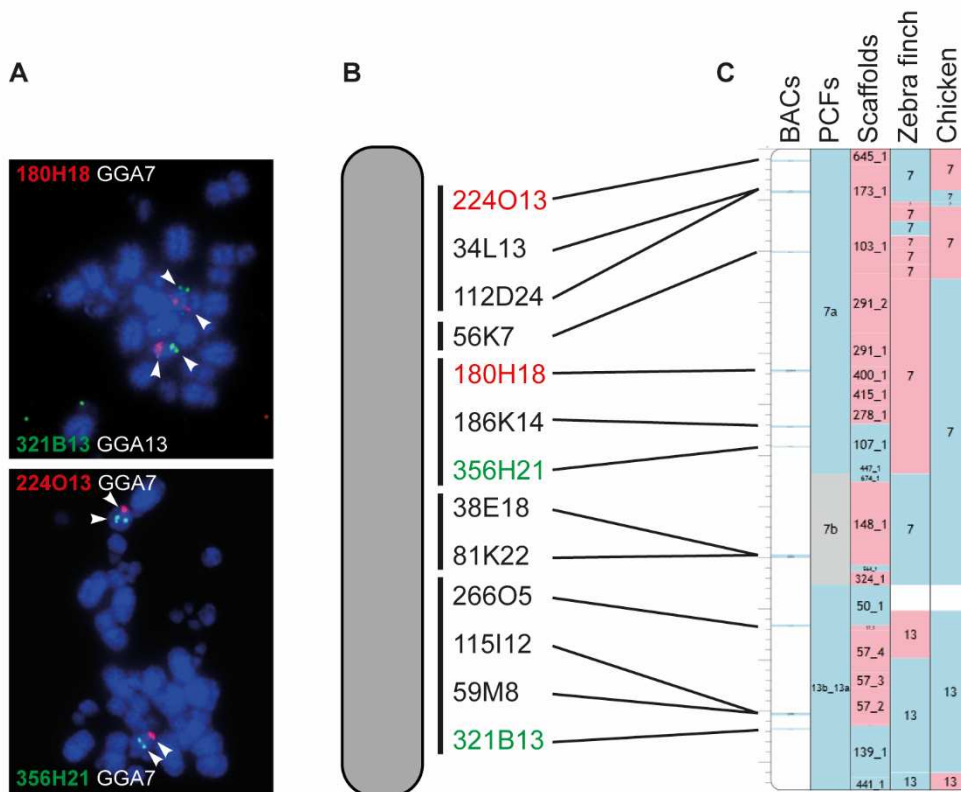
24

25 The objective of this study was therefore to develop a novel, inexpensive, transferrable  
26 approach to upgrade two fragmented genome assemblies (pigeon and falcon) to the  
27 chromosome level and to use them to address fundamental biological questions related to  
28 avian genome evolution. The method combines computational algorithms for ordering

- 1 scaffolds into PCFs, verification of scaffolds by PCR and physical mapping directly to
- 2 chromosomes with a universal set of avian bacterial artificial chromosome (BAC) probes.
- 3 Studying a highly rearranged genome (falcon) compared to the avian ancestor sheds light on
- 4 why interchromosomal rearrangements are infrequent in bird evolution.

1 **RESULTS**

2 Our method involves: (1) the construction of PCFs for fragmented assemblies based on the  
 3 comparative and sequence read data implemented in the RACA algorithm; (2) PCR and  
 4 computational verification of a limited number of scaffolds that are essential for revealing  
 5 species-specific chromosome structures; (3) creation of a refined set of PCFs; (4) the use of  
 6 a panel of “universal” BAC clones to anchor PCFs to chromosomes in a high-throughput  
 7 manner (Fig. 1).



8  
 9 **Figure 1.** Methodology for the placement of the PCFs on chromosomes. (A) dual-color FISH  
 10 of universal BAC clones, (B) cytogenetic map of the falcon chromosome 8 (FPE8) with  
 11 indication of the relative positions of the BAC clones along the chromosome, and (C)  
 12 assembled chromosome containing PCFs 7a, 7b and 13b\_13a. Blue blocks indicate positive  
 13 (+) orientation of tracks compared to the falcon chromosome, red blocks indicate negative (-)  
 14 orientation and grey blocks show unknown (?) orientation.

1 **Construction of PCFs from fragmented assemblies**

2 Predicted chromosome fragments were generated for fragmented falcon and pigeon whole-  
 3 genome sequences using RACA (Kim et al. 2013). For falcon, the zebra finch chromosome  
 4 assembly was used as reference (divergence 62 MYA) and the chicken genome as outgroup  
 5 (divergence 96 MYA). We generated a total of 113 PCFs with N50 of 27.44 Mb (Table 1). For  
 6 pigeon (>70 MY divergence from both the chicken and zebra finch), chicken was used as  
 7 reference and zebra finch as outgroup because: a) fewer pigeon scaffolds were split in this  
 8 configuration (Supplemental Table S1) and b) due to the high similarity of pigeon and chicken  
 9 karyotypes (Derjusheva et al. 2004). This resulted in 150 pigeon PCFs with N50 of 34.54 Mb  
 10 (Table 1). These initial PCF sets contained 72 (15.06%) and 78 (13.64%) scaffolds, for falcon  
 11 and pigeon respectively, that were split by RACA due to insufficient read and/or comparative  
 12 evidence to support their structures.

13

14 **Table 1.** Scaffold-based RACA assemblies for peregrine falcon and pigeon.

Statistics	Peregrine falcon			Pigeon		
	Scaffold assembly	Default RACA	Adjusted RACA <sup>1</sup>	Scaffold assembly	Default RACA	Adjusted RACA <sup>1</sup>
No. scaffolds (≥ 10 kb)	723	478	478	1,081	572	572
No. PCFs	NA	113	93	NA	150	137
Total length (Gb)	1.17	1.14	1.14	1.10	1.07	1.07
N50 (Mb)	3.94	27.44	25.82	3.15	34.54	22.17
Fraction of scaffold assembly (%)	NA	97.17	97.17	NA	95.86	95.86
No. scaffolds split by RACA	NA	72 (15.06 <sup>2</sup> )	15 (3.14 <sup>2</sup> )	NA	78 (13.64 <sup>2</sup> )	20 (3.50 <sup>2</sup> )

15 <sup>1</sup>RACA assembly after the use of adjusted coverage thresholds and post-processing of  
 16 scaffolds verified by PCR.

17 <sup>2</sup>Percentage of all scaffolds included in the RACA assembly.

18

19 **Verification of scaffolds essential for revealing species-specific chromosome**  
 20 **architectures**

21 All scaffolds split by RACA contained structural differences between the target and reference  
 22 chromosomes, suggesting their importance for revealing the architecture of target species

1 chromosomes. The structures of these scaffolds were tested by PCR amplification across all  
2 the split regions defined to <6 kb in the target species scaffolds. Of these, 41 (83.67%) and  
3 58 (84.06%) resulted in amplicons of expected length in pigeon and falcon genomic DNA,  
4 respectively (Supplemental Table S2). For the split regions with negative PCR results we  
5 tested an alternative (RACA-suggested) order of the flanking syntenic fragments (SFs). Out  
6 of these, amplicons were obtained for 2/4 in falcon and 7/7 in pigeon, confirming the chimeric  
7 nature of the original scaffolds properly detected in these cases (Supplemental Table S2). To  
8 estimate which of the remaining split regions (>6 kb; 36 in falcon and 40 in pigeon PCFs) were  
9 likely to be chimeric, we empirically identified two genome-wide minimum physical coverage  
10 (Meyerson et al. 2010) levels, one for falcon and one for pigeon, in the SFs joining regions for  
11 which (and higher) the PCR results were most consistent with RACA predictions. If the new  
12 thresholds were used in RACA without additional scaffold verification (e.g., by PCR) or  
13 mapping data, they would lead to splitting of nearly all scaffolds with large structural  
14 misassemblies in falcon and ~6% of them would still be present in pigeon PCFs. The number  
15 of scaffolds containing real structural differences with the reference chromosomes that would  
16 still be split by RACA was estimated as ~56% in the falcon and ~43% in pigeon PCFs  
17 (Supplemental Table S2). To reduce the number of the real structural differences split in the  
18 final PCF set, PCR verification of selected scaffolds and use of independent (cytogenetic)  
19 mapping have been introduced.

20

### 21 **Creation of a refined set of pigeon and falcon PCFs**

22 For new reconstructions the adjusted physical coverage thresholds were used. In addition, we  
23 kept intact those scaffolds confirmed by PCR, but split those shown to be chimeric and/or  
24 disagreeing with the cytogenetic map (see below) resulting in a total of 93 PCFs with N50  
25 25.82 Mb for falcon and 137 PCFs with N50 of 22.17 Mb for pigeon, covering 97.17% and  
26 95.86% of the original scaffold assemblies, respectively (Table 1). The falcon RACA assembly  
27 contained six PCFs homeologous to complete zebra finch chromosomes (TGU4A, 9, 11, 14,



1 17 and 19) while five pigeon PCFs were homeologous to complete chicken chromosomes  
2 (GGA11, 13, 17, 22 and 25). Only 3.50% of the original scaffolds used by RACA were split in  
3 pigeon and 3.14% in falcon final PCFs (Table 1). The accuracy for the PCF assembly was  
4 estimated as ~85% for falcon and ~89% for pigeon based on the ratio of the number of SFs  
5 to the number of scaffolds (Kim et al. 2013).

6

### 7 **Construction of a panel of comparatively anchored BAC clones designed to hybridize** 8 **in phylogenetically divergent avian species and link PCFs to chromosomes**

9 Initial experiments on cross-species BAC mapping using FISH on five avian species with  
10 divergence times between 23 and 96 MY revealed highly varying success rates (21-94%), with  
11 hybridizations more likely to succeed on species closely related to that of the BAC origin (Table  
12 2). To minimize the effect of evolutionary distances between species on hybridizations,  
13 genomic features that were likely to influence hybridization success were measured in  
14 chicken, zebra finch and turkey BAC clones (Supplemental Tables S3, S4). The classification  
15 and regression tree approach (CART; Loh 2011) was applied to the 101 randomly-selected  
16 BAC clones (Table 2). The obtained classification shows 87% agreement with FISH results  
17 (Supplemental Fig. S1). Correlating DNA features with actual cross-species FISH results led  
18 us to develop the following criteria for selection of chicken or zebra finch BAC clones very  
19 likely to hybridize on metaphase preparations of phylogenetically distant birds (>72 MY of  
20 divergence): the BAC had to have  $\geq 93\%$  DNA sequence alignable with other avian genomes  
21 and contain at least one conserved element (CE)  $\geq 300$  bp. Instead of a long CE, the BAC  
22 could contain only short repetitive elements (<1290 bp) and CEs of at least 3 bp long  
23 (Supplemental Fig. S1; Supplemental Table S4). The hybridization success rate with distant  
24 avian species for the set of newly selected clones obeying these criteria was high (71-94%;  
25 Table 2). The success rates for the selected chicken BAC clones only ranged from 90% to  
26 94%. From these chicken clones, 84% hybridized with chromosomes of all avian species in  
27 our set (Supplemental Fig. S2).

28

1 As a final result, we generated a panel of 121 BAC clones spread across the avian genome  
 2 (GGA 1-28 +Z (except 16)) that successfully hybridized across all species attempted. The  
 3 collection was supplemented by a further 63 BACs that hybridized on the metaphases of at  
 4 least one species that was considered phylogenetically distant (i.e. >72 MY) and a further 33  
 5 that hybridized on at least one other species (Fig. 2; Supplemental Table S5).

6

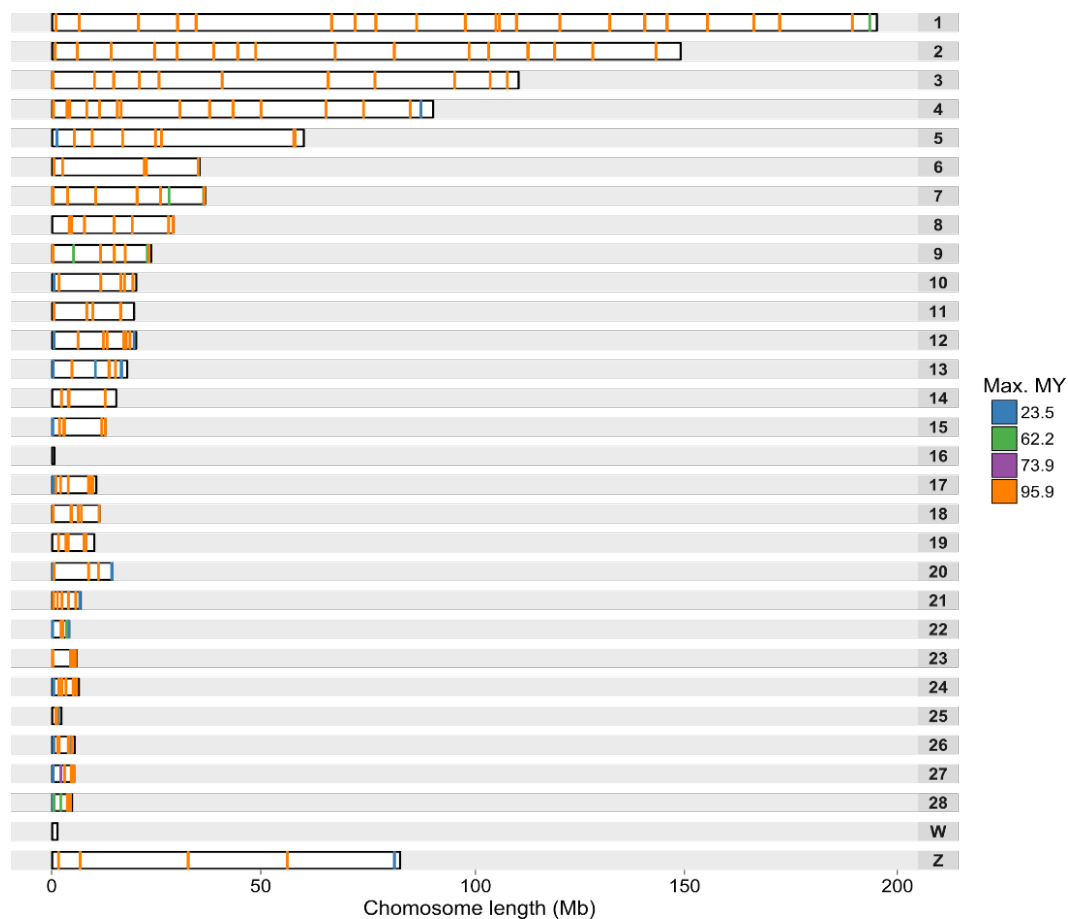
7 **Table 2.** Comparison of zoo-FISH success rate for random and selected set of BAC clones.

	Chicken BAC clones				Zebra finch BAC clones			
	Divergence time (MY)	Success rate (%)			Divergence time (MY)	Success rate (%)		
		Random set N = 53	Selected set N = 99	Ratio		Random set N = 48	Selected set N = 24	Ratio
Chicken	NA	NA	NA	NA	95.88	58.33	75.00	1.29
Turkey	23.47	88.68	100.00	1.13	95.88	54.17	83.33	1.54
Pigeon	95.88	26.42	91.92	3.48	73.87	68.75	70.83	1.03
Peregrine falcon	95.88	47.17	93.94	1.99	62.22	93.75	91.67	0.98
Zebra finch	95.88	20.75	90.91	4.38	NA	NA	NA	NA

8

9 **Physical assignment of refined PCFs on the species' chromosomes**

10 In order to place and order PCFs along chromosomes, BAC clones from the panel described  
 11 above and assigned to PCFs based on alignment results were hybridized to falcon (177  
 12 clones) and pigeon (151 clones) chromosomes (Table 3). The 57 PCFs cytogenetically  
 13 anchored to the falcon chromosomes represented 1.03 Gb of its genome sequence (88% of  
 14 the cumulative scaffold length). Of these, 735.94 Mb were oriented on the chromosomes  
 15 (Table 3; Supplemental Table S6). The pigeon chromosome assembly consisted of 0.91 Gb  
 16 in 60 pigeon PCFs representing 82% of the combined scaffold length. Of these 687.59 Mb  
 17 were oriented (Table 3; Supplemental Table S7). Comparative visualizations of both newly  
 18 assembled genomes are available from the Evolution Highway comparative chromosome  
 19 browser (see Supplemental Results).



1  
2 **Figure 2.** Distribution of universal BAC clones along chicken chromosomes. Each rectangle  
3 represents a chicken chromosome and the lines inside the location of each BAC clone. BAC  
4 clones are colored accordingly to the maximum phylogenetic distance of the species they  
5 successfully hybridized. The distribution of spacing between all these BAC clones is shown  
6 on the Supplemental Fig. S3.

7  
8 **Table 3.** Statistics for the chromosome assemblies of peregrine falcon and pigeon.

Statistics	Peregrine falcon	Pigeon
No. informative BAC clones	177	151
No. PCFs placed on chromosomes	57	60
Combined length (Gb)	1.03	0.91
PCF assembly coverage (%)	90.03	85.23
Scaffold assembly coverage (%)	87.55	81.70
No. oriented PCFs	26	26
Combined length (Mb)	735.94	687.59

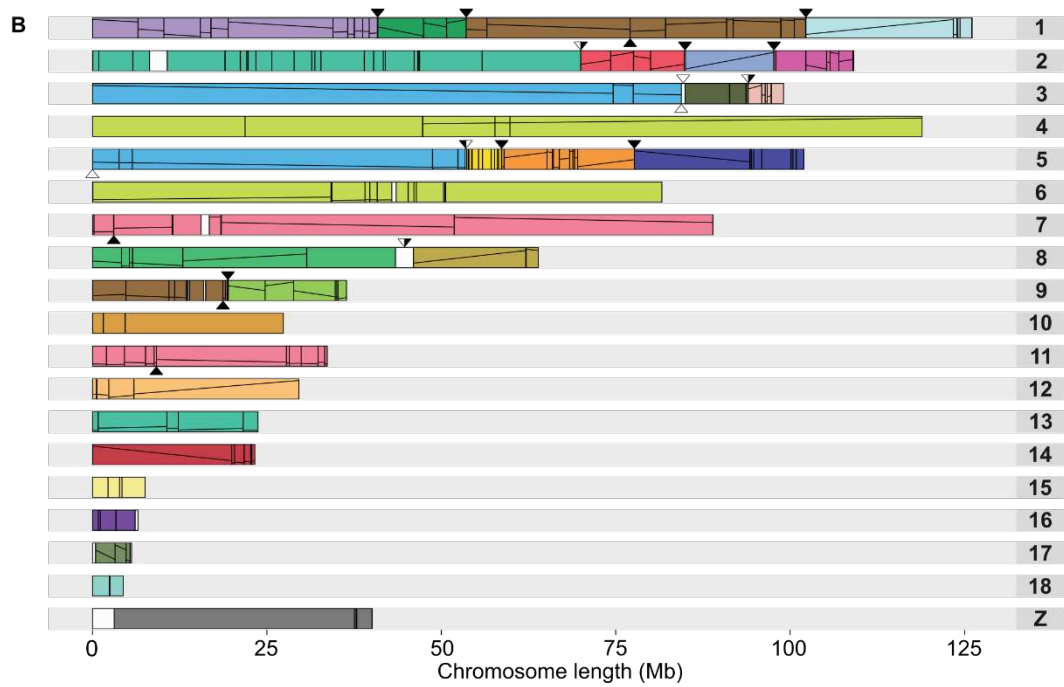
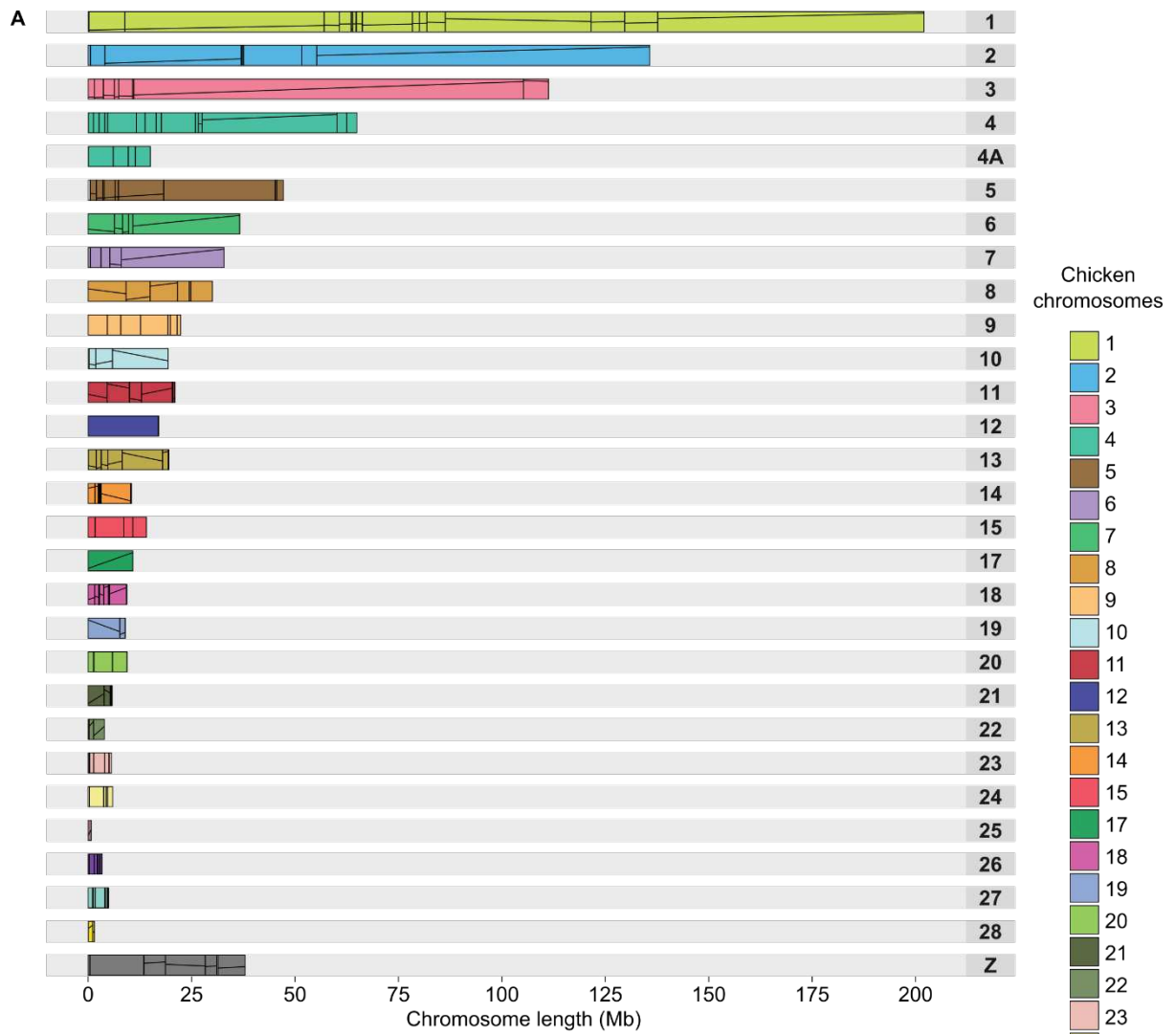
1 **Pigeon chromosome assembly**

2 No deviations from the standard avian karyotype ( $2n=80$ ) were detected for pigeon with each  
3 mapped chromosome having an appropriate single chicken and zebra finch homeologue.  
4 Compared to chicken, the only interchromosomal rearrangement identified was the ancestral  
5 configuration of GGA4 found as two separate chromosomes in pigeon and other birds  
6 (Derjusheva et al. 2004; Hansmann et al. 2009; Modi et al. 2009) (Fig. 3A; [http://eh-](http://eh-demo.ncsa.uiuc.edu/birds)  
7 [demo.ncsa.uiuc.edu/birds](http://eh-demo.ncsa.uiuc.edu/birds)). Nonetheless, 70 intrachromosomal EBRs in the pigeon lineage  
8 were identified (Supplemental Table S8).

9

10 **Falcon chromosome assembly**

11 Homeology between the chicken and the falcon was identified for all mapped chromosomes  
12 with the exception of GGA16 and GGA25 (Fig. 3B; <http://eh-demo.ncsa.uiuc.edu/birds>). In  
13 total, 13 falcon-specific fusions and six fissions were detected (Supplemental Table S8). Each  
14 of the chicken largest macrochromosome homeologues (GGA1 to GGA5) were split across  
15 two falcon chromosomes. Both GGA6 and GGA7 homeologues were found as single blocks  
16 fused with other chicken chromosome material within falcon chromosomes. Among the other  
17 chicken macrochromosomes, only GGA8 and GGA9 were represented as individual  
18 chromosomes. Of the 17 mapped chicken microchromosomes, 11 were fused with other  
19 chromosomes. A total of 69 intrachromosomal EBRs were detected in the falcon lineage  
20 (Supplemental Table S8; Supplemental Results).



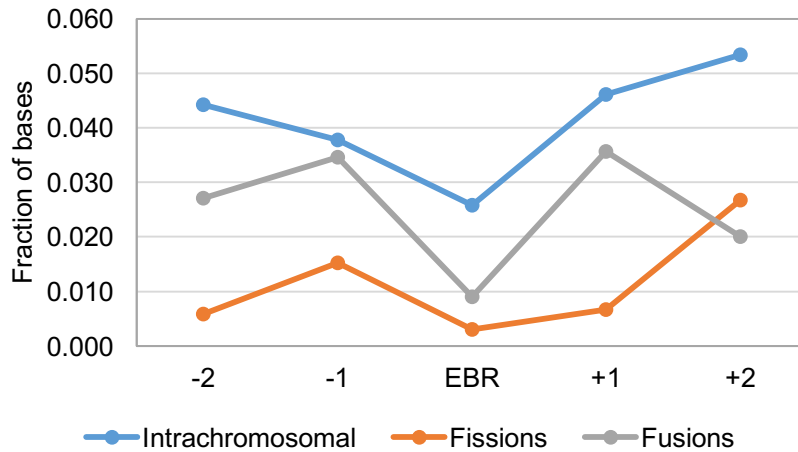
1

1 **Figure 3.** Ideogram of pigeon (A) and peregrine falcon (B) chromosomes. Numbered  
2 rectangles represent chromosomes and colored blocks inside represent regions of homeology  
3 with chicken chromosomes. Lines within colored blocks represent block orientation. Pigeon  
4 chromosomes 1-9 and Z were numbered according to Hansmann et al., 2009 and the  
5 remaining chromosomes according to their chicken homeologues. Falcon chromosomes 1-13  
6 and Z were numbered accordingly to Nishida et al., 2008. The remaining chromosomes were  
7 numbered by decreasing combined length of the placed PCFs. Triangles above the falcon  
8 chromosomes point to the positions of falcon-specific fusions and below chromosomes  
9 demarcate the positions of fissions. Black filling within the triangles point to the EBR  
10 boundaries used in the CNE analysis.

11

### 12 **Fate of CNEs in avian inter- and intrachromosomal EBRs**

13 The falcon chromosome assembly provided us with a set of 19 novel interchromosomal EBRs  
14 not previously found in published avian chromosome assemblies (Fig. 3B; Supplemental  
15 Table S8). To investigate the fate of CNEs in avian EBRs, we calculated densities of avian  
16 CNEs in the chicken chromosome regions corresponding to the chicken, falcon, pigeon,  
17 flycatcher and zebra finch intrachromosomal and interchromosomal EBRs defined to  $\leq 100$  kb  
18 in the chicken genome (Fig. 4; Supplemental Table S9). Avian EBRs had significantly lower  
19 fraction of CNEs than their two adjacent chromosome intervals of the same size each (up- and  
20 downstream ( $p$ -value =  $3.35e-07$ ; Supplemental Table S10)). Moreover, the interchromosomal  
21 EBRs (fusions and fissions) had on average  $\sim 12$  times lower density of CNEs than the  
22 intrachromosomal EBRs ( $p$ -value =  $2.40e-05$ ; Supplemental Table S10). The lowest density  
23 of CNEs was observed in the fission breakpoints ( $p$ -value = 0.04; Fig. 6, Supplemental Table  
24 S10).



1

2 **Figure 4.** Average fraction of bases within conserved non-coding elements (CNEs) in avian  
 3 EBRs and two flanking regions upstream (-) and downstream (+).

4

5 To identify CNE densities and the distribution associated with avian EBRs at the genome-wide  
 6 level, we counted CNE bases in 1 kb windows overlapping EBRs and avian msHSBs >1.5 Mb  
 7 (Farré et al. 2016). The average density of CNEs in the EBR windows was lower (0.02) than  
 8 in msHSBs (0.11). The density of CNEs in the fission EBRs was the lowest observed, zero  
 9 CNE bases ('zero CNE windows'), while in the intrachromosomal EBRs the highest among  
 10 the EBR regions (0.02; Supplemental Table S11). The genome-wide CNE density was 0.09,  
 11 closer to the density observed in msHSBs. Of ~347 Mb of the chicken genome found in the  
 12 'zero CNE windows' 0.5% were associated with EBRs and 15% with msHSBs. To investigate  
 13 if these intervals are distributed differently in the breakpoint and synteny regions we compared  
 14 distances between the 'zero CNE windows' and the closest window with the average msHSB  
 15 CNE density or higher in EBRs, msHSBs, and genome-wide. The median of the distances  
 16 between these two types of windows was the lowest in the msHSBs (~4 kb), intermediate in  
 17 the intrachromosomal (~19 kb) and fusion EBRs (~23 kb), and highest in the fission EBRs  
 18 (~35 kb) (Supplemental Table S12). All these values were significantly different from the  
 19 genome-wide average distance of ~6 kb (p-values <2.2e-16) and also significantly different  
 20 from each other (p-value ≤0.004; Supplemental Table S12; Supplemental Fig. S4).

## 1 DISCUSSION

2 In this study we present a novel approach to upgrade sequenced animal genomes to the  
3 chromosome level. We thereafter generated such assemblies for two previously published but  
4 highly fragmented avian genomes. The resulting chromosome level assemblies contain >80%  
5 of the genomes and, in continuity are comparable to those obtained by combining the  
6 traditional sequencing and mapping techniques (Deakin and Ezaz 2014) but require much less  
7 cost and resource. The design and use of a set of BAC probes intended to work equally well  
8 on a large number of diverged avian species created a resource for physical mapping that is  
9 transferrable to multiple species. Finally, through these new assemblies, we were able to gain  
10 insight into overall genome organization and evolution in birds.

11

12 Molecular and cytogenetic studies to date, suggest that the majority of avian genomes remain  
13 remarkably conserved in terms of chromosome number (in 60-70% of species  $2n \sim 80$ ) and  
14 that interchromosomal changes are relatively rare (Griffin et al. 2007; Schmid et al. 2015).  
15 Exceptions include representatives of *Psittaciformes* (parrots), *Sphenisciformes* (penguins)  
16 and *Falconiformes* (falcons). This study represents the first reconstruction of a highly  
17 rearranged avian karyotype (peregrine falcon). It demonstrates that fusion is the most common  
18 mechanism of interchromosomal change in this species, with some resulting chromosomes  
19 exhibiting as many as four fused ancestral chromosomes. There was no evidence of reciprocal  
20 translocations and all microchromosomes remained intact, even when fused to larger  
21 chromosomes. Recently we suggested possible mechanisms why avian genomes, with  
22 relatively rare exceptions, remain evolutionarily stable interchromosomally and why  
23 microchromosomes represent blocks of conserved synteny (Romanov et al. 2014; Farré et al.  
24 2016). Absence of interchromosomal rearrangement (as seen in most birds) could either  
25 suggest an evolutionary advantage to retaining such a configuration or little opportunity for  
26 change. A smaller number of transposable elements in avian genomes compared to other  
27 animals would indicate that avian chromosomes indeed have fewer opportunities for  
28 chromosome merging using NAHR, explaining the presence of multiple microchromosomes.



1 On the other hand, a strong enrichment for avian CNEs in the regions of interspecies synteny  
2 in birds and other reptiles suggests evolutionary advantage of maintaining established synteny  
3 (Farré et al. 2016), implying that fission events should be rare in avian evolution. In this study,  
4 we present the first analysis of a significant number of interchromosomal EBRs by analysis of  
5 the falcon genome, demonstrating that those rare interchromosomal rearrangements that are  
6 fixed in the avian lineage-specific evolution did indeed appear in areas of a low density of  
7 CNEs. This applies to both fission and fusion events. Our results demonstrate moreover that,  
8 to be suitable for chromosomal fission, the sites of interchromosomal EBRs are restricted  
9 further as they need to be significantly more distant from the areas with high CNE density than  
10 the equivalent intervals found in the regions of multispecies synteny, other EBR types, or on  
11 average in the genome. This might also explain why falcon-specific fission breakpoints appear  
12 to be reused in other avian lineages as intrachromosomal EBRs. Study of intrachromosomal  
13 changes in pigeons, falcons (this study) and Passeriform species (Skinner and Griffin 2012;  
14 Romanov et al. 2014) suggests that these events might have a less dramatic effect on *cis*  
15 gene regulation than interchromosomal events. Indeed, intrachromosomal EBRs appear in  
16 regions of significantly higher CNE density than interchromosomal EBRs. Why then, do  
17 species such as falcons and parrots undergo wholesale interchromosomal rearrangement  
18 (previously reported), but (according to this study) with fission restricted to a few events and  
19 fusion more common? Absence of positive selection for change in chromosome number (or  
20 lack of templates for NAHR) possibly explains why there was little fixation of any  
21 interchromosomal change among birds in general (Bush et al. 1977; Fontdevila et al. 1982;  
22 Burt et al. 1999; Burt 2002), however why this positive selection has been re-introduced (or  
23 barriers to it have been removed) in selected orders is still a matter of conjecture.

24

25 We have previously reported success with the use of high-gene density and low-repeat  
26 content BAC clones for cross-species hybridization (Larkin et al. 2003; Romanov et al. 2011).  
27 In this study, by combining comparative sequence analysis, targeted PCR and optimized high-  
28 throughput cross-species BAC hybridizations we present a chromosome-level assembly

1 approach that is theoretically applicable to any animal genome. The most obvious  
2 phylogenetic Class on which future efforts should be focused is the mammals. To date only  
3 about 20 of the 5,000 extant species have chromosome level genome assemblies (with  
4 primates, rodents and artiodactyls disproportionately overrepresented) but several hundred  
5 currently being assembled to scaffold level by individual projects or consortia such as  
6 Genome10K (Koepfli et al. 2015). Building a mammalian universal BAC set would be a greater  
7 challenge than in birds as mammalian genomes have more repetitive sequences and are  
8 about three times larger thus more BACs would be needed to achieve the same level of  
9 mapping resolution. On the other hand, the development of advanced mapping and  
10 sequencing techniques (e.g., Dovetail, BioNano or PacBio) will eventually provide an  
11 opportunity to replace RACA PCFs with longer and more complete sub-chromosomal sized  
12 superscaffolds or sequence contigs requiring fewer BACs to anchor them to chromosomes.  
13 The availability of large numbers of high-quality mammalian BAC clone libraries from many  
14 species makes our approach more applicable to mammals than to any other animal group. If  
15 we add the fact that our avian BAC set is showing good success rates on lizard and turtle  
16 chromosomes (unpublished results), building chromosomal assemblies for all vertebrate and  
17 ultimately all animal groups supported by universal collection of BACs is a realistic objective  
18 for the near future.

## 1 **METHODS**

### 2 **Avian genome assemblies, repeat masking and gene annotations**

3 The chicken (ICGSC Gallus\_gallus 4.0; Hillier 2004), zebra finch (WUGSC 3.2.4; Warren et  
4 al. 2010), and turkey (TGC Turkey\_2.01; Dalloul et al. 2010) chromosome assemblies were  
5 downloaded from the UCSC Genome Browser (Kent et al. 2002). The collared flycatcher  
6 (FicAlb1.5; Ellegren et al. 2012) genome was obtained from NCBI. Scaffold-based (N50>2  
7 Mb) assemblies of pigeon, falcon, and 16 additional avian genomes were provided by the  
8 Avian Phylogenomics Consortium (Zhang et al. 2014a). All sequences were repeat-masked  
9 using Window Masker (Morgulis et al. 2006) with *-sdust* option and Tandem Repeats Finder  
10 (Benson 1999). Chicken gene (version of 27/04/2014) and repetitive sequence (version of  
11 11/06/2012) annotations were downloaded from the UCSC genome browser (Rosenbloom et  
12 al. 2015). Chicken genes with a single ortholog in the human genome were extracted from  
13 Ensembl Biomart (v.74; Kinsella et al. 2011).

14

### 15 **Pairwise and multiple genome alignments, nucleotide evolutionary conservation** 16 **scores and conserved elements**

17 Pairwise alignments using chicken and zebra finch chromosome assemblies as references  
18 and all other assemblies as targets were generated with *LastZ* (v.1.02.00; Harris 2007) and  
19 converted into the UCSC “chains” and “nets” alignment formats with the Kent-library tools  
20 (Kent et al. 2003; Supplemental Methods). The evolutionary conservation scores and DNA  
21 conserved elements (CEs) for all chicken nucleotides assigned to chromosomes were  
22 estimated using PhastCons (Hubisz et al. 2011) from the multiple alignments of 21 avian  
23 genomes (Supplemental Methods). Conserved non-coding elements obtained from the  
24 alignments of 48 avian genomes were used (Farré et al. 2016).

25

### 26 **Reference-assisted chromosome assembly of pigeon and falcon genomes**

27 Pigeon and falcon PCFs were generated using the Reference-Assisted Chromosome  
28 Assembly (RACA; Kim et al. 2013; Supplemental Methods) tool. We chose zebra finch

1 genome as reference and chicken as outgroup for falcon based on the phylogenetic distances  
2 between the species (Jarvis et al. 2014). For pigeon both chicken as reference and zebra finch  
3 as outgroup and the vice versa experiments were performed as pigeon is phylogenetically  
4 distant from chicken and zebra finch. Two rounds of RACA were done for both species. The  
5 initial run was performed using the following parameters: *WINDOWSIZE=10*  
6 *RESOLUTION=150000 MIN\_INTRACOV\_PERC=5*. Prior to the second run of RACA we  
7 tested the scaffolds split during the initial RACA run using PCR amplification across the split  
8 intervals (see below) and adjusted the parameters accordingly (Supplemental Methods).

9

### 10 **PCR testing of adjacent SFs**

11 Primers flanking split SF joints within scaffolds or RACA predicted adjacencies were designed  
12 using Primer3 software (v.2.3.6; Untergasser et al. 2012). To avoid misidentification of EBRs  
13 or chimeric joints we selected primers only within the sequences that had high quality  
14 alignments between the target and reference genomes and found in adjacent SFs. Due to  
15 alignment and SF detection settings some of the intervals between adjacent SFs could be >6  
16 kb and primers could not be chosen for a reliable PCR amplification. In such cases we used  
17 CASSIS software (Baudet et al. 2010) and the underlying alignment results to narrow gaps  
18 between adjacent SFs where possible. Whole blood was collected aseptically from adult falcon  
19 and pigeon. DNA was isolated using DNeasy Blood and Tissue Kit (Qiagen) following standard  
20 protocols. PCR amplification was performed according to the protocol described in the  
21 Supplemental Methods.

22

### 23 **BAC clone selection**

24 The chromosome coordinates of chicken (CHORI-261), turkey (CHORI-260) and zebra finch  
25 (TGMCBA) BAC clones in the corresponding genomes were extracted from NCBI clone  
26 database (Schneider et al. 2013). We removed all discordantly placed BAC clones (based on  
27 BAC end sequence (BES) mappings) following the NCBI definition of concordant BAC  
28 placement. Briefly, a BAC clone placement was considered concordant when the estimated

1 BAC length in the corresponding avian genome is within [library average length  $\pm 3 \times \textit{standard}$   
2 *deviation*] and BAC BESs map to the opposite DNA strands in the genome assembly. Turkey  
3 and zebra finch BAC clone coordinates were translated into chicken chromosome coordinates  
4 using UCSC Genome Browser *LiftOver* tool (Kent et al. 2002) with the minimum ratio of  
5 remapped bases  $>0.1$ .

6

7 For each BAC clone mapped to the chicken chromosomes various genomic features selected  
8 to estimate the probability of clones to hybridize with metaphase chromosomes in distant avian  
9 species were calculated (Supplemental Table S3) using a custom Perl script or extracted from  
10 gene, repetitive sequence, conserved element and nucleotide conservation score files. The  
11 clones selected for mapping experiments were originally obtained from the BACPAC  
12 Resource Centre at the Children's Hospital Oakland Research Institute and the zebra finch  
13 TGMCBa library (Clemson University Genomics Institute).

14

### 15 **Classification tree**

16 The classification tree was created in R (v.3.2.3; Team 2015) using the classification and  
17 regression tree (CART) algorithm included in the rpart package (v.4.1-10; Therneau et al.  
18 2015). We introduced an adjusted weight matrix setting: the cost of returning a false positive  
19 was twice as high as the cost of a false negative. The tree was visualized with rattle package  
20 (v.4.1.0; Williams 2011).

21

### 22 **Cell culture and chromosome preparation**

23 Chromosome preparations were established from fibroblast cell lines generated from  
24 collagenase treatment of 5- to 7-day-old embryos or from skin biopsies. Cells were cultured at  
25 40°C, and 5% CO<sub>2</sub> in Alpha MEM (Fisher), supplemented with 20% Fetal Bovine Serum  
26 (Gibco), 2% Pen-Strep (Sigma) and 1% L-Glutamine (Sigma). Chromosome suspension  
27 preparation followed standard protocols, briefly mitostatic treatment with colcemid at a final

1 concentration of 5.0 µg/ml for 1 h at 40°C was followed by hypotonic treatment with 75mM KCl  
2 for 15 min at 37°C and fixation with 3:1 methanol:acetic acid.

3

#### 4 **Preparation of BAC clones for fluorescence *in-situ* hybridization (FISH)**

5 BAC clone DNA was isolated using the Qiagen Miniprep Kit (Qiagen) prior to amplification and  
6 direct labelling by nick translation. Probes were labeled with Texas Red-12-dUTP (Invitrogen)  
7 and FITC-Fluorescein-12-UTP (Roche) prior to purification using the Qiagen Nucleotide  
8 Removal Kit (Qiagen).

9

#### 10 **Fluorescence *in-situ* hybridization (FISH)**

11 Metaphase preparations were fixed to slides and dehydrated through an ethanol series (2 min  
12 each in 2xSSC, 70%, 85% and 100% ethanol at room temperature). Probes were diluted in a  
13 formamide buffer (Cytocell) with Chicken Hybloc (Insight Biotech) and applied to the  
14 metaphase preparations on a 37°C hotplate before sealing with rubber cement. Probe and  
15 target DNA were simultaneously denatured on a 75°C hotplate prior to hybridization in a  
16 humidified chamber at 37°C for 72 h. Slides were washed post-hybridization for 30 sec in  
17 2xSSC/ 0.05% Tween 20 at room temperature, then counterstained using VECTASHIELD  
18 anti-fade medium with DAPI (Vector Labs). Images were captured using an Olympus BX61  
19 epifluorescence microscope with cooled CCD camera and SmartCapture (Digital Scientific  
20 UK) system. In selected experiments, we used multiple hybridization strategies, making use  
21 of the Cytocell Octochrome (8 chamber) and Multiprobe (24 chamber) devices. Briefly, labeled  
22 probes were air dried on to the device. Probes were, re-hybridized in standard buffer, applied  
23 to the glass slide (which was sub-divided to correspond to the hybridization chambers) and  
24 FISH continued as above.

25

26

## 1 **EBR detection and CNE density analysis**

2 The multiple alignments of the chicken, zebra finch, flycatcher, pigeon and falcon chromosome  
3 sequences were obtained using progressiveCactus (Paten et al. 2011) with default  
4 parameters. Pairwise synteny blocks were defined using the maf2synteny tool (Kolmogorov  
5 et al. 2014) at 100, 300 and 500 kb resolution. Using chicken as reference genome, EBRs  
6 were detected and classified using the *ad hoc* statistical approach described previously (Farré  
7 et al. 2016). All well-defined (or flanking oriented PCFs) fusion and fission points were  
8 identified from pairwise alignments with the chicken genome. Only the EBRs  $\leq 100$  kb were  
9 used for the CNE analysis. EBRs smaller than 1 kb were extended  $\pm 1$  kb. For each EBR, we  
10 defined two windows upstream (+1 and +2) and two downstream (-1 and -2) of the same size  
11 as the EBR. We calculated the fraction of bases within CNEs in each EBR site, upstream and  
12 downstream windows. Differences in CNE densities were tested for significance using the  
13 Kruskal-Wallis test followed by Mann-Whitney U test.

14

## 15 **Comparing CNE densities in EBRs and msHSBs**

16 Chicken chromosomes (excluding GGA16, W and Z) were divided into 1 kb non-overlapping  
17 intervals. Only windows with  $>50\%$  of their bases with chicken sequence data available were  
18 used in this analysis. All intervals were assigned either to msHSBs  $>1.5$  Mb (Farré et al. 2016),  
19 avian EBRs flanking: fusions, fissions, intrachromosomal EBR, and the intervals found in the  
20 rest of the chicken genome. We estimated the average CNE density for each window type and  
21 also the distance, in number of 1 kb windows, between each window with the lowest CNE  
22 density (0 bp) and the nearest window with the average msHSB CNE density or higher. CNE  
23 densities were obtained using bedtools (v.2.20-1; Quinlan and Hall 2010). Differences in  
24 distances between the two window types in msHSBs and EBRs were tested for significance  
25 using the Kruskal-Wallis test followed by Mann-Whitney U test.

1 **DATA ACCESS**

2 Visualizations of falcon and pigeon genome assemblies are available from the Evolution

3 Highway comparative chromosome browser: <http://eh-demo.ncsa.uiuc.edu/birds>.

4

5 **DISCLOSURE DECLARATION**

6 Authors report no conflict of interests.

7

8 **ACKNOWLEDGMENTS**

9 This work was supported in part by the Biotechnology and Biological Sciences Research

10 Council [BB/K008226/1 and BB/J010170/1 to D.M.L, and BB/K008161/1 to D.K.G].



## REFERENCES

- Andersson L, Georges M. 2004. Domestic-animal genomics: deciphering the genetics of complex traits. *Nat Rev Genet* **5**: 202-212.
- Baudet C, Lemaitre C, Dias Z, Gautier C, Tannier E, Sagot MF. 2010. Cassis: detection of genomic rearrangement breakpoints. *Bioinformatics* **26**: 1897-1898.
- Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**: 573-580.
- Branco MR, Pombo A. 2006. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol* **4**: e138.
- Burt DW. 2002. Origin and evolution of avian microchromosomes. *Cytogenet Genome Res* **96**.
- Burt DW, Bruley C, Dunn IC, Jones CT, Ramage A, Law AS, Morrice DR, Paton IR, Smith J, Windsor D et al. 1999. The dynamics of chromosome evolution in birds and mammals. *Nature* **402**: 411-413.
- Bush GL, Case SM, Wilson AC, Patton JL. 1977. Rapid speciation and chromosomal evolution in mammals. *Proc Natl Acad Sci U S A* **74**.
- Dalloul RA, Long JA, Zimin AV, Aslam L, Beal K, Blomberg Le A, Bouffard P, Burt DW, Crasta O, Crooijmans RP et al. 2010. Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. *PLoS Biol* **8**.
- Deakin JE, Ezaz T. 2014. Tracing the evolution of amniote chromosomes. *Chromosoma* **123**: 201-216.
- Derjushva S, Kurganova A, Habermann F, Gaginskaya E. 2004. High chromosome conservation detected by comparative chromosome painting in chicken, pigeon and passerine birds. *Chromosome Res* **12**: 715-723.
- Driscoll CA, Macdonald DW, O'Brien SJ. 2009. From wild animals to domestic pets, an evolutionary view of domestication. *Proc Natl Acad Sci U S A* **106**: 9971-9978.

- Ellegren H, Smeds L, Burri R, Olason PI, Backstrom N, Kawakami T, Kunstner A, Makinen H, Nadachowska-Brzyska K, Qvarnstrom A et al. 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* **491**: 756-760.
- Fang X, Nevo E, Han L, Levanon EY, Zhao J, Avivi A, Larkin D, Jiang X, Feranchuk S, Zhu Y et al. 2014. Genome-wide adaptive complexes to underground stresses in blind mole rats *Spalax*. *Nat Commun* **5**: 3966.
- Farré M, Narayan J, Slavov GT, Damas J, Auvil L, Li C, Jarvis ED, Burt DW, Griffin DK, Larkin DM. 2016. Novel insights into chromosome evolution in birds, archosaurs, and reptiles. *Genome Biol Evol* doi:10.1093/gbe/evw166.
- Ferguson-Lees J, Christie DA. 2005. *Raptors of the world*. Princeton University Press, Princeton, N.J.
- Fontdevila A, Ruiz A, Ocaña J, Alonso G. 1982. Evolutionary history of *Drosophila buzzatii*. II. How much has chromosomal polymorphism changed in colonization? *Evolution* **36**.
- Griffin DK, Robertson LBW, Tempest HG, Skinner BM. 2007. The evolution of the avian genome as revealed by comparative molecular cytogenetics. *Cytogenet Genome Res* **117**: 64-77.
- Groenen MAM Archibald AL Uenishi H Tuggle CK Takeuchi Y Rothschild MF Rogel-Gaillard C Park C Milan D Megens H-J et al. 2012. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* **491**: 393-398.
- Hansmann T, Nanda I, Volobouev V, Yang F, Scharf M, Haaf T, Schmid M. 2009. Cross-species chromosome painting corroborates microchromosome fusion during karyotype evolution of birds. *Cytogenet Genome Res* **126**: 281-304.
- Harris RS. 2007. Improved pairwise alignment of genomic DNA. Vol Ph.D. The Pennsylvania State University.
- Hillier L. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**: 695-716.

- Hu X, Gao Y, Feng C, Liu Q, Wang X, Du Z, Wang Q, Li N. 2009. Advanced technologies for genomic analysis in farm animals and its application for QTL mapping. *Genetica* **136**: 371-386.
- Hubisz MJ, Pollard KS, Siepel A. 2011. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief Bioinform* **12**: 41-51.
- Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SYW, Faircloth BC, Nabholz B, Howard JT et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**: 1320-1331.
- Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D. 2003. Evolution's cauldron: Duplication, deletion, and rearrangement in the mouse and human genomes. *Proc Natl Acad Sci U S A* **100**: 11484-11489.
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res* **12**: 996-1006.
- Kim J, Larkin DM, Cai Q, Asan, Zhang Y, Ge R-L, Auvil L, Capitanu B, Zhang G, Lewin HA et al. 2013. Reference-assisted chromosome assembly. *Proc Natl Acad Sci U S A* **110**: 1785-1790.
- Kinsella RJ, Kahari A, Haider S, Zamora J, Proctor G, Spudich G, Almeida-King J, Staines D, Derwent P, Kerhornou A et al. 2011. Ensembl BioMart: a hub for data retrieval across taxonomic space. *Database* **2011**: bar030.
- Koepfli K-P, Benedict Paten, Scientists tGKCo, O'Brien SJ. 2015. The Genome 10K Project: A Way Forward. *Ann Rev Anim Biosci* **3**: 57-111.
- Kolmogorov M, Raney B, Paten B, Pham S. 2014. Ragout-a reference-assisted assembly tool for bacterial genomes. *Bioinformatics* **30**: i302-309.
- Larkin DM, Daetwyler HD, Hernandez AG, Wright CL, Hetrick LA, Boucek L, Bachman SL, Band MR, Akraiko TV, Cohen-Zinder M et al. 2012. Whole-genome resequencing of two elite sires for the detection of haplotypes under selection in dairy cattle. *Proc Natl Acad Sci U S A* **109**: 7693-7698.

- Larkin DM, Everts-van der Wind A, Rebeiz M, Schweitzer PA, Bachman S, Green C, Wright CL, Campos EJ, Benson LD, Edwards J et al. 2003. A cattle–human comparative map built with cattle BAC-ends and human genome sequence. *Genome Res* **13**: 1966-1972.
- Larkin DM, Pape G, Donthu R, Auvil L, Welge M, Lewin HA. 2009. Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. *Genome Res* **19**: 770-777.
- Lewin HA, Larkin DM, Pontius J, O'Brien SJ. 2009. Every genome sequence needs a good map. *Genome Res* **19**: 1925-1928.
- Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J, Jordan G, Mauceli E et al. 2011. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**: 476-482.
- Loh W-Y. 2011. Classification and regression trees. *WIREs Data Mining Knowl Discov* **1**: 14-23.
- Mak ACY, Lai YYY, Lam ET, Kwok T-P, Leung AKY, Poon A, Mostovoy Y, Hastie AR, Stedman W, Anantharaman T et al. 2016. Genome-wide structural variation detection by genome mapping on nanochannel arrays. *Genetics* **202**: 351-362.
- Meyerson M, Gabriel S, Getz G. 2010. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet* **11**: 685-696.
- Modi WS, Romanov M, Green ED, Ryder O. 2009. Molecular cytogenetics of the california condor: evolutionary and conservation implications. *Cytogenet Genome Res* **127**: 26-32.
- Morgulis A, Gertz EM, Schaffer AA, Agarwala R. 2006. A fast and symmetric DUST implementation to mask low-complexity DNA sequences. *J Comput Biol* **13**: 1028-1040.
- Murphy WJ, Larkin DM, Everts-van der Wind A, Bourque G, Tesler G, Auvil L, Beever JE, Chowdhary BP, Galibert F, Gatzke L et al. 2005. Dynamics of mammalian

- chromosome evolution inferred from multispecies comparative maps. *Science* **309**: 613-617.
- Nishida C, Ishijima J, Kosaka A, Tanabe H, Habermann FA, Griffin DK, Matsuda Y. 2008. Characterization of chromosome structures of Falconinae (Falconidae, Falconiformes, Aves) by chromosome painting and delineation of chromosome rearrangements during their differentiation. *Chromosome Res* **16**: 171-181.
- Paten B, Earl D, Nguyen N, Diekhans M, Zerbino D, Haussler D. 2011. Cactus: Algorithms for genome multiple sequence alignment. *Genome Res* **21**: 1512-1528.
- Price TD. 2002. Domesticated birds as a model for the genetics of speciation by sexual selection. *Genetica* **116**: 311-327.
- Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD, Sugnet CW et al. 2016. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res* doi:10.1101/gr.193474.115.
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.
- Rhoads A, Au KF. 2015. PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics* **13**: 278-289.
- Romanov MN, Dodgson JB, Gonser RA, Tuttle EM. 2011. Comparative BAC-based mapping in the white-throated sparrow, a novel behavioral genomics model, using interspecies overgo hybridization. *BMC Res Notes* **4**: 211.
- Romanov MN, Farré M, Lithgow PE, Fowler KE, Skinner BM, O'Connor R, Fonseka G, Backström N, Matsuda Y, Nishida C et al. 2014. Reconstruction of gross avian genome structure, organization and evolution suggests that the chicken lineage most closely resembles the dinosaur avian ancestor. *BMC Genomics* **15**: 1-18.
- Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, Dreszer TR, Fujita PA, Guruvadoo L, Haeussler M et al. 2015. The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res* **43**: D670-681.

Schmid M, Smith J, Burt DW, Aken BL, Antin PB, Archibald AL, Ashwell C, Blackshear PJ, Boschiero C, Brown CT et al. 2015. Third report on chicken genes and chromosomes 2015. *Cytogenet Genome Res* **145**: 78-179.

Schneider VA, Chen HC, Clausen C, Meric PA, Zhou Z, Bouk N, Husain N, Maglott DR, Church DM. 2013. Clone DB: an integrated NCBI resource for clone-associated data. *Nucleic Acids Res* **41**: D1070-1078.

Shapiro MD, Kronenberg Z, Li C, Domyan ET, Pan H, Campbell M, Tan H, Huff CD, Hu H, Vickrey AI et al. 2013. Genomic diversity and evolution of the head crest in the rock pigeon. *Science* **339**: 1063-1067.

Skinner BM, Griffin DK. 2012. Intrachromosomal rearrangements in avian genome evolution: evidence for regions prone to breakpoints. *Heredity* **108**: 37-41.

Stringham SA, Mulroy EE, Xing J, Record D, Guernsey MW, Aldenhoven JT, Osborne EJ, Shapiro MD. 2012. Divergence, convergence, and the ancestry of feral populations in the domestic rock pigeon. *Curr Biol* **22**: 302-308.

Teague B, Waterman MS, Goldstein S, Potamouisis K, Zhou S, Reslewic S, Sarkar D, Valouev A, Churas C, Kidd JM et al. 2010. High-resolution human genome structure by single-molecule analysis. *Proc Natl Acad Sci U S A* **107**: 10848-10853.

Team RC. 2015. R: a language and environment for statistical computing (R foundation for statistical computing, Vienna, 2012). <http://www.R-project.org>.

Therneau T, Atkinson B, Ripley B. 2015. rpart: recursive partitioning and regression trees. R package version 4.1–10.

Tucker V, Cade T, Tucker A. 1998. Diving speeds and angles of a gyrfalcon (*Falco rusticolus*). *J Exp Biol* **201**: 2061-2070.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3-new capabilities and interfaces. *Nucleic Acids Res* **40**: e115.

Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Kunstner A, Searle S, White S, Vilella AJ, Fairley S et al. 2010. The genome of a songbird. *Nature* **464**: 757-762.

Williams G. 2011. *Data mining with Rattle and R: The art of excavating data for knowledge discovery*. Springer Science & Business Media.

Zhan X, Pan S, Wang J, Dixon A, He J, Muller MG, Ni P, Hu L, Liu Y, Hou H et al. 2013.

Peregrine and saker falcon genome sequences provide insights into evolution of a predatory lifestyle. *Nat Genet* **45**: 563-566.

Zhang G, Li B, Li C, Gilbert MT, Jarvis ED, Wang J, Avian Genome C. 2014a. Comparative genomic data of the Avian Phylogenomics Project. *GigaScience* **3**: 26.

Zhang G Li C Li Q Li B Larkin DM Lee C Storz JF Antunes A Greenwold MJ Meredith RW et al. 2014b. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**: 1311-1320.



**Isolation of subtelomeric sequences of porcine chromosomes for translocation screening reveals errors in the pig genome assembly**

Journal:	<i>Animal Genetics</i>
Manuscript ID	AnGen-16-06-0193
Manuscript Type:	Full Paper
Date Submitted by the Author:	02-Jun-2016
Complete List of Authors:	O'Connor, Rebecca; University of Kent, Department of Biosciences Fonseka, Gothami; Cytocell Ltd, R&D Frodsham, Richard; Cytocell Ltd, R&D Archibald, Alan; University of Edinburgh, Roslin Institute and Royal (Dick) School of Veterinary Studies; Lawrie, Martin; Cytocell Ltd, R&D Walling, Grant; JSR Genetics, Research & Genetics Griffin, Darren; University of Kent, Department of Biosciences
Keywords:	sub-telomeric, translocation, pig genome, hypoprolificacy, fertility, chromosome, FISH, multiprobe

SCHOLARONE™  
Manuscripts

view



1  
2  
3 1 Isolation of subtelomeric sequences of porcine chromosomes for translocation  
4  
5 2 screening reveals errors in the pig genome assembly  
6  
7  
8 3

9  
10 4 Rebecca E. O'Connor<sup>1</sup>, Gothami Fonseka<sup>2</sup>, Richard Frodsham<sup>2</sup>, Alan Archibald<sup>3</sup>,  
11  
12 5 Martin Lawrie<sup>2</sup>, Grant A. Walling<sup>4</sup>, Darren K. Griffin<sup>1</sup>  
13  
14 6

15  
16  
17 7 1. School of Biosciences, University of Kent, Canterbury, CT2 7AF, UK  
18

19 8 2. Cytocell Ltd, Newmarket Road, Cambridge, UK  
20

21 9 3. The Roslin Institute, R(D)SVS, University of Edinburgh, Division of Genetics  
22 and Genomics, Easter Bush, Midlothian, EH25 9RG, UK  
23  
24 10

25 11 4. JSR Genetics, Southburn, Driffield, North Humberside, YO25 9ED, UK  
26  
27 12  
28  
29 13

30  
31 14 Correspondence to:

32 15 Professor Darren K Griffin  
33

34 16 School of Biosciences,  
35  
36 17 University of Kent,  
37  
38 18 Canterbury CT2 7AF,  
39  
40 19 Tel: 0044 1227 823022  
41  
42 20 Email [d.k.griffin@kent.ac.uk](mailto:d.k.griffin@kent.ac.uk)  
43  
44 21  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 22 **Abstract**  
4

5 23 Balanced chromosomal aberrations have been shown to affect fertility in most  
6  
7 24 species studied often leading to hypoprolificacy (reduced litter size) in domestic  
8  
9  
10 25 animals such as pigs. With an increasing emphasis in modern food production on the  
11  
12 26 use of a small population of high quality males for artificial insemination, the  
13  
14  
15 27 potential economic and environmental costs of hypoprolific boars, bulls, rams etc.  
16  
17 28 are considerable. There is therefore a need for novel tools to facilitate rapid, cost  
18  
19 29 effective chromosome translocation screening. This has previously been achieved by  
20  
21 30 standard karyotype analysis; however this approach relies on a significant level of  
22  
23 31 expertise and is limited in its ability to identify subtle, cryptic translocations. To  
24  
25 32 address this problem we developed a novel device and protocol for translocation  
26  
27 33 screening using subtelomeric probes and fluorescence *in situ* hybridisation. Probes  
28  
29 34 were designed using BACs from the subtelomeric region of the short (p-arm) and  
30  
31 35 long (q-arm) of each porcine chromosome. They were directly labelled with FITC or  
32  
33 36 Texas Red (p-arm and q-arm, respectively) prior to application to a "Multiprobe"  
34  
35 37 device, thereby enabling simultaneous detection of each individual porcine  
36  
37 38 chromosome on a single slide. Initial experiments designed to isolate BACs in  
38  
39 39 subtelomeric regions led to the discovery of a series of incorrectly mapped regions in  
40  
41 40 the porcine genome assembly. Our work therefore highlights the importance of  
42  
43 41 accurate physical mapping of newly sequenced genomes. The system herein  
44  
45 42 described allows for robust and comprehensive analysis of the porcine karyotype, an  
46  
47 43 adjunct to classical cytogenetics that provides a valuable tool to expedite efficient,  
48  
49 44 cost effective food production.  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

46 **Key words:**

47 Chromosome, hypoprolificacy, translocation, genome, karyotype

For Peer Review

## 48 Introduction

49 The domestic pig (*Sus scrofa domestica*) provides 43% of meat consumed worldwide  
50 making it the leading source of meat protein globally (United States Department of  
51 Agriculture 2015). Purebred boars selected for their genetic merit are used at the  
52 top (nucleus) level of the breeding pyramid meaning that any fertility problems in  
53 these animals could significantly reduce litter sizes throughout the breeding  
54 population. This ultimately leads to a reduction in food production and higher  
55 environmental costs per mating animal, issues that are perpetuated further through  
56 an increasing emphasis on artificial insemination (AI) (Merck CM, Kahn S, Line, 2010).

57  
58 Semen used in AI preparations is routinely assessed for parameters that are  
59 considered to be indicative of fertility such as sperm concentration, morphology and  
60 motility. Evidence suggests that these parameters are in fact, not reliable indicators  
61 of prolificacy (Gadea 2005). Indeed, the primary identification of boars that exhibit  
62 hypoprolificacy is deduced from both litter sizes and 'non-return rates', i.e. the  
63 proportion of sows/gilts served by that boar that return to heat (i.e. fail to conceive)  
64 after 21 days. With a gestation length of 115 days and an average litter size of 12  
65 piglets, each sow can produce around 23 slaughter pigs per year assuming there are  
66 no fertility problems (AHDB 2014). In addition, fertility is assessed using farrowing  
67 rates, which indicate how many litters are produced against how many sows were  
68 originally served (ideally >85% (Gadea et al. 2004)). The mating of hypoprolific boars  
69 into the sow population can have a significant effect on non-return rates and litter  
70 sizes, in some cases reducing the number of piglets in a litter by up to 50%. In order  
71 to prevent the perpetuation of reduced fertility, the identification and elimination of

1  
2  
3 72 hypoprolific boars from the breeding population is a priority, particularly given rising  
4  
5 73 global populations and increasing demand for meat products.  
6

7  
8 74

9  
10 75 Balanced chromosomal rearrangements occur frequently in pigs and are seen in as  
11  
12 76 many as 0.47% of AI boars awaiting service (Ducos et al. 2007). Over 130 reciprocal  
13  
14 77 translocations have been identified with chromosomes 1,7, 14 and 15 the most  
15  
16  
17 78 frequently involved (Rothschild & Ruvinsky 2011). Reciprocal translocations  
18  
19 79 adversely affect reproductive performance in pigs by causing a reduction in litter size  
20  
21 80 due to high mortality among early embryos. Approximately 50% of boars exhibiting  
22  
23 81 hypoprolificacy are reciprocal translocation carriers, even though they have a normal  
24  
25 82 phenotype and semen parameters (Rodríguez et al. 2010). Balanced translocations  
26  
27 83 are considered to be the primary reason for hypoprolificacy in pigs due to the  
28  
29 84 generation of unbalanced gametes and subsequent partially aneuploid conceptuses  
30  
31 85 that lead to early loss of zygotes and ultimately litters that are 25-50% smaller than  
32  
33 86 would be expected ((Gustavsson 1990);(Pinton et al. 2000)).  
34  
35  
36  
37  
38  
39

40 87

41 88 Since the latter part of the 20th century several continental European programmes  
42  
43 89 of chromosomal screening have been established, with the largest centre of pig  
44  
45 90 screening being based in the National Veterinary School of Toulouse, France (Ducos  
46  
47 91 et al. 2008). This has led to the identification of a significant number of chromosomal  
48  
49 92 rearrangements in otherwise phenotypically normal boars. However, since this  
50  
51 93 period there has been a reduction in the number of laboratories that perform animal  
52  
53 94 cytogenetics (with approximately 10-15 operating worldwide, mostly in Europe)  
54  
55 95 (Ducos et al. 2008).  
56  
57  
58  
59  
60

1  
2  
3 96  
4  
5 97 Current translocation screening is performed by G-banding and routine karyotyping.  
6  
7 98 While this is simple and cost effective, it requires specialist knowledge of the porcine  
8  
9  
10 99 karyotype and is limited in its ability to detect translocations smaller than 2-3 Mb in  
11  
12 100 size, especially if bands of similar intensity are exchanged. Moreover, even in the  
13  
14 101 best of laboratories, preparations of sub-optimal quality (e.g. yielding few  
15  
16 102 preparations that are difficult to analyse) can occasionally arise. Such is the nature of  
17  
18 103 biological systems and, in these cases, molecular cytogenetics can aid detection  
19  
20 104 protocols. The recent sequencing of the pig genome provided the tools through  
21  
22 105 which molecular cytogenetic resources can be identified and developed for more  
23  
24 106 accurate and unequivocal translocation screening. Results from our own laboratory  
25  
26 107 provided evidence that the strategy of assembling the swine genome BAC-by-BAC  
27  
28 108 ahead of whole genome sequencing provided the ability to select a clone for  
29  
30 109 fluorescence *in-situ* hybridisation (FISH) with 100% confidence that it would map in  
31  
32 110 the predicted chromosomal position. That is, of 71 clones selected, all mapped to  
33  
34 111 the predicted chromosome band (Groenen et al. 2012).  
35  
36  
37  
38  
39  
40 112  
41  
42 113 In humans, Knight et al (1996) demonstrated an approach through which cryptic  
43  
44 114 translocations could be identified in humans using a FISH strategy that involved 24  
45  
46 115 individual hybridizations (one for each chromosome) on a single slide. By hybridizing  
47  
48 116 to the subtelomeric regions of the short (p) and long (q) arms of each chromosome,  
49  
50 117 each in a different colour, any chromosome translocation is clearly visible, even to  
51  
52 118 the untrained eye. This approach has been used extensively in clinical cytogenetics  
53  
54  
55 119 ((Horsley et al. 1998), (Ravnan et al. 2006), (Dawson et al. 2002)) and, to some  
56  
57  
58  
59  
60

1  
2  
3 120 degree in pigs (Mompert et al. 2013). The purpose of the current study was to  
4  
5 121 develop these investigations further to generate a panel of equivalent porcine BACs,  
6  
7 122 extending on the Knight et al study to develop a porcine version of the human  
8  
9  
10 123 system. The aim was to employ a strategy that would significantly increase the speed  
11  
12 124 and accuracy of boar translocation screening, the ultimate objective being the  
13  
14 125 identification and removal of hypoprolific boars from the breeding population. This  
15  
16  
17 126 could potentially improve efficiency, and reduce the cost and environmental  
18  
19  
20 127 footprint of global meat production.  
21  
22 128

## 24 129 **Materials and Methods**

### 26 130 **Chromosome preparations**

28  
29 131 In order to generate the material for screening and identify potential translocation  
30  
31 132 carriers, we established a routine karyotyping service for UK companies wishing to  
32  
33 133 screen their boars for translocations. Blood samples were provided by three of the  
34  
35  
36 134 UKs leading pig breeding companies (JSR Genetics, ACMC and Genus PIC).  
37  
38 135 Heparinized blood samples were cultured for 72 hours in PB MAX Karyotyping  
39  
40 136 medium (Invitrogen) at 37°C, 5% CO<sub>2</sub>. Cell division was arrested by adding colcemid  
41  
42 137 at a concentration of 10.0µg/ml (Gibco) for 35 minutes before hypotonic treatment  
43  
44 138 with 75mM KCl and fixation to glass slides using 3:1 methanol:acetic acid.  
45  
46  
47 139 Metaphases for karyotyping were stained with DAPI in VECTASHIELD® antifade  
48  
49 140 medium (Vector Laboratories). Image capturing was performed using an Olympus  
50  
51 141 BX61 epifluorescence microscope with cooled CCD camera and SmartCapture (Digital  
52  
53 142 Scientific UK) system. SmartType software (Digital Scientific UK) was used for  
54  
55  
56 143 karyotyping purposes after being custom-adapted for porcine karyotyping according  
57  
58  
59  
60

1  
2  
3 144 to the standard karyotype as established by the Committee for the Standardized  
4  
5 145 Karyotype of the Domestic Pig (Gustavsson 1988). All staff were trained in the  
6  
7 146 analysis of porcine chromosomes using the in-house developed program KaryoLab  
8  
9  
10 147 Porc (Payne et al. 2009).

11  
12 148

#### 13 14 15 149 **Selection and preparation of subtelomeric BAC clones for FISH**

16  
17 150 BAC clones of approximately 150kb in size were selected using the Sscrofa Version  
18  
19 151 10.2 NCBI database ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) for each autosome and the X  
20  
21 152 chromosome. A lack of available BACs for the Y chromosome meant that this  
22  
23 153 chromosome was excluded from the study. End-sequenced BACs in the subtelomeric  
24  
25 154 region of the p-arm and q-arm of each chromosome with unique placement in the  
26  
27 155 genome were identified and ordered from both the PigE-BAC library (ARK-Genomics)  
28  
29 156 and the CHORI-242 Porcine BAC library (BACPAC). BAC DNA was isolated using the  
30  
31 157 Qiagen Miniprep Kit, the products of which were then amplified and directly labelled  
32  
33 158 by nick translation with FITC-Fluorescein-12-UTP (Roche) for p-arm probes and Texas  
34  
35 159 Red-12-dUTP (Invitrogen) for q-arm probes prior to purification.

36  
37  
38  
39  
40 160

#### 41 42 43 161 **Development of a novel Multiprobe device for translocation screening**

44  
45 162 Fluorescently labelled probes were diluted to a concentration of 10ng/μl in sterile  
46  
47 163 distilled water along with competitor DNA (Porcine Hybloc, Applied Genetics  
48  
49 164 Laboratories). Each probe combination contained a probe isolated from the distal p-  
50  
51 165 arm (labelled in FITC) and distal q-arm (labelled in Texas Red) from a single  
52  
53 166 chromosome. Where the chromosome is acrocentric, the most proximal sequence  
54  
55 167 was isolated (for simplicity sake, these were individually assigned with the  
56  
57  
58  
59  
60



1  
2  
3 168 chromosome number followed by the letter p in green type and the letter q in red  
4  
5 169 type, as indicated in S1 Fig 1 (supplementary material).  
6  
7

8 170  
9

10 171 The new device was based on the work of Knight et al (1996) using a proprietary  
11  
12 172 Chromoprobe Multiprobe® System device manufactured by Cytocell Ltd, in the UK.  
13  
14 173 Each probe combination (e.g. 1pq) was reversibly air dried on to a square of the  
15  
16 174 device in the orientation indicated in S1 Fig 1. The second part of the device consists  
17  
18 175 of a glass slide subdivided into 24 squares designed to align to the 24 squares on the  
19  
20 176 first part of the device upon which chromosome suspensions were fixed.  
21  
22  
23

24 177  
25

#### 26 178 **Fluorescence *in-situ* hybridisation (FISH)**

27

28  
29 179 Fixed metaphase preparations on the second part of the Multiprobe device (the  
30  
31 180 glass slide) were dehydrated through an ethanol series (2 minutes each in 2xSSC,  
32  
33 181 70%, 85% and 100% ethanol at room temperature). 1µl of formamide based  
34  
35 182 hybridisation buffer (Cytocell Hyb I) was pipetted onto each square of the first part  
36  
37 183 of the device containing the probe to dissolve the probes. The second part (glass  
38  
39 184 slide) was aligned over the first part (containing the rehydrated probes) pressed  
40  
41 185 together and warmed on a 37°C hotplate for 10 minutes. Probe and target DNA were  
42  
43 186 subsequently denatured on a 75°C hotplate for 5 minutes prior to hybridisation  
44  
45 187 overnight in a dry hybridisation chamber in a 37°C water bath. Following  
46  
47 188 hybridization, slides were washed (2 minutes in 0.4 × SSC at 72°C; 30 seconds in 2 ×  
48  
49 189 SSC/ 0.05% Tween 20 at room temperature), then counterstained using DAPI in  
50  
51 190 VECTASHIELD® anti-fade medium. Images were captured using an Olympus BX61  
52  
53 191 epifluorescence microscope with cooled CCD camera and SmartCapture (Digital  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 192 Scientific UK) system. Chromosome preparations from multiple animals were used to  
4  
5 193 verify correct mapping of each BAC.  
6

7  
8 194

9  
10 195 **Results and Discussion**

11  
12 196 **Karyotype analysis**

13  
14  
15 197 Karyotypes were successfully produced via a newly developed in-house service for a  
16  
17 198 total of 230 boars from different breeding populations with an average of 10  
18  
19 199 karyotypes created per boar. Four translocation carriers were identified by classical  
20  
21 200 cytogenetics with no abnormalities identified in the remainder. The translocations  
22  
23 201 were as follows t(1:2); t(7:10) (see Fig 1); t(7:12); and t(13:15).  
24  
25

26  
27 202

28  
29 203 **Development of the Multiprobe device**

30  
31 204 A total of 82 BACs were tested, of which ultimately 45 BACs mapped correctly and 37  
32  
33 205 did not map as anticipated. All FITC labelled probes mapped to the expected locus at  
34  
35 206 or near the p terminus of the chromosome with the exception of the first attempt  
36  
37 207 for a BAC (PigE-134L21) for chromosome 1p (which actually mapped to chromosome  
38  
39 208 8), along with a BAC for chromosome 10p (PigE-231H10) which mapped to  
40  
41 209 chromosome 3 and three BACs originally assigned to chromosome 9p, which  
42  
43 210 mapped elsewhere in the karyotype. After selecting alternative BACs bright green  
44  
45 211 signals were observed at the appropriate end of the chromosome. Surprisingly, 32 of  
46  
47 212 the 51 probes that were originally assigned to the q terminus of specific  
48  
49 213 chromosomes mapped to a place in the genome other than that which was  
50  
51 214 predicted. Of these, 24 clones (75%) mapped to the correct chromosome, but not to  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

215 the q terminus. An example is given in Fig 2 for chromosome 15 and the full list given  
216 in table 1.  
217  
218

For Peer Review

219 Table 1. Incorrectly mapped Porcine BACs and their assignment

Number	Arm	Clone Name	FISH Assignment	Same Chromosome?
1	p	PigE-134L21	8 p-arm	No
1	q	CH242-137C1	10 centromere	No
1	q	CH242-35I10	Multiple	No
1	q	CH242-83P21	7 centromere	No
2	q	CH242-188K23	2 centromere	Yes
2	q	CH242-230M23	2 centromere	Yes
2	q	CH242-441A1	2 centromere	Yes
2	q	PigE-117G14	2 p-arm	Yes
3	q	CH242-265K24	3 p-arm	Yes
3	q	PigE-221G14	3 p-arm	Yes
3	q	PigE-264D16	3 p-arm	Yes
5	q	CH242-133F9	5 p-arm	Yes
5	q	CH242-288F8	5 p-arm	Yes
5	q	PigE-127K14	5 p-arm	Yes
5	q	PigE-178M22	5 p-arm	Yes
7	q	CH242-272F22	7 centromere	Yes
7	q	CH242-518F14	7 centromere	Yes
7	q	PigE-208I10	3 q-arm	No
7	q	PigE-230H8	7 centromere	Yes
7	q	PigE-75E21	7 mid q-arm	Yes
9	p	CH242-215O14	9 centromere	Yes
9	p	CH242-44O5	9 centromere	Yes
9	p	CH242-178L4	9 centromere	Yes
10	p	PigE-231H10	3 p-arm	No
10	q	CH242-237D22	10 centromere	Yes
10	q	CH242-36D16	10 q-arm + extra signal on 1q	Yes
10	q	PigE-60N24	1 centromere	No
11	q	PigE-199B10	11 p-arm	Yes
11	q	PigE-232N19	11 p-arm	Yes
15	q	PigE-108N22	15 mid q-arm	Yes
16	q	CH242-4G9	16 p-arm	Yes
16	q	PigE-124C22	16 p-arm	Yes
16	q	PigE-173H6	16 p-arm	Yes
17	q	PigE-112L22	10 centromere	No
18	q	PigE-141I21	6 p-arm	No
X	q	CH242-447L20	X p-arm	Yes
X	q	PigE-214O4	13 centromere	No

220

1  
2  
3 221 The results therefore indicated that probes assigned to the q-arm were frequently  
4  
5 222 incorrectly mapped, with the majority of probes mapping to the correct  
6  
7 223 chromosome but the incorrect locus. Correctly mapping q-arm probes were  
8  
9  
10 224 eventually assigned by choosing BACs (using an *in-silico* approach) that were  
11  
12 225 assigned to larger, fully mapped contigs closest to the q-terminus.  
13  
14  
15 226  
16  
17 227 Ultimately a device was developed and tested rigorously that gave bright, punctate  
18  
19 228 signals (one green, one red) for each chromosome. Examples of the signals on  
20  
21 229 chromosome 1 in a chromosomally normal preparation are given in Fig 3. The newly  
22  
23 230 developed Multiprobe strategy was applied to 20 chromosomally normal  
24  
25  
26 231 preparations and each translocation carrier in order to confirm the cytogenetic  
27  
28 232 diagnosis. The device confirmed the diagnosis of the following translocations t(1:2);  
29  
30 233 t(7:10) (Figs 1 and 4); t (7:12); t(13:15). Moreover no abnormalities were seen in the  
31  
32  
33 234 other preparations (see below). A full list of subtelomeric BACs that give bright  
34  
35 235 signals on the appropriate chromosome arms is shown in table 2.  
36  
37  
38 236  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

237

238 **Table 2: Correctly Mapping BACs for each porcine chromosome arm.**

Chromosome	Arm	Clone Name	Chromosome	Arm	Clone Name
1	p	CH242-248F13	10	q	CH242-517L16
1	q	CH242-151E10	11	p	PigE-211E21
2	p	PigE-8G19	11	q	CH242-239O11
2	q	CH242-294F6	12	p	PigE-253K5
3	p	PigE-168G22	12	q	PigE-124G15
3	q	CH242-315N8	13	P	PigE-197C11
4	p	PigE-131J18	13	q	PigE-179J15
4	q	PigE-85G21	14	p	PigE-137C12
5	p	PigE-74P10	14	q	PigE-167E18
5	q	CH242-63B20	15	p	PigE-90C11
6	p	PigE-238J17	15	q	CH242-170N3
6	q	CH242-510F2	16	p	PigE-149F10
7	p	PigE-52L22	16	q	CH242-42L16
7	q	CH242-103I13	17	p	CH242-70L7
8	p	PigE-2N1	17	q	CH242-243H19
8	q	PigE-118B21	18	p	PigE-253N22
9	p	CH242-65G4	18	q	PigE-202I11
9	q	CH242-411M8	X	p	CH242-19N1
10	p	CH242-451I23	X	q	CH242-305A15

239

240

241 A further boar that had previously been diagnosed as karyotypically normal was re-

242 referred for analysis using the Multiprobe device which revealed a chromosome

243 translocation between chromosomes 5 and 6 that we missed by classical karyotyping

244 (Fig 5). Further analysis with chromosome painting for porcine chromosomes 5 and 6

245 on this boar revealed a cryptic translocation with the distal portions of the two

246 chromosomes exchanged (Fig 6). Karyotyping was limited by sub-optimal quality of

247 the original chromosome preparation, however results produced using the FISH

1  
2  
3 248 approach clearly identified the translocation despite the poor preparation and the  
4  
5 249 small size of the translocation.  
6  
7 250  
8  
9  
10 251 Results of this study provide proof of principle of an approach that can be used  
11  
12 252 successfully to diagnose chromosomal translocations that directly impact fertility in  
13  
14 253 pigs at a resolution previously difficult to achieve by standard karyotyping. There are  
15  
16  
17 254 three advantages of using this approach over classical karyotyping: The first is that it  
18  
19 255 detects more cryptic translocations than standard karyotyping otherwise would. The  
20  
21 256 boar indicated in this study is an example. Indeed, the fact that a previously  
22  
23 257 undetected cryptic translocation was identified would suggest that the actual  
24  
25 258 number of translocations in the boar breeding population might in fact be  
26  
27 259 significantly higher than previously reported. It is possible that these karyotypically  
28  
29 260 cryptic and unreported translocations are seen more frequently than expected but  
30  
31 261 that the routine use of multiple inseminations per sow may be diluting the effect on  
32  
33 262 the farrowing rates. The boar with a cryptic translocation in this study had a  
34  
35 263 significantly reduced farrowing rate and interestingly also had a significantly lower  
36  
37 264 "born dead" rate suggesting that the translocation in this case results in early  
38  
39 265 embryo loss. It would appear that the production of unbalanced gametes caused by  
40  
41 266 the translocation in question results in embryos which are not compatible with early  
42  
43 267 life causing early embryo mortality in a pattern that is also seen in humans (Tempest  
44  
45 268 & Simpson 2010). In humans, reciprocal translocations arise more frequently *de-*  
46  
47 269 *novo* rather than from being inherited from a carrier parent (Tempest & Simpson  
48  
49 270 2010). It would therefore be reasonable to suggest that the same pattern of familial  
50  
51 271 inheritance applies to pigs and other animals. The *de novo* nature of these  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 272 translocations supports the theory that all boars awaiting service should be screened  
4  
5 273 chromosomally to reduce the risk of using a hypoprolific animal for breeding  
6  
7 274 purposes. In fact, despite over 130 reciprocal translocations being reported in the  
8  
9  
10 275 literature, to date this is the first reported translocation to have occurred between  
11  
12 276 chromosomes 5 and 6 suggesting that this fits that category (Rothschild & Ruvinsky  
13  
14  
15 277 2011). Secondly, as in this case, when preparations are sub-optimal, this approach  
16  
17 278 provides necessary “back-up” to ensure accurate diagnosis. That is, provided FISH  
18  
19 279 signals are clear enough, confident diagnosis can be made on a single metaphase,  
20  
21  
22 280 regardless of the length of the chromosomes.  
23

24  
25 281

26  
27  
28 282 The final issue is that the device permits analysis by individuals who are less well  
29  
30 283 trained in karyotype analysis. Twenty years experience of teaching students to  
31  
32  
33 284 karyotype human and pig karyotypes ((Morris et al. 2007);(Gibbons et al. 2003)) has  
34  
35 285 demonstrated that the technical skills required to produce a karyotype reliably can  
36  
37 286 be variable between individuals and that animal-specific expertise is invaluable.  
38  
39 287 Indeed, although several laboratories have pioneered animal cytogenetics for the  
40  
41 288 purposes of AI boar (and bull) screening there are fewer now than in previous  
42  
43  
44 289 decades despite the continuing need to continue screening in this manner.  
45  
46  
47 290 Nonetheless, it should be made clear that specialist cytogenetic skills are still  
48  
49 291 required to make chromosome preparations reliably in the lab and to perform  
50  
51  
52 292 overall analyses. The scheme developed here should therefore be considered an  
53  
54  
55 293 adjunct to classical cytogenetics, not a replacement for it.  
56  
57  
58  
59  
60



1  
2  
3 294 A second outcome of this study was the revelation that a large number of BACs  
4  
5 295 isolated from the Swine Genome assembly mapped incorrectly. That is, those that  
6  
7 296 were predicted to map to the q-terminus of a particular chromosome that mapped  
8  
9  
10 297 elsewhere on the same chromosome. In many ways this contradicts our previous  
11  
12 298 results in which 100% of the BACs mapped to the predicted chromosomal location  
13  
14 299 (Groenen et al. 2012). The high level of mapping errors found in this study led to  
15  
16 300 further investigation of the clone placement with members of the Swine Genome  
17  
18 301 Sequencing Consortium. It became evident that the problem was the result of some  
19  
20 302 errors in the way in which parts of the draft pig genome sequence were assembled.  
21  
22 303 Specifically, analysis of the BAC sequences revealed that the high error rate was due  
23  
24 304 to misplacement of some of the smaller fingerprint contigs (fpc) within which the  
25  
26 305 BAC was located. These small fpcs did not have full sequence and orientation data  
27  
28 306 when the genome was assembled and it appears that these small poorly mapped  
29  
30 307 contigs were added to the end of the list of contigs for the relevant chromosomes.  
31  
32 308 This resulted in the sequences from the BACs in these poorly mapped contigs being  
33  
34 309 randomly added to the end of the relevant chromosomes, which explains why the  
35  
36 310 error rate was particularly high among BACs chosen to map to the subtelomeric q-  
37  
38 311 arm region. The genome assembly errors found throughout the course of this project  
39  
40 312 highlight the need for caution when choosing BACs for this purpose. In other words,  
41  
42 313 the porcine genome assembly still appears to have assembly flaws, despite being  
43  
44 314 initially considered to be one of the best assembled. These assembly errors are  
45  
46 315 particularly apparent when looking at structural rearrangements and should be  
47  
48 316 taken into consideration when planning future FISH mapping exercises, both for  
49  
50 317 BACs in the pig genome and when investigating the genomes of other animal species  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 318 (e.g. cattle, sheep). The errors highlighted in this paper have been passed to the  
4  
5 319 Swine Genome Sequencing Consortium and their results will be incorporated in an  
6  
7 320 improved pig genome assembly due to be released in 2016. With the rapid  
8  
9  
10 321 expansion in the number of newly sequenced animal genomes being published,  
11  
12 322 along with corresponding BAC libraries for many, the possibility of assembly errors  
13  
14  
15 323 should be an important consideration for future similar studies.  
16

17 324

18  
19 325 Now that a full set of porcine subtelomeric probes has been isolated and applied in  
20  
21 326 the manner described, screening efficiency can be improved by allowing the analysis  
22  
23 327 of the full chromosomal complement on one slide. Given the nature of  
24  
25  
26 328 translocations and their impact on fertility in pigs, the simple, rapid identification of  
27  
28  
29 329 (cryptic or otherwise) translocations will facilitate the detection and subsequent  
30  
31 330 removal of affected animals from the breeding population at an early stage. This has  
32  
33 331 the potential to lead to long-term improved productivity, delivering meat products in  
34  
35 332 a more cost effective and environmentally friendly way to a growing population. The  
36  
37  
38 333 widespread use of artificial insemination and the large market for superior boar  
39  
40 334 semen being sold to both small and large scale pig breeding operations suggests that  
41  
42  
43 335 improvements in productivity impact not just the large commercial breeders but also  
44  
45 336 the smaller farmers where reduced wastage may be more critical.  
46

47 337

48  
49  
50 338 Finally, the application of these subtelomeric FISH probes for translocation screening  
51  
52 339 is not necessarily limited to screening for translocations in pigs. Artificial  
53  
54 340 insemination is also widely used in cattle breeding with a high premium placed on  
55  
56  
57 341 bull semen of superior genetic merit. With sufficient alterations (i.e. incorporating  
58  
59  
60

1  
2  
3 342 cattle subtelomeric BACs) the device could be adapted to this and other species. In  
4  
5 343 addition the increasingly widespread use of embryo transfers in cattle would suggest  
6  
7 344 that the cow and the bull should both be screened for chromosomal translocations.  
8  
9 345 In fact, the cattle karyotype is more difficult to analyse reliably because of a diploid  
10  
11 346 number of 60, largely made up of similar sized acrocentric chromosomes. The cattle  
12  
13 347 karyotype therefore lends itself to the use of a FISH based screening approach such  
14  
15 348 as is described here, as does the largely acrocentric sheep karyotype ( $2n=54$ ).  
16  
17 349 Lessons regarding genome assembly learnt from this exercise would suggest that a  
18  
19 350 cautionary approach be taken when identifying BACs for this purpose and that a  
20  
21 351 combined *in-silico* and experimental (wet lab) approach is crucial in the development  
22  
23 352 of similar tools.  
24  
25  
26  
27  
28  
29  
30

### 31 354 **Conclusions**

32  
33 355 FISH based translocation screening technique developed in this study is a powerful  
34  
35 356 and reliable approach to translocation screening with great potential to be adapted  
36  
37 357 to other species.  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 360 **Acknowledgements:**  
4

5 361 The authors would like to thank Alan Mileham for his help with initial concept  
6  
7 362 development and Edward Sutcliffe for his assistance with blood samples. The work  
8  
9  
10 363 was funded by a Knowledge Transfer Partnership grant awarded to DK Griffin and  
11  
12 364 Cytocell <https://connect.innovateuk.org/web/ktp>. The funder provided support in  
13  
14  
15 365 the form of salaries for GF, but did not have any additional role in the study design,  
16  
17 366 data collection and analysis, decision to publish, or preparation of the manuscript.  
18

19  
20 367

21 368 **Abbreviations:**  
22

23  
24 369 BAC – Bacterial artificial chromosome; FISH – Fluorescence *in situ* hybridisation; SSC  
25  
26 370 – *Sus scrofa*  
27

28  
29 371

30  
31 372 **Competing interests statement:**  
32

33 373 Martin Lawrie, Richard Frodsham and Gothami Fonseka are employees of Cytocell  
34  
35 374 Ltd; Grant Walling is an employee of JSR Genetics. Both companies could, in theory,  
36  
37 375 profit from publication of this manuscript. Rebecca O'Connor, Darren Griffin and  
38  
39 376 Alan Archibald declare no conflict of interests except in that they have collaborated  
40  
41 377 with these companies to generate successful co-funded grant applications. At time  
42  
43 378 of writing, the device described above is not a commercial product marketed by  
44  
45 379 Cytocell, however it is reasonable to suggest that it may become one in the future.  
46  
47

48  
49 380

50  
51 381 All institutional and national guidelines for the care and use of laboratory animals  
52  
53 382 were followed.  
54

55  
56 383  
57  
58  
59  
60

384 **References**

- 385 Dawson, A.J. et al., 2002. Cryptic chromosome rearrangements detected by  
386 subtelomere assay in patients with mental retardation and dysmorphic features.  
387 *Clinical genetics*, 62(6), pp.488–494.
- 388 Ducos, A. et al., 2007. Chromosomal control of pig populations in France: 2002–2006  
389 survey. *Genetics Selection Evolution*, 39(5), pp.583–597.
- 390 Ducos, A. et al., 2008. Cytogenetic screening of livestock populations in Europe: an  
391 overview. *Cytogenetic and genome research*, 120(1-2), pp.26–41.
- 392 Gadea, J., 2005. Sperm factors related to in vitro and in vivo porcine fertility.  
393 *Theriogenology*, 63(2), pp.431–444.
- 394 Gadea, J., Sellés, E. & Marco, M.A., 2004. The predictive value of porcine seminal  
395 parameters on fertility outcome under commercial conditions. *Reproduction in  
396 domestic animals*, 39(5), pp.303–308.
- 397 Gibbons, N.J., Evans, C. & Griffin, D.K., 2003. Learning to karyotype in the university  
398 environment: a computer-based virtual laboratory class (KaryoLab) designed to  
399 rationalize time for the tutor/researcher and to encourage more students to  
400 engage in cytogenetics. *Cytogenetic and genome research*, 101(1), pp.1–4.
- 401 Groenen, M.A.M. et al., 2012. Analyses of pig genomes provide insight into porcine  
402 demography and evolution. *Nature*, 491(7424), pp.393–398.
- 403 Gustavsson, I., 1990. Chromosomes of the pig. *Advances in veterinary science and  
404 comparative medicine*, 34, pp.73–107.
- 405 Gustavsson, I., 1988. Standard karyotype of the domestic pig. *Hereditas*, 109(2),  
406 pp.151–157.
- 407 Horsley, S.W. et al., 1998. Del (18p) shown to be a cryptic translocation using a  
408 multiprobe FISH assay for subtelomeric chromosome rearrangements. *Journal  
409 of medical genetics*, 35(9), pp.722–726.
- 410 Knight, S.J. et al., 1996. Development and clinical application of an innovative  
411 fluorescence in situ hybridization technique which detects submicroscopic  
412 rearrangements involving telomeres. *European journal of human genetics: EJHG*,  
413 5(1), pp.1–8.
- 414 Merck CM, Kahn, S Line E., 2010. *The Merck veterinary manual: Merck & Company,  
415 Incorporated.*
- 416 Mompert, F. et al., 2013. 3D organization of telomeres in porcine neutrophils and  
417 analysis of LPS-activation effect. *BMC cell biology*, 14(1), p.30.
- 418 Morris, W.B. et al., 2007. Practicable approaches to facilitate rapid and accurate  
419 molecular cytogenetic mapping in birds and mammals. *Cytogenetic and genome  
420 research*, 117(1-4), pp.36–42.
- 421 Payne, A.M. et al., 2009. The use of an e-learning constructivist solution in workplace  
422 learning. *International Journal of Industrial Ergonomics*, 39(3), pp.548–553.

- 1  
2  
3 423 Pinton, A. et al., 2000. Chromosomal abnormalities in hypoprolific boars. *Hereditas*,  
4 424 132(1), pp.55–62.  
5  
6 425 Ravnan, J.B. et al., 2006. Subtelomere FISH analysis of 11 688 cases: an evaluation of  
7 426 the frequency and pattern of subtelomere rearrangements in individuals with  
8 427 developmental disabilities. *Journal of medical genetics*, 43(6), pp.478–489.  
9  
10 428 Rodríguez, A. et al., 2010. Reproductive consequences of a reciprocal chromosomal  
11 429 translocation in two Duroc boars used to provide semen for artificial  
12 430 insemination. *Theriogenology*, 74(1), pp.67–74.  
13  
14 431 Rothschild, M.F. & Ruvinsky, A., 2011. *The genetics of the pig*, CABI.  
15  
16 432 Tempest, H.G. & Simpson, J.L., 2010. Role of preimplantation genetic diagnosis (PGD)  
17 433 in current infertility practice. *Int J Infertil Fetal Med*, 1, pp.1–10.  
18  
19 434 The BPEX Yearbook 2013-14, (AHDB), 2014.  
20  
21 435 United States Department of Agriculture, F.A.S., 2015. *Livestock and Poultry: World*  
22 436 *Markets and Trade*.  
23  
24 437  
25 438  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 439 **Figure legends**  
4

5 440 Figure 1. Clone ID PigE-108N22 labelled in Texas Red which should map to the distal  
6  
7 441 end of SSC15 but appears halfway along this acrocentric chromosome. The FITC  
8  
9  
10 442 labeled probe mapped correctly. *Scale bar* 10  $\mu\text{m}$   
11

12 443

13  
14 444 Figure 2: Standard DAPI banded karyotype of a boar carrying a 7:10 Reciprocal  
15  
16  
17 445 Translocation  
18

19 446

20  
21 447 Figure 3. Labelled probes for SSC7 illustrating a reciprocal translocation between  
22  
23 448 SSC7 and SSC10. *Scale bar* 10  $\mu\text{m}$ .  
24  
25

26 449

27  
28  
29 450 Figure 4: FISH image of correctly mapping BAC clones for chromosome 1 tested on a  
30  
31 451 chromosomally normal sample showing clear, punctate signals. *Scale bar* 10 $\mu\text{m}$   
32

33 452

34  
35  
36 453 Figure 5. BAC clones for SSC5 (p-arm labelled in FITC and q-arm labelled in Texas  
37  
38 454 Red) showing a translocation between chromosome 5 and 6. Despite the suboptimal  
39  
40 455 chromosome preparation the translocation is clearly visible. *Scale bar* 10 $\mu\text{m}$   
41

42 456

43  
44  
45 457 Figure 6. Chromosome paints for SSC5 (FITC) and SSC6 (Texas Red) illustrating the  
46  
47 458 cryptic translocation that had been previously undetectable from the karyotype.  
48

49 459 *Scale bar* 10 $\mu\text{m}$ .  
50

51 460

52  
53  
54  
55 461 **Supporting Information**  
56  
57  
58  
59  
60

1  
2  
3 462 S1 Figure 1. Multiprobe device layout of labelled BAC clones by chromosome with a  
4  
5 463 Texas Red labelled probe and FITC labelled probe for each chromosome air dried  
6  
7  
8 464 onto the same square  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43

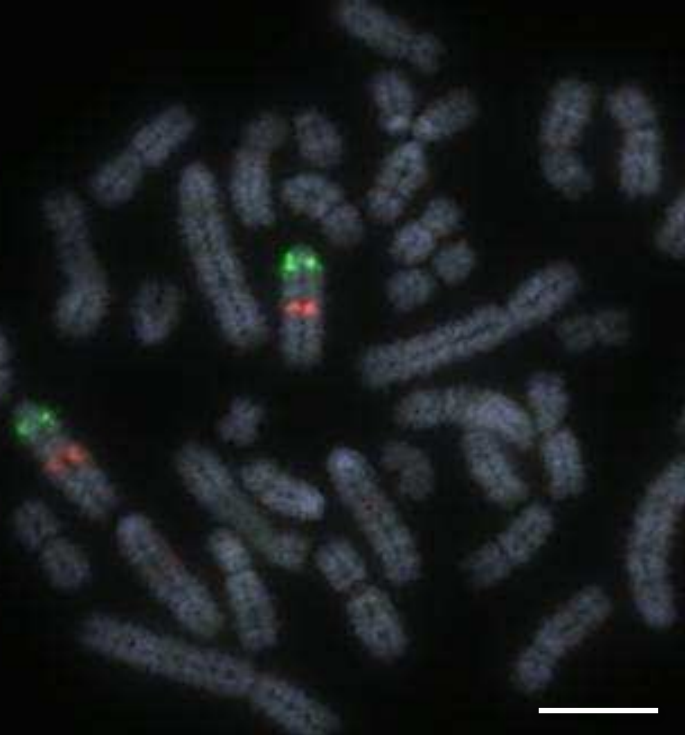
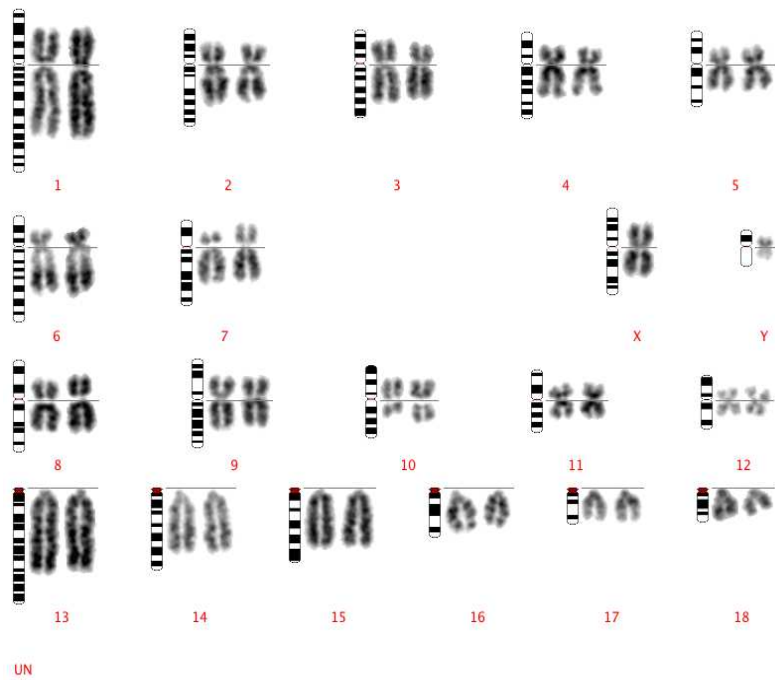


Figure 1: Clone ID PigE-108N22 labelled in Texas Red which should map to the distal end of SSC15 but appears halfway along this acrocentric chromosome. The FITC labeled probe mapped correctly. *Scale bar* 10  $\mu$ m.



28 Figure 2. Standard DAPI banded karyotype of boar  
29 carrying a 7:10 reciprocal translocation  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43

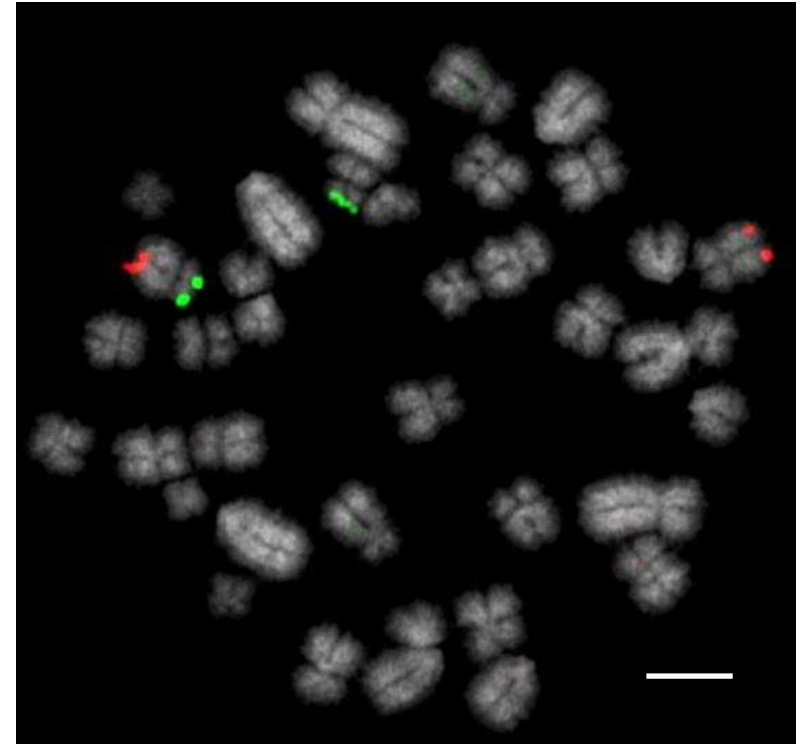
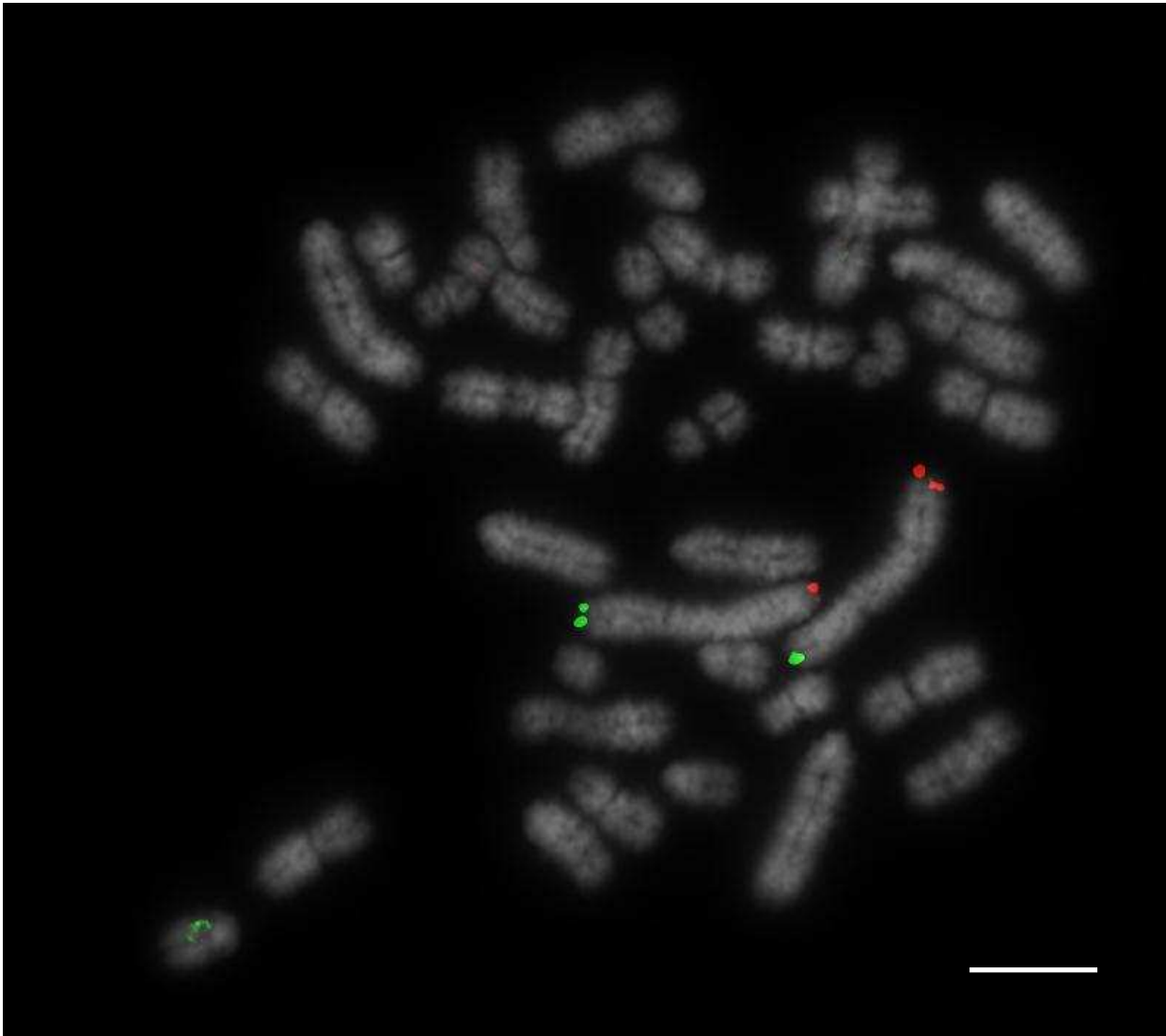


Figure 3. Labelled probes for SSC7 illustrating a  
reciprocal translocation between SSC7 and SSC10.  
Scale bar 10  $\mu$ m.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36



37 Figure 4. FISH image of correctly mapping BAC clones for chromosome 1 tested on  
38 a chromosomally normal sample showing clear, punctate signals. *Scale bar* 10µm.  
39  
40

41  
42  
43

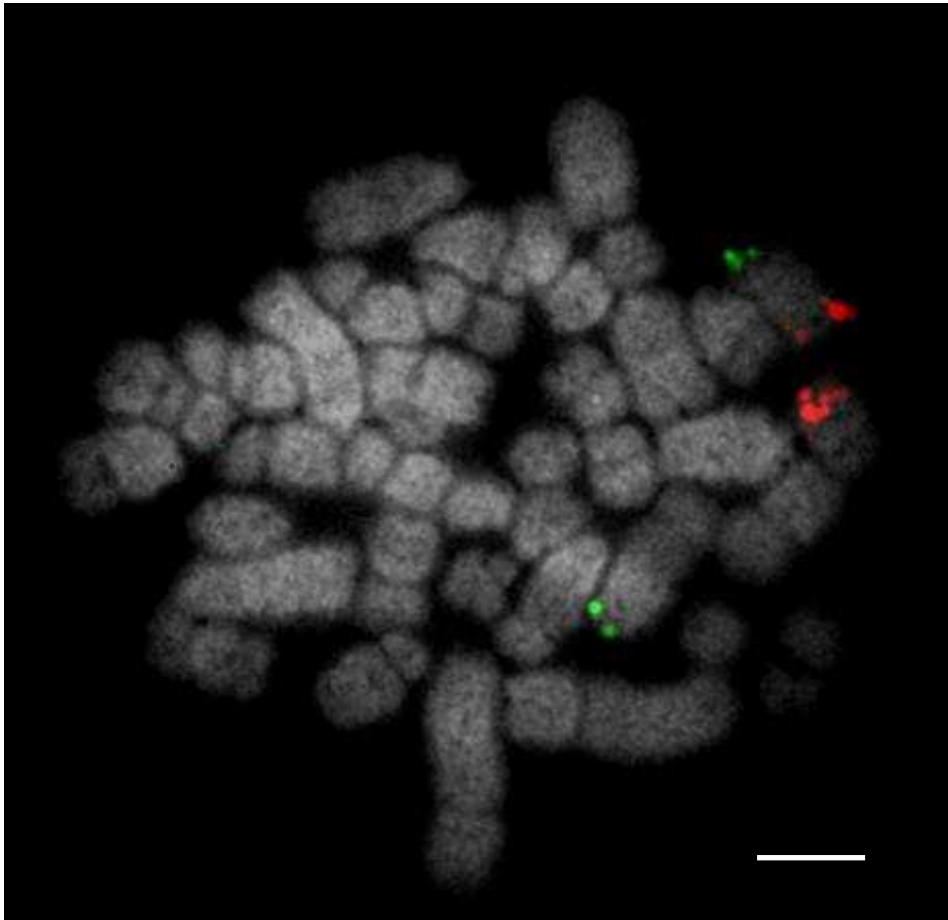


Figure 5. BAC clones for SSC5 (p-arm labelled in FITC and q-arm labelled in Texas Red) showing a translocation between chromosome 5 and 6. Despite the suboptimal chromosome preparation the translocation is clearly visible. *Scale bar 10 $\mu$ m*

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43

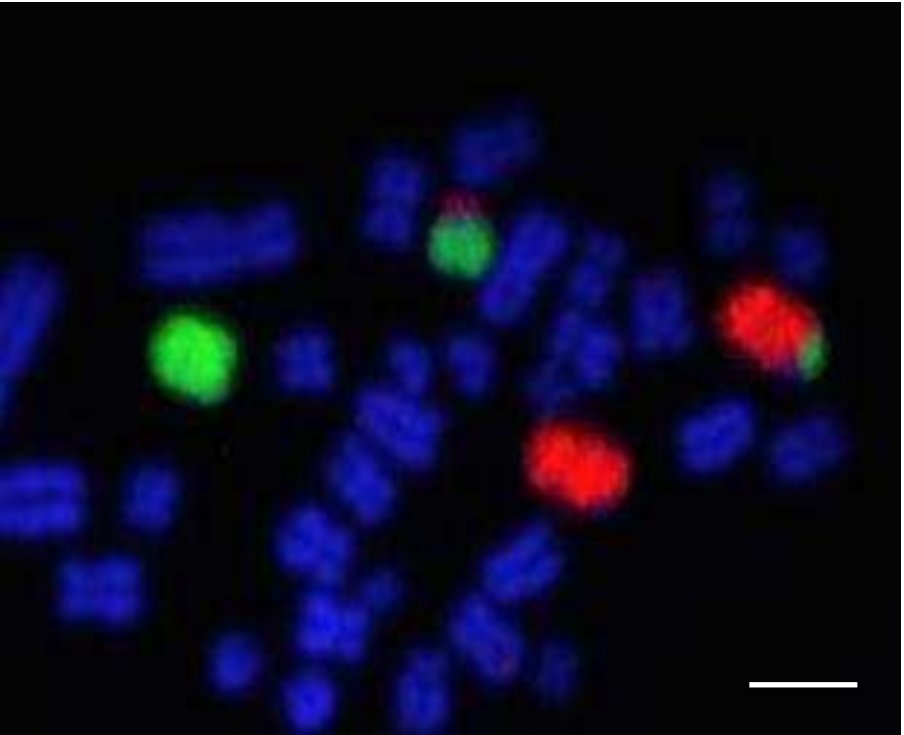


Figure 6. Chromosome paints for SSC5 (FITC) and SSC6 (Texas Red) illustrating the cryptic translocation that had been previously undetectable from the karyotype. *Scale bar* 10 $\mu$ m.



Figure 1. Multiprobe device layout of labelled BAC clones by chromosome with a Texas Red labelled probe and FITC labelled probe for each chromosome air dried onto the same square

# 21st International Colloquium on Animal Cytogenetics and Gene Mapping

(21st ICACGM), June 7–10, 2014, Ischia, Naples, Italy

Published online: 23 July 2014  
© Springer Science+Business Media Dordrecht 2014

## ABSTRACTS

*Edited by: L. Iannuzzi, A. Perucatti, A. Iannuzzi, A. Pauciullo, V. Genualdo, D. Incarnato, L. Keller (CNR-ISPAAAM, Naples, Italy)*

## Plenary Session

### L1

#### Molecular cytogenetics in veterinary diagnosis and research

Malcolm A. Ferguson-Smith (maf12@cam.ac.uk)  
Department of Veterinary Medicine, Cambridge University, Madingley Road, Cambridge CB3 0ES, UK

Fifty years ago Ingemar Gustavsson made the first observation of a chromosome abnormality in a farm animal. The common rob1/29 translocation in cattle has been associated with reduced fertility, prompting efforts at eradication. Since then many other chromosome abnormalities have been identified in domestic species, including sex chromosome abnormalities in race horses, and these have been discussed at many meetings of the ICACGM. Interest in diagnostic veterinary cytogenetics has grown alongside research into comparative genomics and karyotype evolution of farm animals. The current place of molecular cytogenetics in both diagnosis and research in this field is discussed here in several demonstration projects, including artificial

insemination, the fertility of mules and infertility in farm and companion animals due to sex chromosome disorders. Chromosome-specific painting probes, and especially 7-colour FISH probes, have been valuable additions to classical techniques in the resolution of problems associated with high diploid numbers and difficult to distinguish acrocentrics in animal cytogenetics.

### L2

#### Chromosomes, genome analysis and a transforming landscape of applications in the twenty-first century

Bhanu P. Chowdhary (BChowdhary@cvm.tamu.edu)  
College of Veterinary Medicine & Biomedical Sciences,  
Texas A&M University, College Station, Texas 77845

Chromosome analysis has been the center-point for nuclear genome analysis for a long time—perhaps over a century. While initially it provided a peek into the structure and organization of the chromosome, it later led to the discovery of chromosome abnormalities and their impact on phenotypes. Also, it allowed increased understanding of the potential causes for various diseases. However, since the advent of a range of gene mapping and genome analysis techniques beginning early 1990s, time and again it has been suggested that the scope and utility of chromosome analysis will decline and fade into oblivion. Understandably, the “Golden Era” of chromosome analysis may be over, however, some basic aspects of analysis coupled with molecular techniques are indispensable and

that determines the male gender, many other genes, responsible for inducing DSDs, were identified. It's useful to specify how the cytogenetic analysis has often represented a valid methodology for the identification of these genes. Some animal species are excellent models for the identification and study of these genes since the sexual diseases that show are entirely similar to those present in the human species but characterized by a higher frequency due to the lower rate of abortion against them. In this lesson we will present the current knowledge on the subject and on the genes recognized as responsible for sexual disorders. We will also discuss how animals can be a valuable tool to deepen this knowledge that has yet unexplored aspects.

## O1

### Analysis of male infertility: a case study in pigs

H. Barasc<sup>3,2,1</sup>, N. Mary<sup>1,2,3</sup>, A. Ducos<sup>3,2,1</sup>, S. Ferchaud<sup>4</sup>, I. Raymond Letron<sup>5</sup>, M. Yerle<sup>1,2,3</sup>, H. Acloque<sup>1,2,3</sup>, A. Pinton<sup>1,2,3</sup> (a.pinton@envt.fr)

<sup>1</sup>INRA, UMR 1388 Génétique, Physiologie et Systèmes d'Élevage, F-31326 Castanet-Tolosan, France; <sup>2</sup>Université de Toulouse INPT ENSAT, UMR 1388 Génétique, Physiologie et Systèmes d'Élevage, F-31326 Castanet-Tolosan, France; <sup>3</sup>Université de Toulouse INPT ENVT, UMR 1388 Génétique, Physiologie et Systèmes d'Élevage, F-31076 Toulouse, France; <sup>4</sup>GenESI Génétique, Expérimentation et Système Innovants Poitou Charentes F-17700 Saint-Pierre-d'Amilly, France; <sup>5</sup>Université de Toulouse, INP, ENVT, UMS 006, Département des Sciences Biologiques et Fonctionnelles, Laboratoire d'Histopathologie, F-31076 Toulouse, France

Infertility is a significant problem in humans, affecting up to 15 % of couples. Male (co-)factors, leading mostly to spermatogenesis failure, are involved in almost 50 % of the cases. Male infertility is also of major interest in farm-animal populations. On the one hand, a reduction in male fertility can be responsible for major economic losses at the farm level. Otherwise, due to the importance of the male pathway in the creation and dissemination of genetic progress, male infertility can lead to a strong reduction of the efficiency of genetic selection programs. Origins of infertility are still unknown in more than 90 % of the cases in Human but they may be genetic or environmental causes. We recently

developed a research program aiming at deciphering the putative genetic mechanism explaining the bad semen quality parameters observed in boars routinely controlled before reproduction, thanks to cytogenetic, array-CGH and array-painting analyses.

Preliminary results will be presented with a particular attention for an oligo-astheno-terato-spermic boar carrying an asymmetric reciprocal translocation involving chromosomes SSC1 and SSC14. CNVs research, meiotic pairing, recombination and segregation analyses, as well as breakpoints characterization have been carried out and the corresponding results will be presented.

## O2

### Identification of Chromosomal Translocations in Pigs using FISH with Subtelomeric Probes and the development of a novel screening tool for their application

R.E. O'Connor<sup>1</sup>, G. Fonseka<sup>1,2</sup>, D.K. Griffin<sup>1</sup> (ro84@kent.ac.uk)

<sup>1</sup>University of Kent, Department of Biosciences, Canterbury, Kent CT2 7NH; <sup>2</sup>Cytocell Ltd 3–4 Technopark Newmarket Road Cambridge CB5 8PB

Reciprocal chromosome translocations have established to affect fertility in pigs leading to reduced litter sizes and hypoprolificacy. With an increasing emphasis in the commercial pig breeding industry on using a small population of boars for artificial insemination, the potential economic costs of using hypoprolific boars are significant. At present screening for translocations is only performed by karyotyping which, while technically straightforward, requires animal specific expertise for karyotype analysis, which can be unattractive to the industry. The use of subtelomeric probes and fluorescence in situ hybridisation (FISH) eliminates the need for this level of expertise whilst also offering greater accuracy and the ability to identify cryptic translocations. At present, however a universal FISH based screening test for porcine translocations has yet to be developed.

Probes were designed that map to the subtelomeric regions of each chromosome arm to enable detection using FISH. BACs were identified from the subtelomeric region of the p-arm and q-arm of each porcine chromosome and directly labelled with Texas Red or FITC (p-arm and q-arm respectively) prior to



fluorescence microscopy and image capturing using SmartCapture 3 software (Digital Scientific UK).

Clear signals were obtained from each subtelomeric probe. These were tested on normal animals and animals that exhibit translocations, providing preliminary evidence that this technique is a valid tool for the identification of translocations that affect fertility in pigs.

When combined with a tool originally developed for humans to enable the simultaneous detection of all porcine chromosomes on one slide (Multiprobe™ Device), the speed and cost of chromosomal analysis for translocations that affect fertility will be greatly improved, therefore offering significant benefits to animal genetic research and the animal breeding industry.

### O3

#### The incidence of translocations in young breeding boars in Canada

A. Quach<sup>1</sup>, T. Revay<sup>1</sup>, M. Macedo<sup>1</sup>, S. Wyss<sup>2</sup>, B. Sullivan<sup>2</sup>, W.A. King<sup>1</sup> (tquach@uoguelph.ca)

<sup>1</sup>University of Guelph, OVC Biomedical Sciences, N1G2W1 – Guelph- Canada; <sup>2</sup>Canadian Centre for Swine Improvement Inc., K1A0C6 – Ottawa, Canada

The objective of the project was to carry out the first systematic screening program for chromosomal abnormalities in young breeding boars in Canada. To date, a total 300 young boars from 4 different breeds (Duroc, Landrace, Pietrain and Yorkshire) were karyotyped by G-banding. Four previously unreported reciprocal translocation including rcp(1;5), rcp(3;4), rcp(8;13) and rcp(7;15) and one previously reported Robertsonian translocation rob(13;17) were found. Consequently, the frequency of chromosome abnormalities in this study was 1.67 %. By extending the sampling to other members of the pedigree, it was determined that rcp(3;4) and rob(13;17) were inherited from their dams and rcp(8;13) was a “de novo” event. Comparing with the herd average, average litter size of rcp(3;4), rcp(8;13) and rcp(7;15) translocation carrier boars was noted to be reduced (24 %, 24 % and 38 %, respectively) while for carriers of rob(13;17), it was only slightly reduced (9 %). Interestingly, for rcp(3;4), the overall reduction in litter sizes for female carriers was substantially lower (only 4 %) compared to male carriers (24 %). Chromosome analysis of live offspring from 2 full litters of carrier boars showed a 20 and 40 % transmission rate

to progeny for rcp(7;15) and rob(13;17), respectively. More studies need to be carried out to further investigate the effects of these translocations. (Research support was obtained from NSERC, Agriculture and Agri-Food Canada, and the Canada Research Chairs program).

### O4

#### Mix of two chromosomal aberrations in a newborn calf 2n=60,XX, t(11;25)(q11;q14-21)

A. Iannuzzi<sup>1</sup>, A. Perucatti<sup>1</sup>, A. Pauciullo<sup>1</sup>, V. Genuardo<sup>1</sup>, D. Incarnato<sup>1</sup>, L. Pucciarelli<sup>1</sup>, L. De Lorenzi<sup>2</sup>, G. Varricchio<sup>3</sup>, D. Matassino<sup>3</sup>, P. Parma<sup>2</sup>, L. Iannuzzi<sup>1</sup> (alessandra.iannuzzi@cnr.it)

<sup>1</sup>Laboratory of Animal Cytogenetics and Gene Mapping, National Research Council (CNR), ISPAAM, Naples, Italy; <sup>2</sup>Department of Animal Science, Agricultural Faculty of Sciences, Milan, Italy; <sup>3</sup>ConSDABI, Sub-national Focal Point of FAO (Mediterranean Biodiversity), Benevento, Italy.

A newborn calf of the Agerolese breed underwent cytogenetic investigation because presented hyperflexion forelimbs, red eyes and inability to stand up. Anamnesis revealed the mother, phenotypically normal, was carrier of a t(11;25)(q11,q14-21). The newborn died after a few weeks and no internal alterations were found by veterinarian after the post mortem examination. The mother presented, after a cytogenetic investigation, a reciprocal translocation between chromosome 11 and 25 and the presence of two ders: der11 and der25, for the position of corresponding centromere. On the other hand, the veal revealed a different chromosomal aberration in comparison to her mother. In fact, after R-banded karyotype, the calf showed both chromosomes 25, one chromosome 11 and one der (der25). FISH analysis was performed with the same BAC clones used to detect the translocation in the mother: BAC142G06 mapped on the proximal region of both BTA25 and der25; BAC513H08 mapped to BTA 25q22dist; BAC533C11 mapped to the proximal region of BTA11 and der25. Finally, we confirmed both the localization of the breakpoints on band q11 (centromere) of chromosome 11 and q14-21 of chromosome 25, and the loss of the der11. In this way, it is showed a different cytogenetic aberration in the veal: a partial trisomy of chromosome 25 and a partial monosomy of chromosome 11. We have

Nebrodi mountains where this breed is raised in Sicily. SCE-test applied on 42 pigs from Casertana breed (22 males and 20 females) and 19 pigs from Siciliana breed (8 males and 11 females) revealed no statistical differences between the SCE-mean number in Casertana pig ( $7.13 \pm 3.20$ ) than that ( $6.87 \pm 3.12$ ) achieved in Siciliana pig. Statistical differences were found between males ( $7.26 \pm 3.38$ ) and females ( $6.59 \pm 2.90$ ) of Siciliana pig breed, as well as between females of Casertana ( $7.24 \pm 3.26$ ) and Siciliana ( $6.59 \pm 2.90$ ) breeds, while no statistical differences were found between males of the breeds, as well as between males and females of Casertana breed.

*Acknowledgements.* The study has been supported by project “RARECA, PSR, Misura 214 e2” of Campania region and “CISIA-GenePig” project, National Research Council (CNR) of Italy.

### P36

#### Cytogenetic analyses in rabbits feed in presence of Verbascoside: SCE-test

V. Genualdo<sup>1</sup>, A. Perucatti<sup>1</sup>, A. Iannuzzi<sup>1</sup>, A. Pauciullo<sup>1</sup>, L. Pucciarelli<sup>1</sup>, C. Iorio<sup>1</sup>, D. Incarnato<sup>1</sup>, M. Palazzo<sup>2</sup>, D. Casamassima<sup>2</sup>, L. Iannuzzi<sup>1</sup> (viviana.genualdo@ispaam.cnr.it)

<sup>1</sup>National Research Council (CNR), ISPAAM, Laboratory of Animal Cytogenetics and Gene Mapping, Naples, Italy; <sup>2</sup>Department of Agriculture, Environment and Food (AAA), University of Molise, Campobasso, Italy.

Phenylpropanoid glycosides (PPG), like other phenolic compounds, are powerful antioxidants. Beside phenolic compounds, verbascoside, shows the highest scavenger activity in the PPG and has high antioxidant power in comparison with other phenolic compounds. Previous studies by using in vitro exposure of human blood lymphocytes to verbascoside reported a significant increasing of chromosome fragility compared to control. In the present study four homogeneous groups of rabbits (six animals per group) were used to test in vivo the verbascoside by feeding the animals without Verbascoside and Licopeno (*control – group A*), with lycopene (*5 mg/Kg of feeding, group B*), with verbascoside (*5 mg/Kg of feeding, group C*) with verbascoside and lycopene (*5 mg/Kg of feeding each, group D*). Peripheral blood cultures were performed in

three different times: at 0, 40 and 80 days of the experiment. Two types of cell cultures were performed: without (normal cultures) for the AC-test (chromosome and chromatid breaks) and with BrdU (10 µg/ml), the latter added 26 h before harvesting, for the SCE-test. In the present study only data from SCE-test are presented. Mean number of SCEs were generally lower at both 40 and 80 days in groups B, C and D, compared with the same groups at zero day. In particular, they were statistically ( $P < 0.01$ ) lower at 40 and 80 days when using lycopene. In conclusion, on the basis of SCE-test applied on cells of rabbits treated in vivo with verbascoside or/and with lycopene, no chromosome fragility increasing were observed in cells of rabbit feed with verbascoside. However, a final conclusion will be done when data from AC-test will be available.

### Evolutionary and Comparative Cytogenetics

#### L11

#### Avian cytogenetics goes functional

D. K Griffin<sup>1</sup>, M. Farre<sup>2</sup>, P. Lithgow<sup>1</sup>, R. O'Connor<sup>1</sup>, K. Fowler<sup>1</sup>, M. Romanov<sup>1</sup>, D. Larkin<sup>2</sup> (d.k.griffin@kent.ac.uk)

<sup>1</sup>School of Biosciences, University of Kent, Canterbury, UK; <sup>2</sup>Department of Comparative Biomedical Sciences, Royal Veterinary College, University of London, NW1 0TU, London, UK.

Whole chromosomes (and sub-chromosomal homologous synteny blocks (HSBs)) have great significance in molecular studies of genome evolution. In birds, our ability to define chromosomes and HSBs precisely has however been impeded by a near intractable karyotype and so has focused primarily on comparative molecular cytogenetics (zoo-FISH) of the largest chromosomes (1–10+Z). Availability of multiple avian genome sequence assemblies has however allowed us, for the first time, to identify chromosomal syntenies across species. In recent work we have made use of comparative maps for 20+ avian genome assemblies (plus out-groups) and presented them on “Evolution Highway” an open-access, interactive freely available comparative chromosome browser designed to store and visualise comparative chromosome maps.

This browser (<http://evolutionhighway.ncsa.uiuc.edu>) is used to visualize comparative genome organization and to identify and visualize the different types of evolutionary breakpoint regions (EBRs) in chromosomes, e.g., lineage specific, ordinal, superordinal, and reuse. Comparative analysis of all available genomes is providing insight into the mechanisms of chromosome change through correlation of EBRs with transposable elements and non-allelic homologous recombination. Gene ontology analysis is revealing interesting correlations with avian specific phenotype and function. Focus on six genomes (chicken, turkey, duck, zebra finch, ostrich and budgerigar) with both the largest N50s and supporting molecular cytogenetic information, has allowed us to assemble a putative ancestral avian karyotype and identify the key changes that led to the gross genome organization of representatives in the major avian clades (Palaeognathae, Galliformes, Anseriformes and Neoaves). We describe, for the first time, numerous inter-chromosomal rearrangements in a Paleognathaeous bird (the ostrich), plus rearrangements in the budgerigar (Psittaciformes) and 15 other species. Intra-chromosomal evolutionary change in all species studied, can be derived, most parsimoniously, by a series of inversions, inter-chromosomal rearrangements by fissions and fusions. Increased chromosome rearrangement is associated with differentiation in certain clades, with the most intrachromosomal changes (primarily inversions) occurring in the zebra finch (Passeriformes) since its divergence from its sister group, the Psittaciformes 54MYA. This is coincident with the evolution of passerine-specific phenotypes e.g. vocal learning. Results also suggest that the Galloanserae (especially chicken) underwent the fewest changes compared to the ancestral karyotype; notably these birds appear, from fossil evidence, to be the most similar to ancient avian ancestors. We thus present the most comprehensive analysis of chromosomal rearrangements in birds to date and draw novel conclusions about their mechanisms of origin and association with avian-specific phenotypic features.

**L12**

### **Evolution and molecular dynamics of centromeres in the genus *Equus***

E. Giulotto ([elena.giulotto@unipv.it](mailto:elena.giulotto@unipv.it))  
Dipartimento di Biologia e Biotecnologie, Università di Pavia, Italy

The centromere is the locus directing chromosome segregation at cell division. The mechanism by which centromere identity is specified on chromosomal DNA sequences has been deeply enigmatic, with a clear dependence on the epigenetic inheritance of the centromeric histone, CENP-A. While a degree of autonomy of centromere placement along the chromosome has been established by studies of human neocentromeres and observation of evolutionary centromere repositioning, a role for DNA sequence in driving centromere location remains to be elucidated. The typical association of mammalian centromeres with extensive arrays of highly repetitive satellite DNA, has so far hampered a detailed molecular dissection of centromere function and evolution.

In previous work, we discovered that, in the genus *Equus* (horses, asses and zebras), centromere repositioning during evolution was exceptionally frequent and that satellite DNA and centromeres are often uncoupled in this genus. We then described the first native satellite-free centromere discovered in a mammal, that of horse chromosome 11; using a combination of molecular and cytogenetic approaches we recently demonstrated that the precise positioning of this native mammalian centromere is highly variable, even on the two homologous chromosomes in a single individual. These results corroborate the hypothesis that CENP-A is the principal determinant of centromere identity, but they make a much deeper point: CENP-A location along the DNA polymer is not fixed but rather exhibits a diffusion-like behavior.

We are now characterizing a number of satellite-less centromeres in asses and zebras; preliminary observations on the molecular organization of centromeres, based on the exploitation of this powerful model system, will be presented.

**O18**

### **New evolutionary differences between cattle and goat**

L. De Lorenzi<sup>1</sup>, J. Planas<sup>2</sup>, E. Rossi<sup>3</sup>, P. Parma<sup>1,4</sup> ([pietro.parma@unimi.it](mailto:pietro.parma@unimi.it))

<sup>1</sup>Department of Agricultural and Environmental Sciences, Milan University, Milan, Italy; <sup>2</sup>Department of Systems Biology, Vic University, Vic, Spain; <sup>3</sup>Medical Genetics, University of Pavia, Pavia, Italy; <sup>4</sup>CNR-ISPAAAM, Laboratory of Animal Cytogenetics and Gene Mapping, Naples, Italy.

# 13<sup>th</sup> Annual Preimplantation Genetic Diagnosis International Society (PGDIS) Meeting Abstracts

© Springer Science+Business Media Dordrecht 2014

## Compiled by

Professor Darren K Griffin,  
School of Biosciences  
University of Kent  
Canterbury CT2 7NJ, UK

---

## Invited speaker presentations

### S1. The future of PGD: A vision Cohen J<sup>1</sup>

Reprogenetics LLC and ART Institute of Washington, Livingston, New Jersey, USA

### S2. Novel insights into chromosome abnormalities and fertility Turner J<sup>1</sup>

MRC National Institute of Medical Research, Mill Hill, London, UK

### S3. Human aneuploidy: A conspiracy of sex, age, and the environment Hunt P<sup>1</sup>

School of Molecular Biosciences and Center for Reproductive Biology, Washington State University

### S4. Combined genome-wide crossover maps and chromosome segregation outcomes reveal crossover assurance and a novel segregation pattern in single, adult human oocytes Ottolini C<sup>1,2\*</sup>, Newnham L<sup>3\*</sup>, Calpalbo A<sup>4\*</sup>, Natesan SA<sup>5</sup>, Joshi HA<sup>5</sup>, Herbert AD<sup>3</sup>, Gray S<sup>3</sup>, Griffin DK<sup>2</sup>, Rienzi L<sup>4</sup>, Ubaldi F<sup>4</sup>, Handyside A<sup>1,2</sup>, and Hoffmann E<sup>3</sup>

London Bridge Fertility, Gynaecology and Genetics Centre, London UK

Department of Biosciences, University of Kent, Canterbury, UK

MRC Genome Damage and Stability Centre, School of Life Sciences, University of Sussex, Brighton, UK

G.E.N.E.R.A., Rome, Italy

BlueGnome Ltd, Illumina, UK

### S5. Chromosomal mosaicism in preimplantation embryos Wilton LJ<sup>1</sup>

Preimplantation Genetics, Melbourne IVF, East Melbourne, Australia

### S6. Blastocyst stage aneuploidy screening Capalbo A<sup>1</sup>

G.E.N.E.R.A., Clinica Valle Giulia, Rome

### S7. Single cell RNA-seq analysis of preimplantation embryos and its implication on PGD Fan G<sup>1</sup>

Department of Human Genetics, David Geffen School of Medicine, UCLA, Los Angeles, California, USA

### S8. Morphokinetics and ploidy Campbell A<sup>1</sup>

CARE Fertility Group, John Webster House, Lawrence Drive, Nottingham Business Park, Nottingham, NG8 6PZ

### S9. Preventing the transmission of mitochondrial DNA disease Taylor RW<sup>1</sup>

Wellcome Trust Centre for Mitochondrial Research, Newcastle University, Newcastle upon Tyne, UK

### S10. Art in ART (artwork inspired by IVF and PGD) Glover G

### S11. Alternative molecular approaches for aneuploidy detection and IVF outcome prediction Fragouli E<sup>1,2</sup>

Reprogenetics UK, Oxford, UK

Nuffield Department of Obstetrics and Gynaecology, University of Oxford, Oxford, UK

blastulation (OR=1.78 (CI: 1.11–2.87);  $p<0.05$ ; OR=4.69 (CI: 2.89–7.61),  $p<0.0001$ ). Average blastulation rate of biopsied embryos did not differ to non-biopsied embryos (182/461 (39.48 %) vs. 145/315 (46.03 %);  $p>0.05$ ). No difference was found in either chemical (46/79 (58.23 %) vs. 28/48 (58.33 %);  $p>0.05$ ) or clinical (38/79 (48.10 %) vs. 26/48 (54.17 %);  $p>0.05$ ) pregnancy rates between the PGS and non-PGS groups, respectively. Implantation rates were similar ( $41.56\pm 0.586$  vs.  $34.72\pm 0.712$ ;  $p>0.05$ ) between the two groups. A significantly lower amount of embryos were transferred in PGS cycles than in non-PGS cycles ( $1.46\pm 0.006$  vs.  $1.88\pm 0.010$ ;  $p<0.05$ ). We have found that blastulation occurs in a higher frequency in euploid embryos, but blastulation does not predict normality on the chromosome level. In our settings, half of the diagnosed blastocysts were aneuploid. The developmental potential of embryos does not seem to be altered following cleavage stage embryo biopsy, similar blastulation rates were found compared to non-biopsied embryos. Also, when euploid embryo is available the same pregnancy and implantation rates can be achieved as in non-PGS cycles with a lower number of embryos transferred. In vitro fertilization (IVF) combined with PGS provides a viable option for patients having multiple failed IVF cycles, advanced maternal age or recurrent pregnancy loss. Also, unnecessary embryo transfer and embryo freezing can be avoided when PGS strengthen embryo selection. Although, it has to be noted that PGS carries a risk for cancellation of embryo transfer.

### **P34. Use of time-lapse imaging to investigate the impact of embryo biopsy on morphokinetic criteria**

Sadraie M<sup>1</sup>, Bolton VN<sup>2</sup>, Thornhill AR<sup>1,3</sup>, Griffin DK<sup>1</sup>

<sup>1</sup>School of Biosciences, University of Kent, Canterbury, UK

<sup>2</sup>Guy's and St. Thomas' NHS Trust, London, UK

<sup>3</sup>Illumina, Cambridge, UK

Previous studies have suggested that removing blastomeres by cleavage stage embryo biopsy does not have an adverse effect on subsequent development. Others have challenged this notion however proposing that there is a possibility that the biopsy technique can damage the embryo sufficiently to negate any potential

benefits of chromosome aneuploidy screening. The main aim of this study was to investigate the impact of the cleavage stage biopsy technique on human embryo morphology as determined by time-lapse imaging. The availability of time-lapse devices allows visualisation of developing embryos in a controlled environment without any disruption to culture conditions. In this study an 'Embryoscope', a time lapse device from Unisense Fertilitech was used. This device is composed of an incubator with time-lapse microscopy, and an embryo viewer workstation. The Embryoscope time-lapse system provides images through different focal planes at 20 min intervals. Images that had been previously recorded were annotated according to morphological markers such as the extrusion of the second polar body, pronuclear appearance and disappearance, cleavage checks and the start of cell compaction leading to blastulation. Currently, 850 embryos from preimplantation genetic screening cycles have been annotated and an annotation policy has been produced. Analysis of these data is currently ongoing.

### **P35. Evaluation of aneuploidy of chromosomes 1, 16, 12 and 18 in boar sperm samples**

Sadraie M<sup>1</sup>, Fowler KE<sup>1</sup>, O'Connor RE<sup>1</sup> and Griffin DK<sup>1</sup>

<sup>1</sup>School of Biosciences, University of Kent, Canterbury, Kent, UK

High rates of chromosomal abnormalities in boar sperm may be correlated to decreased fertility in boars but not necessarily to decreased fertilising potential of any given sperm. To date only a limited number of aneuploidy studies on boar semen have been reported. In this study we have made efforts to optimise fluorescence in situ hybridisation (FISH) for use with boar spermatozoa to test for chromosomal abnormality levels. FISH may serve as a useful tool to assess boar fertility, as well as complementing morphological and functional assessments with genomic screening. The main aim of this study was to optimise dual colour FISH to test boar sperm chromosomal abnormalities. Bacterial artificial chromosomes (BACs) specific to chromosomes of interest were selected, grown and labelled by nick translation. A multicolour FISH technique was developed to detect aneuploidy in the sperm of boars using DNA probes specific



for small regions of chromosomes 1, 16, 2 and 18. Altogether, 2,032 sperm cells from 2 boars (Large white breed) were examined. The average frequency of sperm with disomy for chromosomes 1, 16 and 12 were 0.099, 0.099 and 0.097 % respectively. Disomy for chromosome 18 was not observed. The average frequencies of diploidy were 0.099 % for 1-1-16-16 and 0.097 % for 12-12-18-18. There was no significant difference between rates of disomy and diploidy and the rate of disomy did not differ significantly by chromosome. This study will be repeated in a further 4 breeds in order to perform a comparative study and minor breed-specific differences were noted.

### **P36. Face to face with the patient: a UK perspective on PGS**

Sage K<sup>1</sup>

<sup>1</sup>The London Bridge Centre/London Women's Clinic, UK

Preimplantation Genetic Diagnosis (PGD) has become a widely accepted treatment for couples at risk of having children with serious genetic conditions. Preimplantation Genetic Screening (PGS) however, has much less acceptance in the UK. Despite emerging data supporting the use of PGS using newer technologies able to detect aneuploidy of all chromosomes in gametes or embryos, there is still a lack of robust evidence from randomized control trials. This "second generation" PGS remains overshadowed by the disappointing outcomes of the "first generation" PGS performed using fluorescent in-situ hybridization (FISH); a technology that screened a limited number chromosomes, was prone to technical errors and demonstrated no improvement in pregnancy rates after 10 years. A 2013 Cochrane review of Assisted Reproductive Technology (ART) stated that PGS is an "ineffective intervention" as it does not show an improvement in live birth rates. The report references a PGS review from 2006 based on FISH technology. The HFEA website information for patients last updated the PGS pages in June 2009, and on these report average success rates for the year 2008, again, based on FISH studies. The HFEA maintains that "centres are required to validate the use of PGS (i.e. demonstrate there is evidence) for each category of patients they offer it to (e.g. advanced maternal age, recurrent implantation failure,

recurrent pregnancy loss and male factor infertility)". Against this background of uncertainty about the efficacy of PGS, aneuploidy screening is increasing in private practice in the UK, driven partly by patient demand and by clinicians reviewing emerging data from clinics collaborating with laboratories to perform their own RCTs to provide the robust data. Patients embarking on PGS treatment need to fully understand the risks, limitations, range of outcomes and potential benefits of the screening and this service can be effectively provided by genetic counselling. The genetic counseling role in ART is developing as increasing genomic advances are beginning to penetrate into clinical practice. Genetic counselors can provide education and support for patients, clinicians and embryologists. In this session I discuss the expanding role of the genetic counselor in ART and specifically in PGS. Preparing patients for aneuploidy screening is essential so that couples and individuals have a realistic understanding of PGS screening in context of their medical and fertility history. Establishing patients' expectations prior to starting treatment can be invaluable in how they adapt to often negative outcomes which can affect their future reproductive decisions. Offering continuing support through the treatment cycle is also an integral part of the genetic counselling process. Additionally, interpreting aneuploidy results, reviewing embryology and providing recommendations pre-embryo transfer can often be challenging especially if the results are inconclusive. Critical time-constrained decision making can be alleviated by preparing patients for the range of possible outcomes prior to treatment start. Clinics offering PGS add an additional workload on embryology teams and genetic counselors can facilitate decision making with patients prior to embryo transfer. The genetic counselor role as a PGD/PGS coordinator is central to managing an effective Genetic Service within the context of ART.

### **P37. Reciprocal translocation and inversion carriers have higher risk of partial aneuploidies than Robertsonian translocation carriers in Preimplantation Genetic Diagnosis (PGD) cycles**

Sandalinas M<sup>1</sup>, Garcia-Guixé E<sup>1</sup>, Jiménez-Macedo A<sup>1</sup>, Arjona C<sup>1</sup>, Balias E<sup>1</sup>, Alsina E<sup>1</sup> and Giménez C

<sup>1</sup>Reprogenetics Spain, Barcelona, Spain

## 20th International Chromosome Conference (ICCXX)

50th Anniversary, University of Kent, Canterbury, 1st–4th September 2014

Darren K. Griffin · Katie E. Fowler · Peter J. I. Ellis ·  
Dean A. Jackson

© Springer Science+Business Media Dordrecht 2015



Dear Colleagues—**Welcome to ICCXX**

On behalf of the International Chromosome and Genome Society (ICGS), in September 2014 we welcomed several hundred delegates to the beautiful city of Canterbury. The programme was distinguished as always by its high scientific interest and contained ample opportunity for social interaction.

It is 50 years since Cyril Darlington first initiated the (then “Oxford”) Chromosome Conferences and this

meeting was the 20th to be held. Dubbed by his biographer (Oren Solomon Harman) as “the man who invented the chromosome”, Cyril Dean Darlington was born in Chorley, Lancashire in 1903. He was educated at Mercer’s School, Holborn, 1912–17, St. Paul’s School, 1917–20 and then, in 1920–23, came to Kent to study at Wye College, Ashford—just down the road from the conference venue. In 1923 he began an association of more than 30 years with the John Innes Institute, starting as a volunteer but later becoming head of Cytology in 1937 and Director in 1939. It was at the John Innes that he did much of his groundbreaking work on chromosomes, augmented by expeditions overseas and through collaboration with many distinguished British, American and Russian colleagues. He resigned in 1953 and accepted the Sherardian Professorship of Botany at Oxford where he took a keen interest in the Botanic Garden, creating “the Genetic Garden.” He vigorously promoted the cause of teaching genetics in the University, retiring in 1971 and remaining in Oxford where he continued to study and publish prolifically until his death in 1981.

Darlington’s legacy is that he was the world’s leading expert on chromosomes of his time and one of the leading biological thinkers of the twentieth century. He sought to answer nature’s biggest biological questions such as how species arise and how variation occurs. Often suffering rebuke, isolation, and obscurity along the way, he lived through Nazi atrocities, the Cold War, the molecular revolution, eugenics, the Lysenko controversy, the Civil Rights movement, the formation of the welfare state and the differing social views of man’s place in the natural world. Darlington’s work provoked him to ask questions

---

D. K. Griffin · K. E. Fowler · P. J. I. Ellis  
School of Biosciences, University of Kent,  
Canterbury CT2 7NJ, UK

D. A. Jackson  
Life Sciences, The University of Manchester,  
Carys Bannister Building, Dover St., Manchester M13 9PL,  
UK

## S31: 'The Avian Genome Explosion'

Tom M Gilbert

Centre for GeoGenetics and Lundbeck Foundation Pathogen Palaeogenomics Group, Natural History Museum of Denmark, University of Copenhagen

Over the past 3 years a global consortium has sequenced and collated a dataset of 48 avian genomes, chosen to represent at least one species of all avian orders. Thanks to the expertise present among the collaborators, this unique dataset has been mined in the phylogenetic and evolutionary genomic context, in such a way that over 25 manuscripts are currently in peer review. I introduce the history and rationale behind this project, provide a rapid tour through some of the highlights of the dataset, and end by describing what developments interested parties can expect over the next year.

## S32: Avian Chromonomics goes functional

Darren K Griffin<sup>1</sup>, Marta Farre<sup>2</sup>, Pamela Lithgow<sup>1</sup>, Rebecca O'Connor<sup>1</sup>, Katie Fowler<sup>1</sup>, Michael N Romanov<sup>1</sup>, Denis Larkin<sup>2</sup>

<sup>1</sup>School of Biosciences, University of Kent, Canterbury, UK; <sup>2</sup>Department of Comparative Biomedical Sciences, Royal Veterinary College, University of London, NW1 0TU, London, UK

Whole chromosomes (and sub-chromosomal homologous synteny blocks (HSBs)) have great significance in molecular studies of genome evolution. In birds, our ability to define chromosomes and HSBs precisely has however been impeded by a near intractable karyotype and so has focused primarily on comparative molecular cytogenetics (zoo-FISH) of the largest chromosomes (1–10+Z). Availability of multiple avian genome sequence assemblies has however allowed us, for the first time, to identify chromosomal syntenies across species. In recent work we have made use of comparative maps for 20+ avian genome assemblies (plus out-groups) and presented them on "Evolution Highway" an open-access, interactive freely available comparative chromosome browser designed to store and visualise comparative chromosome maps. This browser (<http://evolutionhighway.ncsa.uiuc.edu>) is used to visualize comparative genome organization and to identify and visualize the different types of evolutionary breakpoint regions (EBRs) in chromosomes, e.g., lineage specific, ordinal,

superordinal, and reuse. Comparative analysis of all available genomes is providing insight into the mechanisms of chromosome change through correlation of EBRs with transposable elements and non-allelic homologous recombination. Gene ontology analysis is revealing interesting correlations with avian specific phenotype and function. Focus on six genomes (chicken, turkey, duck, zebra finch, ostrich and budgerigar) with both the largest N50s and supporting molecular cytogenetic information, has allowed us to assemble a putative ancestral avian karyotype and identify the key changes that led to the gross genome organization of representatives in the major avian clades (Palaeognathae, Galliformes, Anseriformes and Neoaves). We describe, for the first time, numerous inter-chromosomal rearrangements in a Paleoganthaeous bird (the ostrich), plus rearrangements in the budgerigar (Psittaciformes) and 15 other species. Intra-chromosomal evolutionary change in all species studied, can be derived, most parsimoniously, by a series of inversions, inter-chromosomal rearrangements by fissions and fusions. Increased chromosome rearrangement is associated with differentiation in certain clades, with the most intrachromosomal changes (primarily inversions) occurring in the zebra finch (Passeriformes) since its divergence from its sister group, the Psittaciformes 54MYA. This is coincident with the evolution of passerine-specific phenotypes e.g. vocal learning. Results also suggest that the Galloanserae (especially chicken) underwent the fewest changes compared to the ancestral karyotype; notably these birds appear, from fossil evidence, to be the most similar to ancient avian ancestors. We thus present the most comprehensive analysis of chromosomal rearrangements in birds to date and draw novel conclusions about their mechanisms of origin and association with avian-specific phenotypic features.

O1: Human and mouse artificial chromosomes (HAC/MAC), and their characteristics

Oshimura M<sup>1</sup>, Kazuki Y<sup>1</sup>, Uno N<sup>1</sup>, Hoshiya H<sup>2</sup>

<sup>1</sup>Chromosome Engineering Research Center, Tottori University, Tottori, Japan; <sup>2</sup>Cell and Developmental Biology, University College London, London, UK

Human and mouse artificial chromosomes (HAC/MAC) are exogenous mini-chromosomes artificially generated mainly by either a 'top-down approach' (engineered creation) or a 'bottom-up approach' (de



sequencing, the genetic and physical maps are not available for the majority of the de novo sequenced genomes. To overcome this problem for assemblies that employ long-insert libraries (5–40 Kbp) we recently developed the reference-assisted chromosome assembly (RACA) algorithm (Kim et al., 2013). This method relies on both the raw sequencing data (reads) and comparative information; the latter is obtained from alignments between the target (de novo sequenced), a closely related (reference) and more distantly related (outgroup) genomes.

Using RACA followed by the manual FISH or PCR verification steps we are reconstructing the chromosome organisation of 19 bird species sequenced by the G10K community. We use the publically available chicken (*Gallus gallus*) and zebra finch (*Taeniopygia guttata*) chromosome assemblies as either reference or outgroup for each reconstruction depending on their phylogenetic relationships with each target species. Initially, we established the optimal RACA parameters for a bird chromosome assembly reconstruction using the duck (*Anas platyrhynchos*) and budgerigar (*Melopsittacus undulatus*) super-scaffolds assembled with the support from physical maps. This step allowed us to test the reliability of RACA reconstructions for bird genomes. Due to a higher evolutionary conservation of the bird karyotype compared to the mammalian one, we have achieved ~97 % accuracy of scaffold adjacencies in our predicted chromosome fragments compared to the ~93–96 % accuracies reported for mammals (Kim et al., 2013). We detected ~4–28 % of scaffolds in different target bird genomes that are either chimeric or containing genuine lineage-specific evolutionary breakpoint regions. Some of these scaffolds will be selected for follow up PCR or FISH verifications. All RACA reconstructions will become publicly available from our Evolution Highway comparative chromosome browser <http://evolutionhighway.ncsa.uiuc.edu/birds/> and will be further utilised to study connections between the chromosome evolution, adaptation and phenotypic diversity in birds and other vertebrates.

O22: Reconstruction of the putative Saurian karyotype and the hypothetical chromosome rearrangements that occurred along the Dinosaur lineage

O'Connor RE<sup>1</sup>, Romanov MN<sup>1</sup>, Farré<sup>2</sup> M, Larkin DM<sup>2</sup>, Griffin DK<sup>1</sup>

<sup>1</sup>University of Kent, School of Biosciences, Canterbury, Kent CT2 7NJ; <sup>2</sup>Department of Comparative Biomedical Sciences, Royal Veterinary College, University of London, London, NW1 0TU

Dinosaurs hold a unique place both in the history of the earth and the imagination of many. They dominated the terrestrial environment for around 170 million years during which time they diversified into at least 1000 different species. Reptilia, within which they are placed is one of the most remarkable vertebrate groups, consisting of two structurally and physiologically distinct lineages—the birds and the non-avian reptiles, of which there are 10,000 and 7,500 extant species respectively. The dinosaurs are without doubt the most successful group of vertebrate to have existed. They survived several mass extinction events before finally non-avian dinosaurs were defeated 66 million years ago in the Cretaceous-Paleogene extinction event, leaving the neornithes (modern birds) as their living descendants. Aside from the huge phenotypic diversity seen in this group, the birds and non-avian reptiles interestingly display similar karyotypic patterns (with the exception of crocodylians); with the characteristic pattern of macro and micro chromosomes, small genome size and few repetitive elements, suggesting that these were features exhibited in their common ancestor.

In this study, the availability of multiple reptile genome sequences (including birds) on an interactive browser (Evolution Highway) allowed us to identify multi species homologous synteny blocks (msHSBs) between the putative avian ancestor (derived from six species of extant birds), the Lizard (*Anolis carolinensis*) and the Snake (*Boa constrictor*). From these msHSBs we were able to produce a series of contiguous ancestral regions (CARs) representing the most likely ancestral karyotype of the Saurian (ancestor of archosaurs and lepidosaurs) that diverged from the mammalian lineage 280 mya. From this we have hypothesised the series of inter and intra-chromosomal rearrangements that have occurred along the dinosaur (archosaur) lineage to the ancestor of modern birds (100 mya) and along the lepidosaur lineage to the modern snake and lizard using the model of maximum parsimony.

Our study shows that relatively few chromosomal rearrangements took place over this period with an average of one inter or intra-chromosomal (translocations and inversions respectively) rearrangement occurring approximately every 2 million years. The majority

of these rearrangements appear to be intra-chromosomal suggesting an overall karyotypic stability, which is consistent with that of that of modern birds. Our results support the hypothesis that the characteristically avian genome was present in the saurian ancestor and that it has remained remarkably stable in the 280 million years since. It is credible therefore to suggest that this ‘avian-style’ genome may be one of the key factors in the success of this extraordinarily diverse animal group.

P1: The analysis of transcriptional regulation by cohesin and its loader with semi-in vitro reconstitution methods

Akiyama K, Bando M, Shirahige K

Research Center for Epigenetic Disease, Institute for Molecular and Cellular Biosciences, The University of Tokyo, Tokyo, Japan

Sister chromatid cohesion (SCC) is crucial to ensure accurate chromosome segregation during mitosis. The cohesin complex mediates SCC, and recent studies show cohesin and NIPBL/Mau2 complex, a loader protein required for the loading of cohesin onto chromatin, as important player in transcriptional regulation and chromatin architecture. Discoveries of mutations in subunits of cohesin and NIPBL in human developmental disorders, so-called cohesinopathies, reveal crucial roles for cohesin in development, cellular growth, and differentiation. However, it is still unclear how cohesin and its loader work in the transcriptional regulation. To reveal the complicated mechanisms played by cohesin and its loader in transcriptional regulation, we applied in vitro Pre-initiation complex (PIC) and Early Elongation Complex (EEC) assembly systems. In this system, we used the biotin-labeled DNA template, which contained 5xGAL4 DNA binding motifs, adenovirus late promoter sequence and a part of luciferase gene. After binding of activator protein, GAL4-VP16 recombinant protein, to this DNA, PIC and EEC assembly were induced by addition of the nuclear extract from HeLa cells. Each component of protein complex formed on template DNA was monitored by Western blotting. We showed that PIC factors, mediators, general transcriptional factors and RNA polII, were recruited to the template, which depended on the activator-binding. Further, we observed cohesin- and NIPBL/Mau2-binding to the template, and their recruitments also depend on the activator binding. Interestingly, cohesin seemed to get more stably bound after addition of activator.

Furthermore, when we treated lysate with 5,6-dichloro-1- $\beta$ -D-ribofuranosyl-benzimidazole (DRB), a cyclin-dependent kinase (CDK) inhibitor, we found that DNA binding of NIPBL and Mau2 is dramatically enhanced. And we performed immunoprecipitation by NIPBL antibody in PIC and EEC assembly condition, and analyzed the interacting proteins by LC-MS/MS. As a result, NIPBL interacted with mediators only under the activator-binding condition. Taken together, we propose that cohesin-loader and cohesin together regulates step that controls activation of mediators and paused RNA polII nearby promoter.

P2: Crossing experiments reveal gamete contribution into appearance of di- and triploid hybrid frogs in *Pelophylax esculentus* population systems

Dedukh DV<sup>1</sup>, Litvinchuk SN<sup>2</sup>, Rosanov JM<sup>2</sup>, Shabanov DA<sup>3</sup>, Krasikova AK<sup>1</sup>

<sup>1</sup>Department of Biology, Saint-Petersburg State University, Saint-Petersburg, Russia; <sup>2</sup>Institute of Cytology, Russian Academy of Sciences, Saint-Petersburg, Russia; <sup>3</sup>Department of Biology, V.N. Karazin Kharkiv National University, Kharkiv, Ukraine

Speciation through hybridization is connected with appearance of interspecies hybrids which can survive and reproduce owing to changes in their gametogenesis. In animals, these changes lead to appearance of clonal animals, which for successful reproduction usually depend on parental species and lack of recombination during gamete formation. Polyploidization can resolve these problems and may lead to emergence of new species. *Pelophylax esculentus* complex (complex of European water frogs) represents one of the appropriate models for studying interspecies hybridization and processes of polyploidization. Hybrid nature of the *P. esculentus* (RL genotype, 2n=26) was confirmed after crossings of two parental species *P. ridibundus* (RR genotype, 2n=26) and *P. lessonae* (LL genotype, 2n=26). Nevertheless absence of one parental species (*P. lessonae*) and abundance of triploid hybrid frogs (RRL and LLR genotypes, 3n=39) in population systems at the East of Ukraine challenged us to understand how di- and triploid hybrids can appear and prosper in population systems where hybrids exist only with *P. ridibundus* (R-E type population system). To answer this question we performed cytogenetic analysis of tadpoles appeared after artificial crossing experiments of

contrast can be related to the different involvement of sex-specific sex chromosomes in female meiosis subjected to the female meiotic drive under male versus female heterogamety. Essentially, the male-specific Y chromosome is not involved in female meiosis and is therefore sheltered against the effects of the female meiotic drive affecting the X chromosome and autosomes. Conversely, the Z and W sex chromosomes are both present in female meiosis. Nonrandom segregation of these sex chromosomes as a consequence of their rearrangements connected with the emergence of multiple sex chromosomes would result in a biased sex ratio, which should be penalized by selection. Therefore, the emergence of multiple sex chromosomes should be less constrained in the lineages with male rather than female heterogamety. Our broader phylogenetic comparison across amniotes supports this prediction. We suggest that our results are consistent with the widespread occurrence of female meiotic drive in amniotes.

**P8: Pds5 recruits Esco1 to establish sister chromatids cohesion**

Minamino M<sup>1</sup>, Ishibashi M<sup>1</sup>, Nakato R<sup>1</sup>, Sutani T<sup>1</sup>, Tanaka H<sup>1</sup>, Kato Y<sup>1</sup>, Negishi L<sup>2</sup>, Hirota T<sup>3</sup>, Bando M<sup>1</sup>, Shirahige K<sup>1</sup>

<sup>1</sup>Laboratory of Genome Structure and Function, Research Center for Epigenetic Disease, Institute of Molecular and Cellular Biosciences, The university of Tokyo, Tokyo, Japan.; <sup>2</sup>Laboratory of Cancer Stem Cell Biology, Research Center for Epigenetic Disease, Institute of Molecular and Cellular Biosciences, The university of Tokyo, Tokyo, Japan. <sup>3</sup>Cancer Institute of the Japanese Foundation for Cancer Research (JFCR), Tokyo, Japan.

Sister chromatids cohesion is mediated by cohesin and is essential for accurate chromosome segregation. The cohesin subunit SMC1, SMC3 and Rad21 form a huge tripartite ring within which sister DNAs are thought to be entrapped. This event requires the establishment of cohesion via the acetylation of SMC3 by Esco1 and Esco2 acetyltransferase in human. These two proteins function in partially non-redundant manner, and behave differently from each other during the cell cycle. The Esco2 protein is only detected in S phase and is degraded by the proteasome after DNA replication. In contrast, Esco1 binds to chromosomes throughout the cell cycle and undergoes phosphorylation during mitosis. Despite

the critical role of two Esco proteins in establishment of cohesion, the regulation of these proteins is largely unexplored. Recent studies have reported not only that fission yeast Pds5 interacts with Eso1 in a yeast two-hybrid assay, but also that Pds5 is essential for SMC3 acetylation at least in yeast and in mouse embryonic fibroblasts (MEFs). It is therefore suggested that elucidating the functional mechanism of Pds5 would provide a clue to understand the regulatory mechanism of Esco1 and Esco2. Here, we show that Pds5, a cohesin regulatory subunit bound to Rad21, is essential for Esco1 to establish cohesion via SMC3 acetylation whereas it would be dispensable for Esco2 activity. Correspondingly, Pds5 interacts exclusively with Esco1, and this depends on Esco1 unique domain. Replacement of endogenous Esco1 by its mutant unable to interact with Pds5 reveals the requirement of this interaction for the establishment of cohesion. We further demonstrate that Esco1 localizes to cohesin binding sites during interphase, which requires the Esco1-Pds5 interaction. These results indicate that Pds5 recruits Esco1 to cohesin binding sites to establish sister chromatids cohesion. Moreover, our systematic mass spectrometric analysis of Esco1 identifies 7 sites as the mitosis-specific phosphorylation sites, among which 1 site resides within the Pds5-binding region. We found that this phosphorylation depends on Aurora B kinase, and that it attenuates the interaction of Esco1 with Pds5. We therefore propose that Aurora B suppresses Esco1 activity during mitosis.

**P9: Evaluation of aneuploidy of autosome chromosomes in boar sperm samples**

Sadraie M, Fowler KE, O'Connor RE, Griffin DK  
School of Biosciences, University of Kent, Canterbury, Kent, CT2 7NJ, UK

High rates of chromosomal abnormalities in boar sperm may be correlated to decreased fertility in boars but not necessarily to decreased fertilising potential of any given sperm. To date only a limited number of aneuploidy studies on boar semen have been reported. In this study we have made efforts to optimise fluorescence in situ hybridisation (FISH) for use with boar spermatozoa to test for chromosomal abnormality levels. FISH may serve as a useful tool to assess boar fertility, as well as complementing morphological and functional assessments with genomic screening. The main aim of this study was to optimise dual colour FISH to test boar

sperm chromosomal abnormalities. Bacterial artificial chromosomes (BACs) specific to chromosomes of interest were selected, grown and labelled by nick translation. A multicolour FISH technique was developed to detect aneuploidy in the sperm of boars using DNA probes specific for small regions of autosome chromosomes (1–18). Altogether, 22,226 sperm cells from 3 breeds (Large white, Landrace and Hampshire) were examined. The average frequencies of sperm with disomy for all autosome chromosomes were 0.099 %. The average frequencies of diploidy for all autosome chromosomes were 0.189 %. There was no significant difference between rates of disomy and diploidy and the rate of disomy did not differ significantly by chromosome. This study will be repeated in a further three breeds in order to perform a comparative study and minor breed-specific differences were noted.

P10: Mitotic and meiotic chromosomes of *Tityus mattogrossensis* and *Tityus silvestris* (Scorpiones, Buthidae)

Mattos VF<sup>1</sup>, Carvalho LS<sup>2,3</sup>, Carvalho MA<sup>4</sup>, Schneider MC<sup>5</sup>

<sup>1</sup>Departamento de Biologia, Universidade Estadual Paulista, Rio Claro, SP, Brazil; <sup>2</sup>Universidade Federal do Piauí, Floriano, PI, Brazil; <sup>3</sup>Programa de Pós-Graduação em Zoologia, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil; <sup>4</sup>Departamento de Biologia e Zoologia, Universidade Federal de Mato Grosso, Cuiabá, MG, Brazil; <sup>5</sup>Departamento de Ciências Biológicas, Universidade Federal de São Paulo, Diadema, SP, Brazil

Species of the subgenus *Tityus* (*Archaeotityus*) are small and highly pigmented scorpions. In this work, two species of *T. (Archaeotityus)*—*T. mattogrossensis* and *T. silvestris*—were cytogenetically studied in order to establish the diploid number, chromosome behaviour during meiosis, distribution of the nucleolar organiser regions (NORs) and verify the occurrence of intra/ interspecific and/or intra/interpopulational chromosomal variations. Chromosome preparations were obtained from testes of adult specimens, which were collected in different localities from the state of Mato Grosso, Brazil: nine males of *T. mattogrossensis* from Chapada dos Guimarães (5♂), Cuiabá (3♂) and Poconé (1♂), and nine males of *T. silvestris* from Cláudia. The slides were submitted to standard staining with Giemsa and silver

impregnation. Chromosomes of both species were holocentric, presented synaptic and achiasmatic behaviour during meiosis I and multivalent complex associations in postpachytene cells. *Tityus mattogrossensis* showed intra and interpopulational variation in diploid number: 2n=20 (3♂ from Chapada dos Guimarães), 2n=19 (2♂ from Chapada dos Guimarães and 3♂ from Cuiabá) and 2n=14 (1♂ from Poconé). In *T. silvestris*, all individuals analysed exhibited 2n=16. In both species, the chromosomes gradually decreased in size. The study of postpachytene nuclei of *T. mattogrossensis* revealed an additional intraindividual variability: two males (2n=19 and 2n=20) exhibited 100 % of the cells with 10 bivalents; two specimens (2n=19 and 2n=20) presented in 100 % of the nuclei examined eight bivalents plus one chain of four elements (8II+CIV); two males (2n=19) showed 41.30 % and 58.70 % of the cells with nine and 10 bivalents, respectively; one specimen (2n=20) exhibited nuclei with nine (75.34 %) and 10 (24.66 %) bivalents; one individual (2n=19) revealed cells with nine bivalents (23.53 %), 10 bivalents (29.41 %), seven bivalents plus one chromosome chain constituted of four chromosomes (7II+CIV) (17.65 %) and eight bivalents plus one chain of four elements (8II+CIV) (29.41 %); one male (2n=14) showed seven (44 %) and eight bivalents (56 %). The postpachytene nuclei of all individuals of *T. silvestris* showed two bivalents plus one chromosomal chain constituted of 12 chromosomes (2II+CXII) (68.91 %) and one bivalent plus one chain of 12 chromosomes (1II+CXII) (31.09 %). Metaphase II cells of *T. mattogrossensis* revealed n=9 and n=10, indicating the regular chromosome segregation during anaphase I, with the exception of one individual (2n=14) that presented n=8. All sample of metaphase II cells of *T. silvestris* showed n=8 chromosomes. In both species, silver-impregnated mitotic metaphase cells revealed NORs on the terminal region of two to four chromosomes. The high variability of diploid number observed in *T. mattogrossensis* had previously been described in only one *Tityus* species, *T. bahiensis*, which belongs to subgenus *Tityus*. Within the subgenus *Archaeotityus*, only three species were cytogenetically analysed, *T. maranhensis* (2n=20), *T. mattogrossensis* (2n=20) and *T. paraguayensis* (2n=16). Nevertheless, the occurrence of multivalent associations in meiosis I is a common feature for species of all subgenera investigated and were probably originated by chromosomal rearrangements of the fission/fusion-type. Financial Support: FAPESP (2013/11840-0; 2011/21643-1).

sequences duplicated from essentially every chromosome in the ancestral karyotype. Although most genes on the B chromosome are fragmented, a few are largely intact and present B-specific variations related to single nucleotide polymorphisms (SNPs). Among these high intact sequences, genes involved with cell cycle (microtubule organization, kinetochore structure, recombination and progression through the cell cycle) were detected. Looking for a better understanding of the gene activity related to the B chromosome of *A. latifasciata*, we conducted whole transcriptome sequencing of brain and muscle tissues of males and females with (B+) and without (B-) B chromosome of *A. latifasciata*. Using Illumina Hi-Seq2000 sequencing we generated approximately 30 millions of 100 bp paired-end reads in each library. The reads were mapped to the cichlid *Metriaclima zebra* reference genome using Tophat software and visualized in the JBrowse browser. Based on SNPs analysis, we identified transcript variants for Separin and Tubulin beta-1 genes derived from the B chromosome. The B-variant transcripts of Tubb-1 gene were observed in brain and muscle of B+ males and females. B-variant transcripts of Separin were observed in B+ samples of brain of both sex, but only observed in female muscle and not from male muscle. Our data together with previous description of cell cycle genes in the Bs of diverse organisms as animal and fungal species, strongly suggest such genes may play a role in driving the transmission of the B chromosome in their hosts.

P61: Avian chromosomes in the lampbrush phase: contribution to developmental biology, cell biology and comparative cytogenetics

Gaginskaya ER

Biological Faculty, Saint-Petersburg State University, Saint-Petersburg, Russia

Our initial study of avian oogenesis has shown a fundamental difference between birds and other vertebrates in functional organization of germinal vesicles. A peculiar type of avian oogenesis is mainly defined by selective repression of ribosomal genes in oocytes and lack of nucleoli (both chromosomal and amplified) throughout the entire diplotene stage in the ovary of adult females. In the meantime, simultaneous widespread transcription of RNA from thousands of promoters results in chromosome transfiguration into the so-called lampbrush

chromosomes with the typical chromomere-loop organization. Lampbrush phase chromosomes have giant sizes (about 30 times longer than the corresponding mitotic chromosomes) and, being clearly differentiated throughout their lengths, provide a good ground for their individual identification and mapping of both cytological and molecular chromosome markers. Through adaptation of the technique of manual chromosome dissection from amphibian oocytes to avian lampbrush chromosomes, a new object was introduced into lampbrushology. The peculiarities of oocyte nuclei in Galliform species (lack of extrachromosomal structures), the features of both avian genomes (small size and low percentage of highly repeated DNA) and karyotypes (10 pairs of macro- and 29–30 pairs of microchromosomes) along with the data on sequenced chicken genome make avian lampbrush chromosomes an efficient tool for high precision gene mapping using immunofluorescence and FISH specific probes. The results of the long-term study of avian lampbrush chromosomes carried out by Saint-Petersburg University team are reviewed. Special focus is made on the avian lampbrush chromosomes as a promising system for an insight into chromosome organization and functioning, as well as on the recent data on identification and characterization of chicken macro- and microchromosomes with an exclusively high resolution.

This research is currently supported by RF Programme “Leading Scientific Schools” (project # 3553.2014.4) and SPbU grant (project # 1.37.153.2014). The research has been conducted using the facilities of “Chromas” SPbU Resource Centre.

P62: Inter and intra chromosomal rearrangements in avian microchromosomes

Lithgow PE<sup>1</sup>, O'Connor RE<sup>1</sup>, Smith D<sup>1</sup>, Fonseka G<sup>1</sup>, Al Mutery A<sup>1, 2</sup>, Rathje C<sup>1</sup>, Frodsham R<sup>3</sup>, O'Brien P<sup>4</sup>, Ferguson-Smith MA<sup>4</sup>, Skinner BM<sup>1,5</sup>, Griffin D K<sup>1</sup>

<sup>1</sup>School of Biosciences, University of Kent, Canterbury, UK; <sup>2</sup>Department of Applied Science, University of Sharjah, United Arab Emirates; <sup>3</sup>Cytocell Ltd, Newmarket Road, Cambridge, UK; <sup>4</sup>School of Clinical and Veterinary Medicine, University of Cambridge, Cambridge, UK; <sup>5</sup>Department of Pathology, University of Cambridge, Cambridge, UK

Avian genome organization is characterised, in part, by a set of microchromosomes that are unusually small in



size and unusually large in number. Although containing about a quarter of the genome, they contain around half the genes and three quarters of the total chromosome number. However, due to a lack of probes effectively hybridising cross species bird microchromosomes have not been extensively studied. Chromosomal rearrangements play a key role in genome evolution, fertility and genetic disease and thus tools for analysis of the microchromosomes are essential to analyse such phenomena in birds. Chicken microchromosomal paint pools were produced following flow sorting of chicken microchromosomes and characterised using chicken BAC probes (CHORI-261). These new molecular tools were used across species for Zoo-FISH on *Anas platyrhynchos*, *Taeniopygia guttata*, *Falco rusticolus*, *Chlamydotis undulate* and *Melopsittacus undulatus*, creating a clearer picture of microchromosomal rearrangements between these avian species. Microchromosome assignment allows more detailed comparison between species. This comparison confirms synteny of micro-chromosomes in most bird species. As predicted fusions were identified in *Falco rusticolus* and *Melopsittacus undulatus*, which have an atypically low diploid chromosome number for an avian karyotype of only  $2n=50$  and  $62$  respectively, indicating extensive chromosomal fusions from the ancestral avian karyotype.

P63: Avian ancestral karyotype reconstruction and differential rates of inter- and intra-chromosomal change in different lineages

Romanov MN<sup>1</sup>, Farre M<sup>2</sup>, Lithgow PE<sup>1</sup>, O'Connor R<sup>1</sup>, Fowler KE<sup>1</sup>, Skinner BM<sup>3</sup>, Larkin DM<sup>2</sup>, Griffin DK<sup>1</sup>

<sup>1</sup>School of Biosciences, University of Kent, Canterbury, UK; <sup>2</sup>Department of Comparative Biomedical Sciences, Royal Veterinary College, University of London, London, NW1 0TU; <sup>3</sup>Department of Pathology, University of Cambridge, Cambridge, UK

In birds, genome is organised into several large chromosomes (macrochromosomes) and many smaller chromosomes (microchromosomes) that usually constitute about 25 and 75 % of the karyotype, respectively. Cytogenetic and molecular cytogenetic evidence suggests that avian karyotype is remarkably stable in evolution, with exception of several clades. To date, at least 21 avian genomes have been sequenced and assembled at the chromosome or scaffold level with N50 greater than 2 Mb, thereby allowing cytogenomic studies of chromosome organisation and change. To understand the comparative organisation and evolution of several avian species, we aligned chromosomes and scaffolds using an interactive genome browser (Evolution Highway), identifying homologous synteny blocks (HSBs) and evolutionary breakpoint regions (EBRs). For ancestral karyotype reconstruction, we focused on six species (chicken, turkey, duck, zebra finch, ostrich, and budgerigar; N50 > 10 Mb) and reconstructed avian ancestor chromosomes using an outgroup (Anole lizard). In particular, we addressed the following biological questions: (1) whether species-specific EBRs could represent recombination hotspots, and (2) whether entire microchromosomes could be considered as blocks of conserved synteny. Our study did not reveal a significant association between EBRs and recombination. With support from molecular cytogenetic mapping, we did find that microchromosomes are characterised by a high interchromosomal conservation in almost all birds studied, except ostrich and parrots (budgerigar). By analysing HSBs in six birds and using a lizard outgroup, we reconstructed a tentative avian ancestral genome and chromosomal rearrangements that occurred in the major avian evolutionary lineages. We identified most intrachromosomal changes (mostly inversions) in the zebra finch clade (Passeriformes) since the time when it diverged from the sister group of parrots (Psittaciformes) 54MYA. Our data also suggest the fewest number of chromosomal changes in the chicken as compared to the dinosaur-like avian ancestor.



**July 10–13, Foz do Iguaçu, Brazil**

## **21st International Chromosome Conference (ICC)**

*A venue that offers a diversity of scientific approaches to chromosome biology and a diversity of wildlife in Iguaçu National Park*

The International Chromosome Conferences (ICC) originated from the Oxford Chromosome Conferences, inaugurated by C.D. Darlington and K.R. Lewis in 1964 and held subsequently in England in 1967 and 1970. The Chromosome Conference grew to an international event with its fourth meeting, held in Jerusalem, Israel in 1972, heralding the beginning of 40 years of technological advances that have expanded our understanding of chromosome biology in model and non-traditional biological systems. Having been hosted in Europe and the United States 16 times since then, this year the ICC will be held across the equator in Foz do Iguaçu, Brazil, on July 10–13, 2016. The event will bring scientists from across the globe to a biannual meeting focused on modern advances in chromosome biology, technology and theory. The Iguaçu National Park, a UNESCO World Heritage Centre, includes the Iguaçu Falls and has been chosen as one of the 'New Natural Seven Wonders of the World'. Home to an

amazing diversity of life, including over 2,000 species of vascular plants, exotic mammals such as tapirs, giant anteaters, howler monkeys, ocelots, and jaguars, in addition to hundreds of different bird species and thousands of different insects, the choice of Foz is an excellent analogy for the diverse approaches and systems chromosome biologists explore, and that will be emphasized throughout this conference.

The 2016 ICC program offers seven sessions, beginning with a session on *Chromosome Structure and Nuclear Architecture*, highlighting the influences and interactions chromosomes have on the three-dimensional space of the nucleus. Session II will focus on *Specialized Chromosomes*, such as sex chromosomes and B chromosomes, whose structure and behavior are often distinguished from that of autosomal chromosomes. *Population and Evolutionary Chromosome Biology*, the third session, covers a synthesis of chromosome biology and

## VII.11

### Characterization of Translocation by Chromosome Sequencing on Flow-Sorted Chromosomes: Robust Methods for Identification of Genomic Breakpoint Junctions

F. Kasai<sup>a</sup>, J. Pereira<sup>b,c</sup>, N. Hirayama<sup>a</sup>, S. Shioda<sup>a</sup>, A. Kohara<sup>a</sup>, M. Ferguson-Smith<sup>b</sup>

<sup>a</sup>Japanese Collection of Research Bioresources (JCRB) Cell Bank, National Institutes of Biomedical Innovation, Health and Nutrition, Osaka, Japan; <sup>b</sup>Department of Veterinary Medicine, University of Cambridge, and <sup>c</sup>Cytocell Ltd., Cambridge, UK

Chromosome translocation is a key feature in chromosome abnormalities and can lead to the formation of a fusion gene. Although it is identified by cytogenetic analysis based on banding patterns or chromosome painting, it is hard to characterize the breakpoint at the sequence level. Chromosome sorting by flow cytometry shows flow karyotypes and enables us to generate chromosome painting probes. Abnormal chromosomes are often found to form weak peaks in the flow karyotypes, allowing distinguishing them from normal alleles. In this study, we sorted derivative chromosomes in a human tumor cell line, Ishikawa 3-H-12, and a dog cell line, MDCK, to characterize their genomes. Approximately 2,000 chromosomes of t(9;14) from the Ishikawa cell line and t(27;X) from MDCK were amplified by a WGA kit used for preparation of the genomic DNA fragment library. Chromosome-specific sequencing was performed by the Ion PGM sequencer. The breakpoint junction in der(9) was identified at 9p24.3 and 14q13.1, with the formation of a fusion gene. Sequence analysis of coding regions around 14q13.1 based on the Ion Ampliseq technology showed the differences of SNP frequencies between the upstream and downstream regions of the breakpoint junction. The genomic breakpoint junctions unique to each cell line can be precisely determined through chromosome sequencing.

E-Mail: k-230@umin.ac.jp

## VII.12

### Analysis of a Noncoding Region of the *SIX3* Gene in Patients with Holoprosencephaly

J. Marino<sup>a,b</sup>, N.A. Bergamo<sup>b</sup>, R.H. Rocha<sup>b</sup>, R.M.O.F. Curado<sup>b</sup>, S.K. Hong<sup>c</sup>, P. Hu<sup>c</sup>, E. Roessler<sup>c</sup>, M. Muenke<sup>c</sup>, L. Ribeiro-Bicudo<sup>b</sup>

<sup>a</sup>UNESP – São Paulo State University 'Julio de Mesquita Filho',

<sup>b</sup>Federal University of Goiás, Goiânia, Brazil;

<sup>c</sup>National Institute of Health, Bethesda, Md., USA

Holoprosencephaly (HPE) is characterized by a defect of the middle line of the embryonic forebrain, when a segmentation failure of the previous neural tube occurs. Mutations in the *SHH*, *ZIC2*, *SIX3* genes were detected and are related to 33% of cases, but few studies about the noncoding region are available. Mutations in the gene *SIX3* are present in 1.3% of HPE cases in humans and are associated with a complex phenotype, varying from a single central incisor to cyclopy. A total of 44 individuals with HPE, registered at the database of the Rehabilitation Craniofacial Anomalies Hospi-

tal, USP, Bauru, were analyzed by next-generation sequencing in the laboratory of the National Human Genome Research – National Institutes of Health, Bethesda, Md., USA. The Illumina Platform HiSeq2000 was used and the analysis made in paired-end reads of 100 bp to analyze noncoding elements of the *SIX3* gene in 15 male and 29 female samples with heterogeneous phenotypes of HPE. All patients have some type of cleft, hypotelorism, and microcephaly. We found 28 variants, 27 SNPs and 1 indel. Twenty alterations have already been described, while 8 other dbSNP have not been described; 24 are located in the intergenic region, 2 in the 3' UTR, and 1 upstream of the gene. Although the mapping of complex disease genes is difficult, an increase in the number of susceptibility genes has been identified as a result of the availability of the complete genomic sequence, dense marker maps, and high yield genotyping platforms. However, in many cases, the true susceptibility variant(s) remain unknown and extremely difficult to identify. The identification which of those millions of variants is functional is important for health and research, and bioinformatics methods are required to assess the probability of functionality based on extensive experimental data.

Financial support: CAPES.

E-Mail: jumarino22@hotmail.com

## VII.13

### Upgrading Molecular Cytogenetics to Study Reproduction and Reproductive Isolation in Mammals, Birds, and Dinosaurs

R.E. O'Connor<sup>a</sup>, J. Damas<sup>b</sup>, M. Farré<sup>b</sup>, M.N. Romanov<sup>a</sup>, H. Martell<sup>a</sup>, G. Fonseka<sup>c</sup>, R. Jennings<sup>a</sup>, L. Kiazam<sup>a</sup>, S. Bennett<sup>a</sup>, J. Ward<sup>a</sup>, A. Mandawala<sup>c</sup>, S. Joseph<sup>a</sup>, R. Frodsham<sup>d</sup>, M. Lawrie<sup>d</sup>, A. Archibald<sup>e</sup>, G.A. Walling<sup>f</sup>, K.E. Fowler<sup>c</sup>, D.M. Larkin<sup>b</sup>, D.K. Griffin<sup>a</sup>

<sup>a</sup>School of Biosciences, University of Kent, Canterbury,

<sup>b</sup>Department of Comparative Biomedical Sciences, Royal Veterinary College, University of London, London,

<sup>c</sup>Canterbury Christchurch University, Canterbury, <sup>d</sup>Cytocell Ltd, Newmarket Road, Cambridge, <sup>e</sup>The Roslin Institute, R(D)SVS,

University of Edinburgh, Division of Genetics and Genomics, Easter Bush, Midlothian, and <sup>f</sup>JSR Genetics, Southburn, Driffield, UK

The past 10–15 years have seen a revolution in the field of genomics, first with the human genome project, followed by those of key model and agricultural species (chicken, pig, cattle, sheep) and, most recently, ~60 de novo avian genome assemblies. The ultimate aim of a genome assembly is to create a contiguous unbroken length of sequence from p- to q-terminus to facilitate studies of gene mapping, trait linkage, phylogenomics, and gross genomic organization/change. Chromosome rearrangements are biologically relevant both in the context of reduction in reproductive capability of individual animals and in the establishment in reproductive isolation as species evolve and diverge. Moreover, a karyotype effectively represents a low-resolution map of the genome of any species. In investigating all these aspects, FISH remains the tool of choice, and this study describes a step change in its use thus: (1) Isolation of sub-telomeric sequences from the pig and cattle



genome assemblies to develop a device for the screening of both overt and subtle chromosome rearrangements. This device worked successfully and was the basis for the development of a routine screening test now used in the pig (and potentially in the future the cattle) breeding industries. The work also facilitated an assay of the integrity of the respective genome assemblies, revealing serious errors in sub-telomeric builds of pig chromosomes. Numerous translocations were detected, most notably a 5:6 cryptic translocation that would not have been detected by classical means. (2) Isolation of evolutionarily conserved sequences from the chicken and zebra finch genome builds to develop similar probes and devices designed to assay for comparative genomics and genome evolution in any avian species. This device worked on the chromosomes of all species attempted and successfully detected chromosomal rearrangements. The hypothesis that certain groups were under constant change was accepted for Psittaciformes species but not Falconiformes. (3) Use of the technology developed in 1 and 2 to complete scaffold-based genome assemblies in several key avian species recently sequenced. We have nearly completed the genome assemblies of peregrine falcon, pigeon, budgerigar, and ostrich genomes at the full chromosomal level and the information was uploaded to Evolution Highway. (4) Use of bioinformatic tools to re-create the overall genome structure (karyotype) of both Saurian and Avian ancestors, then retrace the gross evolutionary changes that have occurred down the dinosaur (and various avian) lineages. Gene ontology analysis of homologous synteny blocks and evolutionary breakpoint regions revealed enrichment for genes involved in chromosome rearrangement (consistent with the formation of the signature fragmented karyotype of birds (and probably dinosaurs), and body size, consistent with the overall gross reduction in size as dinosaurs evolved into birds. Taken together, these results represent significant novel insights into gross genomic organization and rearrangements in extant and extinct terrestrial vertebrates. It has the added benefit of developing both physical and online tools for future use in academic studies and for feeding a growing global population.

E-Mail: D.K.Griffin@kent.ac.uk

#### VII.14

### Domestication and Repetitive DNA Genome Fraction in *Capsicum chinense*

M.V. Romero da Cruz<sup>a</sup>, M. Vaio<sup>b</sup>, E.R. Forni Martins<sup>a</sup>

<sup>a</sup>Department of Botany, Institute of Biology, Universidade Estadual de Campinas, Campinas, Brazil; <sup>b</sup>Department of Plant Biology, Universidad de la República, Montevideo, Uruguay

The chili pepper *Capsicum chinense* belongs to the Solanaceae family. It is a diploid, self-pollinating crop and is closely related to potato, tomato, eggplant, tobacco, and petunia. It is 1 of the 5 domesticated chili peppers with several commercial varieties. The species is native to South America with the center of diversity in the Amazon biome. Many authors have questioned the species status of *C. chinense*, perhaps because it is the least known of the 5 domesticated taxa with respect to the center of origin and probable progenitors. To gain a better understanding of *C. chinense* evolution and domestication, we used a next-generation low-coverage

sequencing of the cultivated pepper *Habanero* (*C. chinense* Jacq.) and the wild *C. chinense* and a graph-based clustering approach for the repeat sequence characterization as implemented in the RepeatExplorer pipeline. In total, we identified that more than 60% of the genome of both the cultivated and wild *C. chinense* was represented by repetitive sequences. Both class I and II transposable elements were present. The predominant type of transposable element was the long terminal repeat (LTR) retrotransposons accounting for 31.25% of the genome. Most of the LTRs were *Gypsy* elements. An accumulation of members of the *Caulimoviridae* pararetrovirus family was also observed. A large number of *Caulimoviridae* elements have previously been detected in *C. annum* and other Solanaceae species and might have had a role in the expansion of the pepper genome in both heterochromatic and euchromatic regions. No differences in type or percentage of repetitive sequences were observed between the cultivated and wild forms. This fact suggests that the domestication process in this species did not affect this genome fraction which seems quite conserved.

Financial support: CNPq.

E-Mail: romero.mariav@gmail.com

#### VII.15

### Comparative Male and Female Characterization and Expression of the *dmrt1* Gene of *Apareiodon* sp. (Characiformes, Parodontidae)

M.O. Schemberger<sup>a</sup>, A.P. Schnepfer<sup>a</sup>, V. Nogaroto<sup>a</sup>, G.T. Valente<sup>b</sup>, É. Ramos<sup>b</sup>, C. Martins<sup>b</sup>, R.F. Artoni<sup>a</sup>, M.C. Almeida<sup>a</sup>, M.R. Vicari<sup>a</sup>

<sup>a</sup>Programa de Pós-Graduação em Biologia Evolutiva, Universidade Estadual de Ponta Grossa, Ponta Grossa, and

<sup>b</sup>Universidade Estadual de São Paulo, Botucatu, Brazil

The DMRT (doublesex and mab-3 related transcription factor) gene family is widely conserved from invertebrates to humans. Vertebrate *Dmrt1* gene expression occurs predominantly in testis and is a strong candidate for male sex-determining gene studies. In this respect, the understanding of the function of *Dmrt1* in sex determination is important for understanding the cascade of sex differentiation. Structural characterization of the *dmrt1* locus and protein prediction of *Apareiodon* sp. was conducted by bioinformatic analyses of male and female genomes sequenced by Illumina HiSeq and PCR amplification of cDNA using specific primers. Expression of *dmrt1* of 8 adult male and female *Apareiodon* sp. (ZZ/ZW) was quantified by qRT-PCR. We found 5 exons of *dmrt1* which contain 887 bp for male and female. The protein has 2 domains, the DM DNA-binding domain and doublesex mab3-related transcription factor 1. Promoter prediction of ~6,000 bp upstream of the gene revealed 8 similar/equal regions between male and female. However, 8 regions were different, characterized by 6 additional regions (insertions) in female and 2 additional regions (insertions) in male. SSA (Signal Search Analyses Server) software detected a TATA box, initiator and GC-box promoters in male, and initiator and GC-box promoters in female. Several relics of transposable elements were found in the promoter region, HatN45\_DR (54 bp) was present in the male GC-box, Mariner-1 SSA (150 bp) in the female GC-box, L2-5\_GA (162 bp) in male