



Kent Academic Repository

Joseph, Sunitha (2017) *Chromosome Evolution and Genome Reconstruction in Falcon Species*. Doctor of Philosophy (PhD) thesis, University of Kent,.

Downloaded from

<https://kar.kent.ac.uk/65667/> 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. 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>).

**Chromosome Evolution and
Genome Reconstruction in
Falcon Species**

Sunitha Joseph

18 September 2017

**A thesis submitted to the University of Kent, UK 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 or Institute of learning.

Sunitha Joseph

18 September 2017

Acknowledgements

I thank and praise God, the Almighty for surrounding me with His blessings throughout my research period and strengthening me at every difficult moment to successfully complete this thesis.

I express my sincere gratitude to His Highness Sheikh Mohammed Bin Rashid Al Maktoum, Vice-President and Prime Minister of the UAE and the ruler of Dubai. His vision inspired me to pursue this thesis, and I am honoured by the privilege of being the first female scientist in our industry sponsored by him.

I wish to express my heartfelt gratefulness to the Director General of CVRL, Dr. Ali Ridha, for his belief in my abilities to pursue this PhD. Dr. Ali's unwavering support ensured that I had the freedom to focus on my study time at Kent University alongside my general duties at the Central Veterinary Research Laboratory.

I thankfully owe my deepest love and credit to Priv. Doz. Dr. habil. Ulrich Wernery, the Scientific Director of CVRL and Mrs. Renate Wernery for being my mentors and for their strong support throughout my career, especially during this thesis venture.

It is my great privilege and honour to have Professor Darren Griffin as my research supervisor. I would like to express my thanks to Professor Griffin, since without his efficient guidance and constant help I would not have been able to achieve all that is presented in this book.

I am also very grateful to Dr. Rebecca O'Connor for sharing her expertise in this field, thereby helping my thesis to take its shape. With gratitude, I remember Dr. Ben Skinner, Dr. M.N. Romanov, Rebecca and Lucas, with whose help and support I could complete my thesis project successfully.

I would like to extend my sincere thanks to Mrs. Elizabeth Crichton, Dr. Abdullah Al Mutery, Dr. Jeorg Kinne, Mr. Rajan Babu and Mr. Syed Arshi Uz Zaman for their vital support and assistance in my efforts to make this thesis a reality.

This thesis would not have been possible without the support of many other people such as Mr. Howard Waller, Dr. Melodia, and Mr. Zoltan Lorenti, who helped me to source all required species samples of various falcon breeds.

My heart is full of love and thanks to all of my CVRL colleagues, especially Nissy, Shyna and Rubeena, in reducing my virology diagnostic workload, thereby giving me the time to focus on my thesis work.

Nobody has been more important to me in the pursuit of this project than my family. My parents' love and their encouragement motivated me to pursue my PhD dream. I express the deepest gratitude to my loving husband Bino and my two wonderful children Benetta and Samuel for understanding me. They never doubted my abilities and they would never miss a moment to enhance my confidence at times of self-doubt.

Lastly with a grateful heart I remember everyone who has given direct or indirect help for the successful completion of my thesis.

Dedicated to my wonderful parents,

V.M. Mathai & Molly Mathai

And

Joseph Mathew & Ammini Joseph

Table of Contents

| | |
|---|-----------|
| Declaration..... | i |
| Acknowledgements..... | ii |
| Dedication..... | iii |
| Table of Contents..... | iv |
| List of Figures | viii |
| List of Tables..... | xii |
| Abbreviations..... | xiii |
| Abstract..... | xiv |
| 1 Introduction | 1 |
| 1.1 Falcons | 1 |
| 1.1.1 A Brief History of Falconry | 2 |
| 1.1.2 Morphology..... | 2 |
| 1.1.3 Habitat and Geographical Distribution | 5 |
| 1.1.4 Hybrids | 7 |
| 1.1.5 Relevance to Contemporary Society..... | 9 |
| 1.1.6 Phylogeny of the Falconidae..... | 10 |
| 1.2 Avian Genome Organization..... | 14 |
| 1.2.1 Chromosome Territories..... | 16 |
| 1.2.2 Models for Nuclear Organization..... | 16 |
| 1.2.3 Nuclear Organization In Birds | 17 |
| 1.2.4 Telomeres | 19 |
| 1.3 Avian Genome Sequencing..... | 21 |
| 1.3.1 Genome Sequencing..... | 21 |
| 1.3.2 The Chicken Genome | 24 |
| 1.3.3 The Zebra Finch Genome | 25 |
| 1.3.4 The Turkey Genome..... | 26 |
| 1.3.5 The Duck Genome..... | 27 |
| 1.3.6 Other Sequenced Avian Genomes | 27 |
| 1.3.7 Multiple Avian Genome Sequencing Efforts..... | 28 |
| 1.3.8 Genome 10K Project | 30 |
| 1.3.9 Physical Genome Mapping | 31 |
| 1.4 Avian Genome Evolution | 33 |

| | | |
|-------|---|----|
| 1.4.1 | Avian Chromosomes | 33 |
| 1.4.2 | Chromosomal syntenies | 34 |
| 1.4.3 | Homologous Syteny Blocks And Evolutionary Breakpoint Regions | 35 |
| 1.4.4 | Comparative Genomics and Avian Genome Evolution..... | 36 |
| 1.5 | Methods to Study Chromosomes | 41 |
| 1.5.1 | Classical Cytogenetics | 41 |
| 1.5.2 | Molecular cytogenetic Analysis using Fluorescence <i>in situ</i> Hybridization..... | 44 |
| 1.5.3 | Comparative Genomic Hybridization and Microarrays | 46 |
| 1.5.4 | Comparative Genome Visualization Tools at the Chromosome level | 47 |
| 1.5.5 | Genome Reconstruction using Bioinformatics and FISH | 51 |
| 1.6 | Chromosomal Abnormalities in Individuals and in Evolution..... | 53 |
| 1.6.1 | Numerical Chromosome Aberrations | 54 |
| 1.6.2 | Structural Chromosome Aberrations..... | 54 |
| 1.7 | Avian Chromosomal Studies | 59 |
| 1.7.1 | Avian Karyotyping | 59 |
| 1.7.2 | Chromosome Studies Using Fluorescence <i>in situ</i> Hybridization..... | 60 |
| 1.8 | Cytogenetic and Genomic Studies in Falcons | 64 |
| 1.8.1 | Recent Cross Zoo-FISH study on Peregrine falcon | 69 |
| 1.8.2 | Falcon Genome Sequencing..... | 72 |
| 1.8.3 | Peregrine Genome – Chromosome Mapping | 75 |
| 1.9 | Specific Aims of this Thesis | 77 |
| 2 | Materials and Methods..... | 80 |
| 2.1 | Chromosome Preparation | 80 |
| 2.1.1 | Fibroblast Culture | 80 |
| 2.1.2 | Chromosome Harvesting | 82 |
| 2.1.3 | Blood Culture and DNA Extraction..... | 83 |
| 2.2 | Generation of Labelled FISH Probes | 84 |
| 2.2.1 | Selection of BAC clones..... | 84 |
| 2.2.2 | Preparation of BAC Clones..... | 85 |
| 2.3 | Fluorescence <i>in situ</i> Hybridisation (FISH) | 86 |
| 2.3.1 | Metaphase Slide Preparation and Hybridisation | 86 |
| 2.3.2 | Microscopy..... | 86 |
| 2.3.3 | Image Analysis - Karyotyping..... | 87 |
| 2.4 | Physical Genome Mapping | 87 |
| 2.4.1 | Construction of PCFs using the RACA Algorithm (RVC) | 88 |
| 2.4.2 | Verification of PCFs (RVC) | 88 |

| | | |
|-------|--|-----|
| 2.4.3 | Creation of a Refined Set of Saker PCFs (RVC)..... | 88 |
| 2.5 | Ancestral Karyotype Reconstruction | 88 |
| 2.5.1 | Neognathae Ancestor | 88 |
| 2.5.2 | Genome Rearrangement Analysis | 89 |
| 2.6 | Telomere sequence mapping | 89 |
| 2.7 | Determination of Nuclear Position of Chromosome Territories | 90 |
| 3 | Specific aim 1: To define the overall structure (karyotype) of the Saker falcon genome and make hitherto undiscovered links between the genome assembly and the karyotype, thereby generating a cytogenetic genome map. | 92 |
| 3.1 | Background | 92 |
| 3.2 | Specific Aims | 94 |
| 3.3 | Results..... | 94 |
| 3.3.1 | Specific Aim 1a: To make a high-resolution karyotype and standard ideogram of Saker falcon and to compare it with two other falcons species: Peregrine falcon and Gyrfalcon | 94 |
| 3.3.2 | Specific Aim 1b: To perform comparative FISH with known chromosome probes to define the overall genome structure of Saker and Gyrfalcon in comparison to previously assembled genome sequences such as Peregrine falcon and chicken..... | 101 |
| 3.3.3 | Specific Aim 1c: To map the molecular cytogenetic markers to existing scaffolds (predicted chromosome fragments–PCFs) of the Saker falcon genome assembly and make the data publicly available by an interactive website (Evolution Highway) designed to demonstrate cytogenetic comparisons between individual species..... | 108 |
| 4 | Specific Aim 2: To perform a comparative genomic study of genome evolution in Peregrine falcon, Saker falcon and Gyrfalcon making use of BAC FISH probes that map inter- and intra-chromosomal changes to map the path of chromosome evolution in Falco species. | 113 |
| 4.1 | Background | 113 |
| 4.2 | Specific Aims: | 114 |
| 4.3 | Results..... | 114 |
| 4.3.1 | Specific Aim 4a: To make use of resources (BAC FISH probes) that map inter- and intra-chromosomal changes in at least three falcons species and input them into MGRA2 to generate the likely ancestral karyotype of Neognathe birds | 114 |
| 4.3.2 | Specific Aim 4b: To map the path of chromosome evolution in Falco species by comparison with conserved (e.g. chicken) and other rearranged (e.g. budgerigar) species through the use of the GRIMM tool | 116 |
| 4.3.3 | Specific Aim 4c: To test the hypothesis that Falco genome evolution involved both intra- as well as inter-chromosomal evolution | 123 |
| 4.3.4 | Specific Aim 4d: To test the hypothesis that common mechanisms of chromosomal fusion recur in species with rearranged karyotypes by comparison with a closely related group (<i>Psittaciformes</i>) | 123 |

| | | |
|-------|---|-----|
| 5 | Specific aim 3: To map telomeric sequences in three falcon species to test the hypothesis that remnants of former chromosomal fusions retain their telomeric motifs i.e. appear as interstitial telomeres. To compare with other species that have undergone chromosomal fusions such as the budgerigar and crocodile. | 128 |
| 5.1 | Background | 128 |
| 5.2 | Results..... | 129 |
| 6 | Specific Aim: To test the hypothesis that chromosome that were formerly microchromosomes (but are now fused to larger chromosomes) still “behave” as though they were microchromosomes in terms of their nuclear organization..... | 137 |
| 6.1 | Background | 137 |
| 6.2 | Specific Aims | 138 |
| 6.3 | Results..... | 138 |
| 6.3.1 | Specific Aim 6a: Identification of BAC probes from micro, macro and “former microchromosome” regions of both chicken and Peregrine falcon chromosomes | 138 |
| 6.3.2 | Specific Aim 4b: Mapping the average chromosome position of all selected BACs in metaphase and interphase | 139 |
| 6.3.3 | Specific Aim 3c: Testing the hypothesis that macrochromosomes are peripherally located, microchromosomes are centrally located but “former” microchromosomes still “behave” as though they are microchromosomes and thus occupy a central position | 145 |
| 7 | Discussion | 155 |
| 7.1 | The overall structure (karyotype) of the Saker falcon genome and links to the genome assembly | 156 |
| 7.1.1 | Application for falcon hybrids..... | 162 |
| 7.1.2 | Atypical karyotypes..... | 163 |
| 7.2 | Chromosome evolution in falcons | 167 |
| 7.3 | Mapping of telomeric sequences in falcons and other animals..... | 167 |
| 7.4 | Nuclear organization of “former microchromosome” | 170 |
| 7.5 | Future Work..... | 173 |
| 7.6 | Concluding Remarks | 175 |
| 8 | References: | 177 |
| 9 | Appendix | 199 |

List of Figures

| | |
|---|----|
| Figure 1-1: Peregrine falcon..... | 3 |
| Figure 1-2 : Saker falcon | 4 |
| Figure 1-3: Gyrfalcon | 4 |
| Figure 1-4: Global distribution of the Peregrine falcon indicating whether, summer or winter visitor, breeding resident or passage visitor. (https://www.beautyofbirds.com/peregrinefalcons.html)..... | 6 |
| Figure 1-5: Global distribution of the Saker falcon indicating whether resident, breeding or feeding Saker (http://www.avibirds.com/html/falcons/Saker_Falcon.html#.WhrpgUqWY2w).. | 6 |
| Figure 1-6: Global distribution of the Gyrfalcon indicating native and/or nesting compared to rare/occasional visitor (http://www.oiseaux.net/maps/gyrfalcon.html) | 6 |
| Figure 1-7: Falcon hybrid (Gyr x Peregrine) | 8 |
| Figure 1-8: Phylogeny and classification of the Falconidae based on morphological and mitochondrial cyt-b data (Griffiths 1999; Griffiths <i>et al.</i> 2004)..... | 11 |
| Figure 1-9: Phylogenetic relationship (with Gallus as out-group) between different members of the family Falconinae and crested caracara adapted from (Seibold <i>et al.</i> 1993)..... | 12 |
| Figure 1-10: Phylogenetic relationship between Falconiforms with other major avian groups (Jarvis <i>et al.</i> 2014). | 13 |
| Figure 1-11: Representative G-banded karyotypes for (a) Chicken and (b) Spectacled owl. | 15 |
| Figure 1-12: Schematic representation of radial arrangements of large (red) and small (green) chromosomes in the cell nucleus of human and chicken (Habermann <i>et al.</i> 2001). | 17 |
| Figure 1-13: Chicken chromosome paints for larger (red) and smaller (blue) macrochromosomes and for microchromosomes (green) on chicken metaphase (left) and distribution of chromosome territories in chicken interphase (right) (Habermann <i>et al.</i> 2001). | 18 |
| Figure 1-14: Chromosomal location (GGA 9 and GGA W) of mega-telomere in the chicken (O'Hare and Delany 2009)..... | 20 |
| Figure 1-15: Schematic representation of the Sanger sequencing process (Kircher and Kelso 2010). | 22 |
| Figure 1-16: Overview of the RACA Algorithm (Kim <i>et al.</i> 2013)..... | 32 |
| Figure 1-17: Diploid chromosome numbers (2n) in vertebrate groups. (a) mammals, (b) birds, (c) reptiles and (d) fishes (Ellegren 2010). | 33 |
| Figure 1-18: Schematic representation of relative sizes of the ancestral chromosomes 1-10 + Z (adapted from Griffin <i>et al.</i> 2007). The chromosome number that is stated below is the chromosome in black text with the chicken orthologue in red text and in brackets. As all sizes are relative, no scales are appropriate. | 34 |
| Figure 1-19: Schematic representation of how homologous synteny blocks (HSBs) and evolutionary breakpoint regions (EBRs) are defined in the reference genome by pair-wise comparison adapted from Farre <i>et al.</i> (2011). | 36 |
| Figure 1-20: Archaeopteryx lithographica fossil, the oldest avian ancestor. | 37 |
| Figure 1-21: Phylogeny of birds. Time calibrated phylogeny of 198 species of birds inferred from Bayesian analysis (Prum <i>et al.</i> 2015)..... | 39 |
| Figure 1-22: Human chromosome banding revealed by different staining techniques. (a) G-banding; (b) Q-banding; (c) R-banding; (d) C-banding (O'Connor 2008a)..... | 42 |
| Figure 1-23: G-band ideogram of human chromosome 11 at (from left to right) 350, 550 and 850 band resolutions (Bickmore 2001)..... | 43 |
| Figure 1-24: The principles of fluorescence <i>in situ</i> hybridization (FISH) (https://www.semrock.com/fish.aspx). | 45 |

Figure 1-25: Representation of Array CGH (Theisen 2008). Widely used for detecting amplifications and deletions in tumours (as in this diagram) it can also be used to detect copy number variation (CNV) between individuals, strains and species. 47

Figure 1-26: Conserved synteny in human and mouse genome. Human chromosomes; each colour denotes a particular mouse chromosome (Lander *et al.* 2001). 48

Figure 1-27: Multispecies comparative chromosome alignment against human chromosome 16. Grey blocks denote HSBs with the chromosome numbers (Murphy *et al.* 2005). 50

Figure 1-28: PhyloView of the paired-like homeobox2b (PHOX2B) gene with the human gene as the reference gene and showing high density of conserved synteny between the human and several mammals, also with chicken (Louis *et al.* 2013). 51

Figure 1-29: Single chromosomal rearrangements 55

Figure 1-30: A combination of FISH and bioinformatics data showing the total number of inversions in six avian species as they diverged from the ancestor (Romanov *et al.* 2014). 56

Figure 1-31: Reciprocal translocation illustrating exchange of segment of chromosomes 4 and chromosome 8. 57

Figure 1-32: Schematic representation of Robertsonian translocation between human chromosome 13 and 14. 57

Figure 1-33: Different combinations of ancestral chromosomes fused together as chromosome numbers decreased in parallel within anuran and salamander lineages. Fission and fusion were active in framing the generalized avian karyotype (Voss *et al.* 2011). 58

Figure 1-34: Seven-colour chromosome painting of the stone curlew metaphase (Nie *et al.* 2009). 61

Figure 1-35: Comparative analysis of marker order on chicken chromosome 4 (GGA4) and its zebra finch orthologs, TGU4 and TGU4A. The central part of the figure was created by aligning whole-chromosome sequences using the program GenAlyzer (Volker *et al.* 2010). 63

Figure 1-36: Karyotypes of the (A) Peregrine falcon and (B) Prairie falcon, and (C) Gyrfalcon. (Schmutz and Oliphant 1987). 65

Figure 1-37: Hoechst-banded karyotypes of the Common kestrel (A), Peregrine falcon (B) and Merlin falcon (C) with the assignment of homology with chicken chromosomes delineated by chromosome painting with chicken probes (Nishida *et al.* 2008). 67

Figure 1-38: Chromosome hybridization patterns with chicken chromosome 1, 4 and 6 paints (GGA1, GGA4, GGA6) and a paint pool of 19 microchromosomes (GGAmicro) to PI-stained metaphase chromosome spreads of *Falco tinnunculus* (a-d), *Falco peregrinus* (e-h) and *Falco columbarius*. (Nishida *et al.* 2008). 68

Figure 1-39: Peregrine chromosomes (FPE) with their chicken (GGA) homologs (O'Connor 2016). 70

Figure 1-40: Gyrfalcon karyotype (Al Mutery 2011). 71

Figure 1-41: NJ tree based on the proportion of shared microsatellite alleles among hierofalcon individuals (Nittinger *et al.* 2007). 72

Figure 1-42: Screenshot of sample window of the sequencing analysis software version (Al Mutery 2011). 75

Figure 1-43: Cytogenetic and PCF mapping of Peregrine falcon chromosome 5 (FPE 5) using FISH; (a) Evolution highway alignment of zebra finch, chicken and Peregrine falcon genome along with the PCFs produced by RACA and the BACs that map in this region, (b) Cytogenetic map of BACs in the correct orientation on FPE 5, (c) Physical mapping of BACs to FPE 5 using FISH and (d) Ideogram of FPE 5 homologous to chicken (O'Connor 2016). 76

Figure 2-1: A captured nucleus image with converted to RGB planes before the application of the macro (left) and after (right) with 5 rings of equal area formed. 90

Figure 3-1: The karyotype (2n = 52) of Saker falcon. (a) DAPI and propidium iodide -stained metaphase spread (b) female karyotype..... 95

Figure 3-2: Standard ideogram of Saker falcon as devised in this study. Bands are divided into “light” (i.e. bright on propidium iodide, pale on DAPI), “dark” (dark on propidium iodide, bright on DAPI) and “grey” (pale in both). 96

Figure 3-3: The karyotype (2n = 52) of Gyrfalcon. (a) DAPI and propidium iodide -stained metaphase spread (b) female karyotype..... 97

Figure 3-4: Standard ideogram of Gyrfalcon. Bands are divided into “light” (i.e. bright on propidium iodide, pale on DAPI), “dark” (dark on propidium iodide, bright on DAPI) and “grey” (pale in both)..... 98

Figure 3-5: The karyotype (2n = 50) of Peregrine falcon. (a) DAPI and propidium iodide-stained metaphase spread (b) male karyotype. 99

Figure 3-6: Standard ideogram of Peregrine falcon. Bands are divided into “light” (i.e. bright on propidium iodide, pale on DAPI), “dark” (dark on propidium iodide, bright on DAPI) and “grey” (pale in both)..... 100

Figure 3-7: Comparative map of the Saker falcon (FCH), Gyrfalcon (FRU) and Peregrine falcon (FPE) chromosomes with the position of all BACs (as calculated by the fractional length from the p-terminus of the chromosome). The red lines indicate inverted segments..... 101

Figure 3-8: Cytogenetic and PCF mapping of Saker falcon (*Falco cherrug*) chromosome 1 (FCH-1); (a) Part of the PCFs produced by RACA for Saker falcon Chromosome 1 visualised on Evolution Highway; (b) Physical mapping of BACs to FCH-1 using FISH; (c) Precise cytogenetic mapping of BACs on FCH-1 according to their orientation; (d) Ideogram showing chicken (GGA) homologs to FCH-1. 110

Figure 3-9: Ideogram representation of Saker falcon (*Falco cherrug*) chromosomes and their chicken (GGA) homologs. Colours show the orthologous chicken chromosomes. 111

Figure 3-10: Screenshot from Evolution Highway showing the PCFs of Saker falcon genome representing chromosome 2 (: <http://eh-demo.ncsa.uiuc.edu/birds-test/#/SynBlocks>). 112

Figure 4-1: For the ancestor to the chicken, the GRIMM algorithm mapped 6 changes (4 fusions, 1 translocation and 1 inversion (reversal)). 117

Figure 4-2: For the ancestor to the pigeon, the GRIMM algorithm mapped 7 changes (3 fusions, 1 translocation and 3 inversions (reversals))..... 118

Figure 4-3: For the ancestor to the budgerigar, the GRIMM algorithm mapped 18 changes (7 fusions, 7 translocations and 4 inversions (reversals))..... 119

Figure 4-4: For the ancestor to the Peregrine falcon, the GRIMM algorithm mapped 30 changes (11 fusions, 1 fission, 10 translocations and 8 inversions (reversals)). 120

Figure 4-5: For the ancestor to the Saker (and hence Gyr) falcon, the GRIMM algorithm mapped 25 changes (9 fusions, 7 translocations and 9 inversions (reversals)). 122

Figure 4-6: Homology maps between chicken (GGA) and (a) budgerigar (*Melopsittacus undulatus*) (MUN) (b) cockatiel (*Nymphicus hollandicus*) (NHO). Colors show the orthologous chicken chromosomes. 125

Figure 4-7: Homology maps between chicken (GGA) and the ostrich (*Struthio camelus*) (SCA). Colors show the orthologous chicken chromosomes..... 126

Figure 5-1: FISH with PNA probe on the metaphase chromosomes of the Peregrine falcon showing different hybridization pattern of telomere probes on chromosomes. Chromosomes were counter-stained with DAPI..... 129

Figure 5-2: (a) and (b) Telomeric signals with intense fluorescence at specific chromosomes of the Peregrine falcon, captured with different exposure time to reduce the signal intensity.

Chromosomes were counter-stained with DAPI. (c) Same image is converted to black and white. 130

Figure 5-3: (a) and (b) FISH with telomeric PNA probe on the metaphase chromosomes of the Saker falcon showing intense fluorescence at specific chromosomes, captured with different exposure time to reduce the signal intensity. Chromosomes were counter-stained with DAPI. 131

Figure 5-4: (a) and (b) FISH with telomeric PNA probe on the metaphase chromosomes of the Gyrfalcon showing intense fluorescence at specific chromosomes, captured with different exposure time to reduce the signal intensity. Chromosomes were counter-stained with DAPI. (c) Same image is converted to black and white. 132

Figure 5-5: FISH with PNA probe on the metaphase chromosomes of the Nile crocodile (*Crocodylus niloticus*) showing telomere signal on all chromosomes. Chromosomes were counter-stained with DAPI. (b) Same image is converted without signal. 133

Figure 5-6: FISH with telomeric PNA probe on the metaphase chromosomes of the budgerigar (*Melopsittacus undulates*) showing strong fluorescence at specific chromosomes. Chromosomes were counter-stained with DAPI. (b) Same image is converted to black and white. 134

Figure 5-7: FISH with telomeric PNA probe on the metaphase chromosomes of the ostrich (*Struthio camelus*). Arrows indicate interstitial telomere sequences. Chromosomes were counter-stained with DAPI. (b) Same image is converted to black and white. 135

Figure 6-1: FISH hybridization results of selected probe sets to GGA and FPE metaphase chromosomes. 140

Figure 6-2: DAPI stained chicken (a) and Peregrine falcon fibroblast interphase nucleus (b) showing signals with dual colour FISH probe set 3 and 4; (c) and (d) same images after shell analysis, dividing the nucleus with five concentric shells of equal area showing two different signals within the nucleus. 144

Figure 6-3: Chart showing the mean overall position of chromosomes (from periphery, 0, to interior, 4) in chicken and Peregrine falcon nucleus. 145

Figure 6-4: Signal distribution for macrochromosomes in chicken (GGA) and Peregrine falcon (FPE). Orange indicates GGA and blue indicates FPE. 146

Figure 6-5: Signal distribution for microchromosomes in chicken (GGA) and Peregrine falcon (FPE). Green indicates GGA and brown indicates FPE. 148

Figure 6-6: Signal distribution for macrochromosomes in chicken (GGA) and Peregrine falcon (FPE). In FPE, they were formerly macrochromosomes but now are fused with microchromosomes. Orange indicates GGA and violet indicates FPE. 150

Figure 6-7: Chart showing the overall position of microchromosomes that are fused to macrochromosomes in Peregrine falcon and their corresponding chromosome position in chicken. 152

Figure 6-8: Signal distribution for microchromosomes in chicken (GGA) and Peregrine falcon (FPE). In FPE, they were formerly microchromosomes but now are fused with macrochromosomes. Green indicates GGA and purple indicates FPE. 153

Figure 7-1: Saker falcon karyotype (resolution) comparison of previous studies with present study (a) Amaral *et al.* 2003, (b) current study. 157

Figure 7-2: Schematic representation of proposed device showing different hybridization pattern on Peregrine falcon chromosome (FPE) and Saker falcon chromosome (FCH). 162

List of Tables

| | |
|--|-----|
| Table 1-1: Adapted from The IUCN Red List of Threatened Species. Version 2017-2. < www.iucnredlist.org >..... | 1 |
| Table 1-2: Hybridization between falcon species. Adapted from (Heidenreich 1997). Of the ones listed below Gyr-Saker hybrids are the most common..... | 7 |
| Table 1-3: Performance comparison of sequencing platforms of various generations (Rhoads and Au 2015)..... | 23 |
| Table 1-4: List of vertebrate species with whole genome published (OBrien <i>et al.</i> 2014)..... | 30 |
| Table 1-5: Examples of FISH applications (Riegel 2014). | 46 |
| Table 1-6: Karyotype characterizations of Falconiformes. Adapted from (Amaral and Jorge 2003). (Mac) macrochromosomes, (Mic) microchromosomes (-) absence of distinction between mac and mic..... | 65 |
| Table 1-7: Homologous chromosomes and chromosome segments between chicken and three Falco species as detected by chromosome painting using chicken macrochromosome probes. From (Nishida <i>et al.</i> 2008). | 69 |
| Table 3-1 : Genome statistics of chromosome level assemblies of Saker falcon (<i>Falco cherrug</i>) (Zhan <i>et al.</i> 2013)..... | 109 |
| Table 4-1: Neognathae ancestor configuration according to the output of MGRA2. | 115 |
| Table 4-2: Comparison of the number of conserved syteny segments (GGA 1-9) and fusion combinations in the budgerigar, cockatiel, ostrich and Saker/Gyrfalcon corresponding to chicken autosomes. | 127 |
| Table 5-1: Summary of telomeric sequences mapping results in three falcon species, budgerigar, crocodile and ostrich..... | 136 |
| Table 6-1: BACs details used in different probe sets and their corresponding hybridizing chromosomes in GGA (chicken) and FPE (Peregrine falcon). | 139 |

Abbreviations

| | |
|-------|---|
| BAC | Bacterial Artificial Chromosome |
| CAR | Contiguous ancestral region |
| CITES | Convention on International Trade of Endangered Species of Wild Fauna and Flora |
| DAPI | 4',6-diamidino-2-phenylindole |
| EBR | Evolutionary Breakpoint Region |
| EH | Evolution Highway |
| FCH | <i>Falco cherrug</i> |
| FCO | <i>Falco columbarius</i> |
| FISH | Fluorescence <i>in situ</i> Hybridization |
| FITC | Fluorescein isothiocyanate |
| FPE | <i>Falco peregrinus</i> |
| FRU | <i>Falco rusticolus</i> |
| FTI | <i>Falco tinnunculus</i> |
| GGA | <i>Gallus gallus</i> |
| GRIMM | Genome Rearrangements in Man and Mouse |
| HSB | Homologous Synteny Block |
| ITS | Interstitial Telomeric Sequences |
| IUCN | International Union for Conservation of Nature |
| KCl | Potassium Chloride |
| LB | Luria Bertani |
| MEM | Minimum Essential Medium |
| MGRA | Multiple Genome Rearrangement and Ancestors |
| msHSB | Multispecies Homologous Synteny Block |
| MUN | <i>Melopsittacus undulates</i> |
| Ma | Million years ago |
| NaCl | Sodium chloride |
| NGS | Next-generation sequencing |
| NHO | <i>Nymphicus hollandicus</i> |
| PBS | Phosphate Buffered Saline |
| PCF | Predicted Chromosome Fragment |
| PCR | Polymerase Chain Reaction |
| PNA | Peptide nucleic acid |
| RACA | Reference Assisted Chromosome Assembly |
| RVC | Royal Veterinary College, London |
| SF | Syntenic fragments |
| SNP | Single nucleotide polymorphism |
| SSC | Sodium saline citrate |

Abstract

Falcons and falconry have become an essential part of life in the Middle East since ancient times. In the United Arab Emirates (UAE) itself, the number of trained falcons ranges from 8,000 to 10,000. Over the last five years, falcon racing, a traditional sport, has gained momentum in the UAE where captive falcons are competing for huge prize money. A proportion of the UAE economy goes into their care and conservation e.g. through establishing falcon hospitals with modern facilities for disease treatment and breeding as well as centers for diagnosis and research. Being the national bird of the UAE, any research on falcons is of significant interest for the country. Most of the world's falcon species are in decline. Moreover, Saker falcons are classified as 'endangered' according to the IUCN Red List of Threatened Species. From the chromosomal perspective falcons are very interesting, as they represent birds that have undergone significant genome rearrangement compared to the "norm" of $2n \sim 80$.

'Molecular cytogenomics' in birds includes karyotyping, cross species comparisons, nuclear organization, BAC mapping, physical mapping and telomeric DNA profiling. This thesis makes use of the above approaches to define chromosome evolution and genome organisation in falcon species with the following results:

Firstly, successful conventional characterization of the Saker, Peregrine and Gyrfalcon karyotypes ($2n=50-52$) was achieved producing improved karyotypes and ideograms than those previously published. Comparative genomic analyses among these three-species using molecular cytogenetic approaches revealed differences between peregrine and the other two species, but none between Saker falcon and Gyrfalcon. Also, this study has supported upgrading the fragmented Saker genome assembly to chromosome level using a novel approach hitherto only published for the Peregrine falcon (and pigeon).

Secondly, a comparison of genome-wide BAC-based studies and bioinformatic analysis Multiple Genomes Rearrangement Algorithm 2 (MGRA2) revealed the chromosomal changes (inter- and intra-) that led to the falcon lineage. Also, the present study

established that common mechanisms of chromosomal fusion do not recur in two different groups of species with rearranged karyotypes (falcons and parrots).

This thesis also provided an overview of the telomeric DNA profile in the three species of interest. It established that the highly rearranged karyotypes studied (plus those of the budgerigar and crocodile) do not appear to possess interstitial telomeres at evolutionary fusion points. Also, this study demonstrated the existence of megatelomeres in falcon species, their nature differing between the Peregrine and the other two species studied.

Finally, this thesis produced the first detailed description of nuclear organization in a bird species (Peregrine falcon) other than the *Galloanserae*. Non-fused macro and microchromosomes behave the same way in chickens and falcons. This implies that the same general nuclear organization mechanisms are present in falcons as well as in chickens, ducks and turkeys whose last common ancestor existed around 89 million years ago. Most notably, fused microchromosomes in the Peregrine falcon retain the same nuclear organization pattern despite being fused to a larger chromosome. The findings from this study give insight into the basic nature of chromosome territory patterns in bird species with highly rearranged karyotypes.

Overall, results presented in this thesis provide significant insight into genome organization and evolution in the *Falco* genus, revealing previously undetected levels of chromosomal synteny between three species important to the UAE. Results generated here have also made a significant contribution to the chromosome-level genome assembly of the Saker falcon, providing tools for further study of avian species both within and beyond the falcon group.

1 Introduction

1.1 Falcons

Falcons are one of the most spectacular avian species in the world having particular significance in the Middle East region for hunting and conservation. A proportion of the United Arab Emirates (UAE) economy goes into their care and conservation e.g. through establishing falcon hospitals with modern facilities for disease treatment, breeding centers as well as diagnosis and research. The Peregrine falcon (*Falco peregrinus*) is reported to be the world's fastest bird (Zhan *et al.* 2013). Being the national bird of the U.A.E., any research on the falcon is of significant interest for the country. Most of the world's falcon species are in decline. Therefore, they are categorized in the Red List of threatened species (IUCN) as 'endangered', 'near threatened' and 'vulnerable' (Table 1-1). However, only Saker falcons are classified as 'endangered' according to the IUCN Red List of Threatened Species, whereas Gyrfalcons and Peregrine falcons are now classified as of 'least concern' (IUCN 2017).

Table 1-1: Adapted from The IUCN Red List of Threatened Species. Version 2017-2. <www.iucnredlist.org>.

| Name | IUCN Red List Category |
|--|------------------------|
| Saker falcon (<i>Falco cherrug</i>) | Endangered |
| Gyrfalcon (<i>Falco rusticolus</i>) | Least Concern |
| Peregrine falcon (<i>Falco peregrinus</i>) | Least Concern |
| Prairie falcon (<i>Falco mexicanus</i>) | Least Concern |
| Red-naped shaheen (<i>Falco pelegrinoides</i>) | Least Concern |
| Bat falcon (<i>Falco ruficularis</i>) | Least Concern |
| Laggar falcon (<i>Falco jugger</i>) | Near Threatened |
| Merlin (<i>Falco columbarius</i>) | Least Concern |
| American kestrel (<i>Falco sparverius</i>) | Least Concern |

S. Joseph

Research covering a wide range of aspects of their biology including morphology, ecology, physiology, genetics and genomics is essential to preserve them from extinction and maintain their cultural relevance.

1.1.1 A Brief History of Falconry

Falconry is an ancient sport which is generally believed to have originated on the Central Asian plateau, an area with the highest concentrations of birds of prey for falconry, like Saker falcons, Peregrine falcons, Lanner falcons as well as eagles (Remple and Gross 1993). During the political isolation of the East from the West (Western Europe), falconry advanced separately in both regions. China was an important falconry country and falconry spread to Korea where the first quarantine for falcons was established. Islam's greatest expansion between 632-936 A.D. favored the spread of falconry across the world (Wernery 2017).

Arabian falconry has a long history and today, the Arabian Peninsula and the Gulf states possess the highest number of Middle Eastern falconers. In the United Arab Emirates itself, the number of trained falcons ranges from 8,000 to 10,000 (Amirsadeghi 2008) and 25,000 in the entire Arabian Peninsula (Binothman 2016). The falcon being the national symbol, the heritage of Arabian falconry is strongest in the UAE compared to other countries.

1.1.2 Morphology

Falcons are the fastest of all bird species; their strong muscles, long wings and diurnal habit to hunt birds of prey are some of the characteristic features that distinguish them from other raptors (Zhan *et al.* 2013). They display powerful stocky beaks in comparison to their head, having a unique structure on the cutting edge of the upper mandible known as a 'tomial tooth' which corresponds to a notch on the cutting edge of the lower mandible. Unlike hawks and eagles, the feet of falcons lack the vice-like gripping features and the individual toes of falcons tend to be longer than in other raptors. Falcons hunt by two methods-the 'stoop and glancing blow' or the mid-air 'chase and grab'(Remple and Gross 1993). Falcons fly at remarkable speeds. Peregrine falcons can dive at a speed of 320 km/hr which makes this iconic bird the world's fastest animal (Remple and Gross 1993; Tucker 1998; Ponitz *et al.* 2014). Adult falcons weigh between 100 grams (the

smallest Kestrels) and 2 kilograms (the largest Arctic Gyrfalcons). In proportion to body size, female falcons are usually larger and weigh more than male falcons (Remple and Gross 1993).

1.1.2.1 Peregrine falcon (*Falco peregrinus*)

An adult Peregrine falcon is identified by sooty black feathers on the head and neck with a thick 'malar 'stripe extending down from the eye ("beard"). The cheek, throat and underparts are white to sienna orange. Their flanks display dark brown to black and the upperparts of the back and wings are slate-bluish grey barred with dark brown. Flight feathers incline towards dark slate-brown barred with lighter slate-grey; the tail end is colored white to beige and the skin of the cere, legs, feet and surrounding the eye is bright yellow to orange with no feathers (Figure 1-1) (Remple and Gross 1993).

Figure 1-1: Peregrine falcon



1.1.2.2 Saker falcon (*Falco cherrug*)

Variation in plumage coloration is more common in this species than in any other species of falcon. Most of the Sakers have buff-colored underparts which are streaked, blobbed or spotted with dark brown (Figure 1-2). Head feathers are buff colored streaked with brown, but markings do vary (Remple and Gross 1993).

Figure 1-2 : Saker falcon



1.1.2.3 Gyr falcon (*Falco rusticolus*)

The Gyr falcon is the largest bird in the genus *Falco*. Several plumage color variants exist among these large Arctic Falcons which are mainly white (Figure 1-3), silver, grey and blackish brown (Johnson and Burnham 2011; Johnson *et al.* 2012). Some birds have almost pure white plumage on the underparts with fine markings, while in some the color is various shades of grey brown to sooty slate-grey (Zhan *et al.* 2012).

Figure 1-3: Gyr falcon



1.1.3 Habitat and Geographical Distribution

1.1.3.1 *Peregrine falcon (Falco peregrinus)*

As the name means 'wanderer', this species has an extensive habitat ranging from the hot tropics to cold, wet marine habitats and in arid hot or cold deserts. They are highly migratory in Northern temperature and Arctic zones as shown in (Figure 1-4). Peregrines feed primarily on birds ranging from tiny song birds to geese and herons also rats, rabbits insects, reptiles and fish (del Hoyo *et al.* 2004).

1.1.3.2 *Saker falcon (Falco cherrug)*

Sakers have an extensive nesting range from steppes, wooded and open as well as in abrupt rocky areas. They have also been observed on plains and foothills, mountains and in high plateau up to 4,700 meters. They feed mainly on small mammals, and birds, especially sandgrouse, gamebirds and pigeons. During winter they are seen in Pakistan, Africa, and parts of the Middle East and China (del Hoyo *et al.* 2004) (Figure 1-5).

1.1.3.3 *Gyr falcon (Falco rusticolus)*

Gyr falcons occupy three basic habitats for breeding - maritime, riverine and montane. They are found in tundras and taigas, from sea-level to at least 1,400 meters (Figure 1-6). They feed mainly on birds and mammals. They hunt by flying low over their prey with high speed and striking it. The prey is fed on the ground (del Hoyo *et al.* 2004). Gyr falcons are poorly adapted for long term survival out of their natural habitat, the arctic region. They are immunologically as well as physiologically weak compared to other falcon species and can rapidly succumb to common temperate environmental pathogens (Remple and Gross 1993)

Figure 1-4: Global distribution of the Peregrine falcon indicating whether, summer or winter visitor, breeding resident or passage visitor.

(<https://www.beautyofbirds.com/peregrinefalcons.html>).

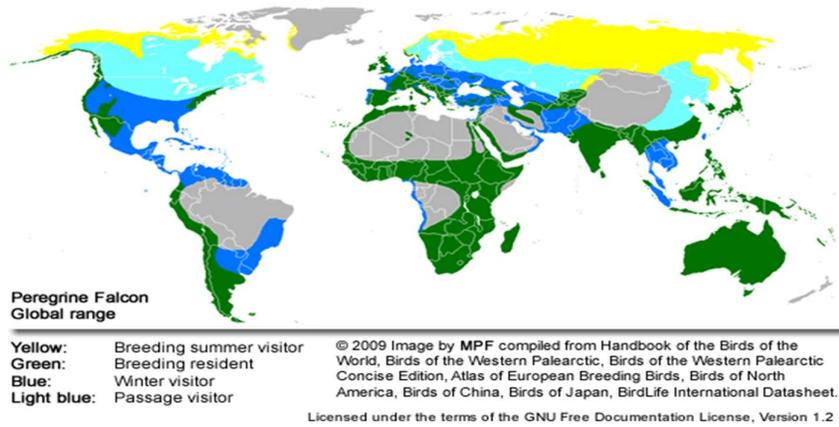


Figure 1-5: Global distribution of the Saker falcon indicating whether resident, breeding or feeding

Saker (http://www.avibirds.com/html/falcons/Saker_Falcon.html#.WhrpgUqWY2w).

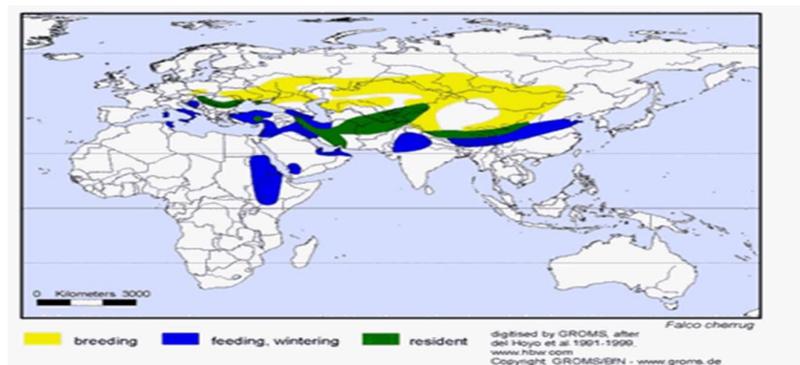
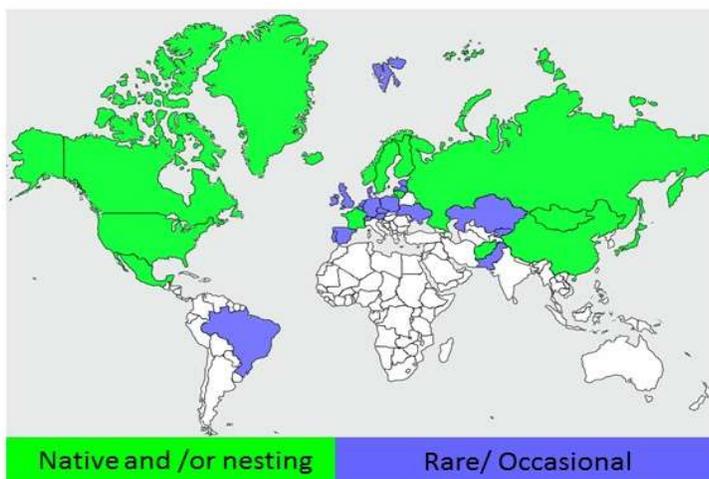


Figure 1-6: Global distribution of the Gyrfalcon indicating native and/or nesting compared to rare/occasional visitor

(<http://www.oiseaux.net/maps/gyrfalcon.html>)



1.1.4 Hybrids

It has been a long tradition among falcon breeders to produce hybrid offspring by crossing different raptor species. Mainly two practices are common-interbreeding subspecies as well as hybridization of birds from different species. Many reports have been published on the successful production of hybrid offsprings by crossing different falcon species (Table 1-2). A wide range of hybrids can be produced between *Falconidae* by artificial insemination. Around 12 species of falcon are used to produce hybrids in captivity in Great Britain by cross breeding in a wide variety of combinations (Fleming *et al.* 2011).

Table 1-2: Hybridization between falcon species. Adapted from (Heidenreich 1997). Of the ones listed below Gyr-Saker hybrids are the most common.

| | | |
|--|---|--|
| Peregrine falcon (<i>Falco peregrinus</i>) | X | Lanner falcon (<i>Falco biarmicus</i>) |
| | X | Saker falcon (<i>Falco cherrug</i>) |
| | X | Prairie falcon (<i>Falco mexicanus</i>) |
| | X | Red-naped shaheen (<i>Falco pelegrinoides</i>) |
| | X | Bat falcon (<i>Falco ruficularis</i>) |
| | X | Gyrfalcon (<i>Falco rusticolus</i>) |
| | X | Laggar falcon (<i>Falco jugger</i>) |
| | X | Merlin (<i>Falco columbarius</i>) |
| | X | American kestrel (<i>Falco sparverius</i>) |
| Gyrfalcon (<i>Falco rusticolus</i>) | X | Red-naped shaheen (<i>Falco pelegrinoides</i>) |
| | X | Saker falcon (<i>Falco cherrug</i>) |
| | X | Merlin (<i>Falco columbarius</i>) |
| Prairie falcon (<i>Falco mexicanus</i>) | X | Gyrfalcon (<i>Falco rusticolus</i>) |
| | X | Red-naped shaheen (<i>Falco pelegrinoides</i>) |
| Common kestrel (<i>Falco tinnunculus</i>) | X | Lesser kestrel (<i>Falco naumanni</i>) |
| | X | Gyrfalcon (<i>Falco rusticolus</i>) |

Most of the falcon hybrids are strong and fertile birds. The F₁ generation is robust and larger than their parents, whilst being less prone to diseases. This is important especially for Gyrfalcons in hot climates.

Fertile falcon hybrids are a potential threat that can alter the gene pool of the wild population by mating with wild falcons. However, many observations have been made on the natural pairing of different species of falcon (Boyd and Boyd 1975; Cade 1982; Eastham and Nicholls 2005). Hybrids produced by falcon pairs which are less-closely

related such as the Gyrfalcon and Peregrine falcon, exhibit weakened fertility (Eastham and Nicholls 2005). In general, male hybrids possess deformed spermatozoa and female hybrids are completely sterile (Heidenreich and Küspert 1992; Eastham and Nicholls 2005). Experiments conducted to determine the fertility status of offspring produced from crossing Peregrine falcons and hierofalcon group concluded that F₂ eggs are fertile but the embryos usually do not survive the incubation period (Heidenreich and Küspert 1992). However, there are no such limitations present within the hierofalcon group, where hybrids are fertile and able to hybridize unlimited number of generations (Figure 1-7). Hierofalcons are superspecies complex of the genus *Falco* comprising Saker falcon (*Falco cherrug*), Lanner falcon (*Falco biarmicus*), Laggar falcon (*Falco jugger*) and Gyrfalcon (*Falco rusticolus*) (Nittinger *et al.* 2007).

Even for experts an exact identification of falcon hybrids is often not possible based on the phenotypic characteristic alone. Recent new laws implemented in Europe, clearly laid down the role of hybrid falcons for the falcon breeder. With these laws, hybridization between captive bred and wild falcons is prohibited (Heidenreich 1997)

Figure 1-7: Falcon hybrid (Gyr x Peregrine)



1.1.5 Relevance to Contemporary Society

Arab falconers trap wild falcons and train them to hunt other birds. These 'passage falcons', at the end of each season, were released back into the wild (Heidenreich 1997). However, times have changed, and hunting is restricted in the UAE due to rapid urbanization which, during the past 20 years has swallowed up most of the desert habitat and altered the entire ecosystem of the plains. Furthermore, strict regulations for each hunting bird have been put in place. Today, falcons are kept in air-conditioned rooms or in free-flying aviaries. This new culture, along with a significantly larger population of falcons in the country, prompted the need for professional health care facilities for falcons. Many modern falcon hospitals have been established in most of the countries in the Middle East region (Barton 2000).

During the 'golden age' of falconry (500 A.D. to 1600 A.D.), trapping and hunting birds was a major business in most of the falconry countries. However, this waned in the 1960s, because the world-wide raptor population had severely declined due to several reasons e.g. the excessive use of pesticides (Porter 1993; Steidl 1991), overhunting and habitat destruction, due to the implementation of new laws that governed the acquisition of raptors. Several Middle East countries have now signed the "Convention on International Trade of Endangered Species of Wild Fauna and Flora" (CITES), therefore only captive-bred birds are allowed into the country to be used in the sport of falconry (Soorae *et al.* 2007). Every hunting falcon must have a CITES ring and a passport. This new rule stopped the decline of the severely depleted wild population of raptors of the world. Despite these new encouraging developments to safeguard the wild falcon population, in 2015, more than 4,000 falcons were illegally brought to the Arabian Peninsula alone (Binothman 2016). Falcon hunting is geographically spread across Eastern China to the Red Sea coast (Barton 2000). The Saker falcon is classified as endangered according to the IUCN Red List of threatened species, whereas Gyrfalcon and Peregrine falcons are now classified as "Least Concern" (IUCN 2017). However, in 2010 UNESCO, the United Nation's cultural agency, decided to include falconry in the 'Convention for Safeguarding of the Intangible Cultural Heritage' list in order to preserve it (Wernery 2017).

Over the last five years, another aspect of falconry has gained momentum. This is “falcon racing” which covers some 500 meters or 1 km distance. This racing is divided into different categories according to the gender, species and age. It has become a very competitive sport in the UAE with enormous prizes for the winning falcon owners. This sport is new and very important sport for young Emiratis. At the same time, it also gives scientists an important role to play for future research in supplying the falcon racing authority with details of genomic purity of falcons to avoid any advantage for falconers using hybrid falcons in racing. Additionally, it partly takes away the hunting pressure from the ‘quarry’ which is the Houbara Bustard (*Chlamydotis undulata*) which has also declined in recent decades due to overhunting and habitat destruction (Remple and Gross 1993).

1.1.6 Phylogeny of the *Falconidae*

Over the past decades, phylogenetic investigations have become necessary for interpreting biological data (Griffiths *et al.* 2004). A well formatted phylogeny can serve as a significant tool for taxonomic and systematic classification. Different approaches and different types of data have been analyzed to create the phylogenetic relationship of the *Falconidae* family; these include osteological, morphological and molecular data (Becker 1987; Boyce 1989; Griffiths 1994; Griffiths 1999; Suschkin 1905).

The current taxonomic classification of the *Falconidae* family started with Suschkin (1905). He investigated skeletons of 140 individual birds of Falconiforms and recognized four subfamilies based on external characteristics only. These are: *Herpetotherinae* for the two neotropical forest-dwelling genera *Herpetotheres* and *Micrastur*; *Polyborinae* for the New World caracas; *Falconinae* for the genus *Falco*; and *Polihieracinae* for the two Old World genera of falconets. Amadon and Bull (1988) recognized two major ‘groups’ of falconids, without subdividing the family which are the *Polyborinae* and the *Falconinae*. The *Polyborinae* includes seven genera: *Daptrius*, *Milvago*, *Polyborus*, *Phalcoboenus* (the caracaras), *Micrastur* (forest falcons), *Herpetotheres* (Laughing Falcon) and *Spizapteryx* (Spot-winged Falconet). The *Falconinae* comprises three genera: *Falco*, *Polihierax* (pygmy falcons) and *Microhierax* (falconets). Reclassification has occurred in the family placing *Herpetotheres* and *Spizapteryx* within the *Falconinae*

and *Micrastur* in its own subfamily (Griffiths 1994). The *Falconidae* is one of the four families of Falconiformes having ~ 64 species. Griffiths (1999) classified 11 genera in the family into two subfamilies (Figure 1-8) which is in accordance to the recent systematic classification (Griffiths 1999). The first sub family, the basal branch, the *Herpetotherinae*, consists of *Micrastur* (the Forest falcons), and *Herpetotheres* (the Laughing falcon). The second sub family, the *Falconinae* is comprised of two tribes, the Falconini, and the Caracarini. The Falconini includes *Polihierax*, *Microhierax* (pygmy- falcons and falconets), *Spizapteryx* (the spot-winged falconet) and the genus *Falco*. The second tribe, the Caracarini is composed of five Neotropical genera (*Daptrius*, *Ibycter*, *Milvago*, *Caracara* and *Phalcoboenus*) (Griffiths *et al.* 2004) as illustrated in Figure 1-9. Seibold *et al.* (1993) analysed the relationship between different members of the family *Falconinae* and crested caracara (*Polyborus plancus*) using mitochondrial cytochrome b gene as the marker gene (Figure 1-9). Griffiths's (1999) study was based on the variation in syringeal morphology and mitochondrial cytochrome-b sequences to summarize falconid phylogeny. The disadvantage of his study was not having concrete support for the basal nodes. To affirm the basal divergence within *Falconidae*, Griffiths *et al.* (2004) conducted a phylogenetic analysis of the nuclear gene *RAG-1* of 15 falconid and 2 outgroup species. The *RAG-1* data concluded the basal evolutionary history of the genera of the *Falconidae*, however, this study alone could not resolve species relationship within the genus *Falco*.

Figure 1-8: Phylogeny and classification of the *Falconidae* based on morphological and mitochondrial cyt-b data (Griffiths 1999; Griffiths *et al.* 2004).

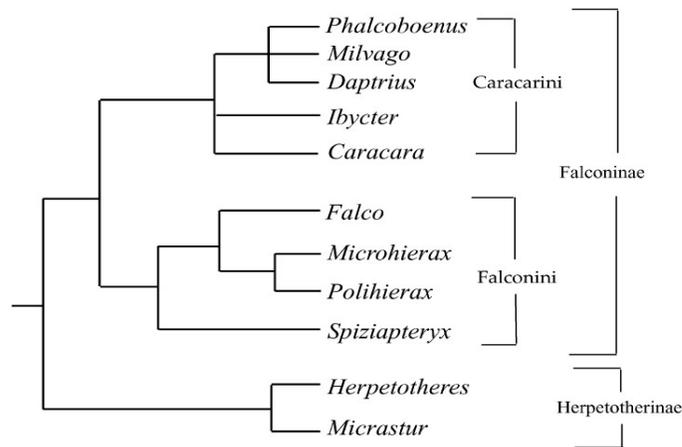


Figure 1-9: Phylogenetic relationship (with *Gallus* as out-group) between different members of the family *Falconinae* and crested caracara adapted from (Seibold *et al.* 1993)

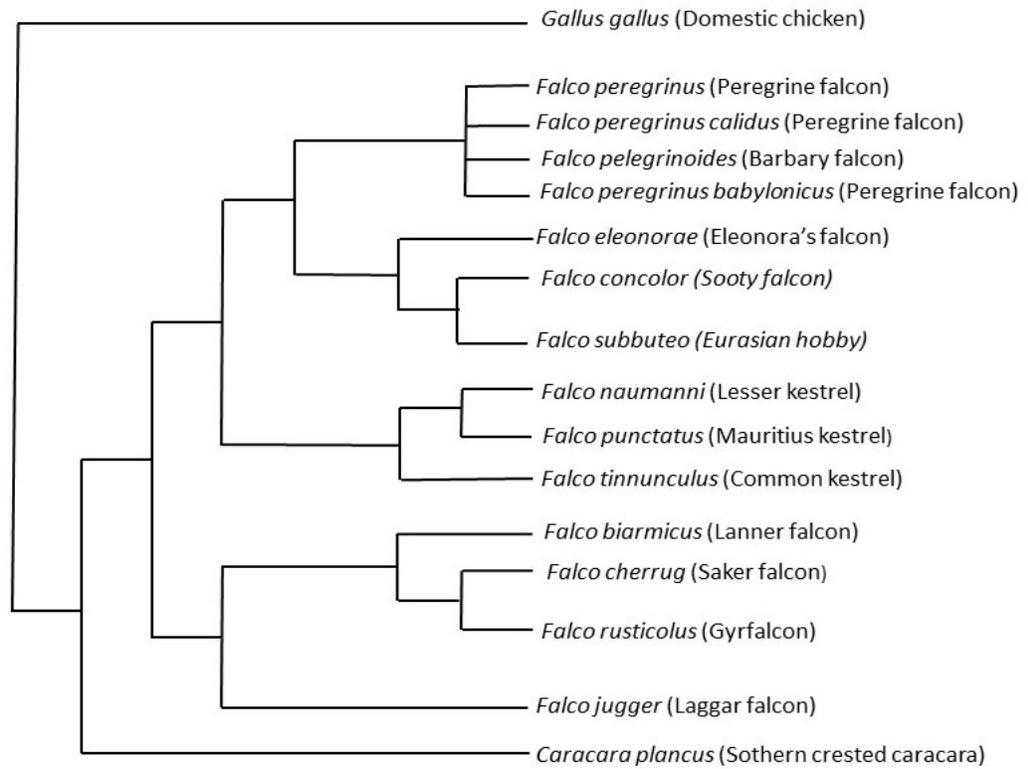
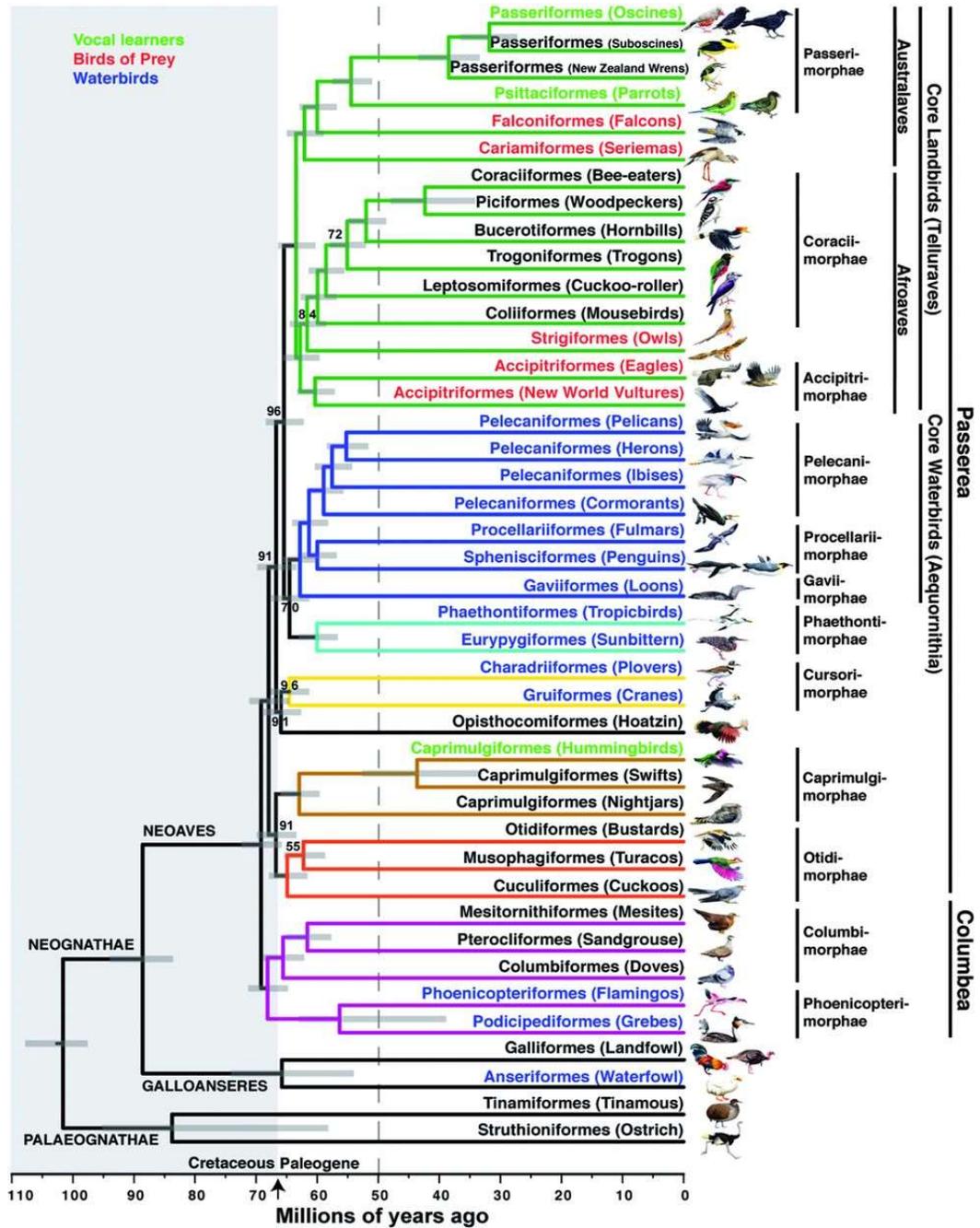


Figure 1-10: Phylogenetic relationship between Falconiforms with other major avian groups (Jarvis *et al.* 2014).



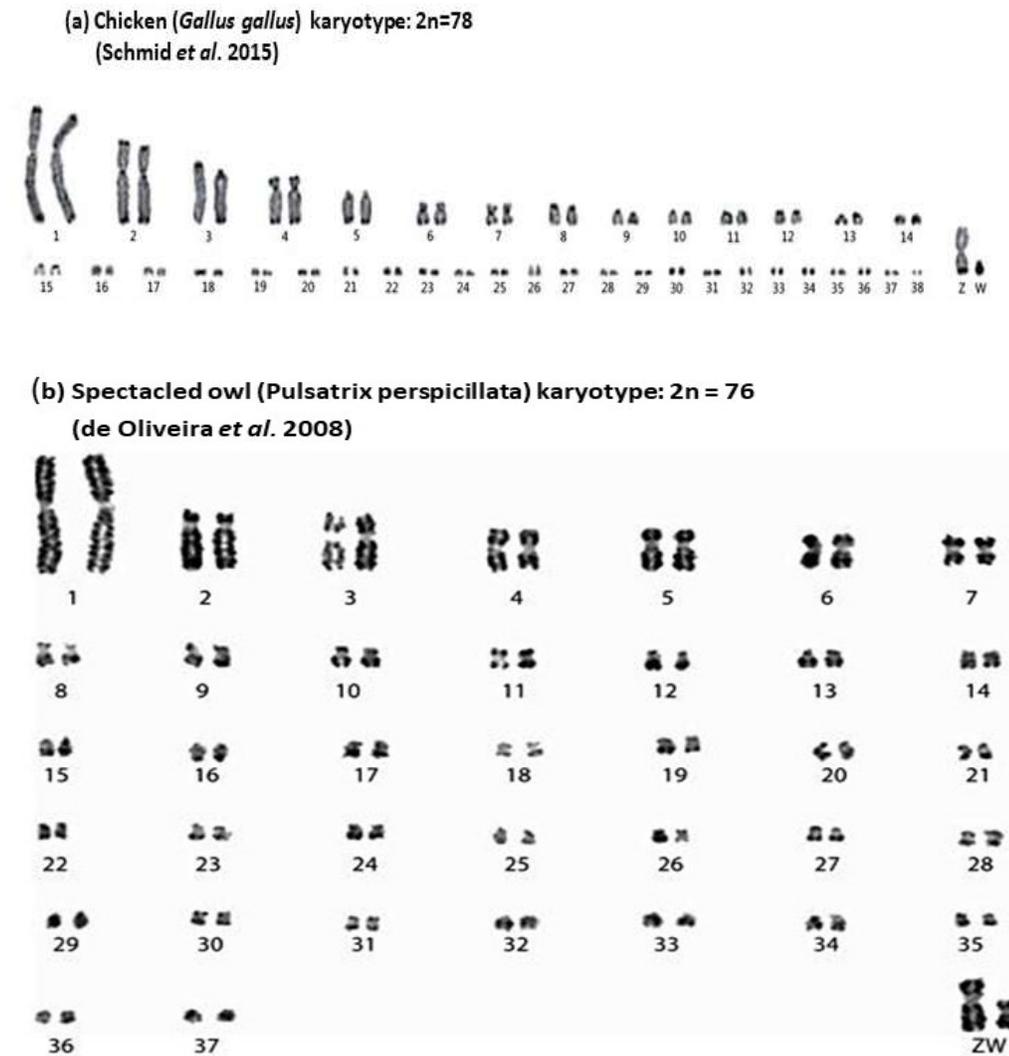
Previous phylogenetic studies suggested that falcons are closely related to other birds of prey, species such as eagles and New World vultures, however the latest phylogenetic study has yielded the most reliable tree of life for birds to date (Hackett *et al.* 2008; Jarvis *et al.* 2014. Prum *et al.* (2015) confirming that falcons are more closely related to parrots and song birds (Figure 1-10). Early genome comparative studies (before 2008) have therefore been performed with a different approach especially in terms of falcon genome evolution.

One of the unique biological features of falcon species is their genome organization (e.g. karyotype). This is dealt with specifically in section 1.7 however, in the following sections 1.2-1.6 it is important to consider avian genome organization and sequencing as a whole. Once we do this the unique nature of falcons' genome become apparent, not least of which is far fewer chromosomes than is typically seen in avian species.

1.2 Avian Genome Organization

Avian genomes are characteristic in the large variation in chromosome size, with most species having ~30 pairs of microchromosomes. Most avian karyotypes typically have large number of chromosomes with a diploid count of $2n \sim 80$ (Figure 1-11). Around 63% of bird species have $2n=74-86$, 24% have a $2n=66-74$; extremes of $2n=40$ and $2n=142$ however exist (Griffin *et al.* 2007). Section 1.7 has more details on avian karyotyping however, for the present, we will consider some general features of avian genome organization.

Figure 1-11: Representative G-banded karyotypes for (a) Chicken and (b) Spectacled owl.



Studies have compared many sequence characteristics, such as the percentage of GC content, CpG island density and genes have been studied with the help of many avian genome sequences. Hillier *et al.* (2004) found that ~38% of chicken CpG islands are conserved in the human genome and gene density correlates negatively with chromosome length which is confirmed in zebra finches (Stapley *et al.* 2008). A study that compared the chicken-turkey intron and coding sequence alignments revealed a clear difference in the effect of evolutionary forces between the different chromosome size classes (Axelsson *et al.* 2005). Microchromosomes showed 18% higher sequence

divergence in introns and a 26% higher rate of synonymous substitutions in coding sequences than macrochromosomes, indicating microchromosomes are prone to germ line mutations.

1.2.1 Chromosome Territories

Since 19th century there has been much research on numerous aspects of nucleus structure. Carl Rabl, an Austrian scientist, was the first to suggest the concept of chromosome territory organization during interphase (Rabl 1885). However, the term chromosome territory (CT) was introduced by Theodor Boveri (Boveri 1909). Boveri claimed that chromosomes try to occupy in a definite position in the nuclear space during interphase. With the introduction of the electron microscope, researchers have concluded that the nucleus is filled with intermingling chromatin fibers (Cremer and Cremer 2010). Using a modified Giemsa stain the visualization of chromosomes in Chinese hamster cells supported the concept that they remain in specific position within an interphase nucleus (Stack *et al.* 1977).

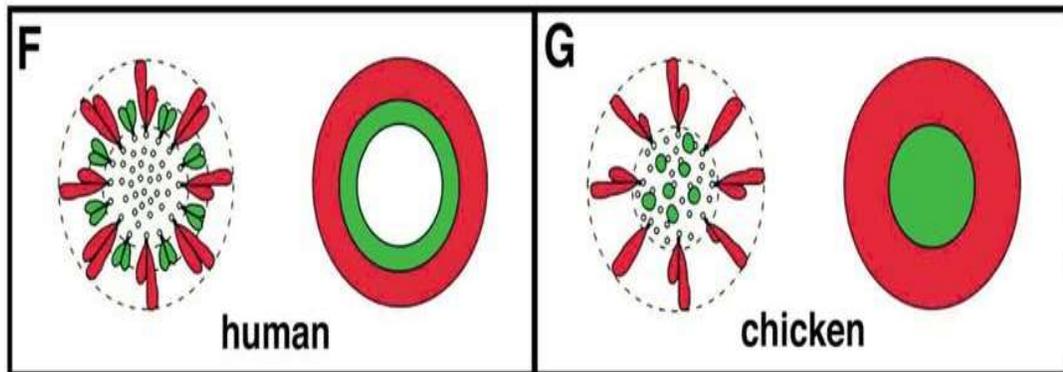
In the mid 1980s, development of FISH (Fluorescent *in situ* hybridization) techniques propelled the possibility of direct visualization of CT with the initial experiment being performed on cell hybrids which harbor one or a few human chromosomes and hamster genome (Schardin *et al.* 1985). Production of chromosome-specific painting probes enable individual chromosomes to be visualized in metaphase and interphase (Fawcett *et al.* 1994). Introduction of three-dimensional (3D) FISH, laser confocal microscopy to perform light optical serial sectioning of nuclei and computational 3D imaging systems made it possible to understand the higher order arrangements of CT (Cremer and Cremer 2010).

1.2.2 Models for Nuclear Organization

Today many models exist to explain how chromosomes are distributed in the interphase nucleus. However, two of the models stand out among the rest with substantial evidence. They are 'gene density based' organization and 'chromosome size based' organization (Skinner *et al.* 2009b). A non-random radial distribution of entire CTs was recognized in human lymphocyte nuclei based on the studies conducted using human

chromosome painting probes of a human chromosome which has the highest gene density and gene-poor chromosomes. Gene-rich chromosomes consistently locate at the interior and the gene poor chromosomes at the periphery (Cremer and Cremer 2001). Additionally, gene-density related radial arrangements have been observed for all human chromosomes and this is evolutionary conserved (Cremer and Cremer 2001). Cremer and Cremer (2010) present data supporting a key role of gene density (2-10 Mb) for the radial position of chromatin in the nucleus (Kozubek *et al.* 2002). Size-based distribution of chromosomes, where the largest are located near the periphery and smaller chromosomes are found toward the interior (Figure 1-12). This arrangement has been studied in both human fibroblast and amniotic fluid cell nuclei which are flat-ellipsoidal in shape (Bolzer *et al.* 2005). However, other factors such as transcriptional activity, replication timing and GC content are linked with nonrandom radial nuclear arrangements of CTs (Grasser *et al.* 2008; Hepperger *et al.* 2008).

Figure 1-12: Schematic representation of radial arrangements of large (red) and small (green) chromosomes in the cell nucleus of human and chicken (Habermann *et al.* 2001).

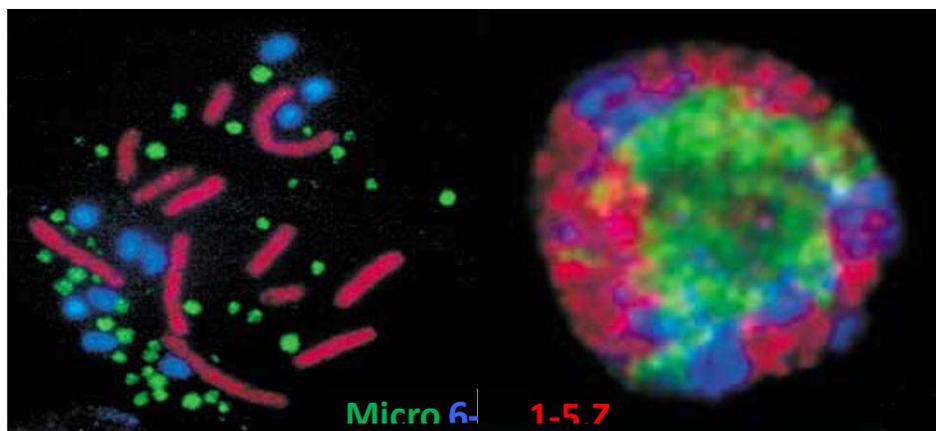


1.2.3 Nuclear Organization In Birds

Nuclear organization studies are few in birds compared to those in humans and other primates (Skinner *et al.* 2009b). Species chromosome territory studies used the chicken as a model. Habermann *et al.* (2001) undertook a detailed study of chromosome territory in the nuclei of chicken fibroblasts and neurons using multicolor chromosome painting, two-dimensional imaging and three-dimensional (3D) reconstruction. Chromosome-specific painting probes were generated to GGA chromosomes 1-10 and Z, and 19 pairs of microchromosomes in different pools. Multicolor FISH results showed a specific radial arrangement pattern in both fibroblast and neuronal cells. Large

chromosomes (1-5 & Z) and medium size chromosome (6-10) were positioned near to the periphery of the nucleus whereas 19 microchromosome pairs were arranged in radial cluster in the inner part of the nucleus (Figure 1-13) (Habermann *et al.* 2001). Taking into consideration that microchromosomes are gene rich, size-related and gene density related nuclear organization models are relevant with the stated result (Skinner *et al.* 2009b). Studies using GC-rich and GC-poor chicken isochores probes on Falconiformes revealed that GC rich isochores are located internal at interphase and GC poor isochores were distributed near to the periphery (Federico *et al.* 2005). This study again shows that gene density related nuclear organization pattern is prominent among birds. Skinner *et al.* (2009b) conducted a detailed study to determine the relative locations of chromosome territory in chicken fibroblast cells using FISH probes from chicken chromosomes 1-28. The same FISH probes were hybridized to turkey and duck to generate comparative genomic data. However, this study suggested that chromosome size based pattern is more appropriate model for chromosome territory distribution in chicken and similar observations were also seen in turkey and chicken. These studies show that a common pattern of genome organization exists in mammals and birds despite their different chromosome size and morphology (Habermann *et al.* 2001; Skinner *et al.* 2009b). To date however few or no studies have analyzed nuclear organization in falcon species.

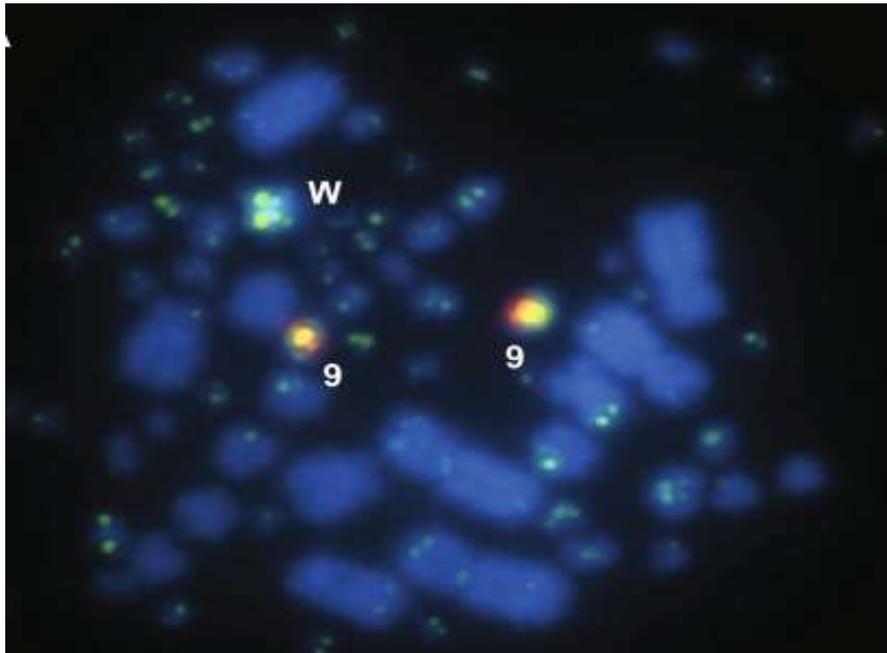
Figure 1-13: Chicken chromosome paints for larger (red) and smaller (blue) macrochromosomes and for microchromosomes (green) on chicken metaphase (left) and distribution of chromosome territories in chicken interphase (right) (Habermann *et al.* 2001).



1.2.4 Telomeres

Telomeres are nucleoprotein complexes consisting of tandemly repeated (TTAGGG)_n sequences at the end of eukaryotic chromosomes (Bauch *et al.* 2012). These specialized, highly conserved, G-rich DNA sequences prevent chromosome end replication problem and provide cellular maintenance (Bolzan and Bianchi 2006). Telomeres were first observed in human chromosomes and thereafter telomeric DNA sequences have been observed in many vertebrates (Nanda *et al.* 2002). Though telomeres are considered as terminal specialized structures, telomeres also occur at non-telomeric sites known as Interstitial Telomeric Sequences (ITS); these can be positioned between the centromere and the actual telomere or can be located near to the centromeres (Ocalewicz 2012). It has been estimated that in most avian species, each diploid genome contains 2-4% telomere sequences (Delany *et al.* 2000). In the chicken, telomeric DNA sequences extends from 0.5kb to about 2Mb which span up to 4% of its diploid genome, whereas, the telomere amount in the human diploid cell is strikingly low (about only 0.3%) (Delany *et al.* 2003). Three classes of telomeric arrays are reported in the chicken genome; they differ in location, age-related stability and size: Class I 0.5 to 10 kb (interstitial), Class II 10 to 40 kb (terminal) and Class III 40 kb to 2 Mb (terminal). The largest telomeric arrays (Class III) are referred as mega-telomeres as analyzed by molecular approaches and via cytogenetic approaches which are shown in (Figure 1-14) (Delany *et al.* 2000; Delany *et al.* 2007; O'Hare and Delany 2009). Nevertheless, it has been shown that most of the Class III telomeric array locate on microchromosomes (Nanda *et al.* 2002).

Figure 1-14: Chromosomal location (GGA 9 and GGA W) of mega-telomere in the chicken (O'Hare and Delany 2009).



Telomere DNA sequence studies in many bird species show different distribution pattern of telomeres as centric, interstitial as well as at the chromosome ends. Many primitive birds (*Paleognathae*) such as the ostrich (*Struthio camelus*), emu (*Dromaius novaehollandiae*), the American rhea (*Rhea americana*) and chicken show different interstitial (TTAGGG)_n hybridization patterns along the entire macrochromosomes (Nanda *et al.* 2002; Ocalewicz 2012). Centrometrically placed telomeric DNA sequences are observed in some birds like the quail (*Coturnix coturnix*), pheasant (*Phasianus colchicus*), eagle owl (*Bubo bubo*) and great grey owl (*Strix nebulosa*). However, the budgerigar (*Melopsittacus undulatus*) does not show any non-telomeric (TTAGGG)_n pattern in spite of having many banded chromosomes (Nanda *et al.* 2002). Nanda *et al.* (2002) also observed that the amount of telomere sequence is much higher in microchromosomes than in macrochromosomes.

Hausmann *et al.* (2003) measured telomere length in five different bird species with distinctly different life spans and suggested that shortening of telomere length corresponds to the lifespan of each bird. Birds with shorter lifespan lose more telomere repeats with age than birds with longer lifespans. Moreover, though Leach's storm-

petrels (*Melopsittacus undulatus*) is a markedly long-lived bird, telomere repeats do not shorten with age suggesting that long-lived organisms evolved from telomere shortening mechanism (Hausmann *et al.* 2003). Studies in birds such as Leach's storm-petrel and the common terns (*Sterna hirundo*) shows that telomeres do not degenerate, but at the same time maintain a high level of telomerase expression throughout their lifespan (Hausmann *et al.* 2005). To date however telomeres of falcons remain under-explored.

1.3 Avian Genome Sequencing

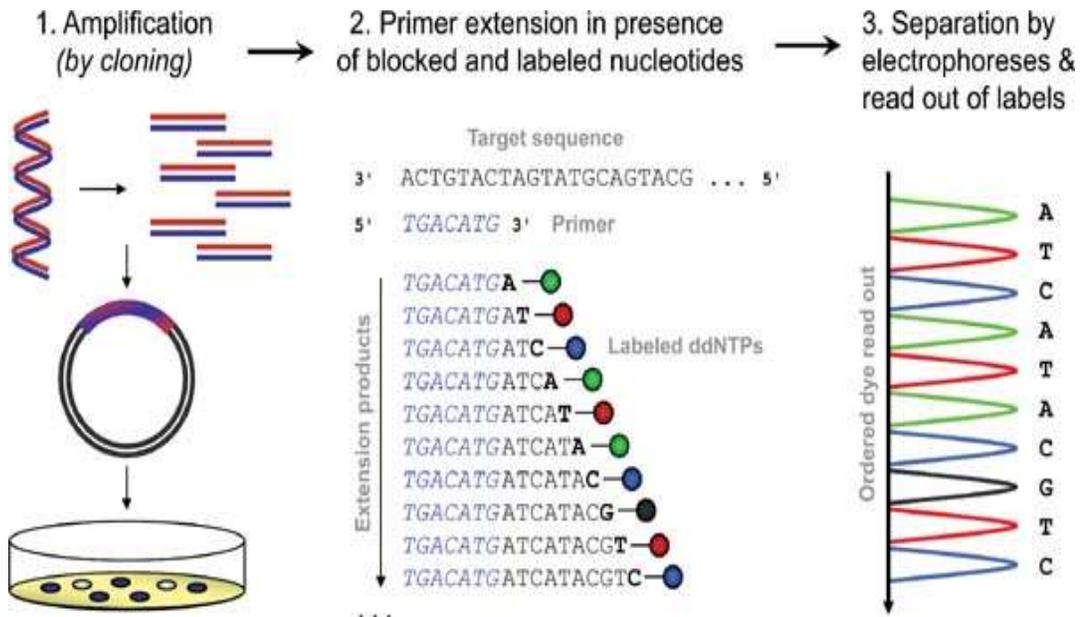
Genome sequencing is a powerful tool that has provided much basic information about nucleotides by thorough analysis of DNA. Entire genome sequencing enables scientists to understand how the genome works as a whole and how genes lead to growth, development and maintenance of an organism. Entire genome sequence knowledge also helps scientists to study the regulatory regions and "nonsense" regions, compare homologous genes across species and identify mutations. DNA sequencing techniques have become popular in many fields from archaeology, anthropology, genetics, biotechnology, molecular biology, forensic science and other relevant fields, leading to new discoveries in many fields which refashion its conceptual foundation (França *et al.* 2002).

1.3.1 Genome Sequencing

Sanger and his colleagues introduced a revolutionary technique for sequencing oligonucleotides with commercially available DNA polymerase (Sanger *et al.* 1977). In brief, target DNA is amplified by cloning in bacterial vectors. Using a mixture of deoxy-nucleotides (dNTPs) and dideoxy-nucleotides (ddNTPs) reverse strand synthesis is performed creating differentially extended molecules (Figure 1-15). These molecules after denaturation are sorted using capillary electrophoresis. Since the beginning of the new millennium, science has exploded with comprehensive information about human development, physiology, medicine and evolution by revealing the human genome sequence (Lander *et al.* 2001). The idea of sequencing the complete human genome was first proposed in 1984 and finished in 2001. The human genome sequencing effort

employed a 'hierarchical shotgun sequencing' approach, also referred to as 'map-based', 'BAC-based' or 'clone-by-clone'.

Figure 1-15: Schematic representation of the Sanger sequencing process (Kircher and Kelso 2010).



The next quantum leap in the history of sequencing started in 2005 when Margulies *et al.* (2005) introduced 454 sequencing, the first next generation sequencing method. With this new invention, the cost of sequencing a human genome was estimated between \$10 and \$25 million. Next generation sequencing (NGS) technologies discarded the use of bacterial vectors, Sanger sequencing and existing assembly approaches are the main source of sequencing technology (Margulies *et al.* 2005). Since then many companies have released commercial platforms with increased bases yield per sequence run. NGS has broadened the existing boundaries of the scientific field and opened up new areas such as the investigation of ancient genomes and metagenomic analyses of environmentally derived samples (Mardis 2008). With the introduction of the second-generation genome sequencing platforms from the world's established biotechnology companies like Roche/454, Illumina, and ABI, scientists have discovered genome sequences from many extinct animals.

Sequencing technology grew rapidly and by 2011 many so called third generation sequencing platforms were released. Outstanding genome sequencing platforms, such as the Ion Torrent Personal Genome Machine (PGM), the Pacific Biosciences (PacBio) RS (Quail *et al.* 2012) and the MinION (Lu *et al.* 2016) were tried on new sequencing technology. Performance comparison of sequencing platforms of various generations is illustrated in the (Table 1-3). New discoveries in genome sequencing have transformed today's biology especially in the field of comparative genomics, mutation discovery, gene expression, DNA bar- coding, metagenomics and epigenomics.

Table 1-3: Performance comparison of sequencing platforms of various generations (Rhoads and Au 2015).

| Method | Generation | Read length (bp) | No. of reads per run | Time per run | Cost per million bases (USD) |
|-----------------------------------|-----------------|------------------|----------------------------|--------------|------------------------------|
| Sanger ABI 3730xl | 1 st | 600-1000 | 96 | 0.5-3 h | 500 |
| Ion Torrent | 2 nd | 200 | 8.2×10^7 | 2 – 4 h | 0.1 |
| 454 (Roche) GS FLX | 2 nd | 700 | 1×10^6 | 23 h | 8.57 |
| Illumina HiSeq 2500 (High Output) | 2 nd | 2 x 125 | 8×10^9 (paired) | 7-60 h | 0.03 |
| Illumina HiSeq 2500 (Rapid Run) | 2 nd | 2 x 250 | 1.2×10^9 (paired) | 1 – 6 days | 0.04 |
| SOLiD 5500 x 1 | 2 nd | 2 x 60 | 8×10^8 | 6 days | 0.11 |
| PacBio RS II: P6-C4 | 3 rd | 1.0-1.5 x 10^4 | $3.5-7.5 \times 10^4$ | 0.5-4 h | 0.40-0.80 |
| Oxford Nanopore MinION | 3 rd | 2-5 x 10^3 | $1.1-4.7 \times 10^4$ | 50 h | 6.44-17.90 |

1.3.2 The Chicken Genome

Being a member of “big 10” sequenced genomes (International Chicken Genome Sequencing Consortium (ICGSC) 2004), the chicken (*Gallus gallus*) is a significant model for studies of health disease and biology and represents the first agricultural bird to have its genome sequenced. Apart from its importance in agriculture, the chicken is a major model organism for human disease. It is used as a model for musculoskeletal disease (e.g. osteomyelitis, muscular dystrophy), infectious diseases and their resistance (e.g. influenza, salmonellosis, adenovirus), immune system disorders such as autoimmunity, haematopoietic disorders (e.g. thrombosis, vitiligo), cardiovascular complaints like cardiomyopathy, atherosclerosis, cancer (leukaemia, melanoma) and even vision-related conditions such as retinal degeneration and myopia (Griffin and Burt 2014).

1.3.2.1 The Genetic Map

Although chicken genetics dates back to the early 1900s, the first “classical” genetic map of the chicken was introduced by F.B. Hutt in 1936. The genetic (linkage) map is formed by the segregation of meiotic products, with distance in centiMorgans (cM). Three different breeds have been used for the construction of chicken genetic maps ; East Lansing (Crittenden *et al.* 1993), Compton (Bumstead and Palyga 1992) and Wageningen (Groenen *et al.* 1998). The map has a current size of 4,200cM with 2,261 loci on 53 linkage groups, 31 of which have been assigned to a particular chromosome (Masabanda *et al.* 2004). Although, microsatellites and amplified fragment length polymorphisms (AFLP) were the main types of markers on the consensus linkage map, single nucleotide polymorphism (SNP) markers became the dominant type of markers (Schmid *et al.* 2005). Sequence maps for the chicken “D group” (chromosomes 33-38) are currently unavailable, however the ultimate aim is to construct linkage and sequence maps for all 39 individual chromosomes in the chicken genome (Griffin and Burt 2014).

1.3.2.2 The Physical Map

A physical map is extremely useful to provide long-range linking of assembly - supercontigs to anchor sequence contigs - to the genetic map and to supply templates for closing gaps in the draft sequence assemblies (Schmid *et al.* 2005).

Though a significant number of large insert libraries based on BAC (bacterial artificial chromosome) clones have been used to generate a physical map of the entire chicken genome (Crooijmans *et al.* 2000; Ren *et al.* 2003), five of these are predominantly used (Burt 2004; Schmid *et al.* 2005). These BAC libraries are generated from an inbred Red Jungle fowl (Ren *et al.* 2003) which was used in whole genome sequencing and also from a White Leghorn bird (Crooijmans *et al.* 2000). Collaborative efforts culminated in the development of a comprehensive BAC contig map covering 95% of the chicken genome (Wallis *et al.* 2004). Over 180,000 BAC clones have been involved to construct a physical map of 260 contigs of which 226 have been allocated to a specific chromosome.

1.3.2.3 Genome Sequencing

First draft genome sequencing was developed from a single female of the inbred line of red jungle fowl (UCD 001) based on the whole genome shotgun approach and supported with the sequences from plasmid, fosmid and BAC-ends (Burt 2004). This sequence exhibits a 6.6-fold coverage of the genome and was assembled using the parallel computation assembly program (PACP). Sequencing generated a genome of 1.05 gigabases (GB), of which 933Mb were anchored to chicken chromosome [International Chicken Genome Sequencing Consortium (ICGSC) 2004]. A major advantage of the chicken genome sequence has been the set of gene predictions. An evidence-based approach (Ensembl) and two comparative ab initio methods (Twinscan and SGP-2) together generated 106,749 predictions of protein-coding exons. Another 85,929 additional exons have been predicted by one or two methods making a total of 20,000-23,000 genes [International Chicken Genome Sequencing Consortium (ICGSC) 2004]. The chicken genome sequence is a milestone in both avian biology and agriculture. It has informed us about the nature of birds and other vertebrates (Burt 2004).

1.3.3 The Zebra Finch Genome

The zebra finch (*Taeniopygia guttata*) belongs to the largest order of birds on earth, the Passeriformes, and represents an important model bird in many fields (Clayton *et al.* 2009; Zann 1996). This model organism has played a significant role in generating insight into the neurobiology of human language. The Zebra finch communicates through learned vocalization sharing this trait with only two distantly related groups of birds

parrots and hummingbirds (Doupe and Kuhl 1999; Jarvis 2004). The genome sequencing was generated from a male zebra finch to describe the Z chromosome. Genome sequencing and assembly were performed using methods described for chicken genome sequencing. 1.2 gigabase (GB) draft assembly was developed and using zebra finch genetic linkage and bacterial artificial chromosome (BAC) fingerprint maps, 1.0 GB has been anchored to 33 chromosomes and three linkage groups. Ensemble predicted 17,475 protein-coding genes from zebra finch genome assembly. Gene expression in the forebrain of zebra finches were analyzed and of the 17, 475 protein-coding gene models, 9,872 (56%) genes were found expressed in the juvenile whereas 10,106 (57%) genes were expressed in the adult bird (Warren *et al.* 2010).

Being the first passerine bird to be sequenced, the zebra finch is an important model organism that has contributed to many fields of biology including neurobiology, ethology, ecology, biogeography and our understanding of evolution. The zebra finch genome has been particularly important for our understanding of the nature and function of different classes of genes, especially the genes involved in vocalization.

1.3.4 The Turkey Genome

Genome sequencing of the turkey (*Meleagris gallopavo*) was performed primarily using NGS platforms, particularly a combination of Roche 454 and Illumina GAI (Dalloul *et al.* 2010). The draft turkey genome sequence serves as the second domestic avian genome to be sequenced and resulted in a genome-level comparison of the two most economically important poultry species. The turkey genome was sequenced from a female turkey, 'Nici' (Nicolas Inbred) originally derived from a commercially prominent breeding line. Nici has been a genome source for developing two BAC libraries (Chaves *et al.* 2009). The sequence data were generated from 454 reads (5 x genome coverage), in GAI reads (25 x coverage) and by traditional Sanger sequencing (\approx 6 x clone coverage). Coverage of the Z and W sex chromosomes was not optimal due to fragmentary marker coverage and hence caused a poor assembly of these chromosomes. Scaffolds and contigs were assembled using a modified form of Celera assembler (Miller *et al.* 2008; Myers *et al.* 2000). The draft turkey genome constitutes approximately 1.1 gigabases, and a total of 28,261 scaffolds having 917 megabytes were anchored to specific

chromosomes (Dalloul *et al.* 2010). The turkey genome has the unique status of being the first for which the major part of the production cost was devoted to analysis and interpretation, instead of generating its sequence. This genome is a valuable resource in comparative genomics including the detection of many SNVs. The turkey genome corroborates the previously known high synteny between turkey and chicken genomes and only thirty predicted rearrangements distinguish their genomes (Dalloul *et al.* 2014). Apart from the low-cost factor of NGS which makes genome analysis feasible (<\$250,000 for the turkey), chromosome assemblies require the combination of multiple data types such as shotgun reads and contigs, genetic linkage maps, BAC maps and cytogenetic assignments.

1.3.5 The Duck Genome

The duck (*Anas platyrhynchos*) is the main reservoir of influenza A virus which can be transferred silently and efficiently to domestic poultry and mammals including humans. Ducks serve as the natural hosts for all currently known 16 haemagglutinin (HA) and 9 neuraminidase (NA) subtypes of influenza A viruses (Olsen *et al.* 2006; Wilcox *et al.* 2011), with the exception of H13 and H16 (Munster *et al.* 2007). (Huang *et al.* 2013). The duck genome sequence was generated from a 10-week old female Beijing duck using whole-genome shotgun sequencing strategy and Illumina Genome Analyser sequencing technology. 77 GB of paired-end reads were constructed with an insert size of 50 base pairs (bp). Short reads were assembled using SOAPdenovo and generated a draft assembly having 78,487 scaffolds and covered 1.1 GB. Forty seven superscaffolds were designed to create chromosomal sequences according to the duck genetic map (Huang *et al.* 2006) and the comparative physical map (Skinner *et al.* 2009a). Furthermore, the duck genome assembly was aligned to seven finished BACs, 240 microsatellite markers (Huang *et al.* 2006) and the 319,996 ESTs which were generated with the duck genome project. These analyses showed a 95% alignment with 7 BACs covering 640 kb on chromosomes 1, 3 and 4.

1.3.6 Other Sequenced Avian Genomes

The genome sequencing technology revolution has led to the release of more bird species genome sequences. These include the budgerigar, *Melopsittacus undulates*

(Koren *et al.* 2012); the pigeon, *Columba livia* (Shapiro *et al.* 2013); the collared flycatcher, *Ficedula albicollis* (Ellegren *et al.* 2012); Puerto Rican parrot, *Amazona vittata* (Oleksyk *et al.* 2012); Black grouse, *Tetrao tetrix* (Wang *et al.* 2014); and large ground finch, *Geospiza magnirostris* (Rands *et al.* 2013). Both Peregrine and Saker falcon genomes have also been sequenced (Zhan *et al.* 2013) however these are dealt with in a later section of this introduction.

1.3.7 Multiple Avian Genome Sequencing Efforts

Zhang *et al.* (2014) produced genome sequences of 45 avian species representing all 32 neognath and two of the palaeognath orders which give better insight into the genetic complexity of birds and explores genomic biodiversity among birds in relation to their phenotypic diversity. The Peregrine falcon, (but not Saker falcon) is included in this analysis. Genome sequences were generated using whole shotgun strategy and the genomes were assembled *de novo* with a scaffold N50 size of 17. Mb to 20kb. A homology-based method was used to annotate protein coding sequences, supported by transcriptome sequencing for some species. This largest genomic overview of a vertebrate class provided information relative to the evolution, diversification and ecological adaptation of avian species and created an information resource in the dataset for further studies related to their evolutionary genomics. Jarvis *et al.* (2014) carried out a genome-scale phylogenetic analysis of 48 species representing all orders of Neoaves aiming to increase knowledge of the history of the modern bird. Aided with an enormous amount of genomic data, this study generated a highly-supported phylogeny spanning a rapid radiation. This study supports the hypothesis of 'big bang' radiations as occurring for neoavian birds after the Cretaceous period to the Paleogene (K-Pg) period of mass extinction, an event of about 66 million years ago (Jarvis *et al.* 2014).

1.3.7.1 Comparative Genomics In Avian Species With Genome Sequences

Comparative genomic analysis was based on genetic and physical maps, until the advent of genome sequencing efforts, providing a device to examine large genomic datasets. Comparative maps between the chicken, mouse and human were constructed based on their genetic and physical maps (Burt 2002). Rapid development of next generation

sequencing techniques offers extraordinary qualities; they can harvest accurate genome sequences which give thorough insight in the field of evolutionary dynamics and phylogenetic studies. Many such studies have intensely initiated research in various biological areas including phylogenetics, population genetics, natural history, nutrition and conservation biology (Seabury *et al.* 2013).

1.3.7.1.1 Genome Size

In the past, our understanding was that avian genomes are significantly smaller than those of mammals. Earlier studies reported that the avian DNA content was 2.82 ± 0.33 pg per cell; this compares with a mammalian DNA content of 8 pg (Hughes and Hughes 1995). Later studies revealed that, among amniotes, birds have the smallest genome size and that many fishes and a few amphibians have even smaller genomes than birds (Zhang *et al.* 2014). The avian genomes range from 0.91 pg in the black-chinned hummingbird (*Archilochus alexandri*) to 2.16 pg in the common ostrich (*Struthio camelus*) with a mean value of $1.36 \text{pg} \pm 0.01$, whereas the genomes of mammals generally range from 1.63 to 8.4 pg (Gregory 2017). It was found that many factors contribute to the small avian genome size. Vertebrates exhibit huge variations in the amount of transposable elements (TEs) and it is well established that these variations significantly account for the evolution of the size of the vertebrate size (Feschotte and Pritham 2007). Therefore, comparative avian genome analysis is important to investigate the role of birds in genome size reduction. Only ~4-10% of the avian genomes consist of interspersed repeats, whereas the mammalian genome is composed of 34-52% interspread repeats (Zhang *et al.* 2014).

Hughes and Friedman (2008) compared gene family size between the chicken and five mammalian genomes (human, dog, rhesus monkey, mouse and rat) to evaluate the relation between the number of gene families and the genome size. Small genome sizes have resulted from the loss of protein coding genes and the researchers have reported that chicken gene families possess significantly fewer numbers of paralogs than the corresponding mammalian families. Having knowledge from the chicken genome only, researchers also suggested that the loss of protein coding genes may have significantly

influenced on the small genome size in birds, along with the reduction in intron size (Zhang *et al.* 2014).

1.3.8 Genome 10K Project

The onset of low cost sequencing technologies has led to a new phase of biological research and ambitious projects like the Genome10K Project (G10K). This was formed in 2009 with the primary goal of sequencing 10,000 vertebrate genomes by a group of scientists across the globe including bioinformaticians and computational scientists. This project also aimed to organize specimens, develop standards for genome assembly and annotate and release the genome data to the public (Koepfli *et al.* 2015). The avian Phylogenomics Consortium has launched the ‘Bird 10K’ project to generate draft genome sequences for all 10,500 extant bird species over the next 5 years. This data will fill numerous gaps in our knowledge in the field of evolution, ecology, population genetics, neurobiology, development and conservation. In addition, it is hoped, this can provide insight into the zoonotic diseases prevalent in birds (Zhang *et al.* 2015). As of 2014, genome sequencing had been published for over 259 vertebrate species (OBrien *et al.* 2014) (Table 1-4); this has included the efforts by G10K and the 61 bird species genome sequences that are available (Koepfli *et al.* 2015).

Table 1-4: List of vertebrate species with whole genome published (OBrien *et al.* 2014).

| | Lineage age (MYA) | Number of species | Number of species with whole genome sequenced |
|--------------|-------------------|-------------------|---|
| Fish | 600 | 31564 | 60 |
| Amphibians | 300 | 6570 | 12 |
| Reptiles | 320 | 9002 | 19 |
| Birds | 150 | 10500 | 61 |
| Mammals | 220 | 5416 | 107 |
| SUM | | 63052 | 259 |

1.3.9 Physical Genome Mapping

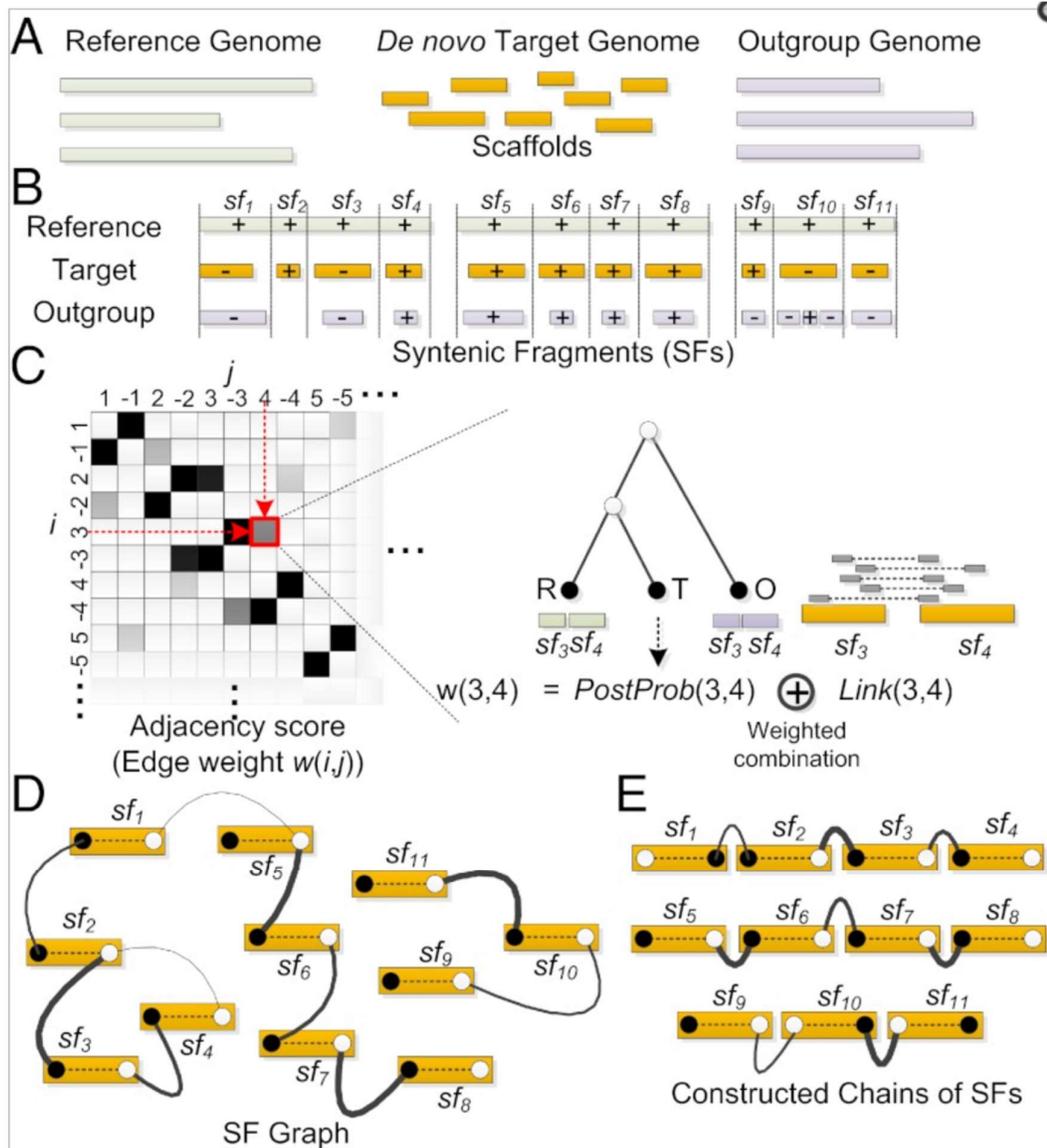
Genome mapping is all about anchoring a group of molecular markers on to their specific position on the genome. Different forms of genetic maps exist; the simplest form of genetic map depicts very little information as it has only two linked loci. Alternatively, a complete physical map portrays the exact physical location of all the genes that exist on the chromosome. In order to establish a long-range physical map of a large genome high resolution mapping procedures are necessary. Restriction mapping, FISH and sequence tagged site (STS) mapping are the three most powerful physical mapping techniques. A total number of 7,600 large-insert clones were mapped on the draft human genome sequence assembly by FISH. The clones used for this project primarily consist of BAC's including clones targeted to contain STS's (Cheung *et al.* 2001).

Modern genomic technology has generated outstanding numbers of animal genome sequences by next generation sequencing data (NGS). A whole genome sequence is essential to recognize chromosome rearrangements and assess evolutionary significance. However, *de novo* assemblies developed by NGS data require strenuous effort to assemble the reads into chromosomes (Kim *et al.* 2011). This is mainly due to the presence of repeat-mediated artifacts which occurs randomly (Tamazian *et al.* 2016). Most of the NGS assembled genomes are made up of sub-chromosomal sized scaffolds (Damas *et al.* 2017). Recently genome sequences of 45 new avian species have been generated using whole-genome shotgun strategy (Zhang *et al.* 2014). Out of 61 available avian genomes (OBrien *et al.* 2014), only 5 (chicken, turkey, duck, zebra finch and collared flycatcher) are assembled to chromosome level. To circumvent this problem, several bioinformatic approaches, e.g., Reference-assisted chromosome assembly (RACA) (Figure 1-16) and Chromosomer were introduced. These techniques use a high quality assembled genome such as human and chicken which are accounted as 'reference genomes'. Hence, it is named as reference-assisted chromosome assembly techniques. Damas *et al.* (2017) developed a novel approach to upgrade sub-chromosomal sized scaffold of the Peregrine falcon and pigeon to the chromosome level which requires much less cost and resources. This study used a set of chicken and zebra finch BACs which had $\geq 93\%$ DNA sequenc alignable with other avian genomes and they

contain at least one conserved element (CE) ≥ 300 base pairs. Their hybridization success rate with distant avian species was found to be high (71-94%).

Figure 1-16: Overview of the RACA Algorithm (Kim *et al.* 2013)

(A) RACA uses a reference, a *de novo* sequenced target (in scaffolds), and one or more outgroup genomes as input data. (B) Construction of syntenic fragments (SFs) by aligning reference and target genome sequences. Plus, minus represent the orientations of the target and outgroup on the reference. (C) Adjacency scores are measured for each SFs. (D) The SF graph is plotted according to adjacency scores. (E) constructed chains of SFs that are extracted by RACA algorithm.

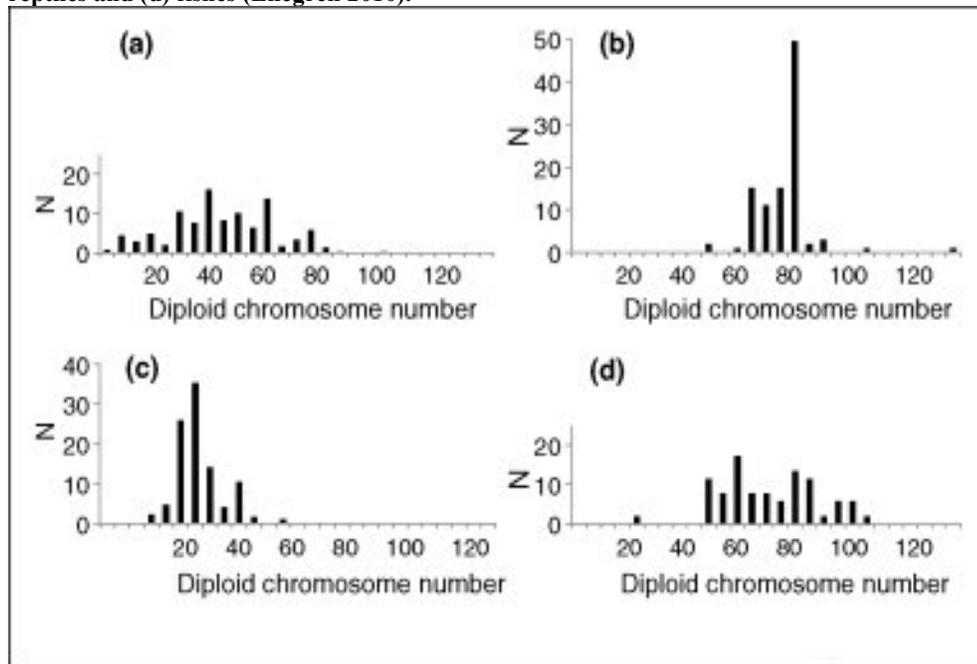


1.4 Avian Genome Evolution

1.4.1 Avian Chromosomes

As mentioned in section 1.1, birds have high diploid numbers and their chromosomes exhibit considerable variation in size than most other vertebrate members (Figure 1-17) (Ellegren 2010). Scaffold-based assembly analysis data together with zoo-FISH data (BACs and chromosome painting) indicate that the avian genome is stable and that a remarkable level of evolutionary equilibrium exists in birds (Griffin *et al.* 2008; Romanov *et al.* 2014). 60-70% of bird species have similar chromosome numbers ($2n \sim 80$), which is significant evidence that the rate of chromosomal rearrangements has been low in birds (Ellegren 2010). This section deals with the nature of the evolutionary chromosomal changes that have occurred in bird lineages.

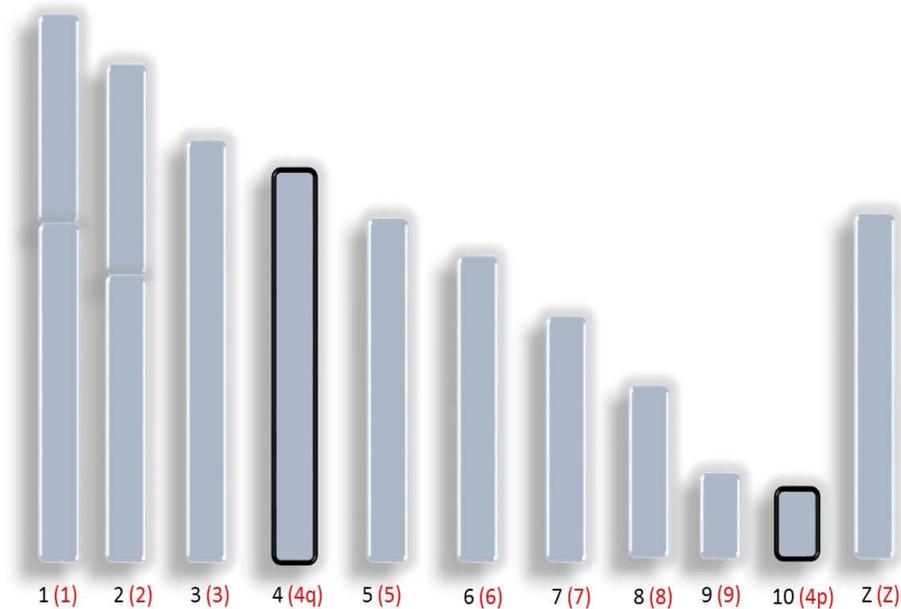
Figure 1-17: Diploid chromosome numbers ($2n$) in vertebrate groups. (a) mammals, (b) birds, (c) reptiles and (d) fishes (Ellegren 2010).



Moreover, based on cytogenetic data attempts that have been made to reconstruct the ancestral avian karyotype (Chromosome 1-10 + Z) (see Figure 1-18) and ancestral organization pattern have demonstrated few interchromosomal changes (Ellegren 2010; Griffin *et al.* 2007). Examples of interchromosomal changes occurring in specific lineages are *Psittaciformes* (parrots), *Falconiformes* (falcons) and *Sphenisciformes* (penguins) (Griffin *et al.* 2007; Schmid *et al.* 2015). Griffin *et al.* (2007) suggest that chicken

orthologues of chromosome 1-3 and 5-10 + Z portray avian ancestral chromosomes with chicken chromosome 4 arising as a result of the fusion of ancestral 4 and a smaller chromosome. Romanov *et al.* (2014) with the help of FISH and bioinformatics tools, further reconstructed ancestral chromosome 1-5 for all birds and chromosome 6-28 & Z for Neognathae and show further that ancestral chromosomes are similar to chicken chromosomes. Lithgow *et al.* (2014) introduced a set of microchromosomal pooled paints and microchromosomal BACs (Damas *et al.* 2017) aiming to identify chicken microchromosome homologs in different species of different orders suggesting an exceptional level of microchromosomal conservation among birds. It also appears that microchromosomes act as distinct units when fused into complex chromosomes such as in parrots and falcons (O'Connor 2016).

Figure 1-18: Schematic representation of relative sizes of the ancestral chromosomes 1-10 + Z (adapted from Griffin *et al.* 2007). The chromosome number that is stated below is the chromosome in black text with the chicken orthologue in red text and in brackets. As all sizes are relative, no scales are appropriate.



1.4.2 Chromosomal syntenies

Comparative genome mapping between the chicken and human showed chromosome synteny between human chromosome 4 and chicken chromosome 4q indicating the

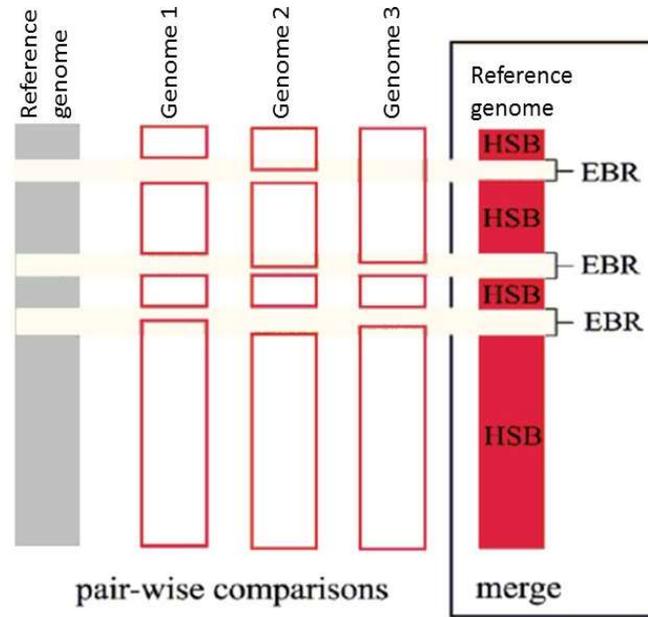
existence of ancestral chromosome 4 at least 310 million years ago (Griffin *et al.* 2007; Raudsepp *et al.* 2002). It has been suggested that chicken chromosome 4p evolved from a fusion of the ancestral chromosome 4 to another ancestral chromosome (Griffin *et al.* 2007; Masabanda *et al.* 2004). However chicken 4p still retains its ancestral chromosome properties of a smaller chromosome such as gene density, recombination rate and CpG island distribution (Masabanda *et al.* 2004). Hybridization with chicken paints (GGA1-9 and Z) on several avian species showed that homoplasmy is prevalent in avian chromosome evolution with the fusion of ancestral chromosome 4 and ancestral chromosome 10 (Skinner and Griffin 2012), whereas ancestral chromosomes 4 and 10 remain independent for most birds in the Order *Anseriformes* (Shibusawa *et al.* 2004), *Casuariiformes* (Guttenbach *et al.* 2003), *Cathartiformes* (Nie *et al.* 2015), *Galliformes* (Shibusawa *et al.* 2004), *Passeriformes* (Guttenbach *et al.* 2003), *Psittaciformes* (de Oliveira *et al.* 2015), *Rheiformes* (Guttenbach *et al.* 2003), *Struthioniformes* (Guttenbach *et al.* 2003), *Tinamiformes* (Nie *et al.* 2015). To date however studies of Falconiforme chromosome change has been limited to the use of chicken microchromosome paints.

1.4.3 Homologous Synteny Blocks And Evolutionary Breakpoint Regions

With the advent of multiple genome sequencing, together with modern sequence analysis tools hypotheses relating to the presence or absence of breakpoints in chromosome evolution (Sankoff 2009) have allowed researchers to point out homologous synteny blocks (HSBs) and evolutionary breakpoint regions (EBRs) (Skinner and Griffin 2012). HSBs have been sustained for millions of years in genomes of several species and the identification of multispecies homologous synteny blocks (MsHSBs) is important to determine their role in evolution (Farre *et al.* 2016). Larkin *et al.* (2009) describes HSB between two species as the part of the genome between minimum of two adjacent markers on the chromosome without interruption. An EBR is defined as the region between two syntenic blocks 4 megabases (Mb) or less in size (Ruiz-Herrera *et al.* 2006) (Figure 1-19). Larkin *et al.* (2009) compared nine mammals and found that in mammals, EBRs are situated in gene-dense regions and they are enriched with copy number variants. Translocations, inversions and fission are observed more often in EBR's sequences than in other parts of the chromosome making these regions prone to breakage (Larkin *et al.* 2009). Farre *et al.* (2016) reported that chromosome breakage in

birds is related to genomic features like transposable elements and conserved non-coding elements, which supports the previous findings in mammals (Larkin *et al.* 2009; Murphy *et al.* 2005). In evolution, chromosome breakage is not random. Breakage are formed from segments which are conserved over millions of years; also, unstable parts of the genome are involved in rearrangements. EBRs appeared to occur in evolutionally active regions where genes are created, increased and destroyed due to multiple molecular mechanisms (Larkin *et al.* 2009).

Figure 1-19: Schematic representation of how homologous synteny blocks (HSBs) and evolutionary breakpoint regions (EBRs) are defined in the reference genome by pair-wise comparison adapted from Farre *et al.* (2011).



1.4.4 Comparative Genomics and Avian Genome Evolution

With ~10,500 extant avian species (Gill and Donsker 2017) and many of them being sequenced, there exists a source of information to understand better the general aspects of behavior, ecology and evolution of birds (Nam *et al.* 2010). Molecular time estimation data suggests the common amniote ancestor for birds and mammals existed 325 million years ago (Ma) and the origin of lepidosaurs occurred approximately 275 Ma (Shedlock and Edwards 2009). The divergence of turtles from Archosaurs is believed to have taken place between 230 to 255 Ma (Chiari *et al.* 2012; Shedlock and Edwards 2009). Phylogenomic studies conducted by Chiari *et al.* (2012) gives compelling evidence of the relationship of turtle with Archosauria as a sister group within Amniota. The

S. Joseph

Archosaur clade includes extinct dinosaurs and pterosaurs, extant crocodiles and birds (Green *et al.* 2014). However, crocodiles and birds are the closest relatives recorded and are separated around 219 Ma (Shedlock and Edwards 2009). Birds originated from dinosaurs about 150 Ma during the Jurassic period (Chiappe and Witmer 2002; Dyke and Kaiser 2011). *Archaeopteryx* is considered as the oldest avian ancestor; it was recovered in 1861 from limestone quarries in southern Germany (Dodson 2000) (Figure 1-20).

Figure 1-20: Archaeopteryx lithographica fossil, the oldest avian ancestor.



Most fossil birds from the Cretaceous period consist of extinct birds and the birds that endured the catastrophic extinctions about 66 Ma, spread out globally into modern birds-Neornithes (Feduccia 2003). The timing of their speciation has been a focus of debate. Jarvis *et al.* (2014) resolved the ambiguity of avian phylogeny and suggested that after the Cretaceous-Paleogene (K-Pg) event, surviving Neornithes rapidly diverged into different species with 36 different lineages over a period of 10-15 million years. Also, worthy of note is that many lineages that evolved from a single Neornithes lineage have survived the global extinction. At the end of the Cretaceous Period in Earth's history there was a worldwide catastrophic, sudden and rapid extinction attributable to an extra-terrestrial impact (Longrich *et al.* 2011).

The first split divided Neornithes into *Paleognathae* (Ratites and Tinamous) and *Neognathae* (all other birds) approximately 100-120 Ma. Ratites and Tinamous are

thought to have diverged around 84Ma. The divergence of *Galloanserae* from the *Neognathae* happened around 88 Ma and the divergence of remaining *Neognathae* members into land and water birds around 60 to 70 Ma which occurred during the K-Pg time (Longrich *et al.* 2011). Recent molecular evidences and fossil records give clear understanding of the evolution of modern birds and separate taxa were well placed in the avian tree. (Ericson *et al.* 2006; Hackett *et al.* 2008; Kimball *et al.* 2013; Jarvis *et al.* 2014). Molecular and contemporary morphological studies by Livezey and Zusi (2007) divide modern birds (*Neornithes*) into *Palaeognathae* (tinamous and flightless ratites), *Galloanseres* [Galliforms (landfowl) and *Anseriformes* (waterfowl)] and *Neoaves* (all other extant birds) (McCormack *et al.* 2013). Genome-scale phylogenetic analysis of 48 species representing all orders of *Neoaves* using phylogenomic methods created a highly-resolved tree which recognized the first divergence in Neoaves into two groups, Passerea and Columbea (Jarvis *et al.* 2014). The present highly resolved total evidence nucleotide tree (TENT) congruence with previously generated smaller-scale multilocus nuclear trees (Ericson *et al.* 2006; Hackett *et al.* 2008; Kimball *et al.* 2013; McCormack *et al.* 2013). However it disproves some relationships in avian phylogenies generated from morphological characters (Livezey and Zusi 2007), DNA-DNA hybridization (Sibley and Ahlquist 1990) and mitochondrial genomes (Pratt *et al.* 2009) by excluding eagles and New World vultures from *Falconiformes*, hornbills and cuckoo-rollers from *Coraciiformes*, picbirds from *Pelecaniformes* and seriemas, bustards, the sunbittern, and mesites from *Gruiformes*. Earlier, Hackett *et al.* (2008) evaluated the avian phylogenetic relationships from 171 avian species and released one of the most unexpected data showing the sister relationship between *Passeriformes* and *Psittaciformes*, with *Falconidae* (falcons) sister to this clade which is supported by the latest phylogenetic studies conducted by Jarvis *et al.* (2014). However, Prum *et al.* (2015) constructed a phylogenetic tree from 198 bird species and 2 crocodylians based on the loci captured using anchored enrichment (Figure 1-21). This genome phylogeny provides additional support for existing comprehensive avian supertrees and comparative evolutionary analyses. It confirms that the ancestor of the diurnal swifts and hummingbirds evolved from a clade which had nocturnal features. It also upholds the hypothesis that land birds have evolved from a raptorial grade. Identification of a new comprehensive water bird-

shorebird clade gave insight into the evolutionary constraint on the ecological diversification of birds.

Figure 1-21: Phylogeny of birds. Time calibrated phylogeny of 198 species of birds inferred from Bayesian analysis (Prum *et al.* 2015)

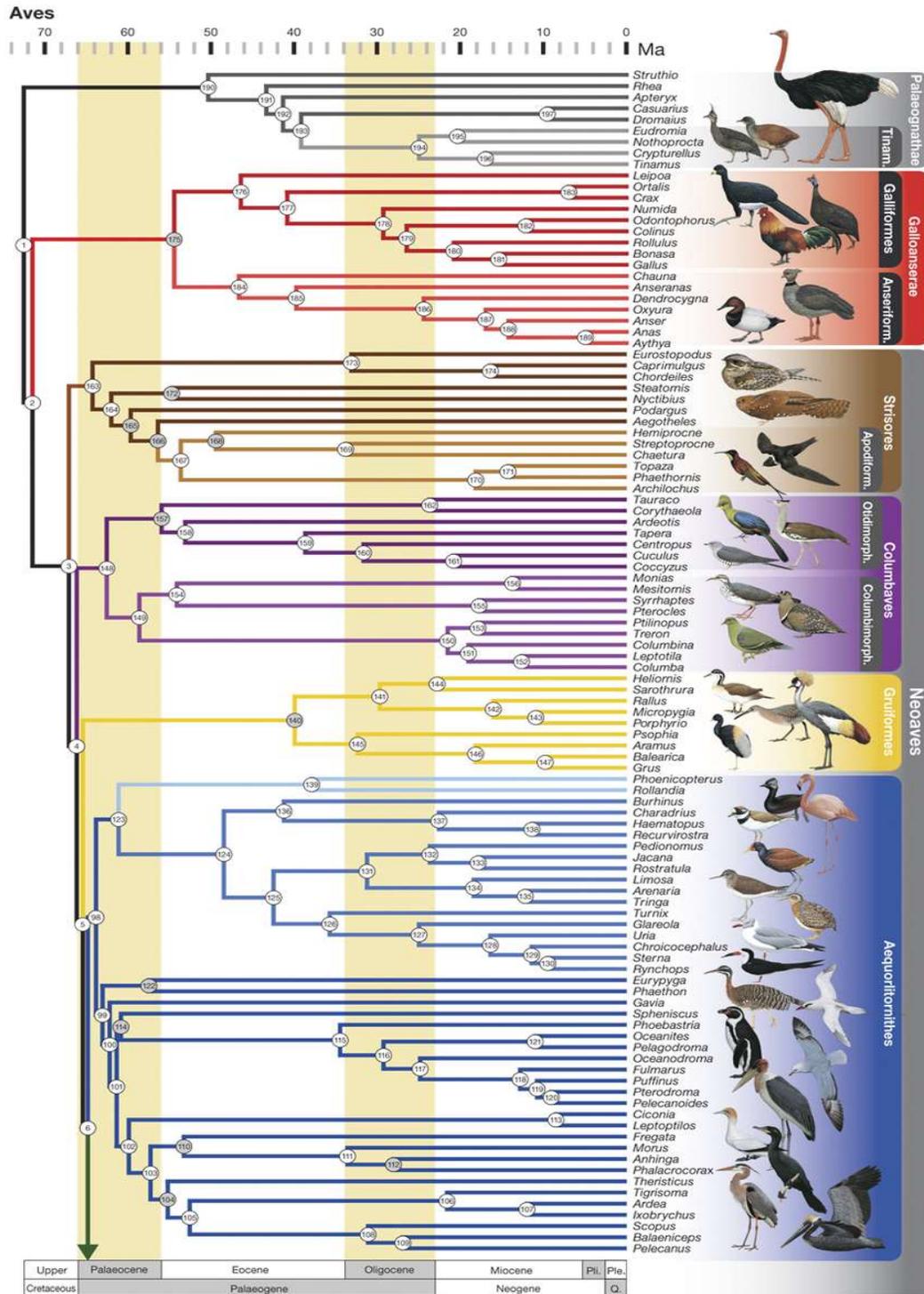
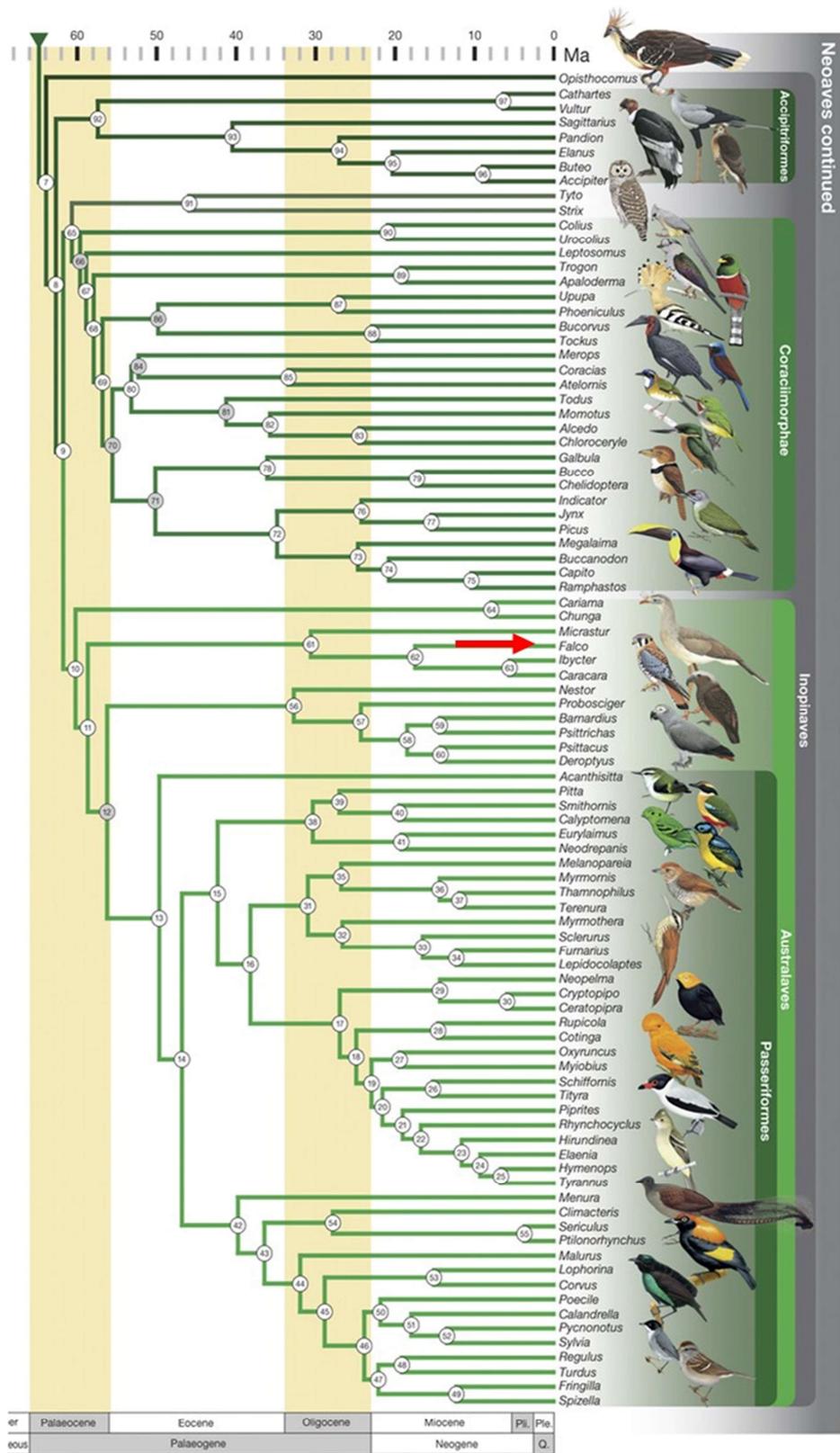


Figure 1-21 (Continued): Phylogeny of birds. Time calibrated phylogeny of 198 species of birds inferred from Bayesian analysis (Prum *et al.* 2015).



1.5 Methods to Study Chromosomes

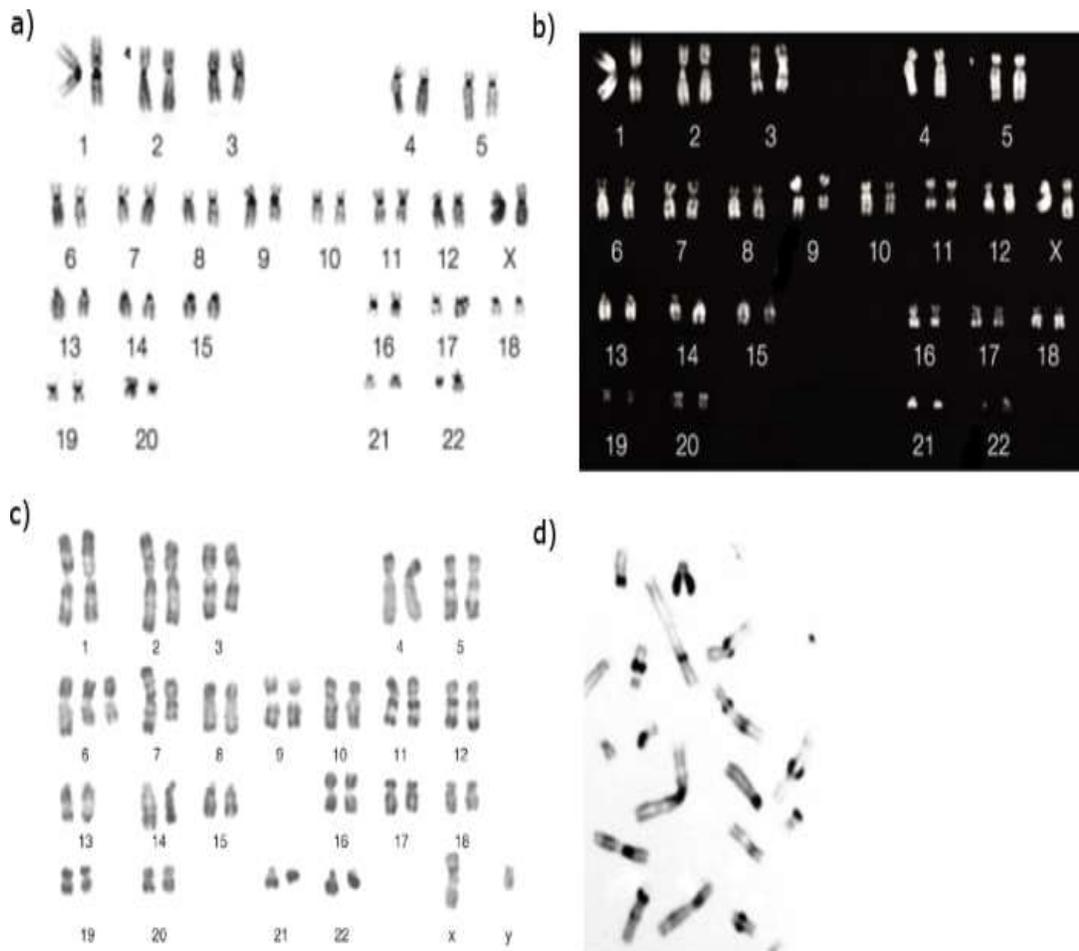
1.5.1 Classical Cytogenetics

Cytogenetics is the study of number and structure of chromosomes. Classical cytogenetics by karyotyping remains as the 'gold standard' method to detect chromosomal alterations in humans. Moreover, a karyotype represents a low-resolution genome map of any species.

1.5.1.1 Standard Banding methods

The primary step of chromosome preparation technique begins with acquiring a large number of dividing cells by the short-term culture of cells obtained from the specimen. Cell division is arrested by addition of colcemid, which disrupt the mitotic spindle. Cell treatment with a hypotonic solution cause their nuclei to swell and the cells to burst. Preservation of the cells is achieved by fixation with Carnoy's solution, a mixture of methanol and glacial acetic acid which arrests cell metabolism. The cells are then dropped on to the slides and can be utilized for various cytogenetic procedures (Moore and Best 2001). Many stains are used to visualize characteristic banding patterns of chromosomes under the microscope. A clear-cut identification of all 24 human chromosomes for the first time was possible with Q-banding (Caspersson *et al.* 1970). Other than Quinacrine, several other fluorescent stains such as Hoescht 33258, diamidazolinophenylindole (DAPI) and daunomycin produce similar fluorescent pattern (Moore and Best 2001). Geimsa (G-) banding uses the common Giemsa stain along with chemical and enzymatic treatment. Intense Geimsa-stained regions corresponds to intense Q-banded fluorescent regions, i.e. AT-rich part of the DNA (Moore and Best 2001). A reverse banding (R-banding) pattern is somewhat contrary to G- or Q-banding patterns. It is the gene-rich chromatin that stains, thus enabling visualization of structural rearrangements. Centromere-banding (C-banding) produces a unique banding pattern with darkly stained constitutive heterochromatin while the rest of the chromatin remains pale colored (Figure 1-22).

Figure 1-22: Human chromosome banding revealed by different staining techniques. (a) G-banding; (b) Q-banding; (c) R-banding; (d) C-banding (O'Connor 2008a).

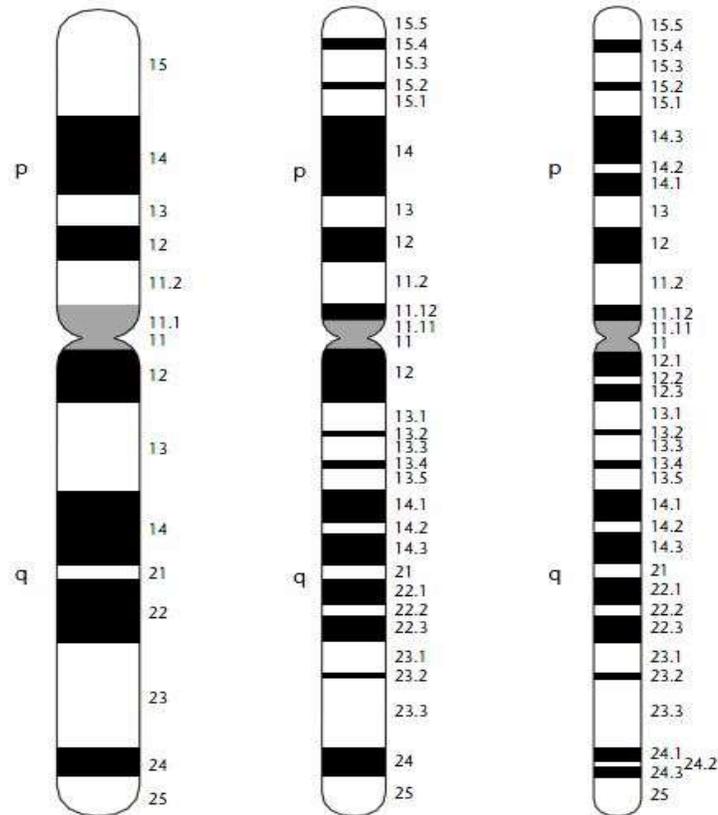


1.5.1.2 Number and Size of Banding

Individual chromosome images are usually arranged in a standardized format called a karyotype. Standard ideograms (cartoon representations) of G-banded chromosomes are considered as standard reference points for chromosome banding. G-bands are usually depicted in black and R-bands in white. Bands are numbered sequentially away from the centromere on both the short (p) and long (q) arms (Bickmore 2001). The quality of an ideogram is generally graded based on the number of bands or 'resolution' present on the ideogram which greatly depends on the stage of mitosis they are in at the preparation of chromosomes. When a low-resolution band is expanded with high resolution techniques, the numbering of each sub-band is given in such a way that the

number is placed next to a decimal point following the first band designation as shown in (Figure 1-23) (Bickmore 2001) .

Figure 1-23: G-band ideogram of human chromosome 11 at (from left to right) 350, 550 and 850 band resolutions (Bickmore 2001).



1.5.1.3 Uses of Chromosome Banding

G- and R- banding are the most commonly used techniques to diagnose a wide range of chromosomal abnormalities in individuals such as chromosome number, translocations of material from one chromosome to another, deletions, inversions or amplification of chromosome segments (Bickmore 2001). The detection of an extra copy of chromosome 21 in patients with Down Syndrome was an important landmark finding in the field of cytogenetics. By pinpointing the location of disease associated genes and their characteristic locations on the chromosomes the diagnosis and prognosis of some cancers became possible (Lejeune *et al.* 1959). The best example of this is the identification of Philadelphia chromosome for the diagnosis of chronic myelogenous

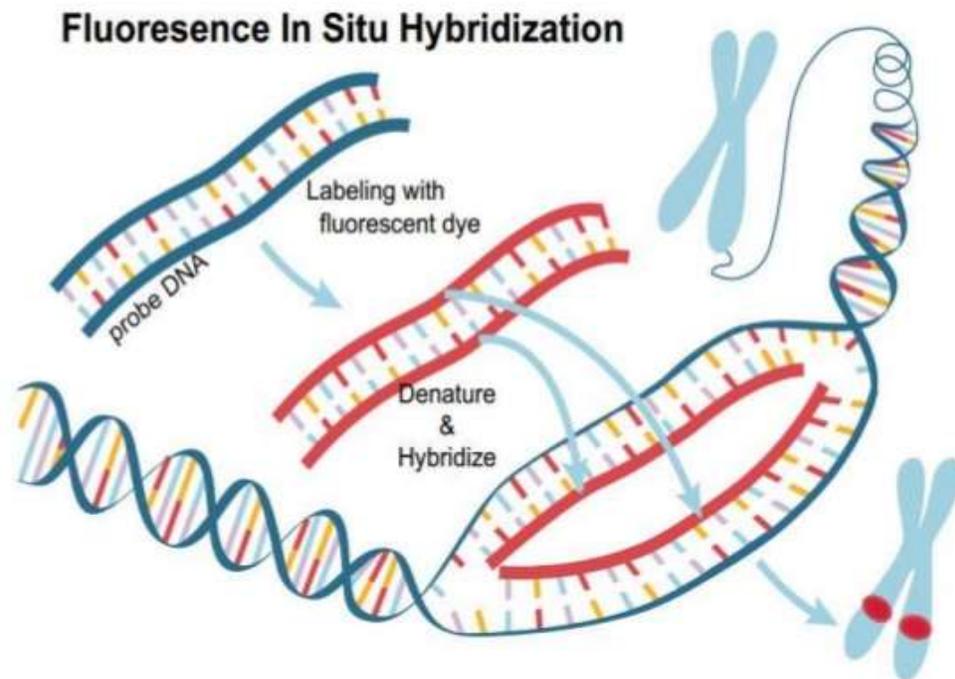
leukaemia (CML) (Sessarego *et al.* 1987). Human genetic disorder diseases like Cri-du-Chat syndrome, Klinefelter syndrome, Turner syndrome and Williams syndrome are the other examples where cytogenetic diagnostic techniques have played a dramatic role.

Even though clinical animal cytogenetics developed along with human cytogenetics in the 1960's, it reached its 'Golden Age' at the end of 1980's. Currently, the estimated number of chromosomal analysis carried out each year worldwide is 8,000 to 10,000, mainly in cattle, pigs, and horses (Ducos *et al.* 2008).

1.5.2 Molecular cytogenetic Analysis using Fluorescence *in situ* Hybridization

Introduction of *in situ* hybridization paved the way for cytogenetics to enter the molecular era (O'Connor 2008b). The principle of fluorescence *in situ* hybridization (FISH) principle (Figure 1-24) is based on the ability of single stranded DNA to anneal complementary DNA. In the case of FISH, interphase cells, metaphase chromosomes or extended chromatin fibers affixed to a slide is the target which is hybridized to a nucleic probe, labelled with a fluorochrome. In general, a wide range of probes are used for FISH applications. They are whole-chromosome painting probes, chromosome-arm painting probes and repetitive centromeric, sub-telomeric and locus-specific probes (Bishop 2010). The best suited FISH probes are produced from BAC clones (bacterial artificial chromosome). The labelled probes anneal to its complementary DNA sequence present in the target which then produces a colored signal at the hybridization site. Colored signals are visualized by fluorescence microscopy (Bishop 2010). For direct detection of signals, most commonly used reporter molecules are FITC, Rhodamine, Texas Red, Cy2, Cy3, Cy5 and AMCA and for the image analysis, COOLED CHARGE COUPLED DEVICE (CCD) cameras along with specific filter sets are used which improves the resolution quality of the image (Speicher and Carter 2005).

Figure 1-24: The principles of fluorescence *in situ* hybridization (FISH)
(<https://www.semrock.com/fish.aspx>).



The flexible feature of the FISH technique has led to the further advancement of several cytogenetic techniques. Thereby, different approaches can be applied based on FISH techniques.

1.5.2.1 Multiple Color FISH Hybridization

FISH has greatly improved the effectiveness of cytogenetics with its appealing aspects such as sensitivity, specificity by introducing several imaging systems for FISH acquisition and the possibility of simultaneous use of one or more probes (Riegel 2014). Multiplex FISH (M-FISH) and spectral karyotyping (SKY) were developed in 1996 making it possible to distinguish 24 human chromosomes in a single hybridization through labelling each chromosome with a different combination of fluorophores (Schrock *et al.* 1996; Speicher *et al.* 1996). Since then, different approaches have been developed for chromosomal identification built on multicolor-FISH (mFISH) techniques. mFISH assays have an indispensable role in precise description of complex chromosomal disorders and mapping of multiple sequences simultaneously (Table 1-5).

Table 1-5: Examples of FISH applications (Riegel 2014).

| Diagnostic | Research |
|--|---|
| The identification of specific chromosome abnormalities | The identification of new non-random abnormalities (by M-FISH or SKY) |
| The characterization of marker chromosomes | Gene mapping and comparative genomics |
| Interphase FISH for specific abnormalities in cases of failed cytogenetics | Identification of novel regions of amplification or deletion |
| Monitoring disease progression | The identification of translocation breakpoints in patients and evolution |
| Monitoring the success of bone marrow transplantation | The study of 3D chromosome organization in interphase nuclei |

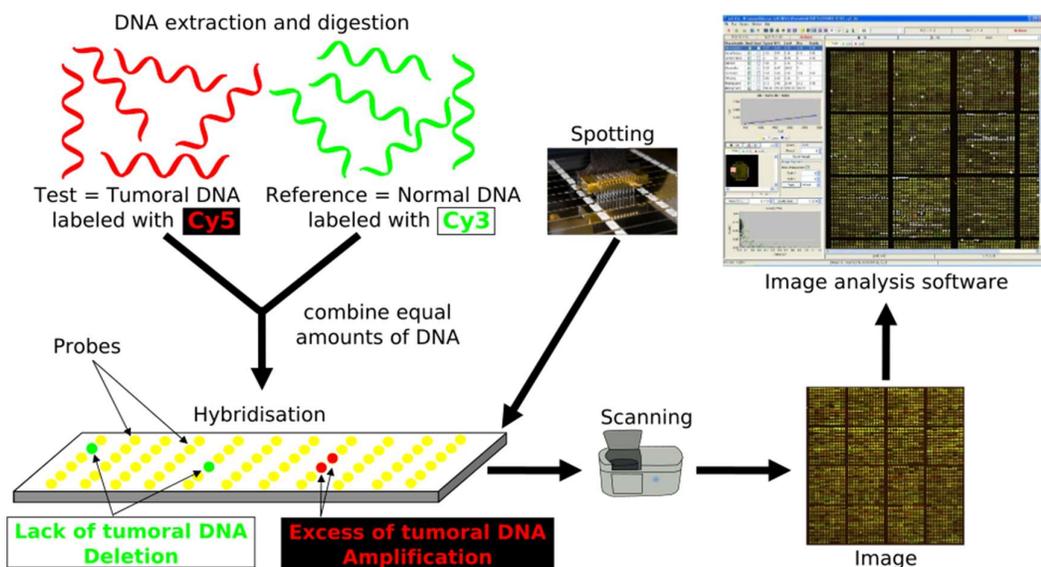
1.5.3 Comparative Genomic Hybridization and Microarrays

Introduction of comparative genomic hybridization (CGH) has increased the clinical impact of FISH techniques, especially in cancer genome identification. As a new chromosome analysis technique it was discovered in 1992 that CGH is an ideal tool for analyzing chromosomes to identify tumour genomes (Kallioniemi *et al.* 1992). In short, instead of high-quality metaphase preparation, in CGH, genomic DNA from test sample and a normal reference sample is used; each is differentially labelled with green or red fluorochrome. These combined probes are hybridized to human reference chromosomes, competing with each other for complementary hybridization sites. Based on the intensity of the fluorescence ratio between green and red which is measured along the chromosomal axis, image analysis software can provide information about the copy number loss or gain of genetic material from the test genome (Weiss *et al.* 1999).

Soon after this discovery, another 'molecular karyotyping' system, microarray (array-CGH) was developed with a resolution of 100 kilobase to 1Mb to evaluate the expression of several genes simultaneously (Skena *et al.* 1995). Here, the hybridization target mainly BAC or plasmid clones, which are robotically printed onto a glass slide in an ordered pattern. The test and reference genome are labelled with different colors and are hybridized to microarray. The relative fluorescence ratio of the test and reference

genome is calculated along the genome using digital imaging which provides data of DNA copy number differences (Skena *et al.* 1995; Theisen 2008). Array-CGH (aCGH) is a powerful tool to detect chromosomal imbalances by identifying CNV's in tumors (Cai *et al.* 2002). The clinical impact of CGH is vital as it has significant role in the evaluation of idiopathic mental retardation, intellectual disability, autistic spectrum disorders, schizophrenia, neuropsychiatric disorders and various birth defects (Vetro *et al.* 2012). Furthermore, aCGH (Figure 1-25) has taken cytogenetics from the microscope to computer and it has the future capability, to establish new cytogenetics approaches with remarkable resolution (Theisen 2008).

Figure 1-25: Representation of Array CGH (Theisen 2008). Widely used for detecting amplifications and deletions in tumours (as in this diagram) it can also be used to detect copy number variation (CNV) between individuals, strains and species.

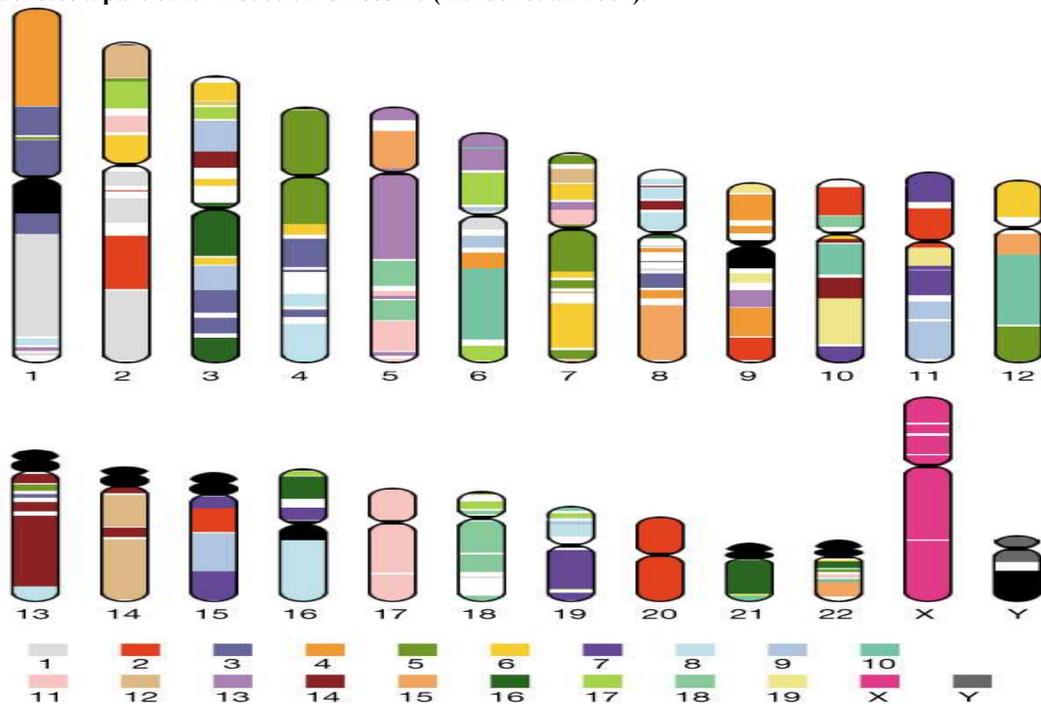


1.5.4 Comparative Genome Visualization Tools at the Chromosome level

Comparative genomics is the field of biological study where genome sequences/cytogenetic data from different organisms are compared to provide better understanding of evolutionary changes among organisms. Apart from evolutionary changes comparative genomic data demonstrate the genes that are common between species and those genes which provide unique character to each organism (Touchman 2010).

Bioinformatic tools along with karyotyping and zoo-FISH experiments are used for a wide spectrum of data comparison ranging from basic comparison such as genome size, number of genes and chromosome number to extreme comparison which includes Gene Ontology , gene families of interest, hot spots of chromosomal breakage/fusion points and high-frequency recombination sites (Chen and Wang 2013; Touchman 2010). Through chromosome level comparison of different genome, one can estimate the nature and degree of conservation of synteny between the species of interest. Chromosome-level comparison between human and mouse genome is shown in Figure 1-26 which shows the synteny between these two mammals.

Figure 1-26: Conserved synteny in human and mouse genome. Human chromosomes; each colour denotes a particular mouse chromosome (Lander *et al.* 2001).

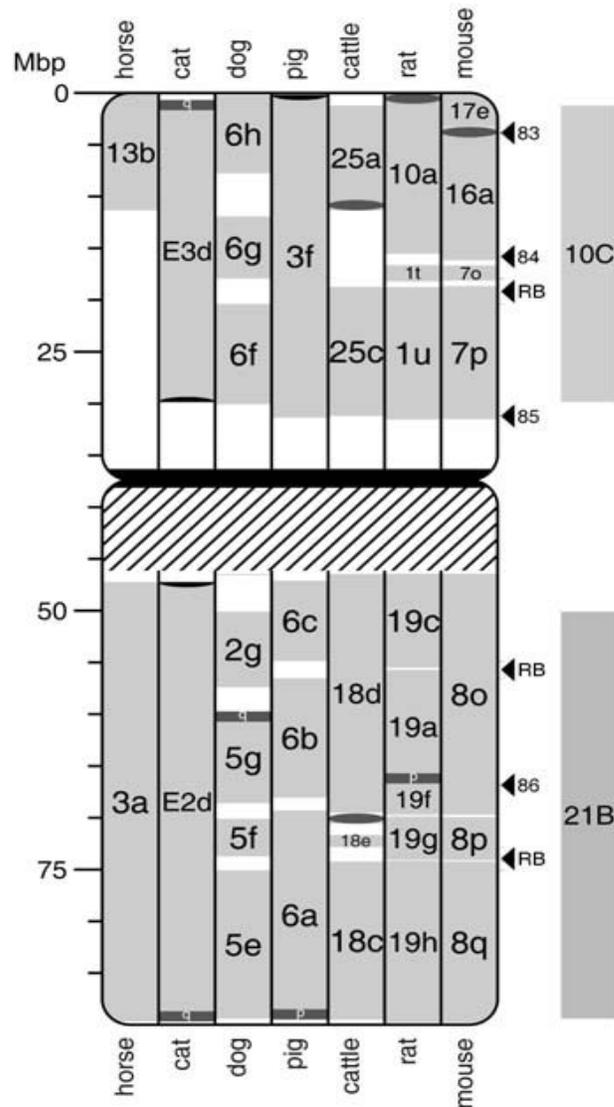


Comparison of distinct domains between distantly and closely related species is attainable by aligning homologous DNA. This approach of comparative genomics is illustrated by aligning a particular human gene (pyruvate kinase: PKLR) with that of the macaque, dog, mouse, chicken and zebra finch. Using computer-based analysis the degree of sequence similarity between different vertebrates has been shown to have been conserved over millions of years (Touchman 2010).

With the availability of an increasing number of whole-genome sequences, comparative analysis of DNA sequences from multiple species belonging to various branches of the evolutionary tree is possible. To assist orthologous-sequence comparison various comparative sequence based visualization tools and database have been established. Most of such tools are publicly accessible resources which have been successfully utilized by many biological investigators. Visualization Tool for Alignment (VISTA) and Percent Identity Plot Maker (PipMaker) are two commonly used comparative genomic tools which primarily convert raw orthologous-sequence data from multiple species into visually illustratable plots (Pennacchio and Rubin 2003).

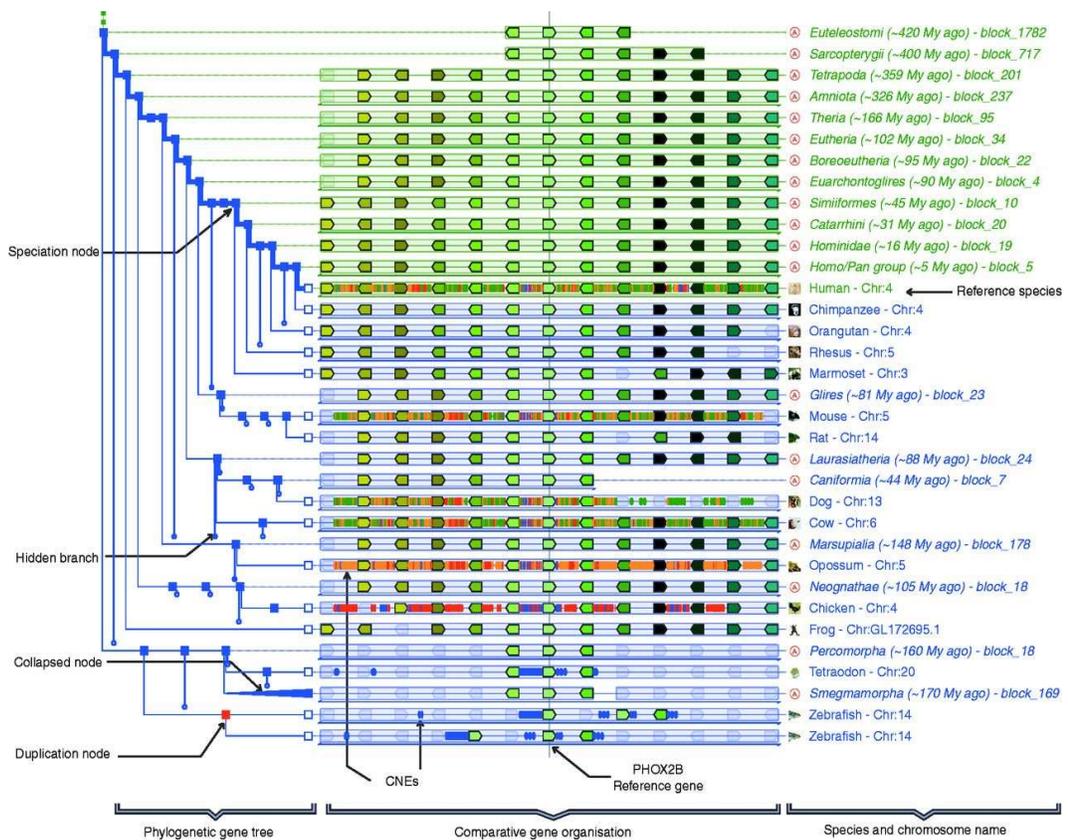
Other good examples of comparative genomic visualization tool are a web service application called Evolution Highway (<http://evolutionhighway.ncsa.uiuc.edu>) and Genomicus (<http://dyogen.ens.fr/genomicus>). Several programs have been developed for reference-assisted chromosome assembly such as Chromosomer (<http://github.com/gtamazian/chromosomer>) and RACA. Evolution Highway allows users to compare multiple species at once with a coverage of millions of base pair to construct evolutionary breakpoint regions and regions of homologous synteny (Damas *et al.* 2017; Romanov *et al.* 2014). This excellent tool has been utilized to identify pairwise HSBs between the genome of the human and six non-primate species to denote the rearrangements between all genomes and their ancestors (Murphy *et al.* 2005) (Figure 1-27).

Figure 1-27: Multispecies comparative chromosome alignment against human chromosome 16. Grey blocks denote HSBs with the chromosome numbers (Murphy *et al.* 2005).



Genomicus is an online genomic visualization tool use to demonstrate gene organization within and between different genomes and allows the identification of differential gene loss and gain, and segmental or genome duplications by analyzing homologous relationships (Louis *et al.* 2013). A screenshot from Genomicus is shown in Figure 1-28.

Figure 1-28: PhyloView of the paired-like homeobox2b (PHOX2B) gene with the human gene as the reference gene and showing high density of conserved synteny between the human and several mammals, also with chicken (Louis *et al.* 2013).



1.5.5 Genome Reconstruction using Bioinformatics and FISH

1.5.5.1 Ancestral Genome Reconstruction

The field of Paleogenomics has made it possible to evaluate genetic variation through time, using ancient DNA which has survived more than 100,000 years. Although such information can resolve many arguments about evolutionary relationships between different species, without nuclear genome data it is unable to infer extinct phenotypes (Shapiro and Hofreiter 2014). Past research mostly promoted interspecies comparison but more recently it has facilitated an understanding of common ancestor of the lineage of interest. Such studies infer the record of genomic rearrangements that occurred during the evolution process of a species. By inferring genome organization and gene content of the ancestral genome, it can provide detailed information about the recent

evolution of the species that evolved from it (Gordon *et al.* 2009). The most recent changes in the genome are often considered important as they indicate the most recent or on going evolutionary process happening for that species (Muffato *et al.* 2010). Precise phylogenetic placement of the taxa of interest is the foremost requirement in reconstructing ancestral genomes followed by the establishment of syntenic data for genome comparison between the related species. Formerly, syntenic data were generated from cytogenetic data, produced by cross species FISH. Today, sequence-based computational approaches are more common (Rajaraman and Ma 2016).

1.5.5.2 Ancestral Reconstruction by Cytogenetic Method (zoo-FISH)

Prior to the development of the zoo-FISH technique, the possibility of studying evolution was restricted to comparative studies of morphology and banding patterns of the chromosomes. The zoo-FISH technique identifies regions of chromosomal homology between species. This approach develops ancestral karyotypes by identifying synteny conservation between species (Muffato *et al.* 2008). Chowdhary *et al.* (1998) was successful in developing the first mammalian karyotype using zoo-FISH and involving eight different species. However, this method is limited mainly due to the lack of hybridization success among evolutionally distant species (Chowdhary *et al.* 1998). Additionally, this method cannot always identify intrachromosomal changes such as inversions and duplications (Wienberg 2004).

1.5.5.3 Ancestral Reconstruction by Bioinformatic Methods

Various disciplines of genomics became active with the increasing numbers of assembled genomes. By analyzing sequenced genomes, a bioinformatic approach can reconstruct ancestral genome. A robust software tool is required to construct the most likely ancestral genome as the number of sequenced genomes having unequal gene content increases (Avdeyev *et al.* 2016). Three different principles have been used in reconstructing ancestral characters from multiple genome sequences. They are mainly based on the different levels of precision which varies from karyotype to gene order to genomic sequences. A computational ancestor can be made by comparing gross genomic rearrangements at the karyotype level. At the gene order level, orthologous

genes are identified across the aligned multiple genome sequences that mainly focus on a particular locus (Louis *et al.* 2013). Whereas, at the sequence level, homologous synteny blocks (HSBs) and evolutionary breakpoint regions are identified across the multiple genomes and allow reconstruction of the ancestral genome by identifying rearrangements in these regions (Larkin *et al.* 2009).

Robust algorithmic tools to reconstruct the common ancestral genomes and to describe evolutionary history between genomes of interest. Several such algorithm based tools are available. They are: the GRAPPA algorithm (Moret *et al.* 2001), MGR tool (Bourque and Pevzner, 2002), inferCARs algorithm (Ma *et al.* 2006), EMRAE algorithm (Zhao and Bourque, 2007), MGRA (Alekseyev and Pevzner 2009) and PMAG (Hu *et al.* 2013). The MGR tool uses an algorithm which has the advantage of minimizing the total number of rearrangements over all the branches of the phylogenetic tree. It was further modified as MGRA (Alekseyev and Pevzner, 2009), which can identify rearrangements in ancestral genomes more precisely. This tool was recently upgraded to MGRA2 where gene gain and losses are considered (Avdeyev *et al.* 2016).

1.6 Chromosomal Abnormalities in Individuals and in Evolution

Variation in the genetic arrangement or a defect in the chromosome can lead to specific physical symptoms of varying severity differing in each individual. Such chromosome abnormalities can be due to various reasons like the attachment of additional genetic material to a chromosome, or the partial or complete loss of material or the defective formation of a chromosome. These chromosomal defects can have adverse effects on the development and function of an individual. Chromosomal abnormalities usually happen during meiosis and fertilization and they can be divided into two major groups; numerical aberrations and structural aberrations. Genetic structure is characteristically unstable and comparative genomic studies among species disclose that small scale to large scale variations are common (Otto 2007). Fifteen percent of eggs that are produced in animal taxa do not hatch due to various reasons like embryo mortality, lack of fertilization, environmental issues and genetic factors (Forstmeier and Ellegren 2010). In humans, approximately 10-30% of fertilized eggs have an abnormal number of

chromosomes which has not been observed in animal species (Hassold and Hunt 2001). Few systematic molecular genetic studies have been conducted on birds to investigate the nature of chromosomal aberrations in this taxon (Forstmeier and Ellegren 2010). Chromosomal changes such as inversions and reciprocal translocations can lead to reproductive isolation and thus cause speciation (Larkin *et al.* 2009). It is reported that the avian karyotype has evolved through several fusion/fission and inversion events (Burt 2002; Schmid *et al.* 2005).

1.6.1 Numerical Chromosome Aberrations

This condition is characterized by an abnormal number of chromosomes either with extra or loss of whole paired chromosomes. According to the difference in number, numerical aberrations can be polyploidy where an extra one or more sets of chromosomes are present in an individual or aneuploidy, where an extra chromosome or a missing chromosome from the egg or sperm results in an abnormal number of chromosomes.

Usually chromosomal aberrations are fatal to the embryo. However, aneuploidy helps interpret sex determination in birds (Kupper *et al.* 2012). There are few reports about chromosomal abnormalities in birds. Chickens with ZWW triploid genotype are usually not viable whereas triploid chicken with ZZZ genotype becomes male phenotype with abnormal sperm development. Triploid chicken with ZZW genotype change from female to male phenotype upon maturity (Lin *et al.* 1995). Forstmeier and Ellegren (2010) have described embryo mortalities in captive zebra finches (*Taeniopygia guttata*) using microsatellite genotyping and identified aneuploidy or polyploidy as the reason for the embryo mortality (3.6%), demonstrating trisomy and triploidy forms of chromosomal abnormality. Yet, Kupper *et al.* (2012) reported that a triploid female Kentish plover (*Charadrius alexandrinus*) reproduced successfully in a natural population.

1.6.2 Structural Chromosome Aberrations

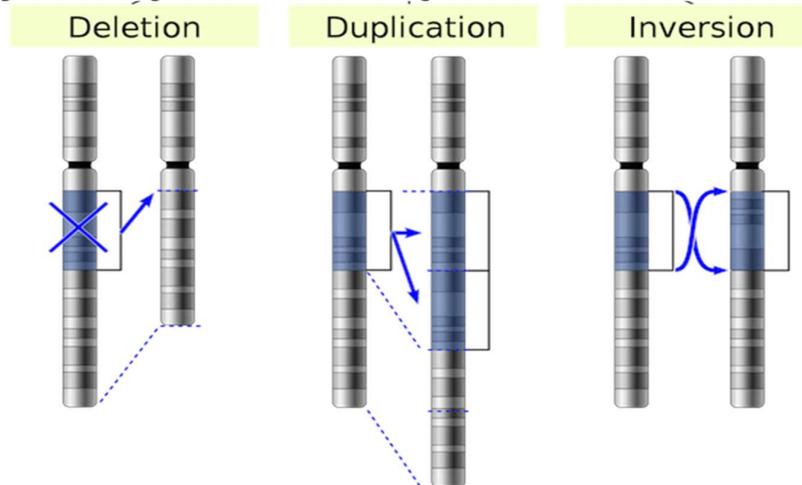
Structural chromosome changes without altering chromosome number can modify genomic information mainly through deletion, duplication, inversion and translocation. These chromosomal rearrangements can either result from a fragile genome region

which breaks resulting in deletion, duplication, inversion or translocation or from errors occurring during cell division. Structural chromosomal changes can affect gene dosage, thus phenotypical alterations occur and the severity depends on the amount of genome involved which can often be drastic (Clancy and Shaw 2008).

1.6.2.1 Deletion, Duplication, Inversions and Translocation

Deletion can occur when a part of a chromosome or a sequence of a DNA is lost during meiosis. Larger chromosomal deletion can be visible on karyotyping; smaller deletions are not visible on karyotyping. The best example of chromosomal deletion in human is Cri du chat syndrome where a piece of chromosome 5 is missing. Neonates with this condition often exhibit characteristic cry of a cat, they are also struck by intellectual disability, microcephaly, hypotonia and delayed development. When chromosomal duplication occurs, an additional chromosomal region is created having different copy numbers of genes (Figure 1-29). Gene duplication has an important role in terms of evolution by providing new genetic materials through producing changes in the genome architecture of an organism; these can lead to diversification of more closely related species (Magadum *et al.* 2013).

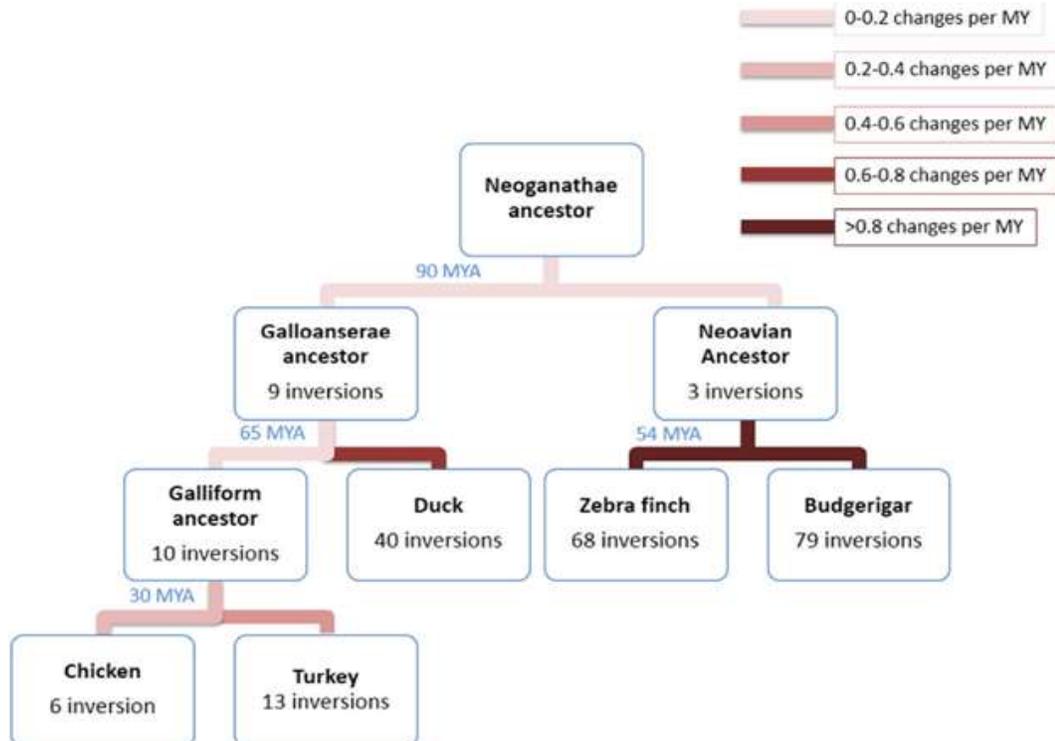
Figure 1-29: Single chromosomal rearrangements



An inversion occurs when a segment of a chromosome breaks off and rejoins the chromosome in the reversed orientation (Figure 1-29). Mainly two classes of inversions are seen: 'pericentric' where inversion includes the centromere and 'paracentric' which does not include the centromere.

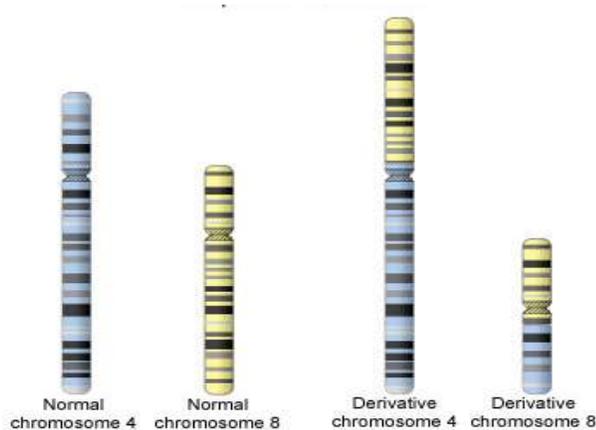
Although the avian genome is considered highly conserved, chromosomal inversions are very often seen in evolution. A review report on avian cytogenetic research which was conducted on 400 species from 58 families of *Passeriformes* reported that the distinguishable genome changes among closely related species are constructed as inversions (Hooper and Price 2015). As more and more avian genome sequences are available, comparative genomics studies reveal that genomic divergence among avian species is mainly due to chromosomal inversions (Hooper and Price 2015). Romanov *et al.* (2014) gave a comprehensive account of avian chromosomal evolution using both bioinformatics and chromosome painting (FISH) and found that the greatest changes (chromosomal inversions) occurred in zebra finches and budgerigars (Figure 1-30). Inversion is important as it has the capability of producing genetic isolation between population and species (Kirkpatrick and Barton 2006). Geographical variation can be a force behind inversion which is used as a local adaptation mechanism in this scenario. During the process of local adaptation different genes are preferred according to the environmental conditions (Kirkpatrick 2010).

Figure 1-30: A combination of FISH and bioinformatics data showing the total number of inversions in six avian species as they diverged from the ancestor (Romanov *et al.* 2014).



Translocations can happen when a segment of the chromosome breaks off and attaches to another chromosome; it can occur in two different forms; reciprocal translocation and Robertsonian translocation. Reciprocal translocation occurs when segments of two non-homologous chromosomes exchange with each other (Figure 1-31). In most cases, genetic information is neither lost or gained and such individuals are healthy but with a high risk of miscarriages (Pourjafari *et al.* 2012).

Figure 1-31: Reciprocal translocation illustrating exchange of segment of chromosomes 4 and chromosome 8.

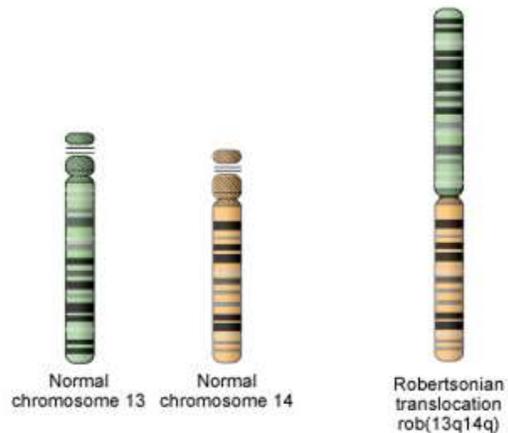


95% of chronic myelogenous leukemia cancer patients have reciprocal translocation between chromosome 9 and 22 where bcr-abl gene is formed on the chromosome 22 (Sessarego *et al.* 1987). Robertsonian translocation is formed from the breakage of two non-homologous acrocentric

chromosomes at the centromere followed by the fusion of their long arm (q-arm) to form a single chromosome (Figure 1-32).

In humans, carriers of a Robertsonian translocation have 45 chromosomes and when a chromosome number 21 is involved can cause Down syndrome.

Figure 1-32: Schematic representation of Robertsonian translocation between human chromosome 13 and 14.

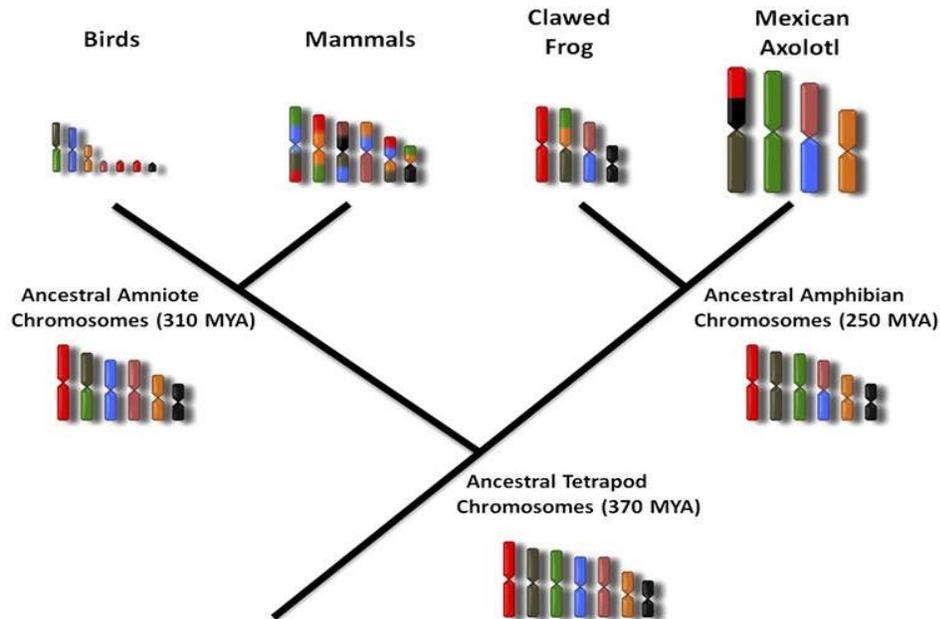


1.6.2.2 Chromosomal Fission and Fusion

The avian karyotype has evolved through independent convergent changes; as chromosome 1 and 2 underwent individual fission in Passeriformes (song bird) and in some Galliformes (land fowl) (Griffin *et al.* 2007). Comparative studies to construct the

evolutionary origin of amniote chromosomes including the amphibians, revealed that the vertebrate genome has evolved through fusion mechanism. Along with this, the origin of chicken chromosomes was reconstructed which again proved an active role of fission, fusion and retention of ancestral chromosomes (Figure 1-33) (Voss *et al.* 2011).

Figure 1-33: Different combinations of ancestral chromosomes fused together as chromosome numbers decreased in parallel within anuran and salamander lineages. Fission and fusion were active in framing the generalized avian karyotype (Voss *et al.* 2011).



1.6.2.3 Copy Number Variations

Chromosomal rearrangements and copy number variants (CNVs) are the major issues of interest in the field of cytogenetics as both play an important role in phenotypic variation, genetic disease and genome evolution. However, molecular processes causing these kind of structural genomic variations are not completely concluded (Volker *et al.* 2010). CNVs are the variations in the DNA fragments $\geq 1\text{Kb}$, defined as “a difference in copy number compared to a reference genome” and are usually identified by Comparative Genome Hybridization (CGH) Array. Medical relevance of CNV is well documented and is thought to be linked with schizophrenia, cancer, autism and other psychiatric disorders (Stefansson *et al.* 2014; Gulsuner and McClellan 2015).

In birds, cross species array-CGH have been conducted in turkey, duck and zebra finch using the chicken as the reference species (Itoh and Arnold 2005; Griffin *et al.* 2008; Skinner *et al.* 2009a; Volker *et al.* 2010). It was found that CNVs are less common in birds than mammals. Skinner *et al.* (2014) hybridized 16 bird species belonging to several key avian clades to a chicken whole genome tiling path microarray using well-established-CGH techniques. This study gives great insight into the key role of CNVs in avian species' variation and genome evolution by analyzing four major hypotheses culminating from previous studies. According to this new study, with improved CNV detection platforms, it is no longer believed that bird CNVs are smaller than mammals. Most of the identified avian CNVs in this study were associated with protein coding genes which indicates that CNVs are functionally significant. Again, this study strongly supports the correlation with intra-chromosomal rearrangements rather than inter-chromosomal rearrangements and also, that the smaller chromosomes are more CNV- dense than macrochromosomes.

1.7 Avian Chromosomal Studies

1.7.1 Avian Karyotyping

The karyotype of an individual or species is vital for any genome-mapping effort where both genetic and physical maps are generated in line with chromosome position (Masabanda *et al.* 2004). The avian karyotype is characterized by having few macrochromosomes and many microchromosomes (~30); their 'so many, so small' pattern makes them unique (Griffin *et al.* 2007). So far there are over 1,000 published avian karyotype papers available in which the most comprehensive account of the chromosome number is given by Christidis (1990) in a study which exceeded 800 species. The majority of birds have around $2n=80$ chromosomes, consisting of 7-10 pairs of large and medium sized macrochromosomes including sex chromosomes and plentiful microchromosomes (Nishida *et al.* 2008). There are few exceptions to this pattern, which include those with a known small diploid number such as the stone curlew (*Burhinus oedicephalus*; $2n=42$), the beach thick knee (*Esacus magnirostris*; $2n=40$), and the trumpeter hornbill (*Ceratogyan bucinator*; $2n=42$), and those with a high diploid chromosome number such as hoopoe (*Upupa epops*; $2n=126$) and the kingfishers ($2n>120$) (Christidis 1990). Microchromosomes, which are numerous and unusually

small cause major challenges for the chromosome biologist classifying chromosomes in detail to investigate differences, if present, between individuals, or among species (Itoh and Arnold 2005; Lithgow *et al.* 2014). Gene density is highest on microchromosomes, at least twice as in macrochromosomes (Smith *et al.* 2000; McQueen *et al.* 1996; Itoh and Arnold 2005).

As mentioned in section 1.5 G-banding is the 'gold standard' method of studying any karyotype. The results have been limited in birds firstly because bands on group A chromosomes are less distinct than in mammals and secondly, some chromosomes (from chromosome 10 onwards) are too small and difficult to visualize by any banding pattern. (Griffin *et al.* 2007). Identification of chromosomes is essential for genetic and physical mapping of any species. To address this problem, chromosome-specific FISH (fluorescence *in-situ* hybridization) probes (both chromosome paints and BACs) for each chicken chromosome were developed (Griffin *et al.* 1999; Masabanda *et al.* 2004) which allowed cross species chromosome painting for several other bird species.

1.7.2 Chromosome Studies Using Fluorescence *in situ* Hybridization

Griffin *et al.* (1999) developed chromosome paints or probes for 27 chicken chromosomes which include all macrochromosomes, several of the larger microchromosomes and a fraction of the smaller microchromosomes. His team used flow cytometry to develop chromosome paints for the larger chromosomes and developed chromosome paints for the smaller ones by microdissecting microchromosomes. Degenerate oligonucleotide-primed polymerase chain reaction (DOP-PCR) was used to amplify and label the isolated 27 chromosomes by both techniques. Later, Habermann *et al.* (2001) utilized these chicken probe sets to study chromosome territory arrangements in chicken nuclei for comparison with the data known for human cells. A complete chicken karyotype, the first (and to date only) in birds was published by Masabanda *et al.* (2004). A definite classification system of chicken chromosomes was generated where the chicken chromosomes are divided into 4 distinct groups; A (1-10 +ZW), B (11-16), C (16-32) and D (33-38).

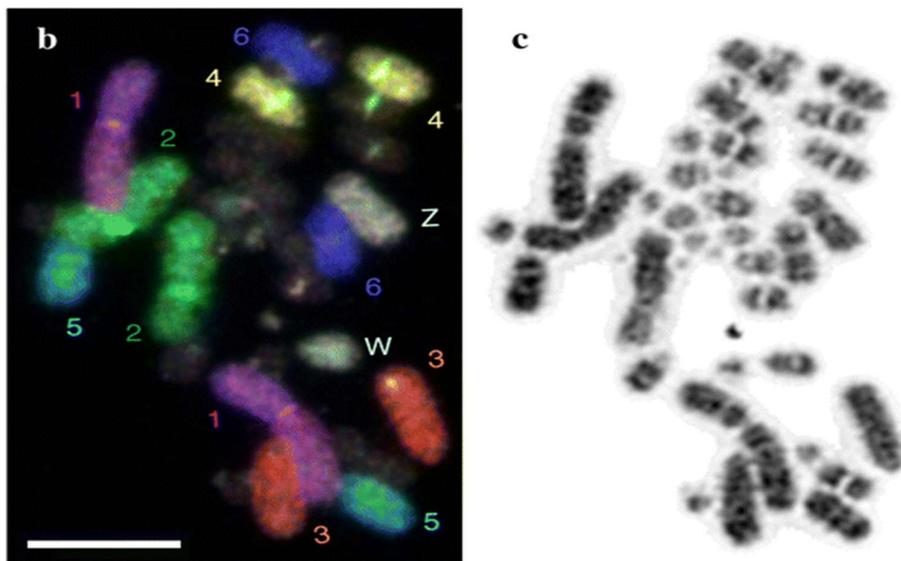
1.7.2.1 Comparative mapping – BAC clones

BAC clones are important tools in FISH techniques as they have the ability to carry inserts of DNA up to 200,000 kb which can detect precise regions in the genome of interest. Availability of reliable chicken BACs (Griffin *et al.* 1999; Lithgow *et al.* 2014) together with chicken paints empowered the establishments of comparative genome mapping between chicken and many avian species. However, cross-species BAC mapping was not successful in a few birds such as the turkey (*Meleagris gallopavo*) and Pekin duck (*Anas platyrhynchos*) (Griffin *et al.* 2008; Skinner *et al.* 2009a). Yet, Damas *et al.* (2017) reported the construction of a panel of chicken BACs (1-28 + Z (except 16)) which has a high hybridization rate for several avian species.

1.7.2.2 Cross species chromosome painting – Zoo FISH

Chromosome paints for comparing genomes of distantly related species worked successfully in numerous avian species by detecting DNA sequence homology between whole chromosomes; this powerful technique is widely used to evaluate karyotype evolution (Itoh and Arnold 2005; Kasai *et al.* 2012; Nanda *et al.* 2007; Nanda *et al.* 2011; Nishida *et al.* 2008; Raudsepp *et al.* 2002; Shetty *et al.* 1999; Shibusawa *et al.* 2002; Seibold-Torres *et al.* 2015) (Figure 1-34).

Figure 1-34: Seven-colour chromosome painting of the stone curlew metaphase (Nie *et al.* 2009).



Zoo-FISH has been adapted to compare genomes between chickens and emus using single chromosome paints from the chicken; it revealed strong homology between 9 macrochromosomes of the emu and the chicken suggesting, chromosome number and morphology between ratites and carinates are conserved (Shetty *et al.* 1999). Raudsepp *et al.* (2002) conducted a similar study in California condors (*Gymnogyps californianus*) which delivered fundamental cytogenetic and comparative information on condor chromosomes. Chicken chromosome paints for macrochromosomes have been extensively used in many avian species like parrots, turkeys, zebra finches, ducks, Guinea fowl and so on. Furthermore, Kasai *et al.* (2012) extended comparative chromosome painting studies in turtles (*Trachemys scripta elegans*) and crocodiles (*Crocodylus niloticus*). The results showed that a high degree of conserved synteny (70% identity) exists among these species though they are distantly related species.

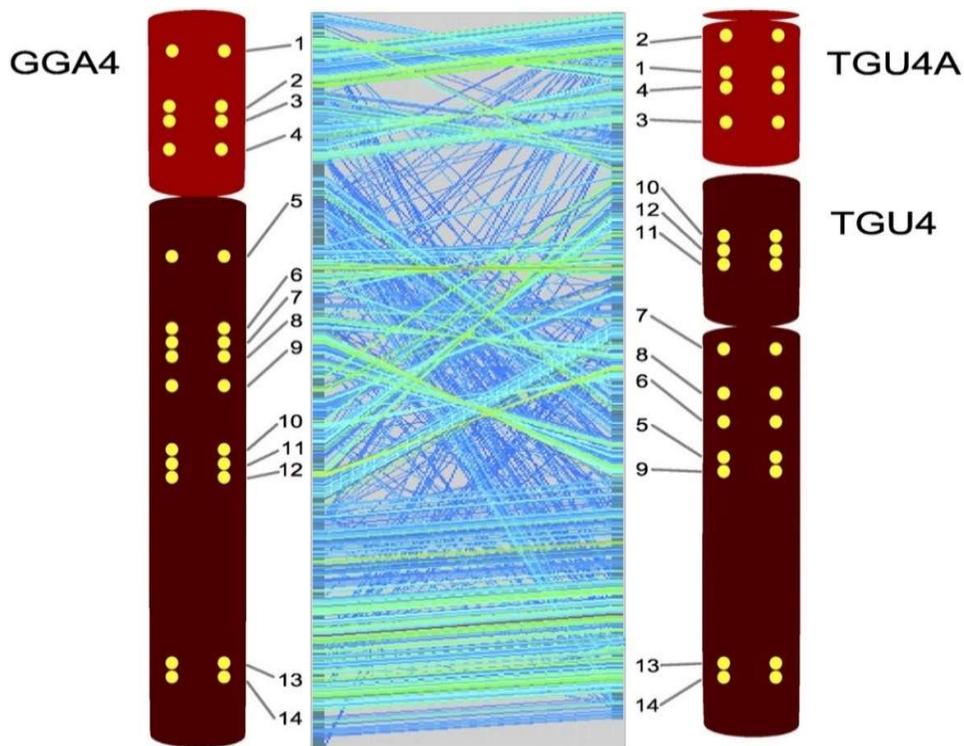
It is evident that the majority of avian comparative genomic studies by molecular cytogeneticists have focused on macrochromosomes. Microchromosomal paints have been used only in few avian species (Griffin *et al.* 1999; Hansmann *et al.* 2009; Nie *et al.* 2009; Nishida *et al.* 2008; Shetty *et al.* 1999; Skinner *et al.* 2009a). Using these paints often hindered interpretation accuracy by recognizing more than one microchromosome by each paint (Lithgow *et al.* 2014). However, Lithgow *et al.* (2014) reported the development of a panel of chicken microchromosomal paint pools and BACs that work on the duck, houbara bustard, goose, Gyrfalcon, budgerigar and zebra finch. Comparative chromosome painting with chicken probes has been successfully hybridized on several avian species from 15 different orders and these studies show that bird karyotypes are highly conserved.

1.7.2.2 Comparative Genomics using Zoo-FISH And Genome Sequence

The availability of more than one genome sequence promoted comparative genomics into a much more detailed level and multi-species alignment became possible (Griffin and Burt 2014). Comparative chromosome mapping has become the favourite tool for cytogeneticists because of its ability to identify inter- and intra-chromosomal rearrangements (Islam *et al.* 2014). In 2010, the first comprehensive comparative maps of the chicken and zebra finch were generated by combining bioinformatics, synteny

analysis and physical mapping by FISH (Volker *et al.* 2010) (Figure 1-35). Alignment of whole-chromosome sequences of orthologous chicken and zebra finch chromosomes revealed 114 intrachromosomal rearrangements in all macro- and microchromosomes and for the first time showed that there is a high degree of conserved synteny in the microchromosomes. To validate the above mentioned results, the authors also did physical mapping by FISH with 131 chicken and 131 zebra finch BACs containing orthologous sequences which confirmed the chromosomal rearrangements. Generally, sequence alignments and FISH mapping gave almost the same results showing reliability of the chicken and zebra finch genome sequence assemblies.

Figure 1-35: Comparative analysis of marker order on chicken chromosome 4 (GGA4) and its zebra finch orthologs, TGU4 and TGU4A. The central part of the figure was created by aligning whole-chromosome sequences using the program GenAlyzer (Volker *et al.* 2010).



Rao *et al.* (2012) constructed the whole duck genome radiation hybrid panel for genome mapping from which chicken/duck comparative mapping became possible. In doing so, several intrachromosomal rearrangements have been identified on microchromosomes which was the first time for these species (Rao *et al.* 2012). When comparative genomics were carried out on chicken, turkey and zebra finch macrochromosomes using

GenAlyser, it was found that intrachromosomal rearrangements were commonplace (Skinner and Griffin 2012). Using an interactive, comparative browser, Evolution Highway and FISH analysis on 20 avian genomes and the Carolina anole lizard (*Anolis carolinensis*) ancestral chromosomes were designed furnishing insight into the genetic recombination role of chromosomal rearrangements (Romanov *et al.* 2014). This analysis suggests that the microchromosomes retains the conserved blocks of synteny and the chicken lineage retains genetical characteristics with the fewest changes compared to dinosaur ancestors.

1.8 Cytogenetic and Genomic Studies in Falcons

This thesis is primarily concerned with falcon species, specifically combining molecular, cytogenetic and genomic studies to understand better the overall genome organization of these fascinating birds. The above sections have largely dealt with advances in chromosomal and genomic studies in animals in general, and particularly in birds. The following sections deal specifically with the state of the art in falcon species.

Attempts to karyotype *Falconidae* species and to investigate their karyological relationship have been documented since 1970 (de Boer 1975). The *Falco* genus is characterized by having lower diploid chromosome numbers ranging from 40-54, with 7-11 pairs of large and medium-sized chromosomes and about 13-16 pairs of microchromosomes. Karyotyping was performed on Peregrine falcons, prairie falcons and Gyrfalcons aiming to design a phylogenetical tree with the help of previously published karyotypes from other falcons (Schmutz and Oliphant 1987). Karyotypes of Peregrine falcon and prairie falcons are similar and differ from Gyrfalcons by 4 chromosomes. However, these karyotypes are of poor quality (Figure 1-36). Amaral and Jorge (2003) reviewed the genetic research summary of 66 species of *Falconiformes* which were analyzed between 1966 to 2001 (Table 1-6). The members in the genus *Falco* show remarkable differences in karyotype, characterised by low diploid numbers (40-52) suggesting translocation and fusion of the microchromosomes into larger chromosomes (Amaral and Jorge 2003). The *Falco* genus comprises 37 or 39 species (Sibley and Ahlquist 1990; del Hoyo *et al.* 2004; Itoh and Arnold 2005) from out of which

10 species have reported karyotype data. They are *Falco columbarius* (2n=40), *Falco mexicanus* (Prairie falcon) (2n=48), *Falco chicquera* (Red-necked falcon), *Falco jugger* (Laggar falcon), *Falco sparverius* (American kestrel), *Falco subbuteo* (Eurasian hobby), *Falco peregrinus* (Peregrine falcon) (2n=50), *Falco rusticolus* (Gyr falcon), *Falco tinunculus* (Common kestrel) (2n=52) and *Falco biarmicus* (Lanner falcon)(2n=52 or 54) (Al Mutery 2011; Belterman and de Boer 1984; de Boer 1990; Longmire *et al.* 1988; Nishida *et al.* 2008; Sasaki *et al.* 1984; Schmutz and Oliphant 1987).

Figure 1-36: Karyotypes of the (A) Peregrine falcon and (B) Prairie falcon, and (C) Gyr falcon. (Schmutz and Oliphant 1987).

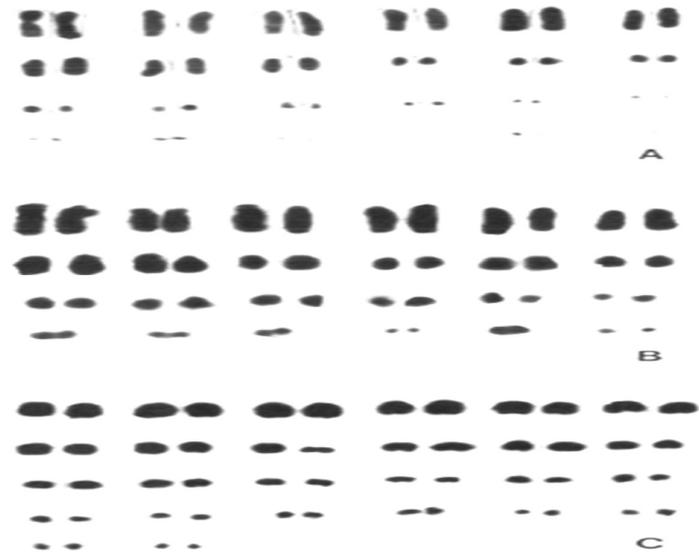


Table 1-6: Karyotype characterizations of Falconiformes. Adapted from (Amaral and Jorge 2003). (Mac) macrochromosomes, (Mic) microchromosomes (-) absence of distinction between mac and mic.

| Species | 2n | Mac | Mic | Main remarks |
|---|----|-----|-----|-----------------------------------|
| Lanner falcon (<i>Falco biarmicus</i>) | 52 | - | - | Z and W not identified |
| Red-headed falcon (<i>F. chiquera</i>) | 50 | - | - | Just the diploid number described |
| Merlin (<i>F. columbarius</i>) | 40 | 20 | 20 | Z and W not identified |
| Laggar falcon (<i>Falco jugger</i>) | 48 | - | - | Z is tentatively. W small ac |
| Prairie falcon (<i>Falco mexicanus</i>) | 48 | - | - | Z and W not identified |

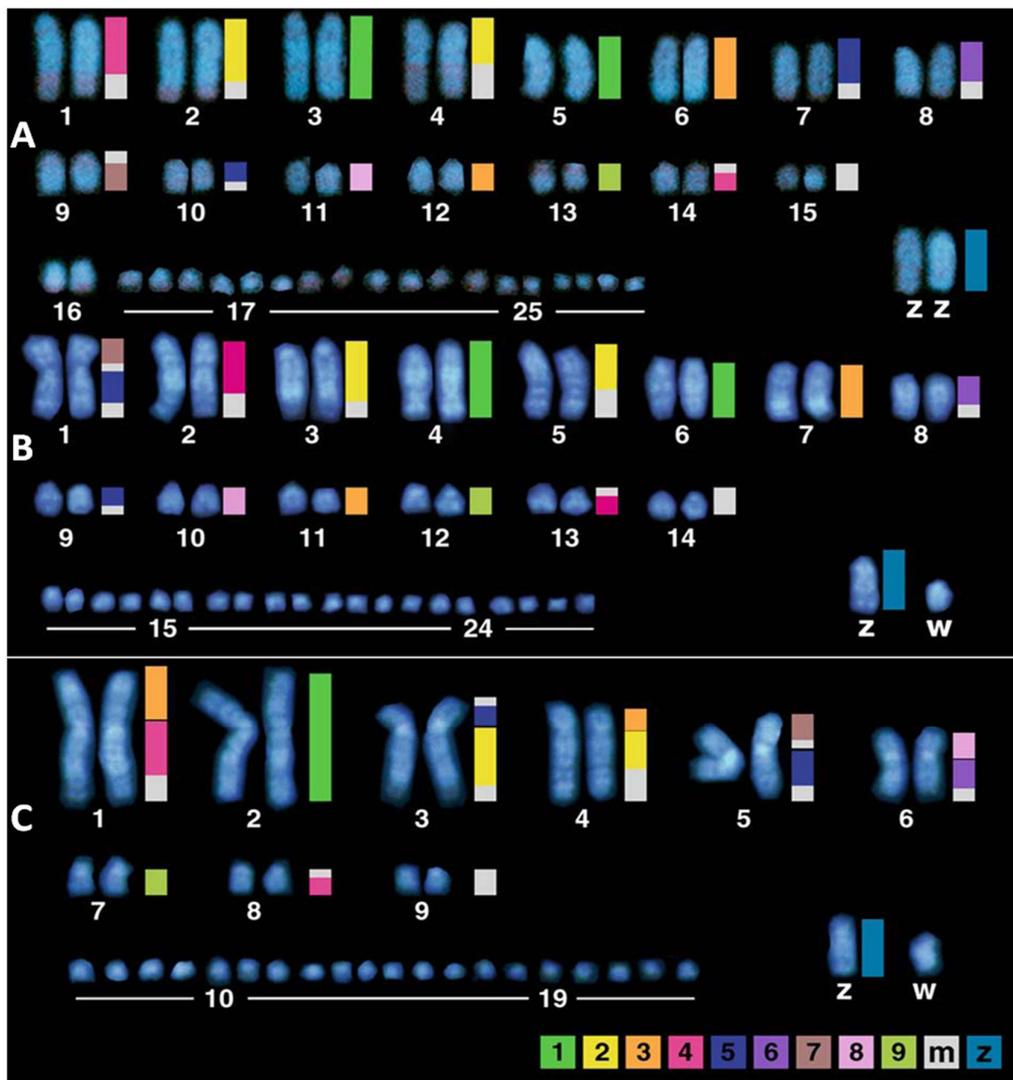
Table 1-6 (Continued): Karyotype characterizations of Falconiformes. Adapted from (Amaral and Jorge 2003). (Mac) macrochromosomes, (Mic) microchromosomes (-) absence of distinction between mac and mic.

| | | | | |
|--|----|----|----|---|
| Peregrine falcon (<i>Falco peregrinus</i>) | 48 | - | - | Identical |
| Gyrfalcon (<i>Falco rusticolus</i>) | 52 | - | - | Identical |
| American kestrel (<i>Falco sparverius</i>) | 50 | 18 | 32 | (2n=80 and 2n=50) |
| Common kestrel (<i>Falco tinnunculus</i>) | 50 | - | - | |
| Mountain caracara (<i>Phalco boenus megalopterus</i>) | 90 | - | - | Identical |
| Yellow-crested caracara (<i>Milvago chimachima</i>) | 84 | - | - | Diploid number is tentatively. Z and W not identified |
| Southern crested caracara (<i>Polyborus plancus</i>) | 84 | - | - | Z and W not identified |
| Osprey (<i>Pandion haliaetus</i>) | 74 | 52 | 22 | |
| Secretary bird (<i>Sagittarius serpentarius</i>) | 80 | 36 | 44 | Discrepancy between diploid number |
| Secretary bird (<i>Sagittarius serpentarius</i>) | 74 | - | - | Number Z and W not identified |

Among falcon species the most comprehensive account to date of the chromosome characterization is given by Nishida *et al.* (2008). Here, molecular cytogenetic characterization of the chromosome components of three *Falco* species, the common kestrel (*Falco tinnunculus*), Peregrine falcon (*Falco peregrinus*) and merlin falcon (*Falco columbarius*) has been performed utilizing chromosome paints of chicken chromosomes 1-9 and Z and a 19 chicken microchromosome-specific paint pool. *Falco tinnunculus* has a karyotype (2n=52) consisting of all acrocentric chromosome except for the submetacentric W chromosome. Karyotyped *Falco peregrinus* has a diploid number of (2n=50), all acrocentric chromosomes except for the one pair of large metacentric macrochromosomes. *Falco columbarius* has a lower chromosome number (2n=40) and

unlike those of other species has six pairs of large bi-armed macrochromosomes (Figure 1-37). By incorporating the karyotype data prepared by Sasaki *et al.* (1984) for *Falco tinnunculus* (FTI) and *Falco peregrinus* (FPE), additionally, from the karyotype data designed by Longmire *et al.* (1988) for *Falco columbarius* (FCO), Nishida *et al.* (2008) identified and arranged chromosomes of these three species from the Hoechst-stained karyotypes.

Figure 1-37: Hoechst-banded karyotypes of the Common kestrel (A), Peregrine falcon (B) and Merlin falcon (C) with the assignment of homology with chicken chromosomes delineated by chromosome painting with chicken probes (Nishida *et al.* 2008).



All chicken probes systematically cross-hybridized to metaphase chromosomes of the three species (Figure 1-38) and homologous chromosomes and chromosome arms between chicken and three *Falco* species are shown in (Table 1-7). Nishida *et al.* (2008) suggest that the ancestral karyotype of *Falco* probably had a diploid number of $2n=52$ or 54, consisting of all acrocentric chromosomes and having a karyotype with all acrocentric chromosomes except for the W chromosome. *Falco tinnunculus* is counted for retaining most of the ancestral status of *Falconinae* karyotypes.

Figure 1-38: Chromosome hybridization patterns with chicken chromosome 1, 4 and 6 paints (GGA1, GGA4, GGA6) and a paint pool of 19 microchromosomes (GGAmicro) to PI-stained metaphase chromosome spreads of *Falco tinnunculus* (a-d), *Falco peregrinus* (e-h) and *Falco columbarius*. (Nishida *et al.* 2008).

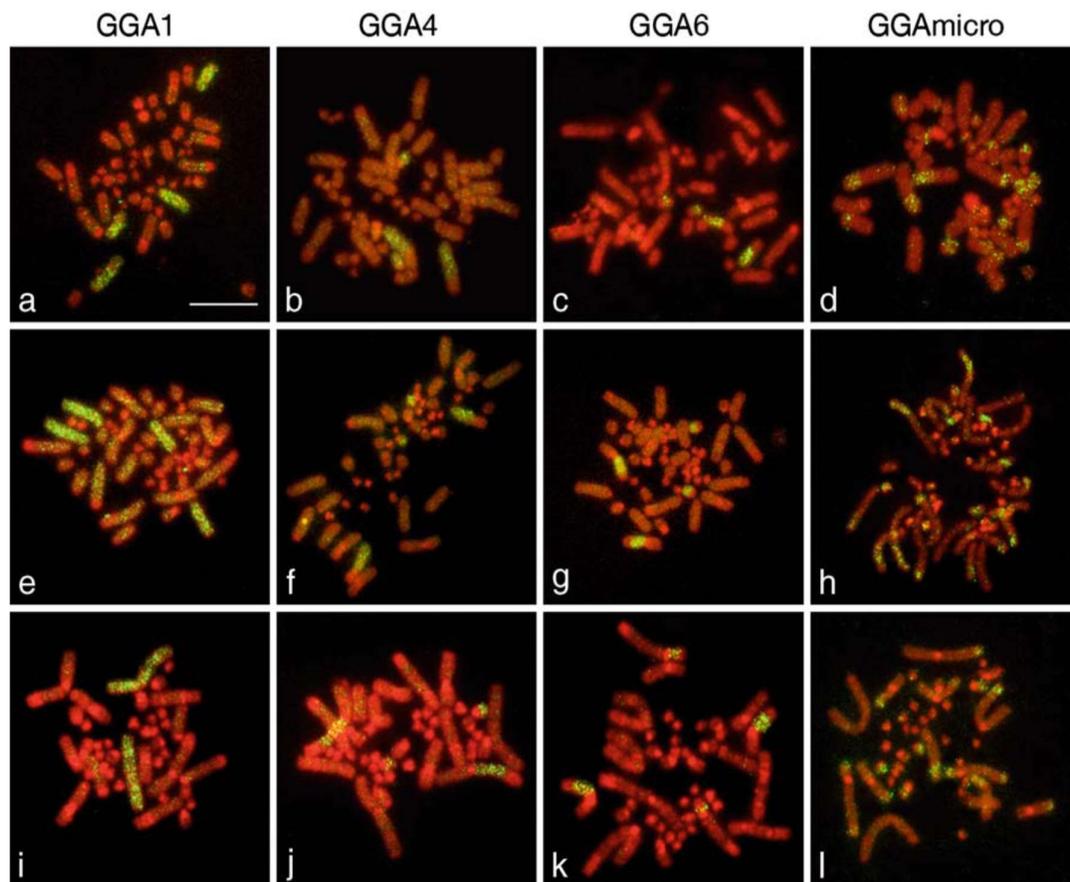


Table 1-7: Homologous chromosomes and chromosome segments between chicken and three *Falco* species as detected by chromosome painting using chicken macrochromosome probes. From (Nishida *et al.* 2008).

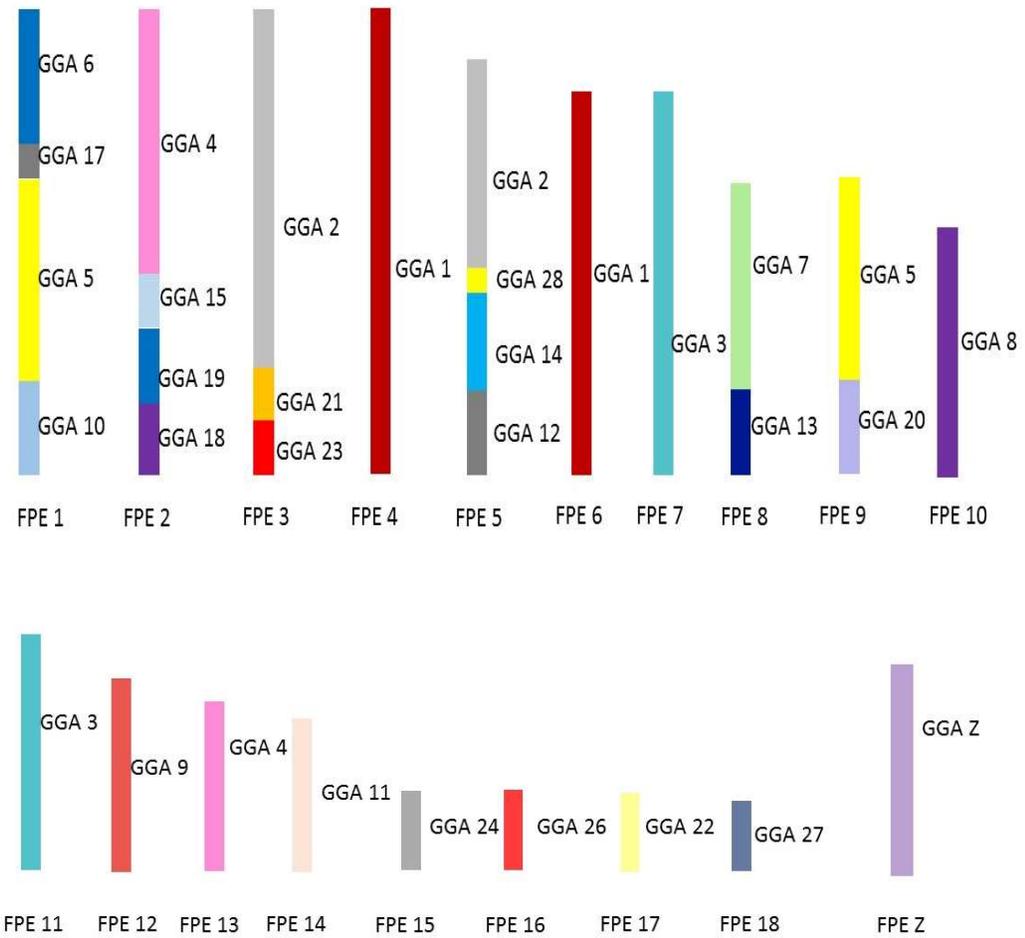
| Species | 2n | Chromosome | | | | | | | | | |
|---|----|------------|---------|---------|---------|---------|----|----|----|----|---|
| <i>Gallus gallus domesticus</i> (Chicken) | 78 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Z |
| <i>Falco tinnunculus</i> (Common kestrel) | 52 | 3 + | 2 + | 6 + | 1 + | 7 + | 8 | 9 | 11 | 13 | Z |
| <i>Falco peregrinus</i> (Peregrine falcon) | 50 | 4 + | 3 + | 7 + | 2 + | 1q + | 8 | 1p | 10 | 12 | Z |
| <i>Falco columbaris</i> (Merlin falcon) | 40 | 2 | 3q + | 1p + | 1q + | 3p + | 6q | 5p | 6p | 7q | Z |

To date however studies are limited to the largest chromosomes (1-9 +Z) and only using whole chromosome paints.

1.8.1 Recent Cross Zoo-FISH study on Peregrine falcon

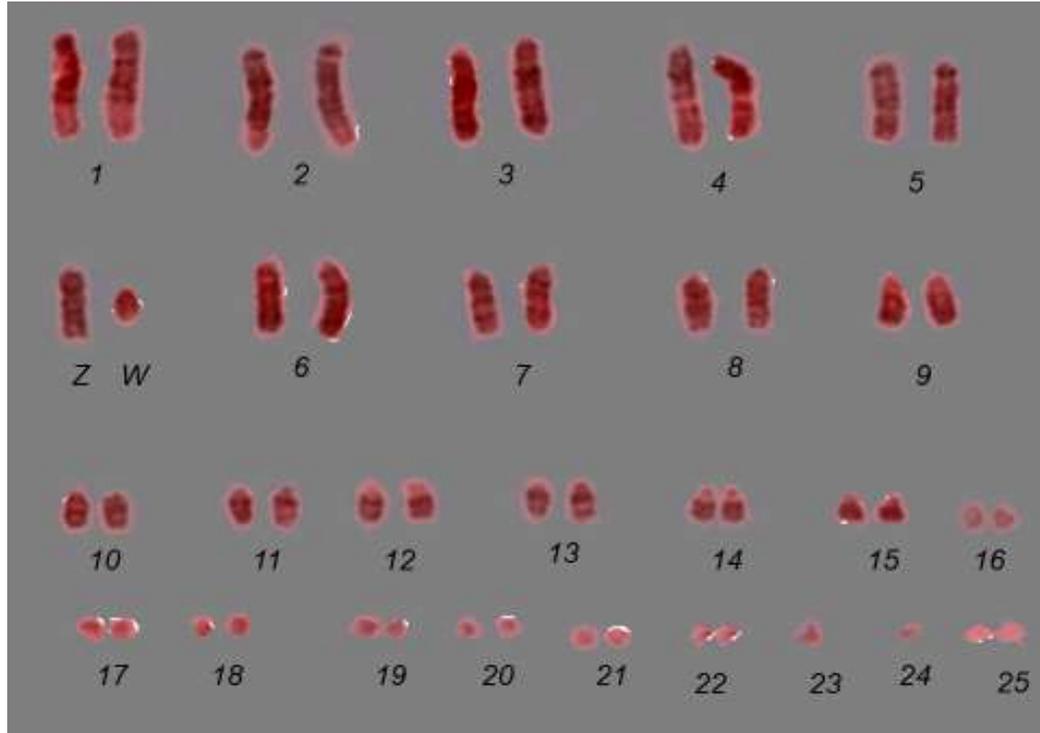
For the first time O'Connor (2016) carried out an extensive homology study between the Peregrine falcon and chicken (Figure 1-39). The zoo FISH karyotype generated with this study correlates exactly the preliminary homology study results between chicken and three falcon species released by Nishida *et al.* (2008). FISH was performed with the selected set of BACs and identified 13 Peregrine specific fusions and 4 fissions when compared to chicken. Peregrine falcon has undergone approximately 68 intrachromosomal rearrangements during the evolution of avian lineages.

Figure 1-39: Peregrine chromosomes (FPE) with their chicken (GGA) homologs (O'Connor 2016).



Moreover, Al Mutery (2011) generated a well banded karyotype (2n=52) and standard ideogram of the Gyrfalcon (*Falco rusticolus*) and raised the possibility of sharing the ancestral status of *Falconinae* karyotypes with *Falco tinnunculus* (Figure 1-40).

Figure 1-40: Gyrfalcon karyotype (Al Mutery 2011).

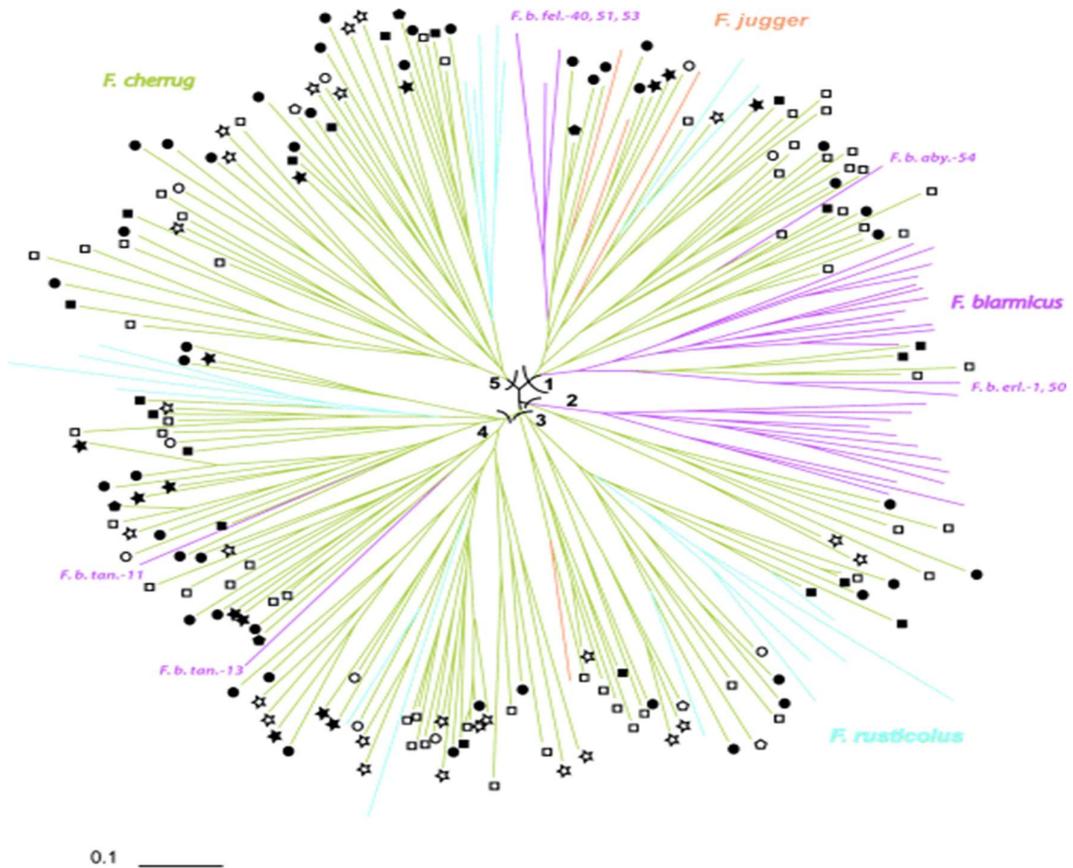


He also established the inter-specific synteny between Gyrfalcons and chickens using chicken paints for chromosomes 1-9+Z by doing zoo-FISH. Mitochondrial DNA sequence data of Gyrfalcon and Saker falcons revealed that they are closely related differing by only one nucleotide in the control region, while the Peregrine is not closely related to this two species (Al Mutery 2011).

Nittinger *et al.* (2007) inferred phylogeography and population genetic structure of the Saker falcon (*Falco cherrug*) by comparing mitochondrial and nuclear DNA markers. This study does not support any subspecific division or even the separation of the two subspecies of the Saker falcon, *Falco cherrug cherrug* and *Falco cherrug milvipes* (del Hoyo *et al.* 1994). This group also compared the patterns of mitochondrial haplotypes along with the variation of microsatellite alleles among the species of the hierofalcon complex (*Falco cherrug* (Saker falcon), *Falco rusticolus* (Gyrfalcon), *Falco biarmicus* (Lanner falcon), *Falco jugger* (Laggar falcon)) and revealed that they lack genetic differentiation (Figure 1-41). From an evolutionary aspect, this result indicates that hierofalcons are a young group.

Figure 1-41: NJ tree based on the proportion of shared microsatellite alleles among hierofalcon individuals (Nittinger *et al.* 2007).

Colour codes for species: green, *Falco cherrug*; purple, *Falco biarmicus*; blue, *Falco rusticolus*; orange, *Falco jugger*. Numbers 1–5 indicate group 1, heterogeneous comprising all four species; group 2 consist of two clusters of *Falco biarmicus* and *Falco cherrug*; group 3 contains mainly *Falco cherrug* and *Falco rusticolus*; group 4 comprises mainly *Falco cherrug*; group 5 mainly consist of *Falco cherrug*. Symbols indicate populations of *Falco cherrug*: black squares, Central Europe-historical samples; white squares, Central Europe-contemporary; black circles, Eastern Mongolia; white circles, Central Mongolia; black asterisks, North Kazakhstan; white asterisks, South Kazakhstan; black pentagons, South Siberia; white pentagons, Central Asian group.



To date however high quality karyotype analysis of the Saker falcon and an appraisal of the comparative genomics of Saker compared to other species only has limited information. Comparison of multiple species however sheds insight into the direction and rate of evolutionary change.

1.8.2 Falcon Genome Sequencing

Peregrine (*Falco peregrinus*) and Saker falcon (*Falco cherrug*) genomes were sequenced to understand evolutionary aspects of predatory adaptations of falcons (Zhan *et al.* 2013). Sequencing of the male Peregrine and male Saker falcon genome has been carried out with the help of next-generation genome sequencing platform, generating 128.07Gb

and 136.21Gb of sequence for *Falco peregrinus* and *Falco cherrug*, respectively. Genome size of both species was estimated at 1.2Gb with a genome coverage of 106.72x for *Falco peregrinus* and 113.51x for *Falco cherrug* (Zhan *et al.* 2013). Genome sequences were assembled using SOAPdenovo (Kim *et al.* 2011) and assessment of the assembled *F. peregrinus* and *Falco cherrug* genomes using fosmid-based Sanger sequencing confirmed >99% (Peregrine) and 97% (Saker) coverage. Protein-coding genes were predicted using homology and *de novo* methods and a second approach, RNA sequencing data was used to process fine gene structure and establish novel genes. As a result of this combined effort, 16,263 genes were predicted for *Falco peregrinus* and 16,204 were predicted for *Falco cherrug* (Zhan *et al.* 2013).

Orthologous genes were identified in the chicken, zebra finch, turkey, Peregrine and Saker when comparative genome analysis was performed using the program TreeFam (Li *et al.* 2006). Falcons possess less repetitive DNA and like zebra finches, the falcon genome has fewer DNA transposable elements and long interspersed nucleotide elements (LINEs). Furthermore, the falcon genome has fewer large segmental duplications and lineage-specific genes than that of the chicken or zebra finch (Zhan *et al.* 2013).

Though falcons have a bigger olfactory bulb ratio than the zebra finch and a comparable ratio to the chicken (Steiger *et al.* 2009), the olfactory receptor gene analysis in falcons showed that they harbor the fewest intact olfactory receptor genes. This finding is contradictory to the previous finding where these two traits have been positively correlated (Steiger *et al.* 2008). Moreover, the major olfactory receptor clade, namely γ -c does not exist in falcons, whereas both chicken and zebra finches show expansions in the olfactory receptor γ -c clade (Steiger *et al.* 2009). This analysis supports their dependency on vision for locating prey (Roper 1999).

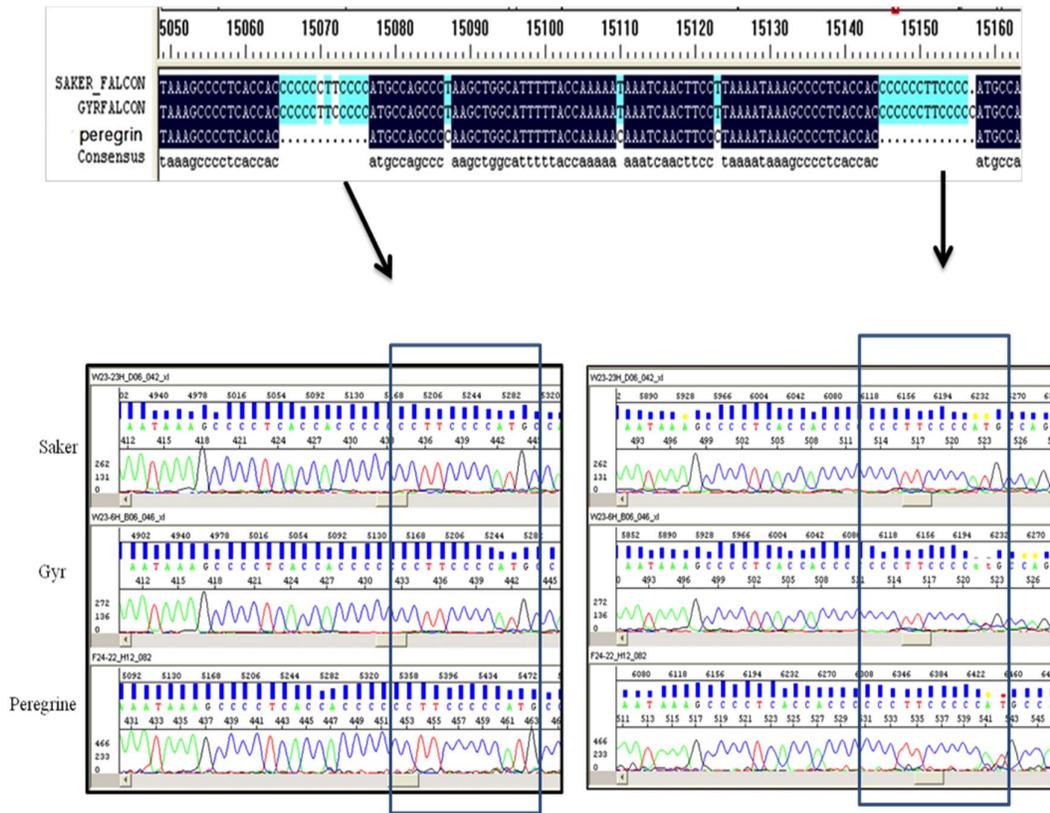
Zhan *et al.* (2013) have identified 879,812 heterozygous SNPs in the Peregrine genome and 761,748 heterozygous SNPs in the Saker genome. However, Saker has a higher heterozygous SNP rate (0.8 heterozygous SNP rate per kb) than the Peregrine (0.7 heterozygous SNP rate per kb). This finding suggests that mutations in the Peregrine

genomes are more evenly distributed due to a more heterogeneous SNP distribution in this species than the Saker. A demographic history of both species has been done using pairwise sequentially Markovian coalescent method (PSMC) (Li and Durbin 2011). For the Peregrine, construction of a demographic history from 2 Ma to 10, 000 years ago was possible, whereas, for the Saker, it was necessary to analyze both the Saker and its ancestral hierofalcon (Nittinger *et al.* 2005). The fossil record shows that the Saker emerged less than 34,000 years ago (Nittinger *et al.* 2005). PSMC showed that both falcon species have experienced at least one bottleneck.

Both falcons used for sequencing are from Eurasian populations. The breeding distribution of the Saker is restricted to the Palearctic and arid environments it primarily inhabits (Nittinger *et al.* 2005). It is assumed that Sakers require greater maintenance of osmotic equilibrium and suffer heat stress more than Peregrines due to their geographical distribution. Genes from two kidney expressed KEGG (Kyoto to Encyclopedia of Genes and Genomes) pathways and other genes involved in homeostasis were studied in the two species (Zhan *et al.* 2013). *Rab 11a* and *GNAS* have major inputs to the V2R water conservation pathway (Kanehisa *et al.* 2010; Tajika *et al.* 2005), and both genes have 2 additional copies in Saker falcon than in Peregrine falcon.

Genome sequencing data of the Peregrine and Saker is a source of information for the future study of evolution and adaptation in birds, especially in raptors. Genome sequencing of Gyrfalcon (*Falco rusticolus*) was presented as a part of a thesis by Al Mutery (2011) for the purpose of scaffolding the genome assembly and identifying the chromosomal fusions during falcon evolution. Using Illumina next generation sequencing 2 libraries with 300bp inserts and one library with 500bp inserts were generated. Two different sets of assembly versions were created with assembly lengths of 1.19Gb and 1.16Gb which are available for BLAST searching via ARK Genomics database. However, efforts to detect the presence and nature of chromosomal fusions which could be a unique phenomenon in falcons were not successful (Figure 1-42).

Figure 1-42: Screenshot of sample window of the sequencing analysis software version (Al Mutery 2011).



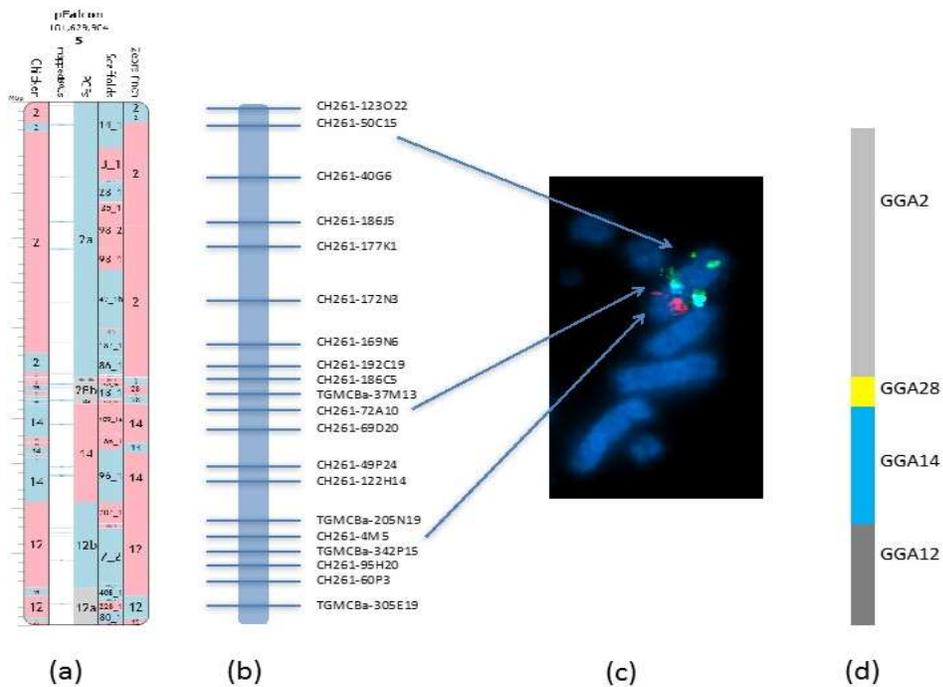
1.8.3 Peregrine Genome – Chromosome Mapping

Genome sequences available for Peregrine and Saker falcons are in the form of sub-chromosomal sized “scaffolds” generated with next-generation genome sequencing technology (Zhan *et al.* 2013). One of the most appreciable insights into falcon genome offered by Damas *et al.* (2017) was the development of a new approach to upgrade the fragmented Peregrine falcon genome assembly to the chromosome level. This is achieved by using bioinformatics approach called the Reference-Assisted Chromosome Assembly (RACA) that produce sub-chromosome sized predicted chromosome fragments (PCFs), combined with the verification of scaffolds by PCR and the physical mapping to chromosome by hybridizing with a universal set of avian bacterial artificial chromosome (BAC) probes (Figure 1-43).

To design PCFs along chromosomes, 177 clones were localized to microchromosomes and macrochromosomes of Peregrine falcons. Comparing homology between the

chicken and falcon showed that in total, 13 fusions and six fissions occurred during evolution. Moreover, out of 17 mapped chicken microchromosomes, 11 were found to be fused with other chromosomes. Sixty-nine intrachromosomal evolutionary breakpoint regions (EBRs) were also observed in this study giving evidence for the rare involvements of interchromosomal rearrangements in bird evolution (Damas *et al.* 2017).

Figure 1-43: Cytogenetic and PCF mapping of Peregrine falcon chromosome 5 (FPE 5) using FISH; (a) Evolution highway alignment of zebra finch, chicken and Peregrine falcon genome along with the PCFs produced by RACA and the BACs that map in this region, (b) Cytogenetic map of BACs in the correct orientation on FPE 5, (c) Physical mapping of BACs to FPE 5 using FISH and (d) Ideogram of FPE 5 homologous to chicken (O'Connor 2016).



The Damas *et al.* (2017) study successfully generated a successful cytogenetic genome map of the Peregrine falcon. To date however this has yet to be achieved in any other falcon species.

1.9 Specific Aims of this Thesis

Given the information presented in this introduction, it is clear that, while there have been a series of advances in falcon genomics and cytogenetics, there are still a number of questions to be answered. For instance, as indicated in section 1.8.3 only one falcon species (Peregrine falcon) has an accurate cytogenetic map (Damas *et al.* 2017) that is anchored to the recently established genome assembly. Without a series of falcon genomes reconstructed to the level of the chromosome it is difficult to determine, in detail, the path of gross genomic (chromosome) change that occurred during the evolution of this group. We know, from basic cytogenetic and zoo-FISH data that the overall falcon genome organization arose, in part, through a series of chromosomal fusions atypical for birds. As indicated in section 1.2.4, however, we do not know if the telomeric motifs that existed before the chromosomes fused in the falcon lineage are still retained on the chromosome and, if so, whether retention of interstitial telomeric repeats is a general feature of species that have undergone chromosomal fusion during evolution. Finally, while we know that chromosomes that were previously microchromosomes (before they fused to larger chromosomes in the falcon lineage) retain many genomic features (e.g. CG content, recombination rates, CpG islands) that characterize microchromosomes (see section 1.4.2) we do not know if these “former microchromosomes” adopt the same central nuclear position (see section 1.2.1) now that they are attached to larger chromosomes.

The purpose of this thesis was therefore first to assemble the genome of the other falcon species with (Saker falcon) to chromosome level, while generating robust overall genome structure references (karyotypes/ideograms) for this and a third species Gyr falcon. With a chromosomally anchored genome assembly (Saker) and an order of syntenic blocks (Gyr falcon – the genome sequence is not sufficiently assembled to perform a chromosome level assembly in this species), it was possible to address the above questions. With this in mind, the aims of this thesis were as follows:

Specific Aim 1: To define the overall structure (karyotype) of the Saker falcon genome and make hitherto undiscovered links between the genome assembly and the karyotype, thereby generating a cytogenetic genome map. This specific aim is covered in chapter 3 of the thesis where karyotypes of all three species (Peregrine falcon, Saker falcon, Gyrfalcon) followed by standard ideograms of the same. For Saker, the cytogenetic map is anchored to the genome assembly.

Specific Aim 2: To perform a comparative genomic study of genome evolution in falcons making use of BAC FISH probes that map inter- and intra-chromosomal changes to map the path of chromosome evolution in *Falco* species. This specific aim is covered in chapter 4 of the the thesis. Using the information generated in the previous chapter, cytogenetic and bioinformatics approaches are used to map the chromosomal changes from the avian ancestor to the three species of interest. In addition the hypothesis that gross falcon evolution involved both significant inter- and intra chromosomal arrangement is tested.

Specific Aim 3: To map telomeric sequences in three falcon species to test the hypothesis that remnants of former chromosomal fusions retain their telomeric motifs (i.e. appear as interstitial telomeres) and compare with other groups that have undergone chromosomal fusions. This specific aim is covered in chapter 5 of the thesis. Here the presence or absence of telomeric motifs is assessed by fluorescence in-situ hybridization (FISH) in the three-falcon species, as well as others known to have undergone chromosomal fusion (including both parrots and crocodiles).

Specific Aim 4: To test the hypothesis that chromosomes that were formerly microchromosomes (but are now fused to larger chromosomes) still “behave” as though they were microchromosomes in terms of their nuclear organization by adopting a central location. The hypothesis tested in this specific aim is covered in chapter 6 of the thesis using standard approaches for measuring the relative nuclear position of each chromosome or chromosome segment. 3D extrapolations of 2D data is performed.

S. Joseph

In each chapter, specific issues not covered in the general introduction are included. The thesis, finally, concludes with a general discussion that brings together the individual themes (chapters) of the work as a whole.

2 Materials and Methods

2.1 Chromosome Preparation

2.1.1 Fibroblast Culture

Avian primary fibroblast cell cultures were prepared from avian tissue samples which included trachea, skin and early stage embryos. Falcon primary fibroblast cell cultures were established only from falcon skin samples. Sampling in this study was reviewed and approved by the Animal Ethic Committee of CVRL, and Ministry of Climate Change and Environment (MOCCA) UAE, according to the Ministerial Decree No. 384 of the year 2008 on the executive by-law of the Federal Law No. 16 of the year 2007 concerning Animal Welfare.

Establishing primary falcon fibroblast culture was an arduous process, mainly due to difficulties of obtaining sources of falcon skin biopsies, the standardization of processes to optimize cell survival and growth and nutritional requirement differences that exist among the species. A requirement for a large amount of chromosome preparation to carry out analyses like FISH mapping was an additional challenge. Gyrfalcon fibroblast culture was easy when compared to other falcon species while Saker fibroblast culture was labour-intensive. Several falcon biopsies were collected and processed to establish primary fibroblast cultures for performing all the analyses listed under section 2.

2.1.1.1 Biopsy Collection

Skin biopsies were collected after induction of anaesthesia by the veterinarian at the Dubai Falcon Hospital, UAE. The biopsy site was prepared using 70% alcohol and a biopsy was collected from the chest region using a 4mm diameter sterile biopsy punch. Samples were placed into Alpha MEM (Fisher) containing 1% Pen-Strep-L-Glutamine (Sigma) and were processed on the same day.

2.1.1.2 Media Preparation

Alpha MEM (Fisher) was the medium of choice to support the growth and maintenance of falcon fibroblast cell cultures and HyClone MEM Alpha Modification medium (GE Healthcare and Life Sciences) was used to grow reptile fibroblast cells. Both media were supplemented with 10-20% Foetal Bovine Serum (Biowest) and 1% Pen-Strep-L-Glutamine (Sigma) and the formulated media were stored at 4°C until needed. All laboratory procedures were conducted in a class II biological safety cabinet.

2.1.1.3 Primary Fibroblast Culture - Skin Tissue

Skin was the preferred tissue sample for the derivation of falcon primary cell cultures and was obtained through biopsies. Tissue samples were placed into a sterile petri dish containing approximately 1ml Hanks Balanced Salt Solution (HBSS) (Gibco) with 1% Pen-Strep Fungizone (Gibco). The tissue was cut into small pieces using a sterile scalpel with a rolling motion in order to avoid pieces with ragged edges. Macerated tissue was transferred to a sterile beaker, washed twice with 3 ml of HBSS and 5ml 1xTrypsin EDTA solution (Sigma) and stirred in a magnetic shaker at 37°C for 30 to 45 minutes. After trypsin digestion, the entire solution was transferred into a 15ml sterile tube. Cells were pelleted in a centrifuge at 2000 rpm for 10 minutes and the cell pellet was re-suspended in 5 ml of growth medium in T25 cm² flask (TPP). Flasks were incubated at 40°C (30°C for reptile cells) under 5% CO₂.

2.1.1.4 Primary Fibroblast Cell Culture - Embryonic Tissue

Under aseptic conditions, embryonated chicken eggs were cleaned with 70% ethanol. The egg was opened by cutting along the air sac, taking the shell off and removing the embryo carefully using sterile forceps. Embryos were placed in a sterile petridish, the limbs and neck were removed and minced into small pieces using sterile scissors. Minced tissues were washed twice with sterile 0.85% NaCl. Approximately 10ml of 1xTrypsin EDTA solution (Sigma) was added to the minced tissues which were stirred on a magnetic stirrer at 37°C for 15 minutes. The cell suspension was filtered through a sterile glass funnel lined with sterile gauze into sterile centrifuge tubes, then centrifuged at 2000 rpm for 5 minutes. The supernatant was decanted, and the cell pellet was washed once with

S. Joseph

sterile 0.85% NaCl and then centrifuged. The cell pellet was re-suspended in complete growth medium.

2.1.1.5 Refreshing prepared

Spent medium was removed from the flask with a sterile pipette and fresh growth medium (as described in section 2.1.1.2) was added to a total volume of 5ml in a T25 cm² flask (TPP) and 10ml in a T75 cm² flask (TPP). Fresh growth medium was added to flasks every other day till the cell monolayer reached confluence.

2.1.1.6 Passaging

Cells were passaged when they reached 100% confluence. Growth medium was discarded from the flask and the cell monolayer was rinsed with 1ml HBSS (2ml for T75 cm²) by gentle rocking of the flasks, and then the solution was aspirated and discarded. 1ml (2ml for T75 cm²) pre-warmed 0.05% Trypsin-EDTA (Sigma) solution was added to the flask to coat the cells and the flask was placed in the incubator at 37°C for 1-2 minutes to facilitate enzymatic detachment of cells. Flasks were checked under the microscope for cell dissociation and the flasks were tapped firmly on the side to dislodge any attached cells. Growth medium (4.5 ml) was added to each flask to re-suspend the cells which were then transferred into a sterile 15ml conical tube and centrifuged at 2000 rpm for 10 minutes. The cell pellet was re-suspended in a minimum amount of complete growth medium and a cell count was determined using a Neubauer chamber. The cell suspension was then diluted in complete growth medium with a cell seeding density of 3×10^3 cells/ml. Five ml of cell suspension was transferred into each T25 cm² flask (10ml for T75 cm² flask) and kept at 37°C and 5% CO₂ in a humidified incubator.

2.1.2 Chromosome Harvesting

Cells at 90-100% confluency were the main selection criteria used for chromosome harvesting when they were not required for passaging. Colcemid (50ul: Gibco) at a concentration of 10µg/ml was added to each T25 cm² flask (100µl added to a T75 cm²) and the cells were incubated at 37°C for 1 hour in a 5% CO₂ incubator. Media from the flasks were then transferred into a 15ml falcon tube (cell culture supernatant) and the

cells were rinsed with 1ml HBSS, which was then transferred back into the same falcon tube. One ml of 0.25% Trypsin-EDTA (Gibco) was added to the flask which was then incubated at 37°C for 2 minutes. After examining the cells using an inverted microscope and when the cells were round and detached from the flask, the cells were re-suspended in cell culture supernatant. The cell suspensions were transferred again into the 15ml falcon tube. Samples were centrifuged for 10 minutes at 1,000 rpm, the supernatant was discarded, leaving approximately 0.5ml in the tube and the cell pellet was re-suspended in the remaining fluid. Cells were exposed to a hypotonic solution (5ml pre-warmed 0.075M KCl added drop-wise followed by incubation at 37°C for 20 minutes) to induce swelling. Three drops of fixative (3:1 methanol: acetic acid) were added to the cells with gentle agitation and the solution was centrifuged 10 minutes at 1,000rpm. The supernatant was discarded and the pellet re-suspended using a Pasteur pipette. The re-suspended pellet was drawn up into a Pasteur pipette, 5ml of fixative was then added to the tube and the cell suspension was dropped gently into the fixative from a distance of approximately 2 inches. The tube was centrifuged for 10 minutes at 1,000 rpm, the fixative process repeated 3 times and the samples were then stored at -20°C in 5ml of fixative.

2.1.3 Blood Culture and DNA Extraction

2.1.3.1 *Peripheral Lymphocyte Culture*

Lymphocyte culture medium was formulated with the following components: 217.5ml of RPMI 1640 medium (Sigma), 25ml Chicken serum (Invitrogen), 5ml Penicillin-Streptomycin (Invitrogen), 2.5ml L-Glutamine (Invitrogen) and 25mg Concanavalin A type IV (Sigma). A density gradient medium, Histopaque (Sigma), was used to separate viable lymphocytes from the whole blood. For this 3ml Histopaque was dispensed into a sterile 15ml centrifuge tube (BD Falcon) and 2-3ml of whole blood, collected in heparinized blood tubes (BD vacutainer) was gently layered over the top. Tubes with blood samples were centrifuged at 400 x g for 30 minutes at room temperature. After centrifugation, the opaque interface was collected containing mononuclear lymphocytes and transferred into a sterile 15ml falcon tube containing 10ml of PBS where they were mixed gently. The cells were pelleted at 250 x g for 10 minutes and

gently aspirated out of the supernatant. Cell pellets were washed with 5ml of PBS, mixed gently and centrifuged at 250 x g for 10 minutes. The supernatant was decanted, and the cell pellet was re-suspended in 10ml of lymphocyte culture medium which was removed from the T25 cm² tissue culture flask. Flasks were periodically gently shaken during their incubation for 72 hours at 40°C under 5% CO₂ in air.

Colcemid (50ul at 5µg/ml) was added to flasks which were then incubated for 1 hour at 40°C under 5% CO₂. The cell suspension was then transferred into a 15ml falcon tube and centrifuged at 400 x g for 5 minutes. The supernatant was discarded and 6ml of pre-warmed KCl (0.075M) solution at 37°C was added drop wise while gently flicking the tube. Cells in KCl solution were allowed to swell at 37°C for 15 minutes. Cells were fixed by the addition of 3 methanol: 1 acetic acid which was added drop wise to a final volume of 14ml while inverting the tube to mix the reagents. Cells were pelleted by centrifugation at 400 x g for 5 minutes, the supernatant was discarded, and the cells re-suspended by flicking the tube to avoid large cell clumps. The cell pellet was washed 3 times in fresh fixative and samples were stored in 5ml of fresh fixative at -20°C.

2.1.3.2 Extraction of Genomic DNA

DNA was extracted from skin tissue samples and EDTA blood from different birds. DNA extraction was carried out using a Qiagen DNEasy Blood and Tissue kit (Qiagen) following the manufacturer's guidelines. However, an additional final elution step to increase the DNA yield was included, where the eluted DNA was run through the same spin column followed by a final centrifugation step.

2.2 Generation of Labelled FISH Probes

2.2.1 Selection of BAC clones

2.2.1.1 BAC Selection – Subtelomeric Avian BACs

BACs in the subtelomeric part of the p-arm and q-arm of each chicken chromosome were selected using the *Gallus gallus* Version 2.1 NCBI database for Chicken BACs and the *Taeniopygia guttata* Version 3.2.4 NCBI database for Zebra Finch BACs

(www.ncbi.nlm.nih.gov). They were purchased from CHORI-261 Chicken BAC library (BACPAC) and the Zebra Finch TGMCBa library (Wageningen).

2.2.1.2 BAC Selection – Selected Avian BACs

BAC selection for cross-species FISH testing was achieved with the collaboration of members of the Larkin laboratory at the Royal Veterinary College, London, UK. To increase the working efficiency of BACs across a diverse range of species, the following specific criteria were selected, which include the proportion of conserved sequence between species, the number of repetitive elements and the GC content of the BAC. The entire process for detection of appropriate BACs consisted of the following steps: a) creation of multi-species alignments; b) identification 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 conducted to recognize 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 which has to be tested on the metaphases of different species in order to check whether the criteria are satisfied; g) building a statistical model based on the results of the training set which has been tested on metaphases of many other avian species. These sets of cross species BACs were successfully hybridized across many avian species including the Peregrine falcon (Damas *et al.* 2017).

2.2.2 Preparation of BAC Clones

Selected Avian BAC clones were provided by Griffin lab, University of Kent, Canterbury, UK for the FISH analysis. Briefly, BACs were cultured in Luria Bertani Agar (LB Agar) and the clone DNA was extracted by QIAprep Spin Miniprep Kit (Qiagen). BACs were labelled by nick translation using FITC and Texas Red.

2.3 Fluorescence *in situ* Hybridisation (FISH)

2.3.1 Metaphase Slide Preparation and Hybridisation

Metaphase cell preparations stored at -20°C were centrifuged at 1,000 rpm for 10 minutes to remove the fixative which was replaced with freshly prepared fixative (3:1 methanol: acetic acid) that was added in small quantity to result in an appropriate cell density to give a reasonable number of metaphases in each microscopic field. Alcohol-cleaned slides were placed over a Coplin jar containing hot water to attain slight humidity for a better cell spread. Chromosome suspension (10µl) was dropped on each half of the slide and allowed to air dry thoroughly. Slides were washed in 2xSSC (Saline-Sodium Citrate) (Gibco) for 2 minutes, dehydrated by serial ethanol washing for 2 minutes each in 70% (v/v), 85% and 100% ethanol and left to air dry. Probe mix was prepared in the following way; 1.5µl of FITC labelled probe, 1.5µl of Texas Red labelled probe, 1µl of chicken Hybloc (Applied Genetics Laboratories), 6µl of Hyb I (Cytocell) hybridisation buffer to a total volume of 10 µl and a probe concentration of 10ng/µl. Air dried slides and 22x22mm coverslips according to the number of samples to be tested were kept on the 37°C hotplate. Probe mix (10µl) was pipetted onto the coverslip and carefully placed on the chromosome spread. Coverslips were sealed to slides with Fixogum Rubber Cement and the probes and the target DNA were denatured by keeping the sealed slides on a 75°C hotplate for 2 minutes. Slides were incubated in a hybridisation chamber at 37°C for 72 hours, allowing the proper binding of probe DNA to the target DNA. Slides were then removed, the coverslips discarded and washed with 2xSSC with 0.05% of Tween 20 (Sigma-Aldrich) at room temperature for 30 seconds which minimised the loss of weakly bound probes. When the slides were still wet, 1-2 drops of Vectashield Antifade Mounting Medium with DAPI (Vectorlab) was added on each half of the slide and a 24x50mm coverslip affixed.

2.3.2 Microscopy

Metaphase images captured using an Olympus BX-61 epifluorescence microscope equipped with a cooled CCD camera and SmartCapture 3 software (Digital Scientific UK). Three different filters were principally used to acquire images - DAPI, fluorescein isothiocyanate and Texas Red filters.

2.3.3 Image Analysis - Karyotyping

For karyotyping to result in a well-defined band on the chromosomes, the following staining procedures were used. Propidium iodide solution (Sigma Aldrich) with a strength of 1.0mg/ml was further diluted to 2% in water and 6µl of diluted propidium iodide was mixed with 200µl of Vectashield antifade mounting medium with DAPI to result in a final concentration of 0.6µg/ml of propidium iodide. Avian metaphase chromosomes which had been fixed in a solution of 3:1 methanol: acetic acid were dropped onto clean slides as described in section 2.3.1. One drop of prepared DAPI/propidium iodide stain was added to the metaphase spread once it was dry. A cover slip was placed, and the images were captured in the Olympus BX61 epifluorescence microscope with a cooled CCD camera and with Smart Capture 3 (Digital Scientific, UK) software. Karyotyping was performed using the SmartType (Digital Scientific, UK) system. Fifty metaphase images were captured for each avian species to create a standard karyotype.

2.4 Physical Genome Mapping

The first procedure for physical genome mapping is to create predicted chromosome fragments (PCFs). This was performed by the Larkin lab at the Royal Veterinary College, London (RVC) followed by the FISH mapping at the University of Kent, UK. The following joint efforts by two universities to carry out physical mapping included: 1) With the help of the RACA algorithm, predicted chromosome fragments (PCFs) were constructed by orienting sequence scaffolds generated by NGS; 2) PCFs were confirmed by PCR and computational verification; 3) A precise set of PCFs was created, after being thoroughly verified; 4) Zoo-FISH was performed to assemble full-length chromosomes using a 'universal set' of BACs which have the ability to hybridize properly across the whole genome of phylogenetically divergent species. I would like to express my thanks to Joana Damas, Dr. Marta Farré and Dr. Denis Larkin for performing the bioinformatics part of the project.

2.4.1 Construction of PCFs using the RACA Algorithm (RVC)

Saker falcon genome sequencing was carried out using Illumina deep sequencing (Zhan *et al.* 2013). Nearly predicted chromosome fragments (PCFs) were constructed using RACA (Kim *et al.* 2013) from the fragmented Saker falcon (*Falco cherrug*) genome assemblies, which were generated using Illumina deep sequencing with greater than 100-fold coverage (Zhan *et al.* 2013). The zebra finch chromosome level genome assembly and the chicken genome assembly were used as closely related references (divergence 62 Ma and 96 mya respectively). RACA constructed 103 PCFs with an N50 of 22.27 Mb using default RACA parameters. A total number of 458 scaffolds were used to generate PCFs which is 97.26% of the total assembly.

2.4.2 Verification of PCFs (RVC)

RACA produced splits in Saker falcon genome scaffolds and in order to verify the structures, polymerase chain reaction (PCR) amplification was performed across the split regions less than 6kb in the target species scaffolds. The resulted split regions with negative PCR were tested with an alternative (RACA- suggested) order of the flanking syntenic fragments (SFs). Remaining split regions greater than 6kb in size were further analyzed to identify chimeric scaffolds; a total of 25 (5%) were identified.

2.4.3 Creation of a Refined Set of Saker PCFs (RVC)

A refined set of PCFs was created using adjusted coverage thresholds. Also, 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 further producing a total of 103 PCFs with N50 22.27 Mb for the Saker falcon covering 97.26% of the original assembly.

2.5 Ancestral Karyotype Reconstruction

2.5.1 Neognathae Ancestor

The following avian genomes were selected to reconstruct the hypothetical Neognathae ancestor: chicken (*Gallus gallus*), pigeon (*Columba livia*), budgerigar (*Melopsittacus undulatus*), Saker falcon (*Falco cherrug*) and Peregrine falcon (*Falco peregrinus*). The ostrich (*Struthio camelus*) was included in this analysis as an outgroup. The 92 BAC clones representing 24 chicken chromosomes and mapped by FISH to the above 6 bird genomes

were selected for the ancestor reconstruction using the Multiple Genomes Rearrangements and Ancestors tool version 2 (MGRA2) (Avdeyev et al. 2016; Alekseyev and Pevzner 2009). Species specific arrangements of chromosomal segments were generated to serve as MGRA2 inputs for individual genomes. From these data, a series of contiguous ancestral regions (CARs) were produced representing the most likely ancestral configuration for the Neognathae ancestor.

2.5.2 Genome Rearrangement Analysis

To reconstruct the chromosomal changes that occurred between the hypothetical Neognathae ancestor and the set of 6 extant species, 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 and Pevzner 2002; <http://grimm.ucsd.edu/>). MGRA2 outputs served as MGR/GRIMM inputs to trace the most parsimonious scenarios for evolutionary changes including the intra- and inter-chromosomal rearrangements that might have occurred from the hypothetical Neognathae ancestor to the 6-extant species.

I am grateful to Dr. Mike Romanov for running the MGRA2 and GRIMM analysis.

2.6 Telomere sequence mapping

Analysis to determine telomere distribution on chromosomes was carried out using a Telomere PNA (Protein Nucleic Acid) FISH kit (DakoCytomation). FISH for detection of telomere sequence was performed according to the manufacturer's instructions. In brief, a metaphase spread was prepared as described in section 2.3.1 and treated with the fluorescein - conjugated peptide nucleic acid (PNA) probe. The metaphase area was covered with a coverslip and the sample DNA was denatured at 80°C for 5 minutes. Hybridization was carried out in the dark at room temperature for 2hrs. Hybridization was followed by a brief wash with a Rinse solution and a post-hybridization wash with a Wash solution at 65°C for 5 minutes. Slides were counterstained with Vectashield

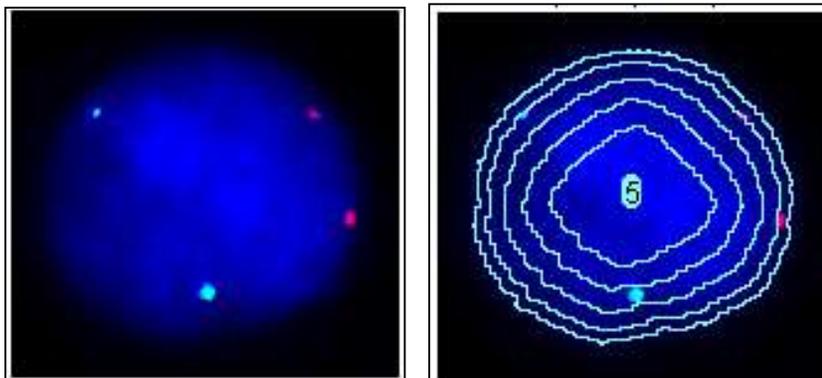
antifade medium with DAPI (Vector Labs, USA). FITC and DAPI images were captured at 100x magnification with an Olympus BX-61 epifluorescence microscope equipped with a cooled CCD camera and appropriate filters. A positive reaction is recognized as a green fluorescence signal at the site of hybridization.

2.7 Determination of Nuclear Position of Chromosome Territories

In order to be able to assess nuclear position of individual BACs from micro, macro and “former micro” – chromosomes, a minimum of 100 nuclei were captured for each, leading to up to 200 signals analysed per probe. The relative nuclear position of chromosomes within the nucleus was measured using an automated method adapted from an approach originally published by Croft *et al.* (1999) as follows:

A macro written for ImageJ (Michael Ellis, Digital Scientific UK) and adapted by Dr. Ben Skinner (ImageJ plugin Nuclear Morphology Analysis version 1.13.5) (https://bitbucket.org/bmskinner/nuclear_morphology/wiki/Home) was used for analysis of relative chromosome positioning. Essentially, this divides each image of a nucleus to separate RGB planes (red and green for FISH signals and blue for DAPI counterstain) and then converts the blue image to a binary mask from which five concentric regions of interest (rings) of equal area are created (Skinner 2009) thus:

Figure 2-1: A captured nucleus image with converted to RGB planes before the application of the macro (left) and after (right) with 5 rings of equal area formed.



The proportion of signal contained within each ring was calculated with reference to the binary mask (Skinner 2009). The output of these results was converted to an excel spreadsheet where the relative position of each signal (or part thereof) is entered. To compensate for the fact that an object that was essentially 3D (roughly spherical) is observed in 2D (after nucleus flattening) the proportion of signal within each shell was normalised against DAPI density (Boyle *et al.* 2001). In other words, a signal in shell 1 (outermost) obtains a relatively higher “score” than one in shell 5 (innermost) as the former can only be peripherally located, whereas the latter could be either centrally located, or peripherally located and subsequently “rolled over” in the flattening process. Normalised versions of each signal were created and an ‘overall’ position for each probe in each nucleus image (Skinner 2009). The median value of the overall positions for all nuclei with a specific probe was taken as the overall position for the probe (Skinner 2009). Since data appeared to be non-normally distributed thus it was non-parametric, median and interquartile ranges were calculated rather than standard error of the mean.

The final output is therefore a histogram of 5 bars from which we asked whether the distribution was non-random or not distinguishable from a random pattern. This was calculated using a χ^2 test (non-random when $p < 0.05$) according to Croft *et al.* (1999) and Boyle *et al.* (2001). Following the strategy established by the two above studies, histograms found to represent non-random distributions were examined by visual inspection to establish which shell (1-5) was predominantly represented. Thus, a histogram in which the highest bar was for shell 1 was representative of a peripheral distribution, if shell 5, a central distribution and shells 2-4 something between these two extremes.

3 Specific aim 1: To define the overall structure (karyotype) of the Saker falcon genome and make hitherto undiscovered links between the genome assembly and the karyotype, thereby generating a cytogenetic genome map.

3.1 Background

As indicated in 1.7.1, the karyotype of each species is important as it represents essentially a low-resolution map of the genome of an organism. Karyotype analysis demonstrates the distinctive features of the chromosomes such as their number, size, symmetry, position of the centromeres and banding patterns. Any deviations from the norm can be indicative of genetic disease including infertility. Size, position of the centromeres and banding patterns are usually represented on an ideogram, which is a pictorial representation of chromosome relative size, centromere position and banding patterns (Mirzaghaderi and Marzangi 2015).

As described in section 1.8, karyotype studies in different species within the genus *Falco* have demonstrated that this group has a significantly lower chromosome number and fewer microchromosomes than the usual bird chromosome pattern. To date, however, there have been very few studies that have created high quality banded chromosomes of falcons and represented them as ideograms (Al Mutery, 2011– Gyrfalcon and Damas *et al.* 2017 – Peregrine falcon). Moreover, there has been no such study at all in Saker falcon and no karyotypic comparison between falcon species.

A recurring theme of this thesis is to reiterate the ultimate need to have a genome sequence which is assembled to whole chromosome level i.e. to generate a chromosomal length array of sequence for each chromosome (see section 1.3.9). As pointed out in section 1.8.3, the available Saker falcon genome sequence is in the form of sub chromosomal-sized “scaffolds” generated with NGS technology (Zhan *et al.* 2013). Such highly fragmented genome assemblies can severely limit the discovery of crucial features of chromosomal evolution (Kim *et al.* 2013). Bioinformatic approaches

such as RACA (Kim *et al.* 2013) can predict chromosome-sized fragments (PCF) from the sub-chromosomal sized “scaffolds” by constructing synteny fragments from *de novo* sequenced target genome (in scaffolds), a reference genome and one or more outgroup genomes. However, PCFs require further genetic or physical mapping to anchor them precisely on the chromosome (Kim *et al.* 2013). A well described karyotype and ideogram is an essential initial step to achieve this.

The purpose of this chapter was therefore to describe for the first time the karyotype of the Saker falcon, improve on the karyotypes and ideograms recently produced for Gyr and Peregrine falcons, and make initial rudimentary comparative genomic analyses ahead of more molecular and sequence-based approaches. Like some other falcons, the Saker falcon is classified as ‘endangered’ according to the International Union for Conservation of Nature (IUCN) Red List of Threatened Species (IUCN 2017).

Once an ideogram has been established, the next obvious step is to increase the resolution of the comparative genomics through comparison with the genome structure of an existing species. Traditionally this has been done using chromosome painting probes (in avian studies, these are almost always derived from chicken - Griffin *et al.* 2007) applied to the metaphases of the species in question. Recently, however, a means of selected avian BACs likely to hybridize to the metaphases of a range of species was described (Damas *et al.* 2017). This was applied to pigeon and Peregrine falcon but, to date, no other species. In this chapter, therefore, the purpose was to provide a molecular cytogenetic map of the Saker falcon.

A final stage in the creation of a chromosome level assembly is to map the existing molecular cytogenetic markers to known scaffolds of a genome sequence. For the Saker falcon, a genome assembly with sufficiently sized scaffolds exists (Zhan *et al.* 2013). For the Gyr falcon, however, only a rudimentary genome assembly with small, numerous scaffolds exist (Al Mutery 2011).

3.2 Specific Aims

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

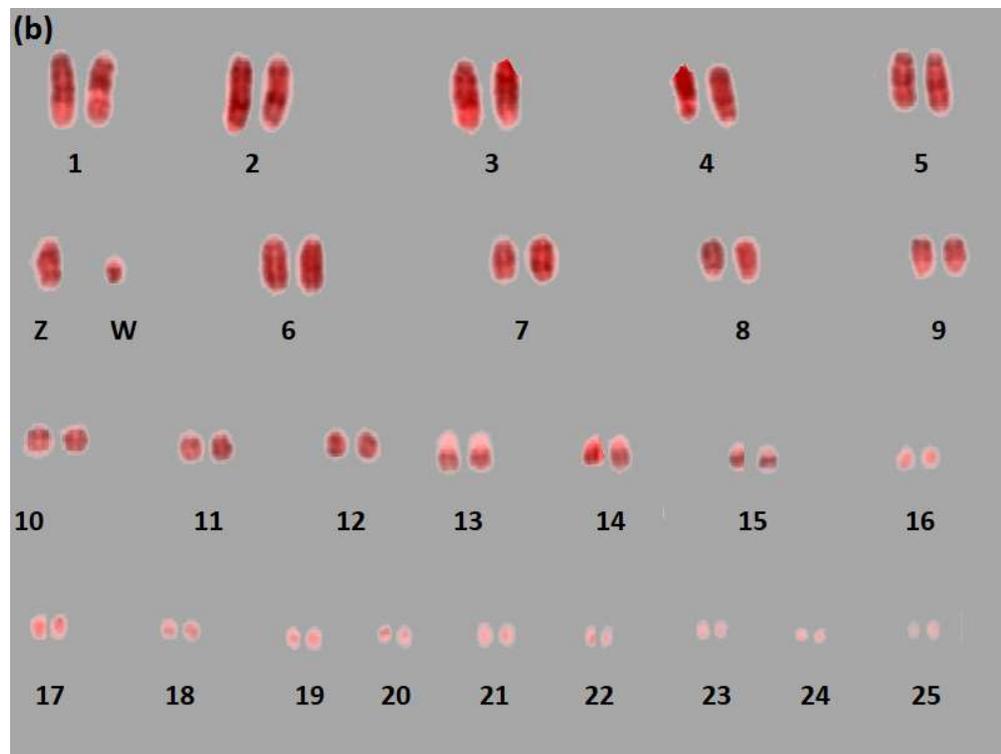
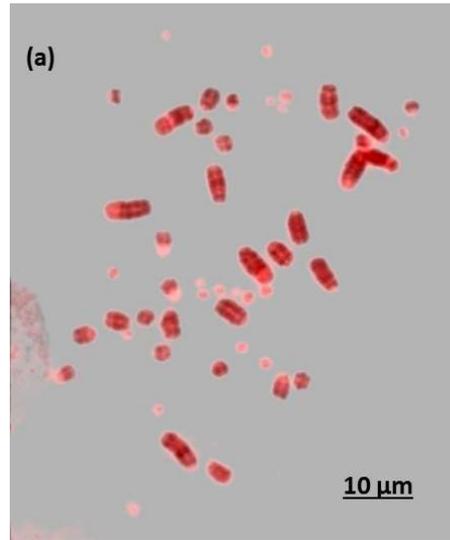
- **Specific aim 1a:** To make a high-resolution karyotype and standard ideogram of Saker falcon and to compare it with two other falcon species: Peregrine falcon and Gyrfalcon
- **Specific aim 1b:** To perform comparative FISH with known chromosome probes to define the overall genome structure of Saker and Gyrfalcon in comparison to previously assembled genome sequences such as Peregrine falcon and chicken
- **Specific aim 1c:** To map the molecular cytogenetic markers to existing scaffolds (predicted chromosome fragments–PCFs) of the Saker falcon genome assembly and make the data publicly available by an interactive website (Evolution Highway) designed to demonstrate cytogenetic comparisons between individual species

3.3 Results

3.3.1 Specific Aim 1a: To make a high-resolution karyotype and standard ideogram of Saker falcon and to compare it with two other falcons species: Peregrine falcon and Gyrfalcon

To determine diploid numbers and to make the karyotype, fifty different metaphase spreads from each falcon species were stained with a combination of DAPI and propidium iodide. One female Saker falcon was used for the karyotype preparation which is shown in (Figure 3-1). The chromosomes ($2n=52$) are displayed according to their relative size.

Figure 3-1: The karyotype ($2n = 52$) of Saker falcon. (a) DAPI and propidium iodide -stained metaphase spread (b) female karyotype.



All chromosomes appeared acrocentric, the karyotype consists of easily distinguishable 1-9 large chromosomes, 10-14 medium sized chromosomes, 14-24 microchromosomes, and a pair of sex chromosomes. For instance, chromosome 1 and 2 have a pale band at the base but the former chromosome has a broader pale band than the latter. Chromosome 4 has a distinct pale band in the center making it easily distinguishable. Chromosome 1- 16 are identifiable, but chromosome 17-25 are microchromosomes which are indistinguishable. Rudimentary measurement analysis allowed the construction of an ideogram for this species (Figure 3-2).

Figure 3-2: Standard ideogram of Saker falcon as devised in this study. Bands are divided into “light” (i.e. bright on propidium Iodide, pale on DAPI), “dark” (dark on propidium iodide, bright on DAPI) and “grey” (pale in both).

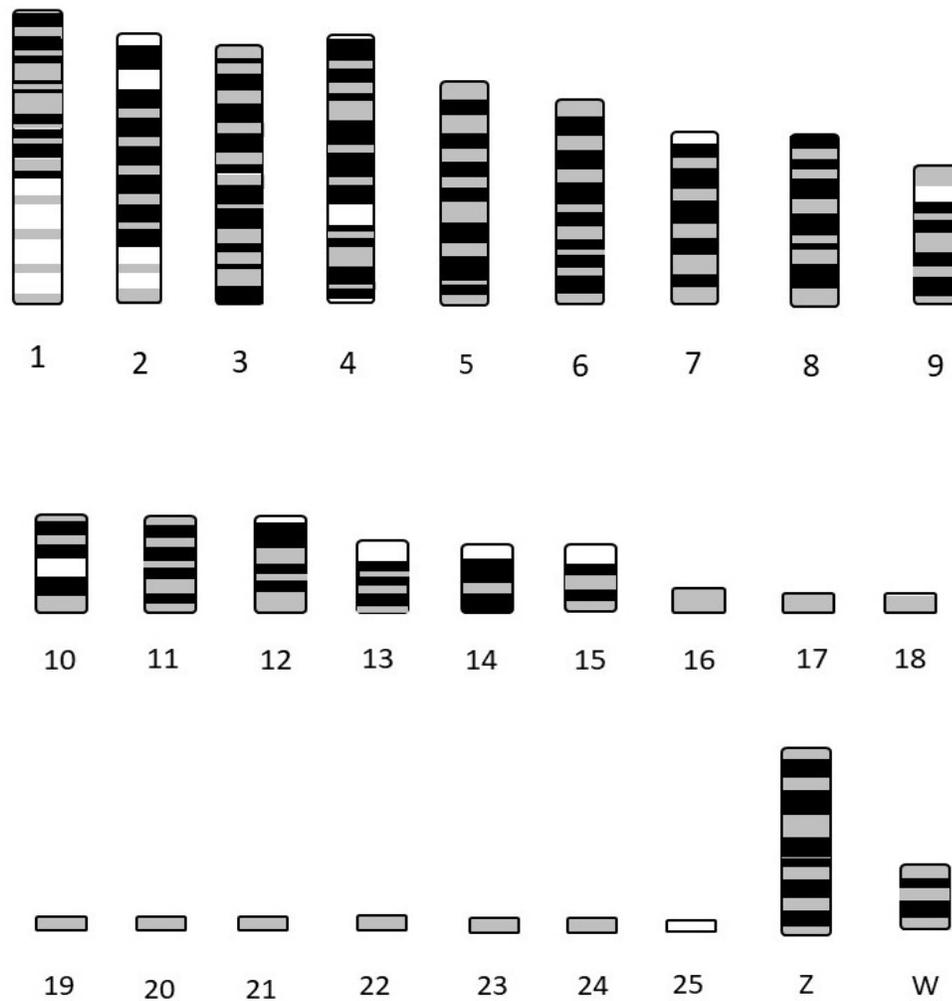
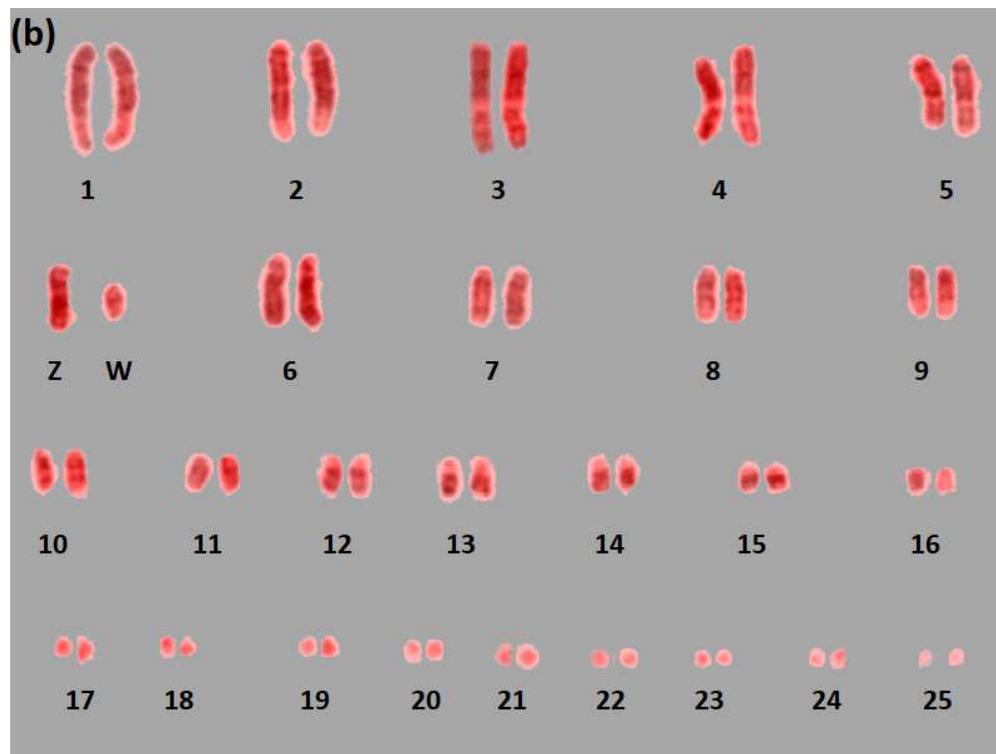
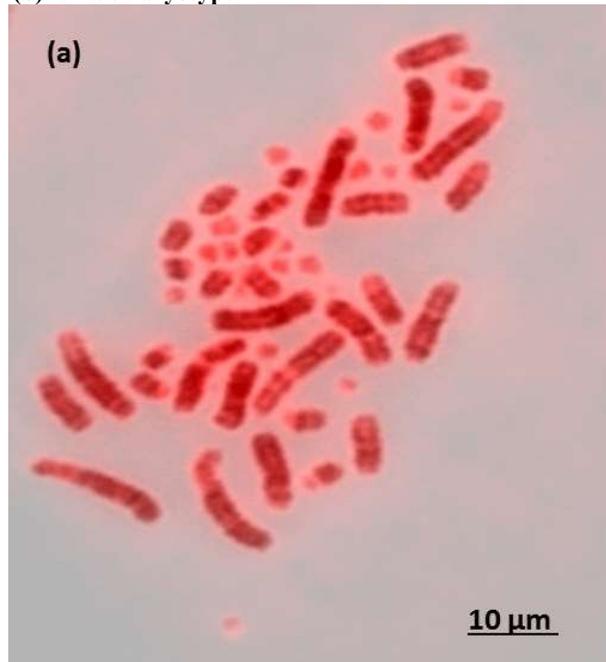


Figure 3-3: The karyotype ($2n = 52$) of Gyrfalcon. (a) DAPI and propidium iodide -stained metaphase spread (b) female karyotype.



One male and one female Gyrfalcon were analyzed for the karyotype preparation (Figure 3-3). The Gyrfalcon karyotype is similar to that of the Saker falcon with a diploid number of $2n=52$ (Figure 3-4).

Figure 3-4: Standard ideogram of Gyrfalcon. Bands are divided into “light” (i.e. bright on propidium Iodide, pale on DAPI), “dark” (dark on propidium iodide, bright on DAPI) and “grey” (pale in both).

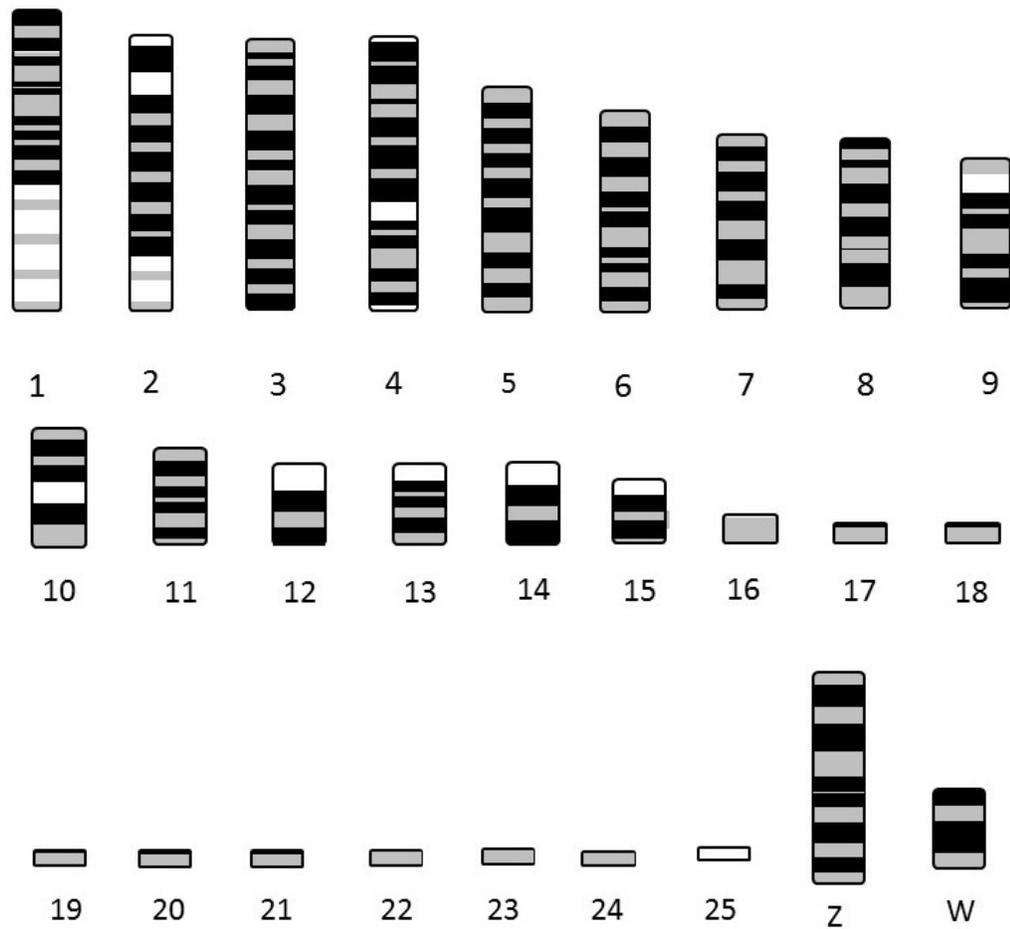
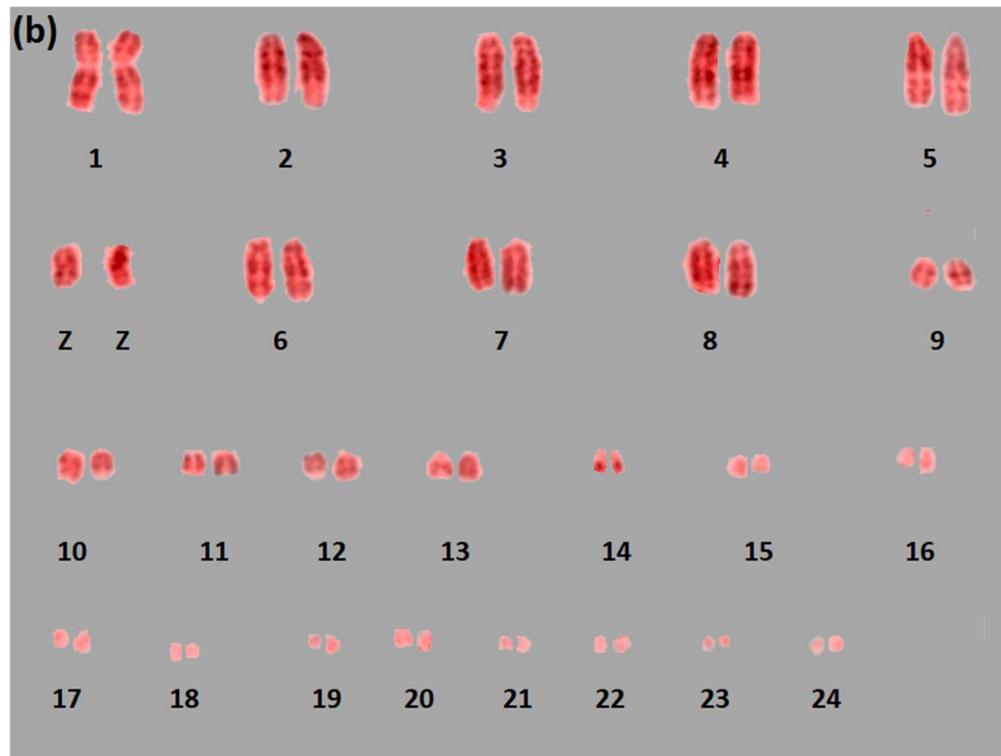
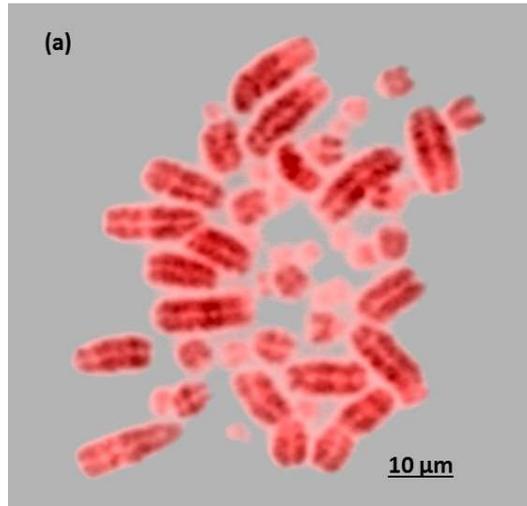
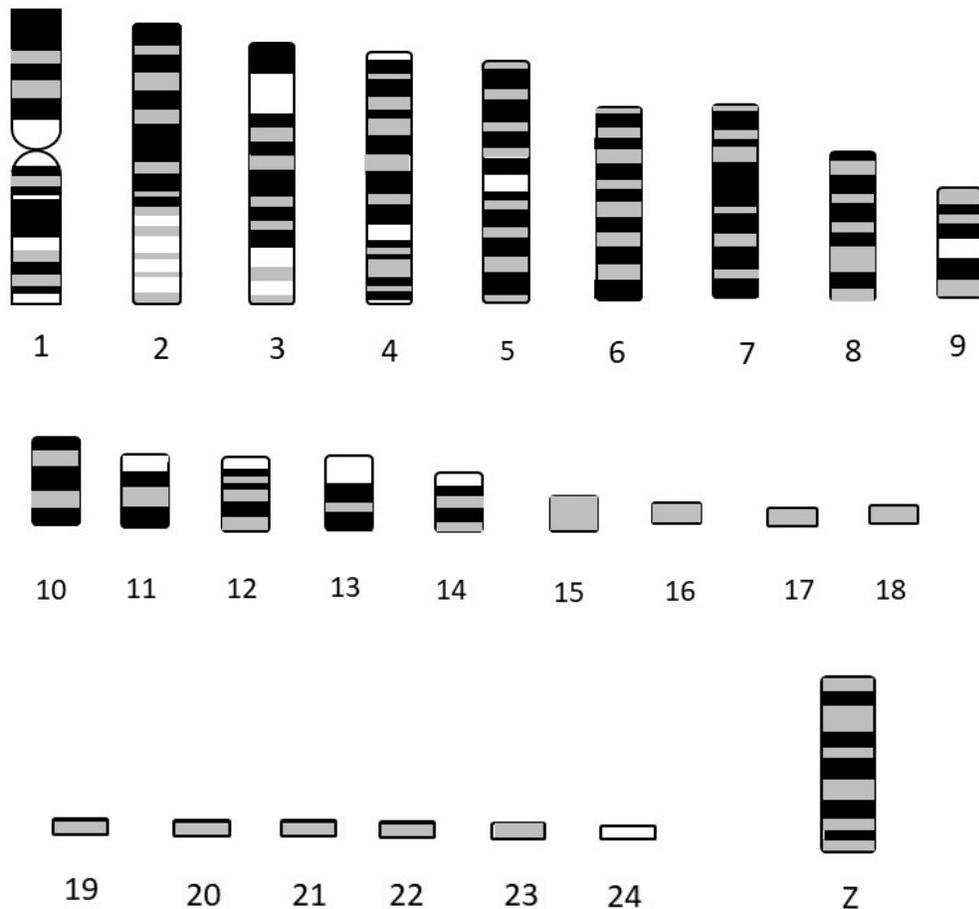


Figure 3-5: The karyotype ($2n = 50$) of Peregrine falcon. (a) DAPI and propidium iodide-stained metaphase spread (b) male karyotype.



In contrast to the Saker and Gyrfalcon, a second karyotype pattern was observed in the Peregrine falcon with a diploid number $2n=50$. The karyotype was analyzed using a male Peregrine falcon (Figure 3-5). Apart from the number of chromosomes, the main difference between both karyotypes is that the Peregrine falcon has a metacentric chromosome 1 with a pale distinct band on the p arm (Figure 3-6). Interestingly, Peregrine falcon chromosomes 2 to 6 are morphologically like chromosomes 1 to 5 of the Saker falcon and Gyrfalcon. However, Peregrine falcon chromosome 4 has a light band which is absent in the corresponding chromosome 3 of Saker and Gyrfalcon.

Figure 3-6: Standard ideogram of Peregrine falcon. Bands are divided into “light” (i.e. bright on propidium iodide, pale on DAPI), “dark” (dark on propidium iodide, bright on DAPI) and “grey” (pale in both).



3.3.2 Specific Aim 1b: To perform comparative FISH with known chromosome probes to define the overall genome structure of Saker and Gyrfalcon in comparison to previously assembled genome sequences such as Peregrine falcon and chicken

To construct a comparative cytogenetic map that would facilitate mapping of inter- and intra- chromosomal changes in Saker falcon, Gyrfalcon and in Peregrine falcon, a panel of 167 BACs were used and were mapped successfully on Saker and Gyrfalcons by FISH. Hybridization success of these BACs, which had been previously proved and mapped on Peregrine falcon, were utilized here (Damas *et al.* 2017). Out of 167 BACs, 129 were localized on macrochromosomes (1-11 and Z) with 38 BACs on microchromosomes. Figure 3-7 shows the standard ideogram of these falcons with the positions of the BACs mapped to their chromosomes.

Figure 3-7: Comparative map of the Saker falcon (FCH), Gyrfalcon (FRU) and Peregrine falcon (FPE) chromosomes with the position of all BACs (as calculated by the fractional length from the p-terminus of the chromosome). The red lines indicate inverted segments.

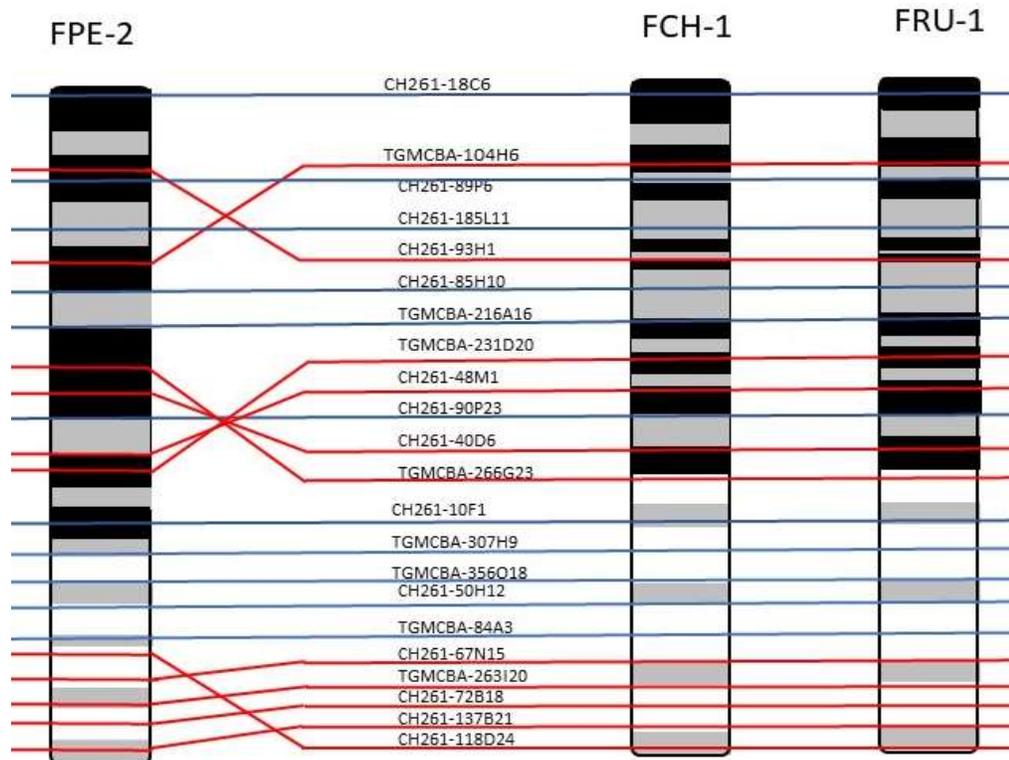


Figure 3-7 (Continued): Comparative map of the Saker falcon (FCH), Gyrfalcon (FRU) and Peregrine falcon (FPE) chromosomes with the position of all BACs (as calculated by the fractional length from the p-terminus of the chromosome). The red lines indicate inverted segments.

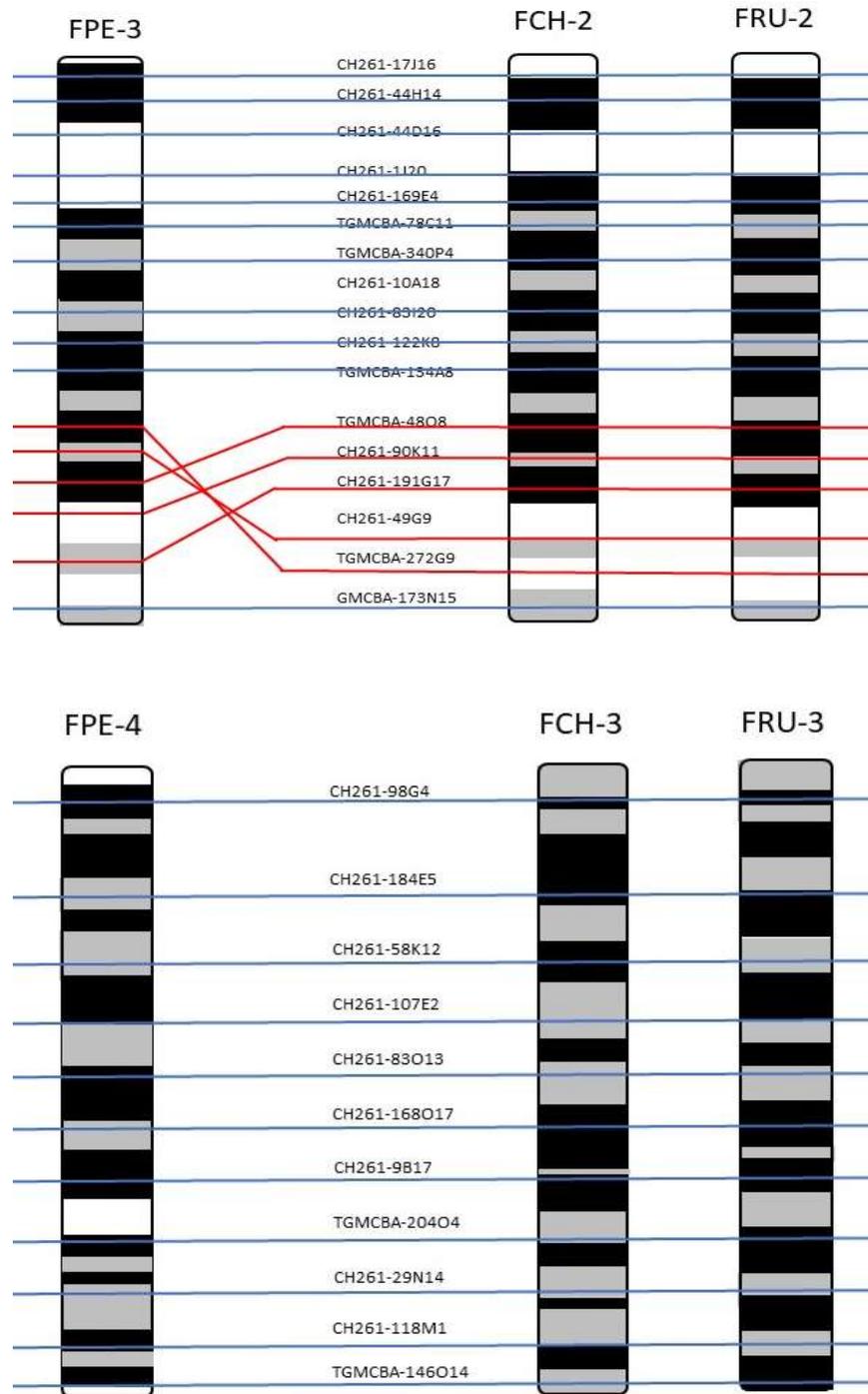


Figure 3-7 (Continued): Comparative map of the Saker falcon (FCH), Gyrfalcon (FRU) and Peregrine falcon (FPE) chromosomes with the position of all BACs (as calculated by the fractional length from the p-terminus of the chromosome). The red lines indicate inverted segments.

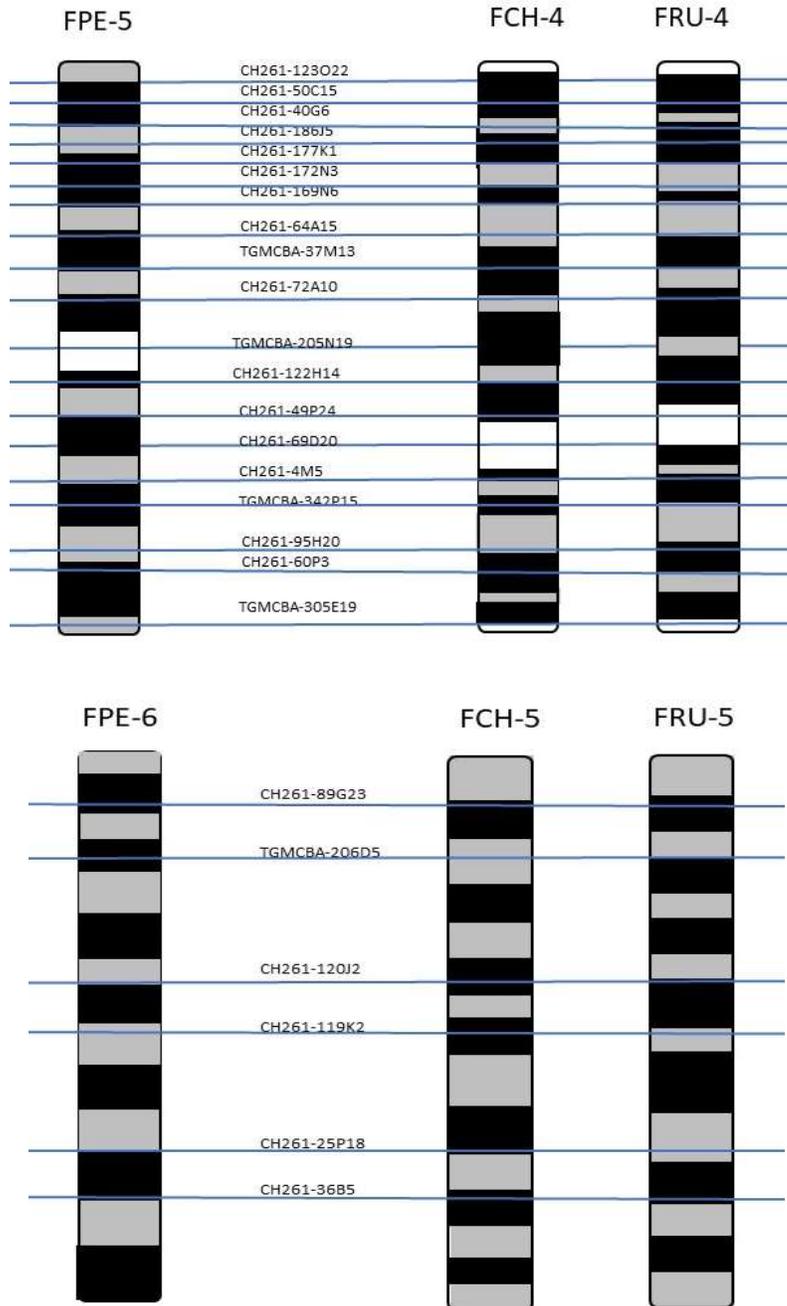


Figure 3-7 (Continued): Comparative map of the Saker falcon (FCH), Gyrfalcon (FRU) and Peregrine falcon (FPE) chromosomes with the position of all BACs (as calculated by the fractional length from the p-terminus of the chromosome). The red lines indicate inverted segments.

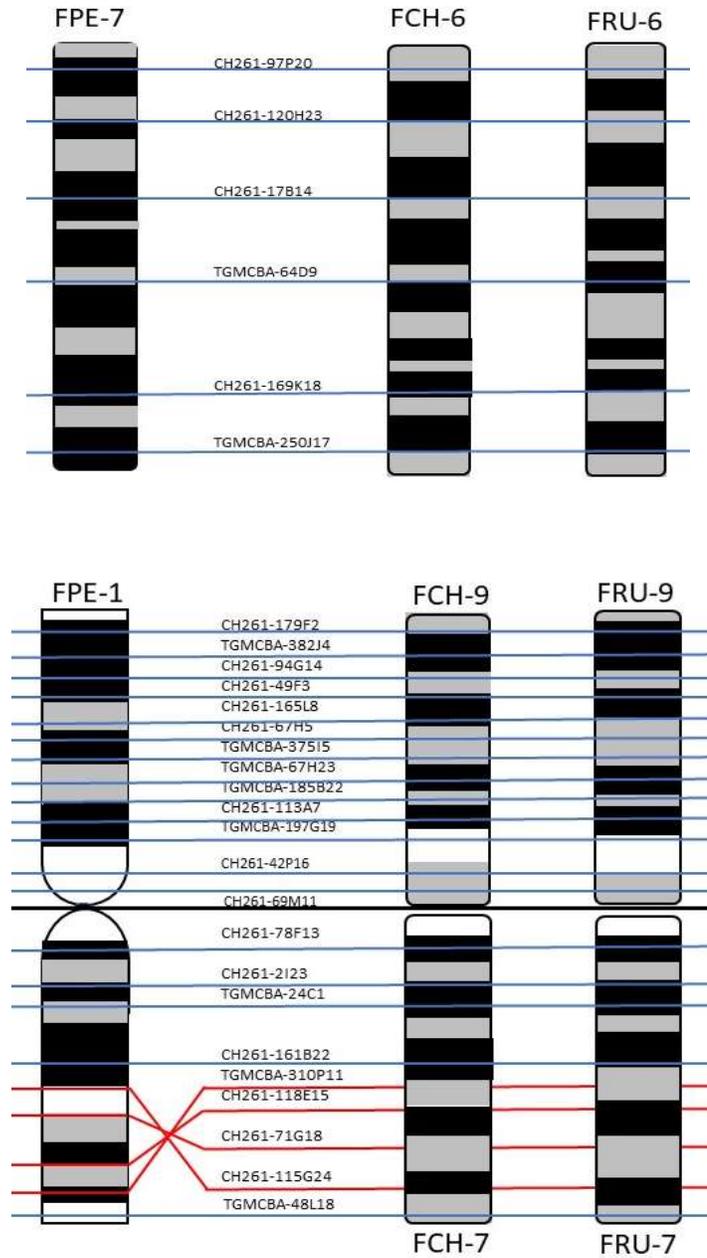


Figure 3-7 (Continued): Comparative map of the Saker falcon (FCH), Gyrfalcon (FRU) and Peregrine falcon (FPE) chromosomes with the position of all BACs (as calculated by the fractional length from the p-terminus of the chromosome). The red lines indicate inverted segments.

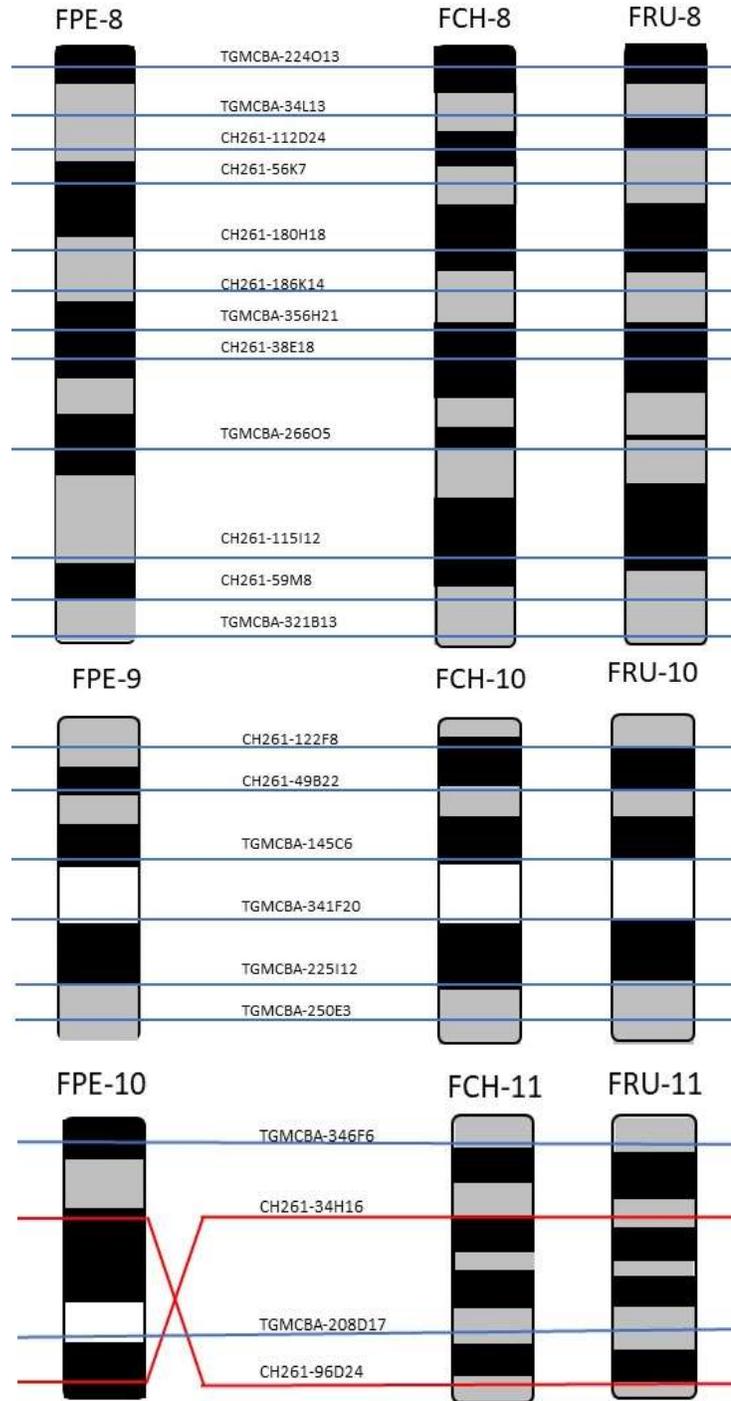


Figure 3-7 (Continued): Comparative map of the Saker falcon (FCH), Gyrfalcon (FRU) and Peregrine falcon (FPE) chromosomes with the position of all BACs (as calculated by the fractional length from the p-terminus of the chromosome). The red lines indicate inverted segments.

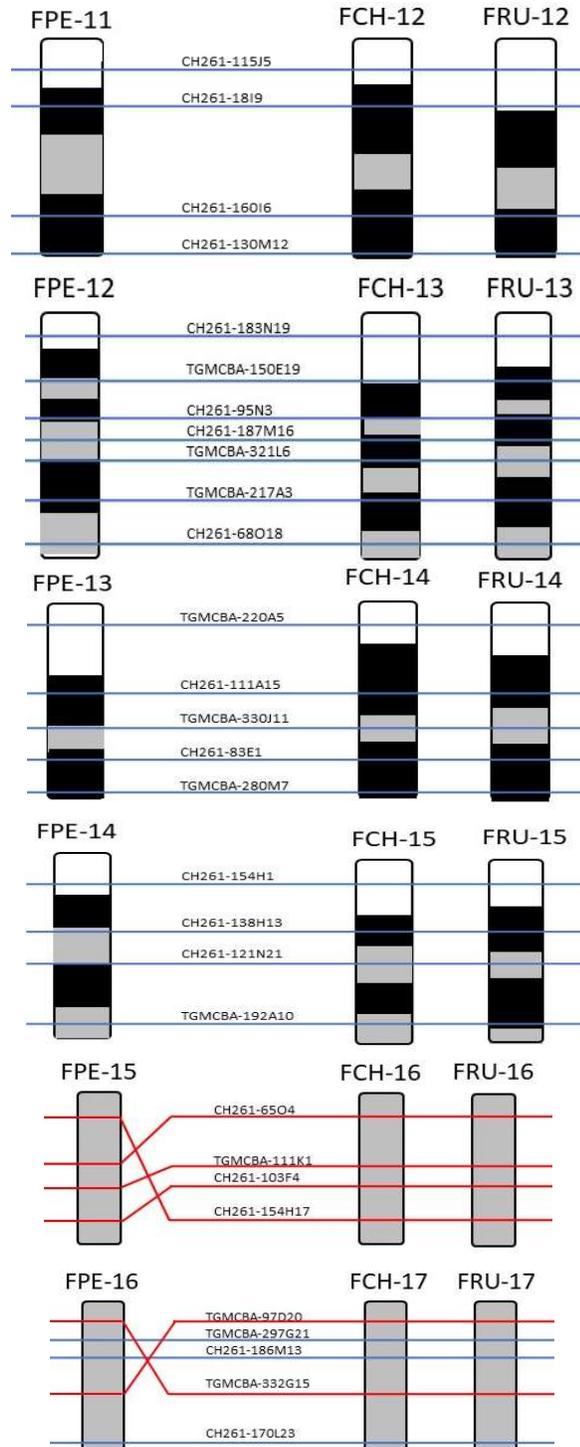
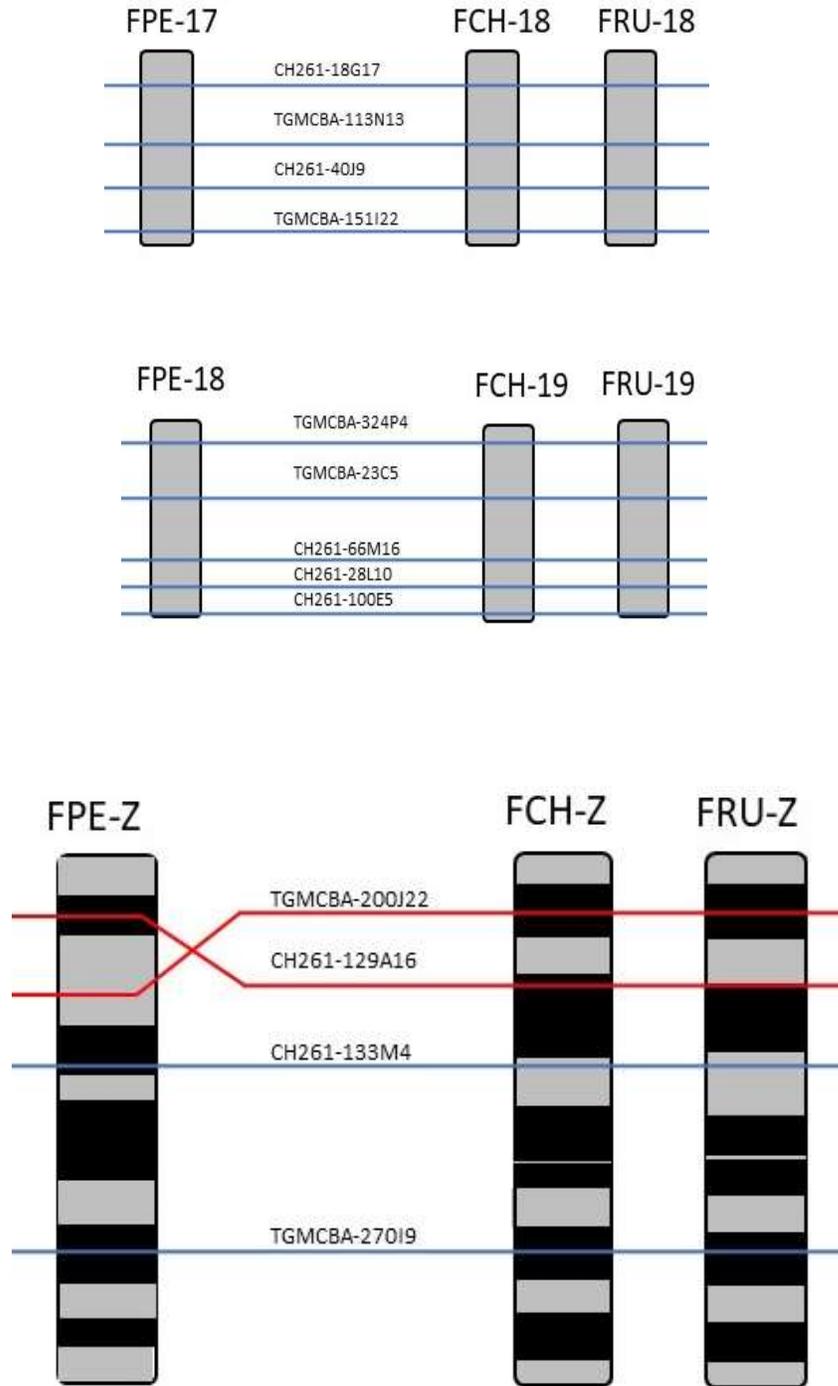


Figure 3-7 (Continued): Comparative map of the Saker falcon (FCH), Gyrfalcon (FRU) and Peregrine falcon (FPE) chromosomes with the position of all BACs (as calculated by the fractional length from the p-terminus of the chromosome). The red lines indicate inverted segments.



S. Joseph

In general, comparative data gives strong support for conserved synteny between FCH, FRU and FPE. Also, a comparative study confirmed that FCH/FRU 7 and 9 are homologous to FPE 1q and FPE 1p respectively; FCH/FRU 1, 2, 3, 4, 5, 6 and 8 are homologous to FPE 2, 3, 4, 5, 6, 7, and 8 respectively. FCH/FRU 10 and 11 to 25 are homologous to FPE 9 and 10 to 24 respectively.

Chromosome locations of the BACs and their gene orders in FCH (Saker falcon) and FRU (Gyr falcon) displayed no apparent intrachromosomal rearrangements between them. This was determined by visual inspection of probe order on the chromosomes and measurement of fractional length from the p-terminus of each signal. Similarly, FPE (Peregrine falcon) chromosome showed a different order (inversion) from FCH and FRU (established by the same approach) (Figure 3-7).

3.3.3 Specific Aim 1c: To map the molecular cytogenetic markers to existing scaffolds (predicted chromosome fragments–PCFs) of the Saker falcon genome assembly and make the data publicly available by an interactive website (Evolution Highway) designed to demonstrate cytogenetic comparisons between individual species

In order to place and order PCFs along Saker falcon chromosomes, a panel of BACs were selected that have been successfully hybridized across Peregrine falcon chromosomes (Damas *et al.* 2017). 167 clones were utilized to assign RACA generated 103 PCFs to their proper position across the genome (Figure 3-8). The anchored 103 PCFs cover 1.13 Gb of its genome sequence, which is 97.26% of the cumulative scaffold length (Table 3-1).

Table 3-1 : Genome statistics of chromosome level assemblies of Saker falcon (*Falco cherrug*) (Zhan *et al.* 2013).

| Original assembly | |
|--------------------------------|---------------------|
| Genome Statistics | Saker falcon |
| No. scaffolds longer 10Kbp | 731 |
| Total length (Gb) | 1.17 |
| N50 (Mb) | 4.16 |
| RACA assembly | |
| Genome Statistics | Saker falcon |
| No. PCFs placed on chromosomes | 103 |
| Combined PCF length (Gb) | 1.13 |
| N50 (Mb) | 2.22 |
| No. Chimeric scaffolds | 25 (5%) |
| No. oriented scaffolds | 458 |
| Scaffold assembly coverage (%) | 97.26 |

A full list of BACs and their coordinates in the Saker falcon PCFs is given in the appendix section, Table 1 (S1).

Also, homology between chicken and Saker falcon was established for all sequenced chromosomes (1-28) except 25 which did not hybridize on Saker falcon chromosomes (Figure 3-9). Also, no probes were available for chicken chromosome 16.

In total, 12 Saker specific fusions and 5 fissions were identified when compared to chicken.

Figure 3-8: Cytogenetic and PCF mapping of Saker falcon (*Falco cherrug*) chromosome 1 (FCH-1); (a) Part of the PCFs produced by RACA for Saker falcon Chromosome 1 visualised on Evolution Highway; (b) Physical mapping of BACs to FCH-1 using FISH; (c) Precise cytogenetic mapping of BACs on FCH-1 according to their orientation; (d) Ideogram showing chicken (GGA) homologs to FCH-1.

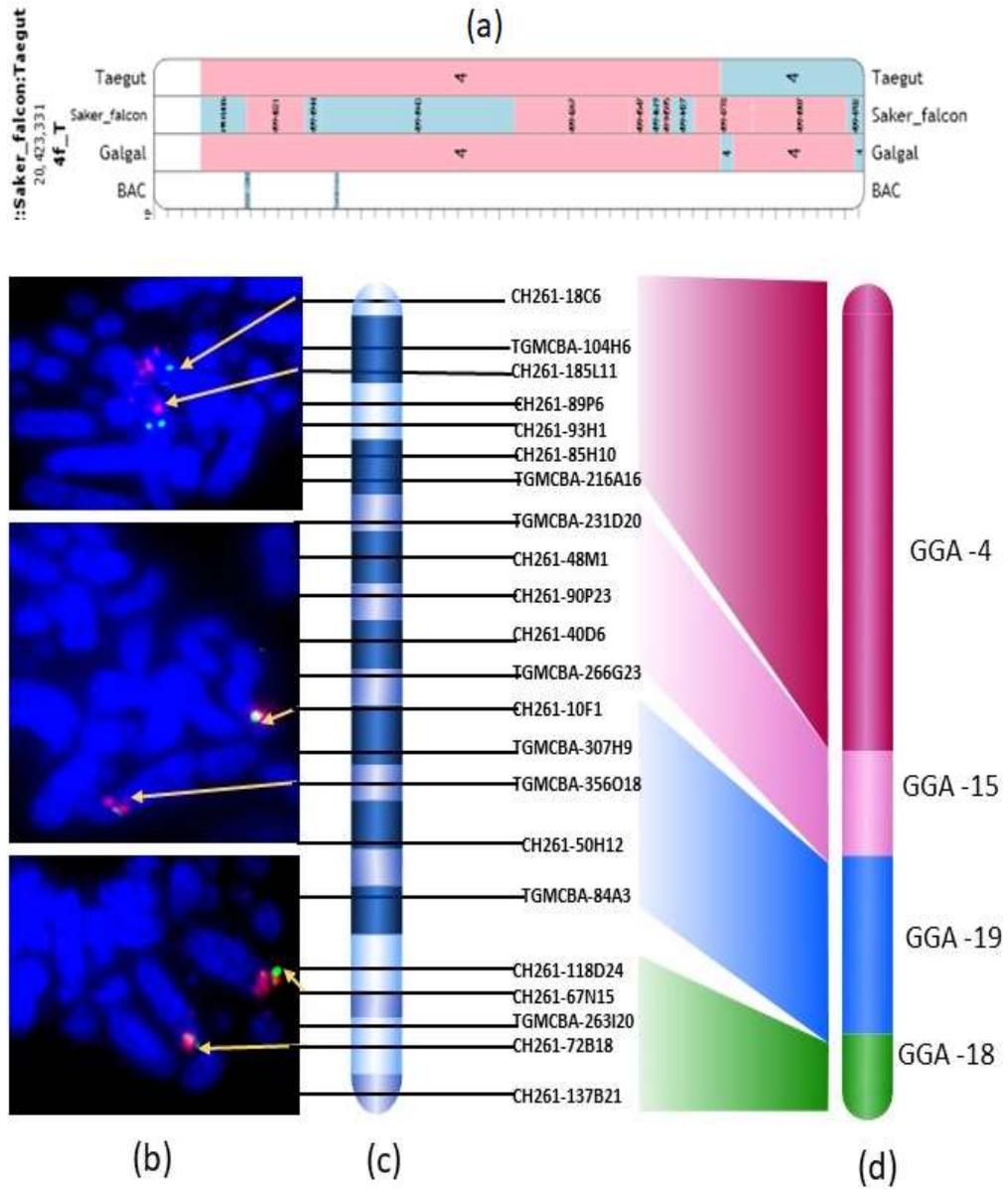
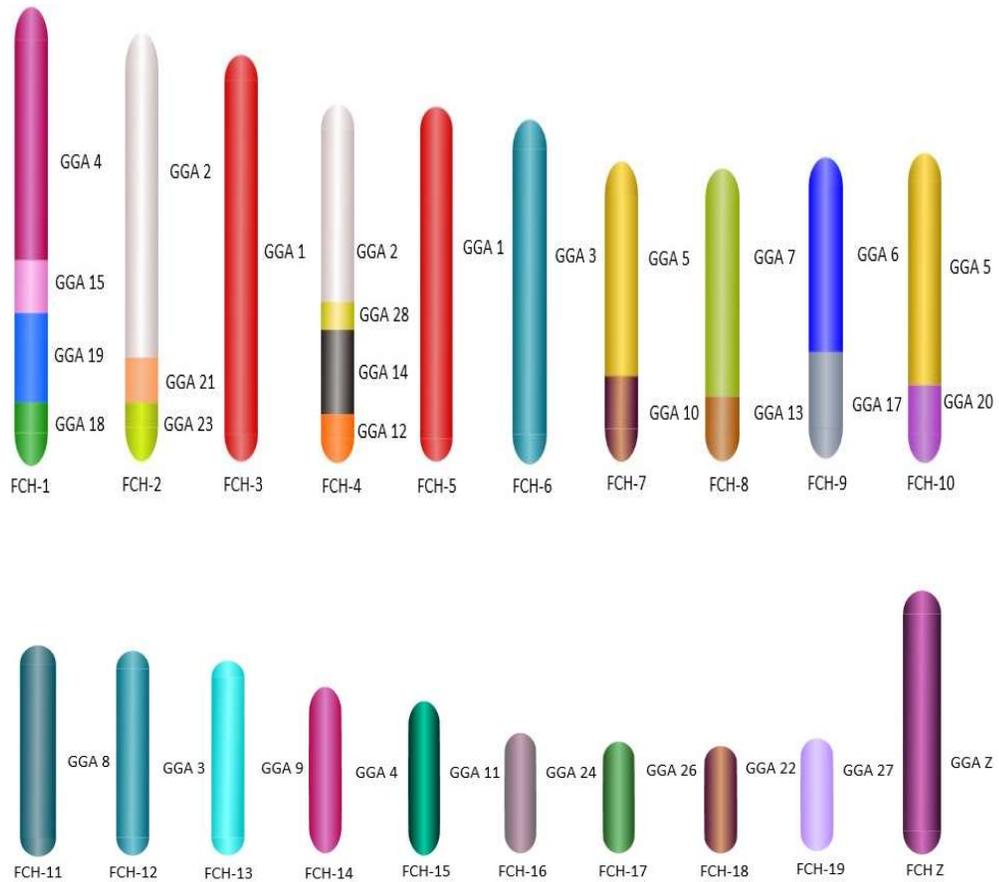
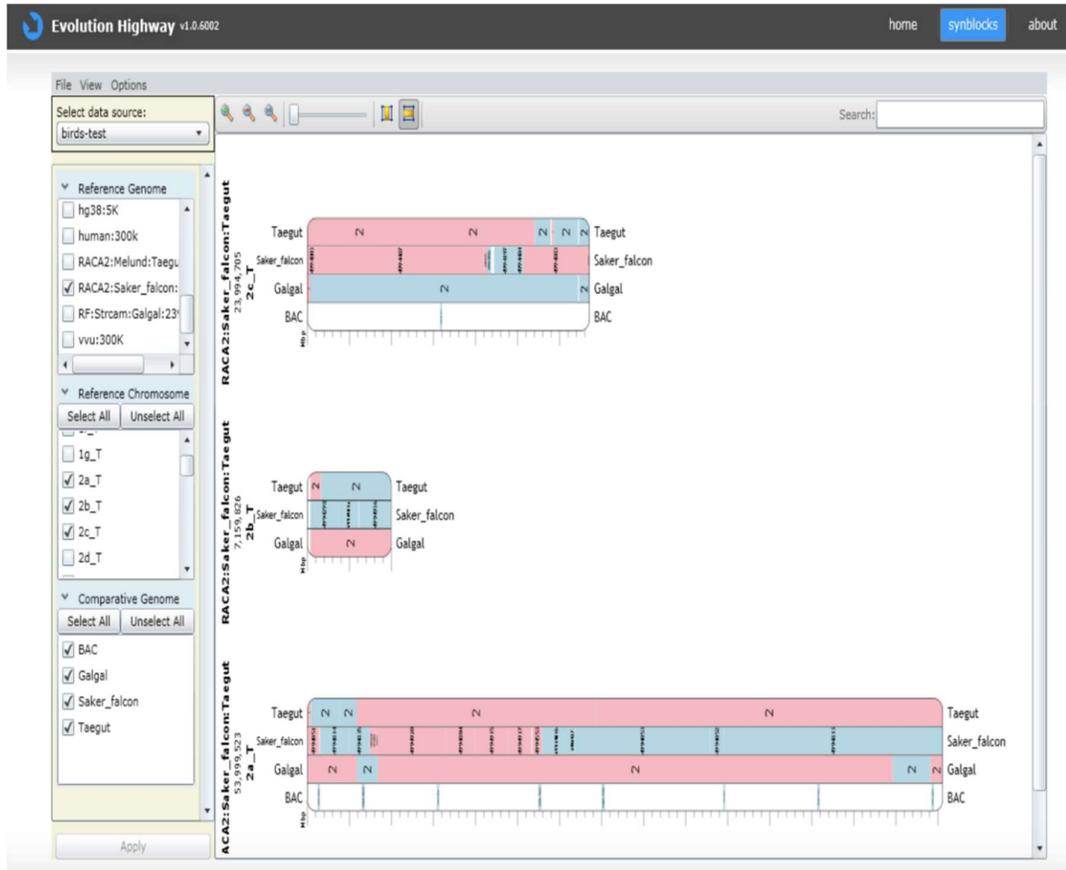


Figure 3-9: Ideogram representation of Saker falcon (*Falco cherrug*) chromosomes and their chicken (GGA) homologs. Colours show the orthologous chicken chromosomes.



The newly assembled Saker falcon (*Falco cherrug*) genome was uploaded to the Evolution Highway comparative chromosome browser for the comparative visualization. This can be visualized under the reference genome name 'saker_falcon' (<http://eh-demo.ncsa.uiuc.edu/birds-test/#/SynBlocks>), an example of which is illustrated in (Figure 3-10). RVC lab ventures in this regard are commendable. Saker falcon chromosome 1-13 and Z are named according to Damas *et al.* (2017), and chromosomes 14-19 are numbered in descending order of the combined assigned PCFs length.

Figure 3-10: Screenshot from Evolution Highway showing the PCFs of Saker falcon genome representing chromosome 2(<http://eh-demo.ncsa.uiuc.edu/birds-test/#/SynBlocks>).



Taken together therefore, the results of this chapter have led to a greater understanding of the overall structure of the genomes of three falcon species but the Saker falcon in particular.

4 Specific Aim 2: To perform a comparative genomic study of genome evolution in Peregrine falcon, Saker falcon and Gyrfalcon making use of BAC FISH probes that map inter- and intra-chromosomal changes to map the path of chromosome evolution in *Falco* species.

4.1 Background

The atypical karyotype of falcons ($2n=40$ to 54) has been mentioned several times in the introduction with a review of the previous cross species chromosome painting data. In the previous chapter, novel results described the comparative genomics of falcons compared to other avian species. This is a powerful tool to determine homologous chromosome regions among different species but can also be used to describe the course of karyotype evolution (Islam *et al.* 2014). By comparisons with multiple species the likely ancestral arrangements can be determined and the path of gross genomic change described. To date, for multiple species in the genus *Falco*, a comparative study has only been constructed using cross-species chromosome chicken paints (Nishida *et al.* 2008). The recent development of a panel of BACs that work reliably across species as well as along the entire chromosomes (Damas *et al.* 2017 and previous chapter), however, permits more detailed cytogenomic studies of chromosome rearrangements in birds. Moreover, the Multiple Genome Rearrangement and Analysis tool (MGR2) (see section 2.5) can be applied to both PCFs using either assembled genomes as a starting point (Romanov *et al.* 2014; Damas *et al.* 2017) or molecular cytogenetic markers derived from cross-species FISH experiments. To the best of my knowledge, MGR2 has yet to be used in the latter context to trace falcon gross genome evolution.

Therefore, the purpose of this chapter was to perform a genome-wide BAC based comparative study in three different falcons; Saker falcon (*Falco cherrug*), Gyrfalcon (*Falco rusticolus*) and Peregrine falcon (*Falco peregrinus*) to reveal chromosomal changes (inter and intra) that led to the falcon lineage. By combining the data with that generated in the lab for other birds (chicken, pigeon, budgerigar, and an ostrich

outgroup), the aim of this chapter was to improve our knowledge of the overall changes happening in birds in general. In other words, to visualize the changes happening in falcons in the context of the rest of avian genome evolution. This led to the following specific aims:

4.2 Specific Aims:

- **Specific aim 2a:** To make use of resources (BAC FISH probes) that map inter- and intra-chromosomal changes in at least three falcon species and input these data into MGRA2 to generate the likely ancestral karyotype of Neognathe birds
- **Specific aim 2b:** To map the path of chromosome evolution in *Falco* species by comparison with conserved (e.g. chicken) and other rearranged (e.g. budgerigar) species through the use of the GRIMM tool (see materials and methods section 2.5) and manual observation
- **Specific aim 2c:** To test the hypothesis that *Falco* genome evolution involved both intra- as well as inter-chromosomal evolution
- **Specific aim 2d:** To test the hypothesis that common mechanisms of chromosomal fusion recur in species with rearranged karyotypes by comparison with a closely related group (*Psittaciformes*) and a more distantly related avian group (*Ratites*)

4.3 Results

4.3.1 Specific Aim 4a: To make use of resources (BAC FISH probes) that map inter- and intra-chromosomal changes in at least three falcons species and input them into MGRA2 to generate the likely ancestral karyotype of Neognathe birds

As mentioned in section 2.5, 92 BACs (numbered 1-92) were mapped to 24 chromosomes of chicken (*Gallus gallus*), pigeon (*Columba livia*), budgerigar (*Melopsittacus undulatus*), Saker falcon (*Falco cherrug*) and Peregrine falcon (*Falco peregrinus*) and ostrich (*Struthio camelus*) outgroup. BACs were originally numbered 1-

92 from the p arm of chicken chromosome 1 to chicken chromosome 28 then the Z. For instance, for the 11 BACs that mapped to chromosome 1, the order in chicken was: pter-1-2-3-4-5-6-7-8-9-10-11-qter. As mentioned in the materials and methods section, the relative position of the BACs on the chromosome was measured both by visual inspection (e.g. was a red signal closer to the p-terminus of the chromosome?) and quantitatively by measuring the relative position of the signal from the p-terminus compared to the length of the whole chromosome.

The Multiple Genomes Rearrangements and Ancestors tool version 2 (MGRA2) (Avdeyev *et al.* 2016; Alekseyev & Pevzner 2009) successfully generated the most likely ancestral configuration for the Neognathae ancestor. This is illustrated in Table 4-1.

Table 4-1: Neognathae ancestor configuration according to the output of MGRA2.

| Neognathae ancestor chromosome number assignment | Chicken (GGA) Chromosome equivalent | Order of BACs in Neognathae ancestral karyotype |
|--|-------------------------------------|---|
| Chromosome 1 (11 BACs) | GGA1 | 1 2 3 4 5 6 7 8 9 10 11 |
| Chromosome 2 (12 BACs) | GGA2 | 12 13 14 15 16 17 18 19 20 21 22 23 |
| Chromosome 3 (10 BACs) | GGA3 | 24 25 26 27 28 29 30 31 32 33 |
| Chromosome 4 (7 BACs) | GGA4q | 40 39 38 41 42 43 44 |
| Chromosome 5 (3 BACs) | GGA5 | 45 46 47 |
| Chromosome 6 (2 BACs) | GGA6 | 48 49 |
| Chromosome 7 (4 BACs) | GGA7 | 51 52 53 50 |
| Chromosome 9 (2 BACs) | GGA9 | 54 55 |
| Chromosome 10 (4 BACs) | GGA4p | 35 34 36 37 |
| Chromosome 11 (2 BACs) | GGA10 | 56 57 |
| Chromosome 12 (1 BAC) | GGA11 | 58 |
| Chromosome 13 (2 BACs) | GGA12 | 60 59 |
| Chromosome 14 (2 BACs) | GGA13 | 61 62 |
| Chromosome 15 (3 BACs) | GGA14 | 65 64 63 |
| Chromosome 16 (4 BACs) | GGA15 | 66 67 68 69 |
| Chromosome 17 (2 BACs) | GGA17 | 70 71 |
| Chromosome 19 (4 BACs) | GGA19 | 72 73 74 75 |
| Chromosome 21 (2 BACs) | GGA21 | 76 77 |
| Chromosome 22 (2 BACs) | GGA22 | 78 79 |
| Chromosome 23 (3 BACs) | GGA23 | 80 81 82 |
| Chromosome 24 (2 BACs) | GGA24 | 83 84 |
| Chromosome 25 (1 BAC) | GGA26 | 85 |
| Chromosome 26 (2 BACs) | GGA27 | 86 87 |
| Chromosome 27 (1 BAC) | GGA28 | 88 |
| Chromosome Z (4 BACs) | GGAZ | 89 90 91 92 |
| NB. No data for GGA8, GGA16, GGA 18, GGA20, GGA25 | | |

Note the similarity to the chicken chromosome configuration.

4.3.2 Specific Aim 4b: To map the path of chromosome evolution in *Falco* species by comparison with conserved (e.g. chicken) and other rearranged (e.g. budgerigar) species through the use of the GRIMM tool

As mentioned in section 2.5, to reconstruct the chromosomal changes that occurred between the hypothetical Neognathae ancestor (above) and the extant 6 species, two approaches were used. The first was a manual approach (interchromosomal changes only). Through visual inspection the number of interchromosomal changes that could be observed when comparing an ancestral rearrangement were as follows:

| | | |
|--|----|-----------------------------|
| Ancestor to Chicken: | 1 | (fusion to form GGA4) |
| Ancestor to Pigeon | 0 | |
| Ancestor to Budgerigar | 14 | (3 fissions and 11 fusions) |
| Ancestor to Peregrine falcon | 16 | (3 fissions and 13 fusions) |
| Ancestor to Saker (and hence Gyr) falcon | 15 | (3 fissions and 12 fusions) |

The second approach required the Genome Rearrangements In Man and Mouse (GRIMM) tool (Bourque and Pevzner 2002; <http://grimm.ucsd.edu/>). It is noteworthy here, that a computer algorithm will follow its own protocols and not necessarily produce the same results that seem obvious to the human eye. Figures 4-1 to 4-5 shows GRIMM algorithm outputs.

Nonetheless, in the following GRIMM outputs, each chromosome is represented in a different colour and the numbers correspond to the BAC order in table 4.1. For instance, in figure 4.1, chromosome 7 (ordered BACs 50-53) is in blue, part of chromosome 10 (BACs 34 and 35) are numbered in green and so on. When the order is thought to be reversed (e.g. in a chromosome inversion) the numbers are represented as negative numbers. Thus, on the first line of figure 4.1, an inversion is inferred from the avian ancestor chromosome 7 (51 52 53 50) to the chicken chromosome 7 (-53 -52 -51 50). Fusions are apparent when the colour changes on the same line e.g. on the second line of figure 4, a fusion is inferred (-34 -35 and -37 -36 in the ancestor to -34 -35 36 37 in the

chicken). Thus, as shown in figure 4.4, from the avian ancestor, the GRIMM output identified 30 changes – 11 fusions, 1 fission and 8 inversions from the ancestor to the Peregrine falcon. Note, that the GRIMM and visual inspection from the cytogenetic results gives a similar number of fissions and fusions in each case.

Figure 4-1: For the ancestor to the chicken, the GRIMM algorithm mapped 6 changes (4 fusions, 1 translocation and 1 inversion (reversal)).

| | | | |
|----------|----------------------|---|--|
| 1 | Reversal | 51 52 53 50 | -53 -52 -51 50 |
| 2 | Fusion | -34 -35 -37 -36 | -34 -35 36 37 |
| 3 | Fusion | -44 -43 -42 -41 -38 -34 -35 36 37 | -44 -43 -42 -41 -38 -37 -36 35 34 |
| 4 | Translocation | -44 -43 -42 -41 -38 -37 -36 35 34 40 39 | -44 -43 -42 -41 -40 -34 -35 36 37 38 39 |
| 5 | Fusion | -44 -43 -42 -41 -40 -34 -35 36 37 38 39 | -44 -43 -42 -41 -40 -39 -38 -37 -36 35 34 |
| 6 | Fusion | -63 -64 -65 | -63 -64 65 |

For this, and other analyses, the three extra fissions compared to the visual inspection may be explained by a “quirk” in the algorithm that tends to “break up” the chromosomes into small fragments if it cannot easily resolve them.

Figure 4-2: For the ancestor to the pigeon, the GRIMM algorithm mapped 7 changes (3 fusions, 1 translocation and 3 inversions (reversals)).

| | | | |
|---|---------------|------------------------------|-----------------------------|
| 1 | Reversal | 51 52 53 50 | -53 -52 -51 50 |
| 2 | Reversal | -47 -46 -45 | -47 45 46 |
| 3 | Reversal | -69 -68 -67 -66 | 68 69 -67 -66 |
| 4 | Fusion | -34 -35 36 37 | -34 -35 -37 -36 |
| 5 | Fusion | 65 -63 -64 | 65 64 63 |
| 6 | Translocation | 40 39 -44 -43 -42 -41 -38 | 40 39 44 -43 -42 -41 -38 |
| 7 | Fusion | 40 39 44 -43 -42 -41 -38 | 40 39 44 38 41 42 43 |

See previous comments about fusions

Figure 4-3: For the ancestor to the budgerigar, the GRIMM algorithm mapped 18 changes (7 fusions, 7 translocations and 4 inversions (reversals)).

| | | | |
|----|---------------|---|--|
| 1 | Reversal | 12 13 14 15 16 17 18 19 20 21 22 23 | -17 -16 -15 -14 -13 -12 18 19 20 21 22 23 |
| 2 | Reversal | -33 -32 -31 -30 -29 -28 -27 -26 -25 -24 | 32 33 -31 -30 -29 -28 -27 -26 -25 -24 |
| 3 | Reversal | 32 33 -31 -30 -29 -28 -27 -26 -25 -24 | 26 27 28 29 30 31 -33 -32 -25 -24 |
| 4 | Reversal | 80 81 82 | 80 -82 -81 |
| 5 | Fusion | -57 -56 60 59 | -57 -56 -59 -60 |
| 6 | Fusion | -47 -46 -45 -65 | -47 -46 -45 65 |
| 7 | Fusion | -47 -46 -45 65 48 49 | -65 45 46 47 48 49 |
| 8 | Fusion | -37 -36 -34 -35 | -37 -36 35 34 |
| 9 | Fusion | -37 -36 35 34 40 39 | -34 -35 36 37 40 39 |
| 10 | Translocation | 1 2 3 4 5 6 7 8 9 10 11 -71 -70 | 1 2 3 -11 -10 -9 -8 -7 -6 -5 -4 -71 -70 |
| 11 | Fusion | -49 -48 -47 -46 -45 65 64 63 | -49 -48 -47 -46 -45 65 -63 -64 |
| 12 | Fusion | 58 -34 -35 36 37 40 39 | 58 -39 -40 -37 -36 35 34 |
| 13 | Translocation | 1 2 3 -11 -10 -9 -8 -7 -6 -5 -4 -71 -70 | 5 6 7 8 9 10 11 -3 -2 -1 -4 -71 -70 |
| 14 | Translocation | -34 -35 36 37 40 39 -58 38 41 42 43 44 | -34 -35 36 37 -38 58 -39 -40 41 42 43 44 |
| 15 | Translocation | 26 27 28 29 30 31 -33 -32 -25 -24 70 71 4 1 2 3 | 26 27 -70 24 25 32 33 -31 -30 -29 -28 71 4 1 2 3 |
| 16 | Translocation | -50 -53 -52 -51 64 63 -65 45 46 47 48 49 | -50 -53 -52 -51 -46 -45 65 -63 -64 47 48 49 |
| 17 | Translocation | -50 -53 -52 -51 -46 -45 65 -63 -64 47 48 49 | -50 -53 -48 -47 64 63 -65 45 46 51 52 49 |
| 18 | Translocation | -50 -53 -48 -47 64 63 -65 45 46 51 52 49 | -50 -52 -51 -46 -45 65 -63 -64 47 48 53 49 |

Figure 4-4: For the ancestor to the peregrine falcon, the GRIMM algorithm mapped 30 changes (11 fusions, 1 fission, 10 translocations and 8 inversions (reversals)).

| | | | |
|----|---------------|---|---|
| 1 | Reversal | -11 -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 | -11 -10 -9 -8 -7 -6 -5 -4 -3 1 2 |
| 2 | Reversal | 51 52 53 50 | -53 -52 -51 50 |
| 3 | Reversal | -53 -52 -51 50 | -53 -52 -50 51 |
| 4 | Reversal | -33 -32 -31 -30 -29 -28 -27 -26 -25 -24 | -33 -32 -31 -30 -29 -28 26 27 -25 -24 |
| 5 | Reversal | -33 -32 -31 -30 -29 -28 26 27 -25 -24 | 30 31 32 33 -29 -28 26 27 -25 -24 |
| 6 | Reversal | 30 31 32 33 -29 -28 26 27 -25 -24 | 30 31 32 29 -33 -28 26 27 -25 -24 |
| 7 | Reversal | 89 90 91 92 | -90 -89 91 92 |
| 8 | Fusion | 48 49 -71 -70 | 48 49 70 71 |
| 9 | Fusion | -51 50 52 53 -62 -61 | -51 50 52 53 61 62 |
| 10 | Fusion | 45 46 47 56 57 | 45 46 47 -57 -56 |
| 11 | Fusion | -75 -74 -73 -72 38 41 42 43 44 | -75 -74 -73 -72 -44 -43 -42 -41 -38 |
| 12 | Fusion | -23 -22 -21 -20 -19 -18 -17 -16 -15 -14 -13 -12 -88 | -23 -22 -21 -20 -19 -18 -17 -16 -15 -14 -13 -12 88 |
| 13 | Fusion | -23 -22 -21 -20 -19 -18 -17 -16 -15 -14 -13 -12 88 -63 -64 | -23 -22 -21 -20 -19 -18 -17 -16 -15 -14 -13 -12 88 64 63 |
| 14 | Fusion | -23 -22 -21 -20 -19 -18 -17 -16 -15 -14 -13 -12 88 64 63 -65 | -23 -22 -21 -20 -19 -18 -17 -16 -15 -14 -13 -12 88 64 63 65 |
| 15 | Fusion | 36 37 35 34 | 36 37 -34 -35 |
| 16 | Reversal | 36 37 -34 -35 | 36 35 34 -37 |
| 17 | Fusion | -23 -22 -21 -20 -19 -18 -17 -16 -15 -14 -13 -12 88 64 63 65 -59 -60 | -23 -22 -21 -20 -19 -18 -17 -16 -15 -14 -13 -12 88 64 63 65 60 59 |
| 18 | Translocation | 76 77 -82 -81 -80 | 76 77 81 82 -80 |
| 19 | Translocation | -75 -74 -73 -72 -44 -43 -42 -41 -38 66 67 68 69 | -75 -74 -73 -72 -67 -66 38 41 42 43 44 68 69 |
| 20 | Fusion | -75 -74 -73 -72 -67 -66 38 41 42 43 44 68 69 | -75 -74 -73 -72 -67 -66 -69 -68 -44 -43 -42 -41 -38 |
| 21 | Translocation | -75 -74 -73 -72 -67 -66 -69 -68 -44 -43 -42 -41 -38 | -75 -74 -73 -72 -67 -66 -69 -68 -44 -43 -42 -41 -40 |

Figure 4-4 (*Continued*): For the ancestor to the Peregrine falcon, the GRIMM algorithm mapped 29 changes (11 fusions, 1 fission, 10 translocations and 8 inversions (reversals)).

| | | | |
|----|---------------|--|--|
| | | 40 39 | 38 39 |
| 22 | Translocation | -75 -74 -73 -72 -67 -66 -69 -68 -44 -43 -42 -41 -40 | -75 -74 -73 -72 -67 -66 -69 -68 -44 -43 -39 -38 |
| | | 38 39 | 40 41 42 |
| 23 | Translocation | -75 -74 -73 -72 -67 -66 -69 -68 -44 -43 -39 -38 | -75 -74 -73 -72 -67 -66 -69 -68 -44 |
| | | 40 41 42 | 38 39 43 40 41 42 |
| 23 | Translocation | 48 49 70 71 | 24 |
| 23 | Translocation | 24 | 25 |
| 23 | Translocation | 25 | 26 |
| 23 | Translocation | 26 | |
| 27 | Fission | | 48 49 70 71 47 -57 -56 |
| 28 | Translocation | -23 -22 -21 -20 -19 -18 -17 -16 -15 -14 -13 -12 88 64 63 65 60 59 -82 -81 -77 -76 | -23 -22 -21 -20 -19 -18 76 77 81 82 -59 -60 -65 -63 -64 -88 12 13 14 15 16 17 |
| 29 | Fusion | -23 -22 -21 -20 -19 -18 76 77 81 82 -80 | -23 -22 -21 -20 -19 -18 76 77 81 82 80 |

Figure 4-5: For the ancestor to the Saker (and hence Gyr) falcon, the GRIMM algorithm mapped 25 changes (9 fusions, 7 translocations and 9 inversions (reversals)).

| | | | |
|----|---------------|--|--|
| 1 | Reversal | 38 41 42 43 44 | 38 -42 -41 43 44 |
| 2 | Reversal | -11 -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 | -11 -10 -9 -8 -7 -6 -5 -4 -3 1 2 |
| 3 | Reversal | -50 -53 -52 -51 | 53 50 -52 -51 |
| 4 | Reversal | 53 50 -52 -51 | 53 52 -50 -51 |
| 5 | Reversal | 24 25 26 27 28 29 30 31 32 33 | 24 25 -27 -26 28 29 30 31 32 33 |
| 6 | Reversal | 24 25 -27 -26 28 29 30 31 32 33 | 24 25 -27 -26 28 29 -33 -32 -31 -30 |
| 7 | Reversal | 24 25 -27 -26 28 29 -33 -32 -31 -30 | 24 25 -27 -26 28 33 -29 -32 -31 -30 |
| 8 | Fusion | 38 -42 -41 43 44 -75 -74 -73 -72 | 38 -42 -41 43 44 72 73 74 75 |
| 9 | Fusion | -23 -22 -21 -20 -19 -18 -17 -16 -15 -14 -13 -12 -88 | -23 -22 -21 -20 -19 -18 -17 -16 -15 -14 -13 -12 88 |
| 10 | Fusion | -23 -22 -21 -20 -19 -18 -17 -16 -15 -14 -13 -12 88 -63 -64 | -23 -22 -21 -20 -19 -18 -17 -16 -15 -14 -13 -12 88 64 63 |
| 11 | Fusion | -23 -22 -21 -20 -19 -18 -17 -16 -15 -14 -13 -12 88 64 63 -65 | -23 -22 -21 -20 -19 -18 -17 -16 -15 -14 -13 -12 88 64 63 65 |
| 12 | Fusion | 36 37 35 34 | 36 37 -34 -35 |
| 13 | Reversal | 36 37 -34 -35 | 36 35 34 -37 |
| 14 | Fusion | -23 -22 -21 -20 -19 -18 -17 -16 -15 -14 -13 -12 88 64 63 65 -59 -60 | -23 -22 -21 -20 -19 -18 -17 -16 -15 -14 -13 -12 88 64 63 65 60 59 |
| 15 | Fusion | 53 52 -50 -51 61 62 | 51 50 -52 -53 61 62 |
| 16 | Translocation | 38 -42 -41 43 44 72 73 74 75 -2 -1 3 4 5 6 7 8 9 10 11 | 38 -42 -41 43 -4 -3 1 2 -75 -74 -73 -72 -44 5 6 7 8 9 10 11 |
| 17 | Translocation | 38 -42 -41 43 -4 -3 1 2 40 39 | 38 -42 -41 43 -39 -40 -2 -1 3 4 |
| 18 | Reversal | 38 -42 -41 43 -39 -40 | 38 -42 -41 39 -43 -40 |
| 19 | Translocation | 76 77 -82 -81 -80 | 76 77 82 -81 -80 |
| 20 | Fusion | 76 77 82 -81 -80 | 76 77 82 80 81 |
| 21 | Translocation | -11 -10 -9 -8 -7 -6 -5 44 72 73 74 75 66 67 68 69 | -11 -10 -9 -8 -7 -6 -5 44 -67 -66 -75 -74 -73 -72 68 69 |
| 22 | Translocation | 76 77 82 80 81 | -17 -16 -15 -14 -13 -12 88 64 63 |

Figure 4-5 (Continued): For the ancestor to the Saker (and hence Gyr) falcon, the GRIMM algorithm mapped 25 changes (9 fusions, 7 translocations and 9 inversions (reversals)).

| | | | | | | | |
|----------------------------------|--|-------------------------------|-------------------------------|--|--|----------------------------------|---------------------|
| | <table border="1"> <tr><td>-59 -60 -65 -63 -64 -88 12 13</td></tr> <tr><td>14 15 16 17 18 19 20 21 22 23</td></tr> </table> | -59 -60 -65 -63 -64 -88 12 13 | 14 15 16 17 18 19 20 21 22 23 | <table border="1"> <tr><td>65 60 59</td></tr> <tr><td>-81 -80 -82 -77 -76 18 19 20 21</td></tr> <tr><td>22 23</td></tr> </table> | 65 60 59 | -81 -80 -82 -77 -76 18 19 20 21 | 22 23 |
| -59 -60 -65 -63 -64 -88 12 13 | | | | | | | |
| 14 15 16 17 18 19 20 21 22 23 | | | | | | | |
| 65 60 59 | | | | | | | |
| -81 -80 -82 -77 -76 18 19 20 21 | | | | | | | |
| 22 23 | | | | | | | |
| 23 Translocation | <table border="1"> <tr><td>38 -42 -41 39 -43 -40</td></tr> <tr><td>66 67 -44 5 6 7 8 9 10 11</td></tr> </table> | 38 -42 -41 39 -43 -40 | 66 67 -44 5 6 7 8 9 10 11 | <table border="1"> <tr><td>38 -42 -41 39 -43 -40 44 -67 -66</td></tr> <tr><td>5 6 7 8 9 10 11</td></tr> </table> | 38 -42 -41 39 -43 -40 44 -67 -66 | 5 6 7 8 9 10 11 | |
| 38 -42 -41 39 -43 -40 | | | | | | | |
| 66 67 -44 5 6 7 8 9 10 11 | | | | | | | |
| 38 -42 -41 39 -43 -40 44 -67 -66 | | | | | | | |
| 5 6 7 8 9 10 11 | | | | | | | |
| 24 Fusion | <table border="1"> <tr><td>38 -42 -41 39 -43 -40 44 -67</td></tr> <tr><td>-66</td></tr> <tr><td>-75 -74 -73 -72 68 69</td></tr> </table> | 38 -42 -41 39 -43 -40 44 -67 | -66 | -75 -74 -73 -72 68 69 | <table border="1"> <tr><td>38 -42 -41 39 -43 -40 44 -67 -66</td></tr> <tr><td>-69 -68 72 73 74 75</td></tr> </table> | 38 -42 -41 39 -43 -40 44 -67 -66 | -69 -68 72 73 74 75 |
| 38 -42 -41 39 -43 -40 44 -67 | | | | | | | |
| -66 | | | | | | | |
| -75 -74 -73 -72 68 69 | | | | | | | |
| 38 -42 -41 39 -43 -40 44 -67 -66 | | | | | | | |
| -69 -68 72 73 74 75 | | | | | | | |
| 25 Translocation | <table border="1"> <tr><td>56 57</td></tr> <tr><td>-47 -46 -45</td></tr> </table> | 56 57 | -47 -46 -45 | <table border="1"> <tr><td>56 57 47</td></tr> <tr><td>-46 -45</td></tr> </table> | 56 57 47 | -46 -45 | |
| 56 57 | | | | | | | |
| -47 -46 -45 | | | | | | | |
| 56 57 47 | | | | | | | |
| -46 -45 | | | | | | | |

4.3.3 Specific Aim 4c: To test the hypothesis that *Falco* genome evolution involved both intra- as well as inter-chromosomal evolution

While the specific details of the changes vary in the “visual inspection” approach the overall message remains the same. Namely that it is evident that both inter and intra chromosomal changes have been involved in *Falco* genome evolution. Specifically, the number of inter-chromosomal rearrangements in *Falco* species is significantly more (15-16) than in most birds (typically no more than 1 or 2) from the avian ancestor. In general terms, the falcons have more chromosome rearrangements than the budgerigar (representative of the *Psittaciformes*) with approximately twice as many (intrachromosomal) inversions.

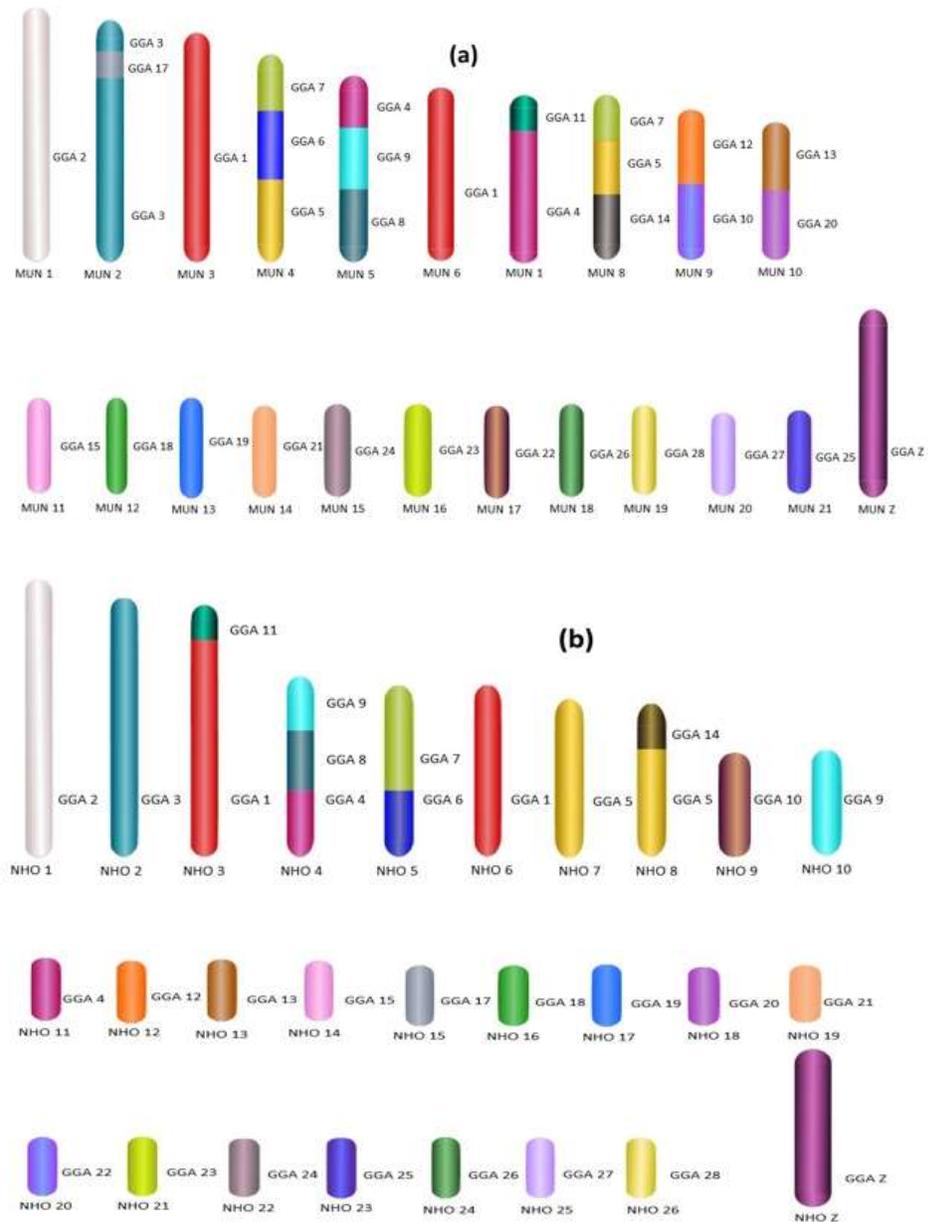
4.3.4 Specific Aim 4d: To test the hypothesis that common mechanisms of chromosomal fusion recur in species with rearranged karyotypes by comparison with a closely related group (*Psittaciformes*)

Given that rearrangements although infrequent in birds in general are common in both *Psittaciform* and *Falconiform* species, is it reasonable to hypothesize that some rearrangements will be shared between the groups, either identical by descent, by homoplasy, or by hemiplasy (see section 1.4.2). Patterns of rearrangement would suggest they arise in such a way to benefit the divergence of the group itself.

In order to address this question, representative of three groups of birds that have previously been reported as having multiple chromosomal rearrangements were studied, namely *Psittaciformes*, *Falconiformes* and *Ratites* (*Ratites* are also included as Romanov *et al.* (2014) previously reported subtle rearrangements, despite having an overall structure of $2n=80$). To this end, a set of BACs as outlined in section 2.2.1.2, was used to investigate the hypothesis that chromosome fusions are in common in species with rearranged karyotypes. For this part of the study, the 2(3) falcons studied above, plus the budgerigar, (*Melopsittacus undulatus*), plus a third *Psittaciform* species not used for the MGRA2/GRIMM study, the cockatiel (*Nymphicus hollandicus*) were investigated.

As detailed in section 3.3.2 hybridization patterns are the same in the Saker falcon and Gyrfalcon, similar in Peregrine falcon (1 fusion to form Peregrine chromosome 1 is the difference) as illustrated in Figure 3.9. For the 2 *Psittaciform* species, results are illustrated in Figure 4-6.

Figure 4-6: Homology maps between chicken (GGA) and (a) budgerigar (*Melopsittacus undulatus*) (MUN) (b) cockatiel (*Nymphicus hollandicus*) (NHO). Colors show the orthologous chicken chromosomes.

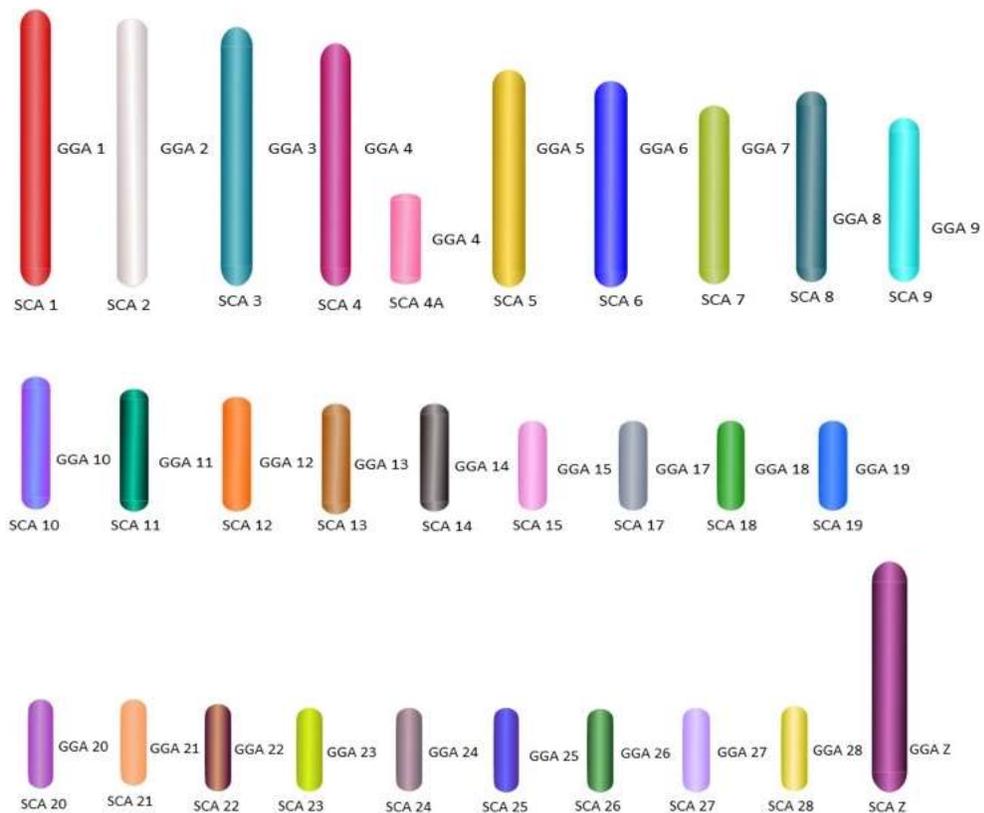


GGA 1, 4, 5 and 7 was syntenic to two segments of MUN while GGA 1, 4 and 5 to 2 syntenic segments of NHO. Neighboring synteny combinations for GGA 5/6/7 in MUN and GGA 6/7 in NHO were also observed. Nevertheless, both have a different synteny combination pattern except 4/8/9 which is common in both species. In addition, MUN exhibit a fusion of homologs of GGA 10, 11, 12, 13, 14 and 17 to macrochromosomes

and homologs of GGA 13 and 20 are fused to form a single chromosome whereas, homologs of GGA 11 and 14 are fused to separate macrochromosomes in NHO. The overall karyotypic rearrangements are shown in Table 4-2. The present findings are in good agreement with the painting studies conducted by Nanda *et al.* (2007) but, here, a greater number of chromosomes were analyzed.

In terms of possible interchromosomal rearrangements in the ostrich, contrary to reports by Romanov *et al.* (2014), this species appeared to retain an ancestral pattern Figure 4-7.

Figure 4-7: Homology maps between chicken (GGA) and the ostrich (*Struthio camelus*) (SCA). Colors show the orthologous chicken chromosomes.



When comparing the 4 species directly, similarities were apparent between the budgerigar and the cockatiel (probably reflecting a shared ancestry). However, only one common fusion (chromosome 5 and 14) was apparent in the three (not ostrich) compared species. The results are illustrated in Table 4-2.

Table 4-2: Comparison of the number of conserved synteny segments (GGA 1-9) and fusion combinations in the budgerigar, cockatiel, ostrich and Saker/Gyr Falcon corresponding to chicken autosomes.

| GGA 1-9 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Total | Fusion combinations |
|---|---|---|---|---|---|---|---|---|---|-------|--|
| Budgerigar (<i>Melopsittacus undulatus</i>) | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 13 | (5/6/7), (4/8/9/), (14/5/7), (3/17), (4/11), (10/12) and (13/20) |
| Cockatiel (<i>Nymphicus hollandicus</i>) | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 2 | 13 | (4/8/9), (6/7), (1/11), and (5/14) |
| Saker falcon (<i>Falco cherrug</i>) and Gyr Falcon (<i>Falco rusticolus</i>) | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 14 | (4/15/19/18), (2/12/14/28) (2/21/23), (6/17), (5/10) (7/13) and (5/20) |
| Ostrich (<i>Struthio camelus</i>) | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 10 | None |

In this study, therefore I reject the hypothesis that (with one possible exception) that there are commonalities between the rearrangements observed for Psittaciform and Falconiform species. In other words, the patterns observed for both groups, although characterized by wholesale chromosome fusion in general, arose by independent mechanisms.

5 Specific aim 3: To map telomeric sequences in three falcon species to test the hypothesis that remnants of former chromosomal fusions retain their telomeric motifs i.e. appear as interstitial telomeres. To compare with other species that have undergone chromosomal fusions such as the budgerigar and crocodile.

5.1 Background

Karyotype evolution in birds such as *Falconiformes* and *Psittaciformes* has been previously reported (see introduction) and expanded upon significantly in this thesis has led to considerable changes in karyotype structure and number mostly through chromosomal fusion events (Nanda *et al.* 2002 and previous chapters of this thesis). Consequently, during speciation and karyotypic evolution telomeres can be lost or gained. As described in section 1.2.4, previously studies birds have a different telomeric DNA profile. This includes interstitial telomeric arrays as well as so called “mega-telomeres.” The prevalent hypothesis holds that highly evolved species may show both interstitial and telomeric (TTAGGG)_n sites due to the fusion of micro-or macrochromosomes in a common ancestor. Also, microchromosomes are enriched with (TTAGGG)_n sequences (Nanda *et al.* 2002; Delany *et al.* 2003) and the mega-telomeres are only hitherto reported on microchromosomes (Delany *et al.* 2007).

Therefore, the purpose of this chapter was to define the telomeric DNA profile in the three different falcon species and to compare the falcon telomeric profile with that of the budgerigar and Nile crocodile (*Crocodylus niloticus*), which also have highly rearranged karyotypes. Mapping of telomeric sequences was carried out by FISH using FITC-labelled commercial (TTAGGG)_n PNA probe (DAKO) to detect telomeric sequences (see material and methods).

5.2 Results

Telomeric probes produced signals on both telomeric ends of all chromosomes in three falcon species. The distribution of hybridization signals on microchromosomes were contrasting: In Peregrine falcons, (TTAGGG)_n sequences with intense fluorescence were observed on several specific chromosomes (Figure 5-1). In contrast, only one pair of chromosome in the Saker falcon (Figure 5-3) and the Gyrfalcon (Figure 5-4) showed strong signals. Patterns of telomeric signals were the same in both the Saker falcon and Gyrfalcon. Occurrences of interstitial telomeric sites were not detected in any chromosomes from all 3-falcon species. As previously observed by Delany and colleagues, (TTAGGG)_n sequences with strong fluorescence (mega-telomeres) were observed only on microchromosomes (Figure 5-2).

Figure 5-1: FISH with PNA probe on the metaphase chromosomes of the Peregrine falcon showing different hybridization pattern of telomere probes on chromosomes. Chromosomes were counter-stained with DAPI.

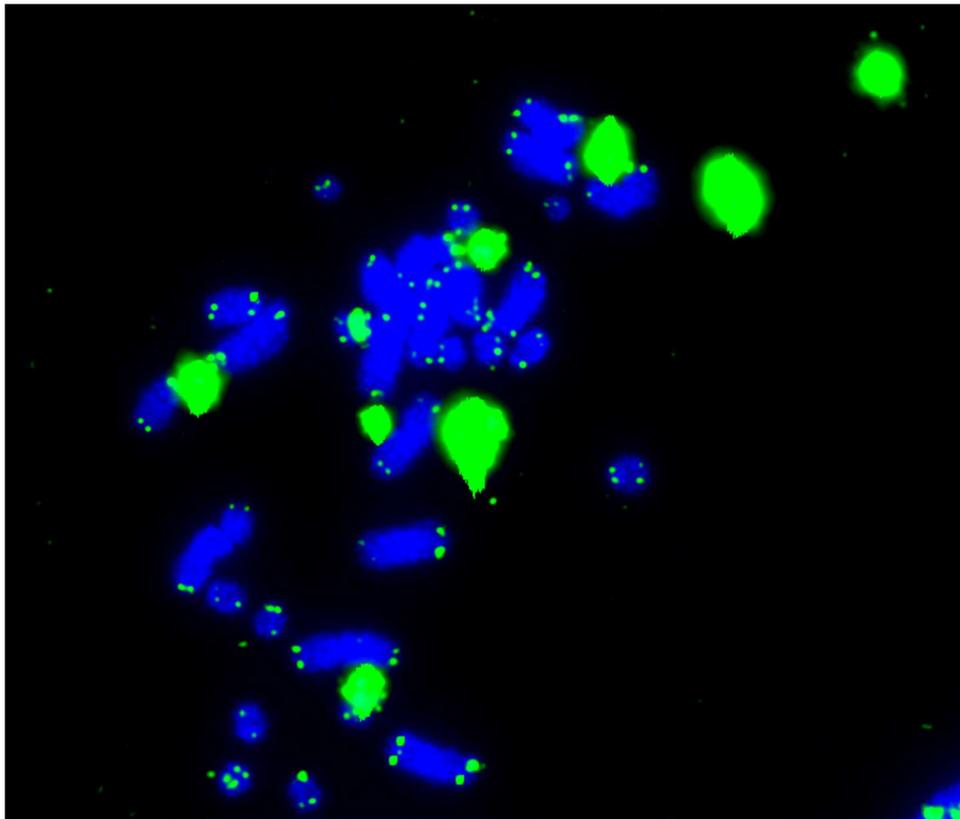


Figure 5-2: (a) and (b) Telomeric signals with intense fluorescence at specific chromosomes of the Peregrine falcon, captured with different exposure time to reduce the signal intensity. Chromosomes were counter-stained with DAPI. (c) Same image is converted to black and white.

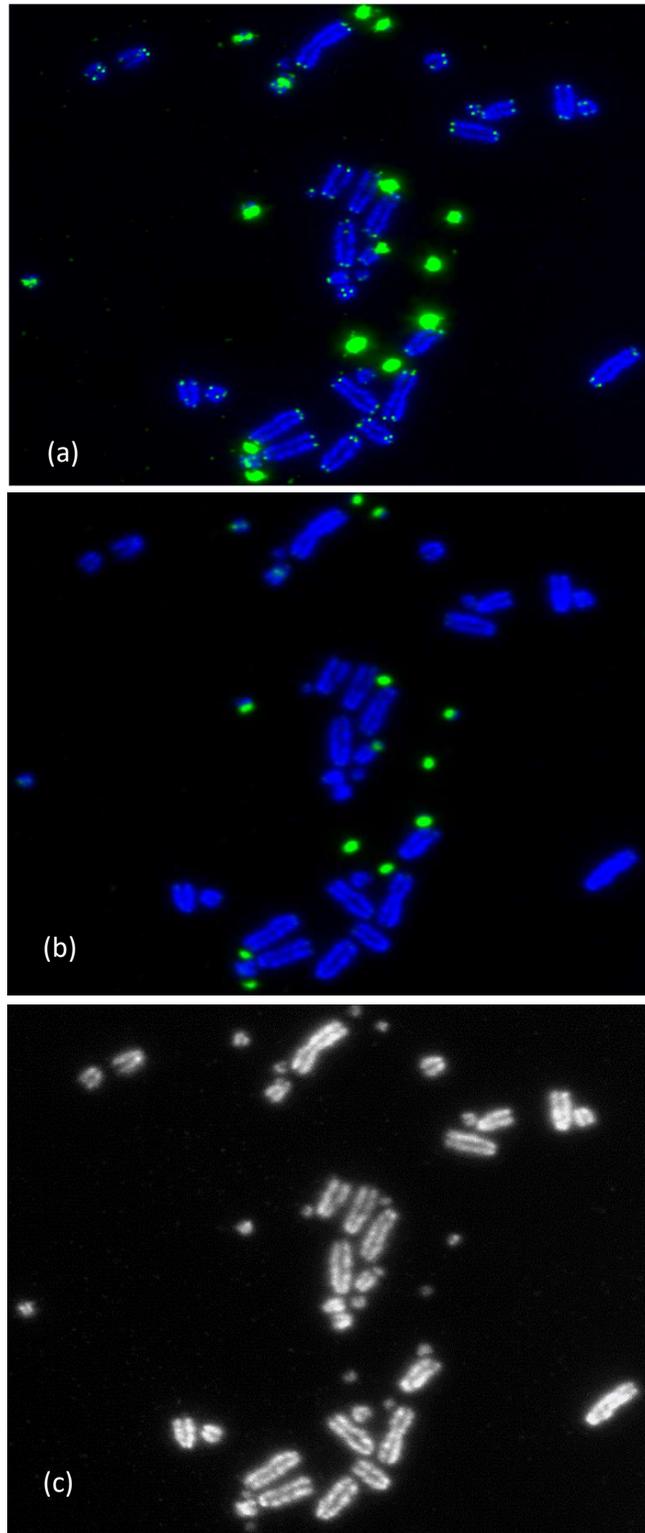


Figure 5-3: (a) and (b) FISH with telomeric PNA probe on the metaphase chromosomes of the Saker falcon showing intense fluorescence at specific chromosomes, captured with different exposure time to reduce the signal intensity. Chromosomes were counter-stained with DAPI. (c) Same image is converted to black and white.

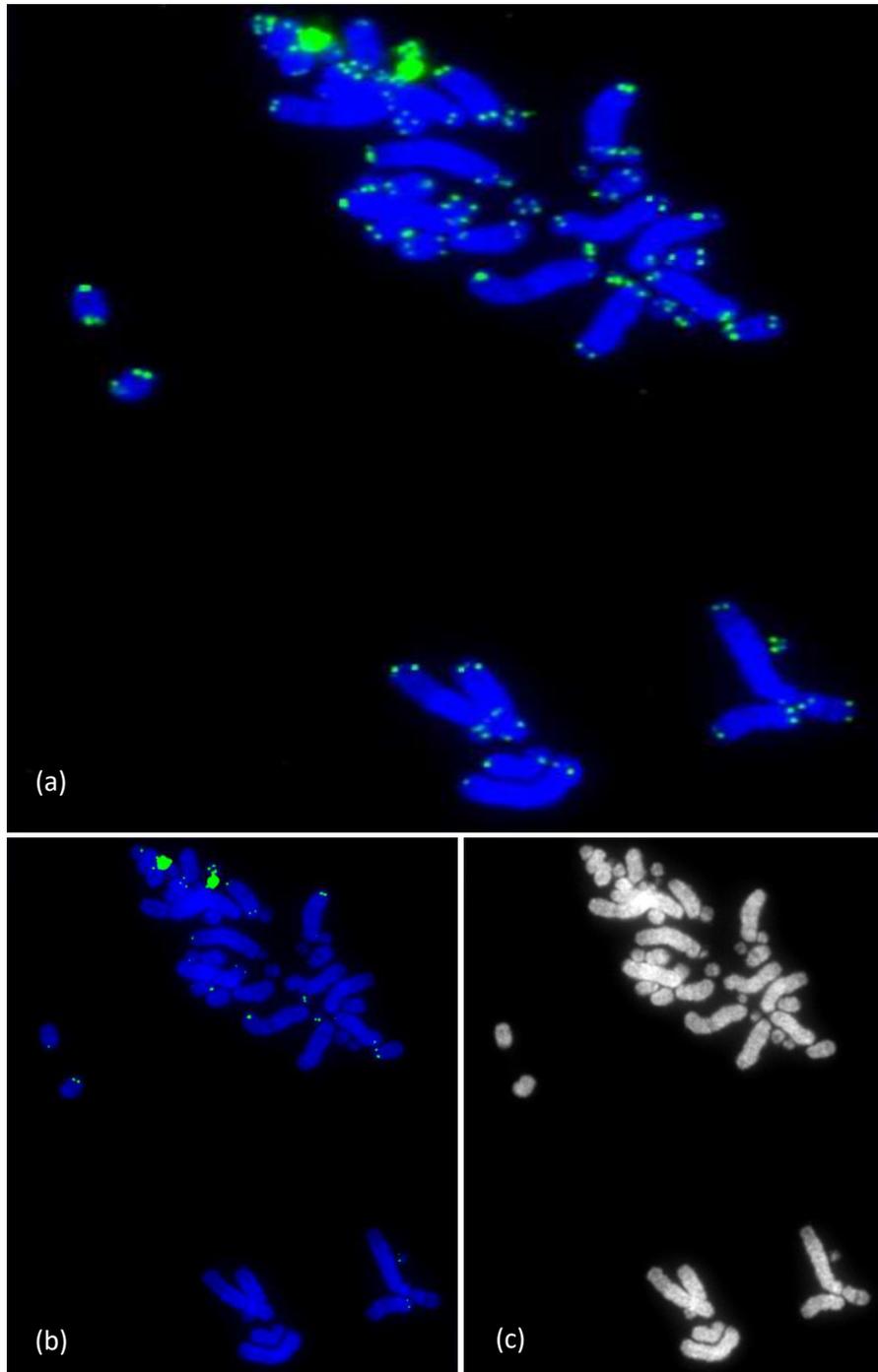


Figure 5-4: (a) and (b) FISH with telomeric PNA probe on the metaphase chromosomes of the Gyrfalcon showing intense fluorescence at specific chromosomes, captured with different exposure time to reduce the signal intensity. Chromosomes were counter-stained with DAPI. (c) Same image is converted to black and white.

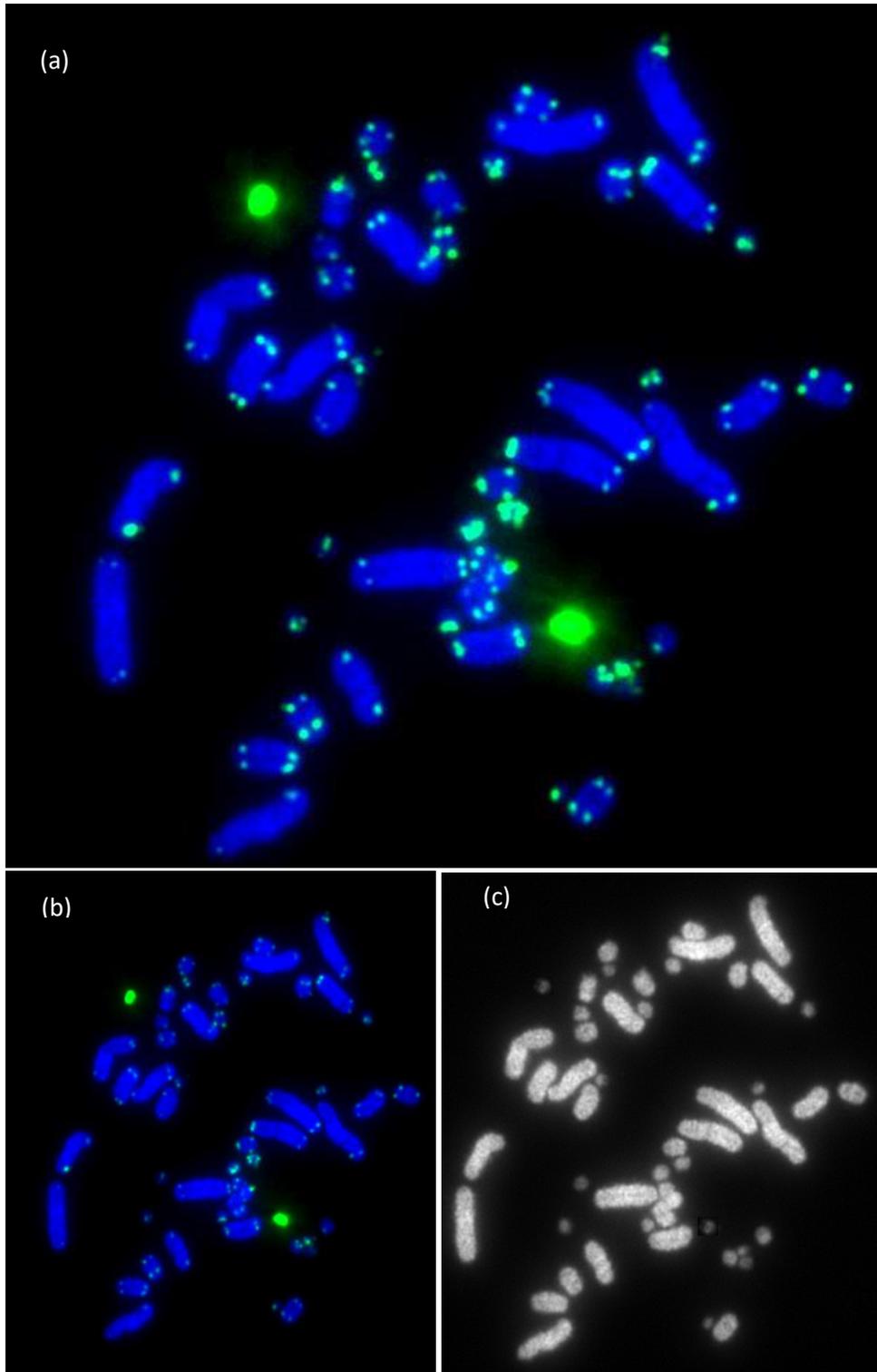
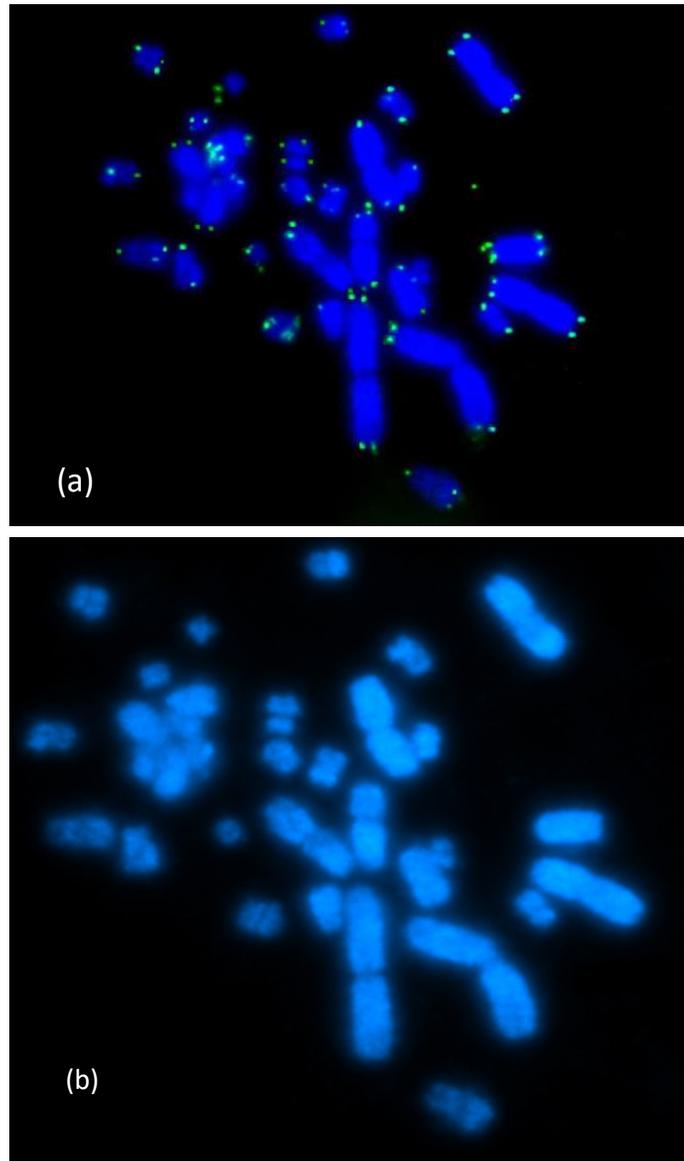
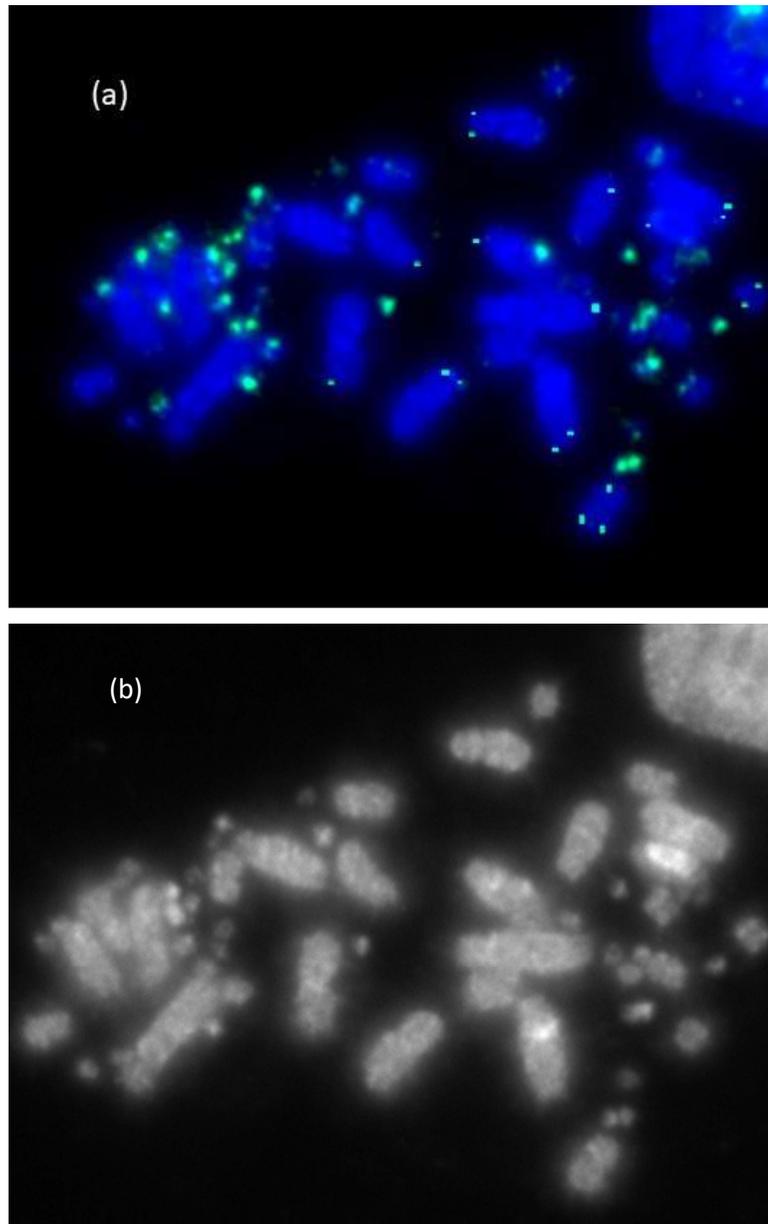


Figure 5-5: FISH with PNA probe on the metaphase chromosomes of the Nile crocodile (*Crocodylus niloticus*) showing telomere signal on all chromosomes. Chromosomes were counter-stained with DAPI. (b) Same image is converted without signal.



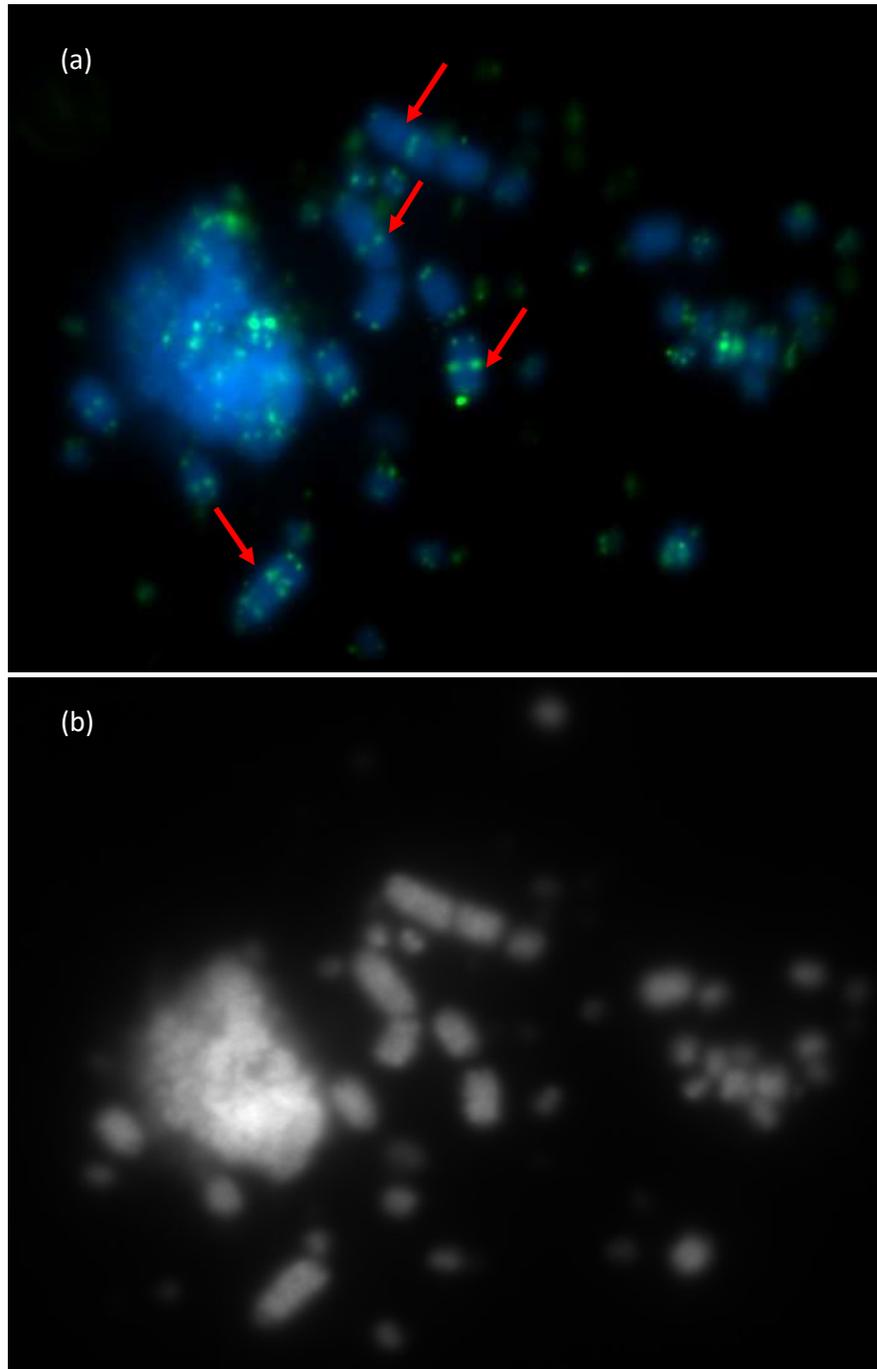
Despite the low chromosome number ($2n=32$), interstitial telomeric signals were not seen in the Nile crocodile (*Crocodylus niloticus*). Telomeric sequences were observed at the termini of all the chromosomes. Unlike in falcon species, telomeric signals on all the chromosomes were evenly distributed without intensity variation (Figure 5-5).

Figure 5-6: FISH with telomeric PNA probe on the metaphase chromosomes of the budgerigar (*Melopsittacus undulates*) showing strong fluorescence at specific chromosomes. Chromosomes were counter-stained with DAPI. (b) Same image is converted to black and white.



Interestingly, the budgerigar (*Melopsittacus undulates*) which has experienced extensive evolutionary rearrangements did not show non-telomeric (TTAGGG)_n sites (Figure 5-6). Telomeric ends of all the chromosomes showed the hybridization signal. However, prominent hybridization signals were observed on the microchromosomes.

Figure 5-7: FISH with telomeric PNA probe on the metaphase chromosomes of the ostrich (*Struthio camelus*). Arrows indicate interstitial telomere sequences. Chromosomes were counter-stained with DAPI. (b) Same image is converted to black and white.



The distribution of telomeric (TTAGGG)_n sequences in the ostrich (*Struthio camelus*) showed several non-telomeric (TTAGGG)_n sites which is shown in (Figure 5-7). There is no evidence however that these correspond to any evolutionary fusion event.

Overall findings of this study are summarized in (Table 5-1).

Table 5-1: Summary of telomeric sequences mapping results in three falcon species, budgerigar, crocodile and ostrich.

| Species | No of mega telomeres (Chromosome pairs) | Number of interstitial telomeres (Chromosome pairs) | Interstitial telomeres at site of chromosome rearrangement? |
|--|---|---|---|
| Peregrine falcon (<i>Falco peregrinus</i>) | 6 | 0 | 0 |
| Saker falcon (<i>Falco cherrug</i>) | 1 | 0 | 0 |
| Gyrfalcon (<i>Falco rusticolus</i>) | 1 | 0 | 0 |
| Budgerigar (<i>Melopsittacus undulatus</i>) | 0 | 0 | 0 |
| Nile crocodile (<i>Crocodylus niloticus</i>) | 0 | 0 | 0 |
| Ostrich (<i>Struthio camelus</i>) | 0 | >6 | 0 |

In this study, I therefore reject the hypothesis that interstitial telomeres appear at the sites of former fusion of the chromosomes, either in falcons, parrots or crocodylians. Indeed, the only interstitial telomeres found were in the ostrich and did not correspond to sites of former fusion.

6 Specific Aim: To test the hypothesis that chromosome that were formerly microchromosomes (but are now fused to larger chromosomes) still “behave” as though they were microchromosomes in terms of their nuclear organization.

6.1 Background

As detailed in section 1.2.1, chromosomes occupy distinct territories in the interphase nucleus. There are two popularly described models: the size-based arrangement of the chromosomes in the interphase nuclei (bigger chromosome at the periphery, smaller in the interior) and the gene density model (gene rich chromosomes central, gene poor peripheral). Nuclear organization of an interphase nucleus can be established by determining the relative nuclear position of chromosome territories in flattened nuclei, adjusting for the 2D nature of the preparation through chromatin density measurements (Foster and Bridger 2005). To date, most studies of nuclear organization in birds have mainly focused on chicken (Habermann *et al.* 2001), duck and turkey (Skinner *et al.* 2009b). In chicken, it has been demonstrated that in embryonic fibroblasts and neurons, the larger chromosomes tend to localize toward the periphery while microchromosomes are more centralized (see section 1.2.3). Given that the smaller chromosomes tend to be more gene dense however, most birds are consistent with either model of nuclear organization.

Taking all available information into account, a description of individual chromosome territory position in the somatic cells of birds with rearranged (non-typical) karyotypes does not exist. Therefore, in this study, the relative nuclear position of chromosome territory was assessed in falcons because of the prevalence of fused chromosomes (micro- and macro- chromosomes). The ideal target for this study was the Peregrine falcon as 11 chicken microchromosomes were fused with other chromosomes in this species (Damas *et al.* 2017) and this study objective included their pattern of nuclear organization. Specifically, the overall genomic characteristics (GC content, CpG islands,

CNV density, recombination rate) of parts of macrochromosomes that were formally microchromosomes before they fused suggests that, at the sequence level, “former microchromosomes” still behave as though they are still the microchromosomes that they once were. Whether this is the case in the context of nuclear organization however remains to be established. With this in mind, this specific aim can be broken down into the following:

6.2 Specific Aims

Specific aim 6a: To identify BAC probes from micro, macro and “former microchromosome” regions of both chicken and Peregrine falcon chromosomes

Specific aim 6b: To map the average chromosome position of all selected BACs on metaphase and interphases, in the latter, inferring 3D position from 2D data using standard protocols (see section 2.7)

Specific aim 6c: To test the hypothesis that macrochromosomes are peripherally located, microchromosomes are centrally located and “former” microchromosomes still “behave” as though they are microchromosomes and thus occupy a central position

6.3 Results

6.3.1 Specific Aim 6a: Identification of BAC probes from micro, macro and “former microchromosome” regions of both chicken and Peregrine falcon chromosomes

Chicken and zebra finch BACs were selected to study the chromosome territory position in Peregrine falcon that has previously proved to hybridize on Peregrine falcon (Damas, *et al.* 2017). Four sets of probes were prepared: (i) Probe sets 1-6 to study the distribution of chromosome territories of macrochromosomes, (ii) Probe sets 7-11 for the microchromosomes, (iii) Probe sets 12-16 to study the distribution of chromosome territories of macrochromosomes which were formerly macrochromosomes but now are fused with microchromosomes in Peregrine falcon (iv) Probe sets 17-21 for microchromosomes which were formerly microchromosomes but

S. Joseph

are now fused with macrochromosomes in Peregrine falcon. Details of the BACs are given in the Table 6-1.

Table 6-1: BACs details used in different probe sets and their corresponding hybridizing chromosomes in GGA (chicken) and FPE (Peregrine falcon).

| Probe set-number | BAC clone name | Label | GGA Chr | FPE Chr | Probe set-number | BAC clone name | Label | GGA Chr | FPE Chr |
|------------------|---|-----------|---------|---------|------------------|---|-----------|---------|---------|
| 1 | CH261-130M12 CH261-16016 CH261-1819 | FITC | 3 | 11 | 11 | CH261-121N21 CH261-138H13 CH261-154H1 | FITC | 11 | 14 |
| 2 | CH261-107E2 CH261-168O17 CH261-184E5 | FITC | 1 | 4 | 12 | CH261-10F1 CH261-50H12 | Texas Red | 4 | 2 |
| 3 | CH261-9B17 CH261-83O13 CH261-58K12 | Texas Red | 1 | 4 | 13 | CH261-2I23 CH261-78F13 | FITC | 5 | 1 |
| 4 | CH261-97P20 CH261-120H23 CH261-17B14 | Texas Red | 3 | 7 | 14 | CH261-44H14 CH261-44D16 | FITC | 2 | 3 |
| 5 | CH261-119K2 CH261-120J2 CH261-25P18 | Texas Red | 1 | 6 | 15 | CH261-40G6 CH261-50C15 | FITC | 2 | 5 |
| 6 | CH261-96D24 TGMCBA-208D17 CH261-34H16 | Texas Red | 8 | 10 | 16 | CH261-185L11 CH261-89P6 | FITC | 4 | 2 |
| 7 | CH261-103F4 CH261-154H17 CH261-65O4 | FITC | 24 | 15 | 17 | CH261-93H1 CH261-85H10 | FITC | 19 | 2 |
| 8 | CH261-18G17 CH261-40J9 TGMCBA-15I122 | FITC | 22 | 17 | 18 | CH261-42P16 CH261-72P11 | Texas Red | 17 | 1 |
| 9 | CH261-170L23 CH261-186M13 TGMCBA2797G21 | FITC | 26 | 16 | 19 | CH261-10A18 CH261-122K8 CH261-83I20 | Texas Red | 21 | 3 |
| 10 | CH261-100E5 CH261-28L10 CH261-66M16 | FITC | 27 | 18 | 20 | CH261-101C8 TGMCBA-37M13 | Texas Red | 28 | 5 |
| | | | | | 21 | CH261-67N15 CH261-72B1 | Texas Red | 18 | 2 |

6.3.2 Specific Aim 4b: Mapping the average chromosome position of all selected BACs in metaphase and interphase

In total, 54 BACs (forming 21 sets) were successfully hybridized on chicken and Peregrine falcon metaphases and interphases (Figure 6-1).

Figure 6-1: FISH hybridization results of selected probe sets to GGA and FPE metaphase chromosomes.

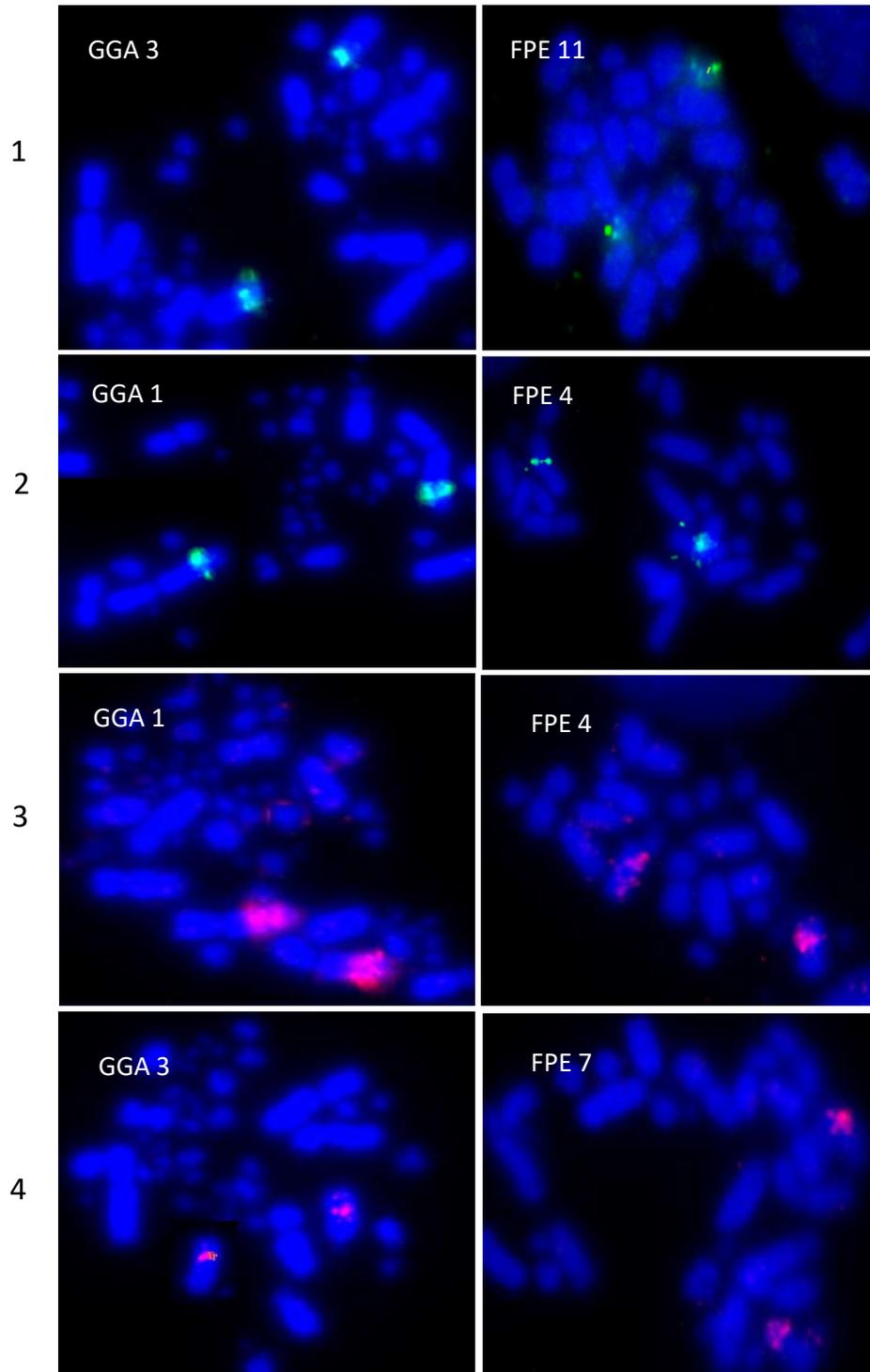


Figure 6-1 (Continued): FISH hybridization results of selected probe sets to GGA and FPE metaphase chromosomes.

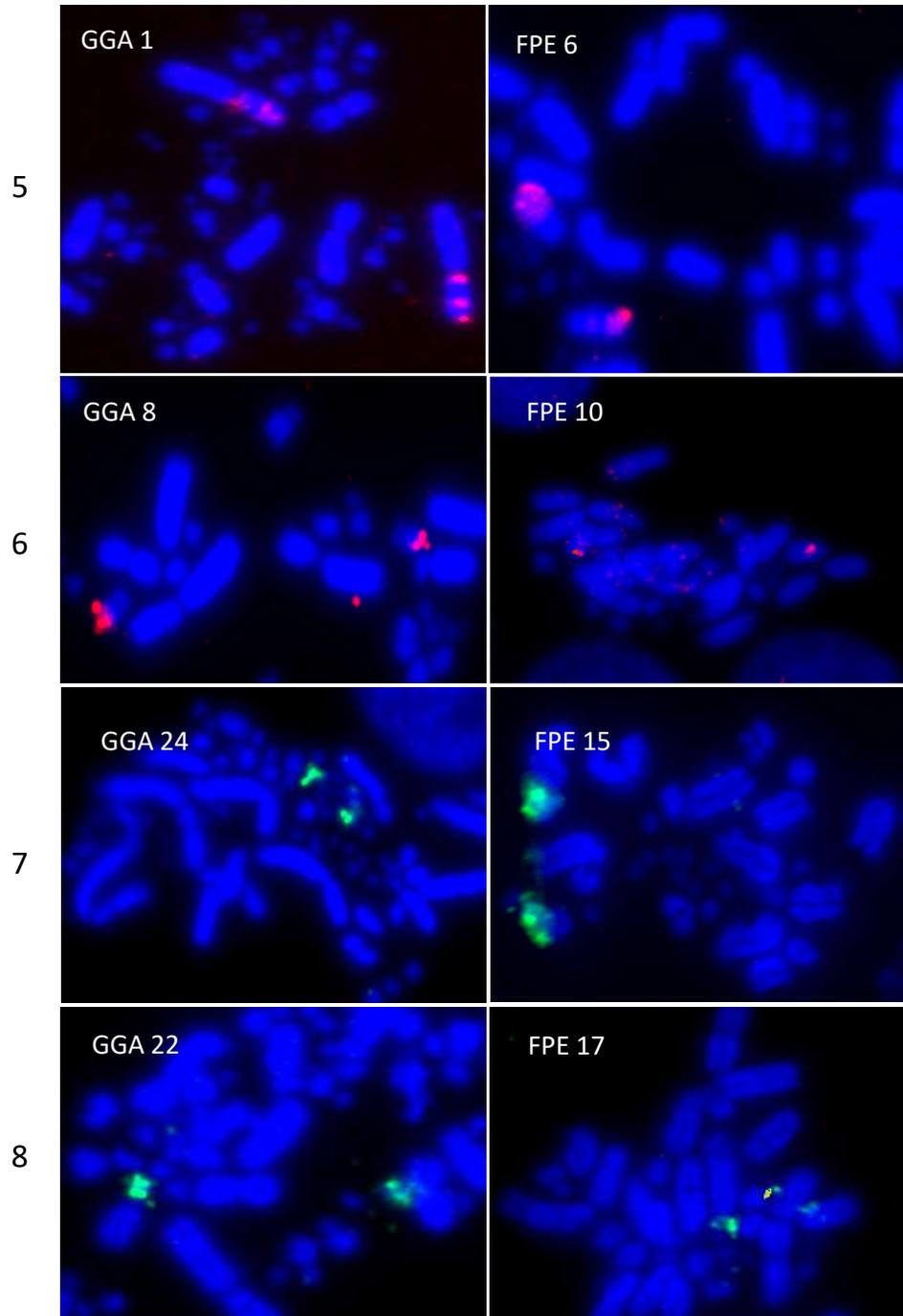


Figure 6-1 (Continued): FISH hybridization results of selected probe sets to GGA and FPE metaphase chromosomes.

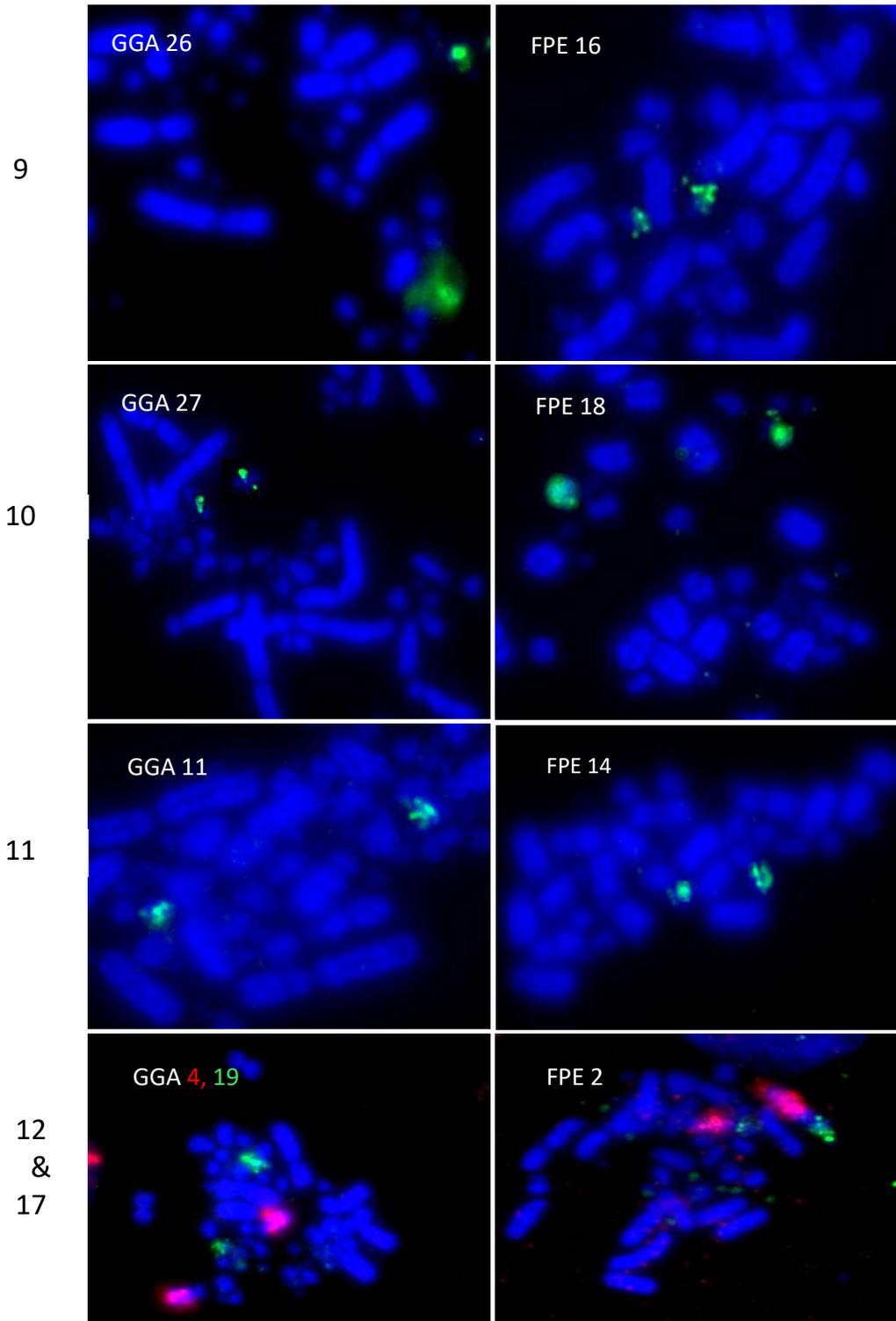
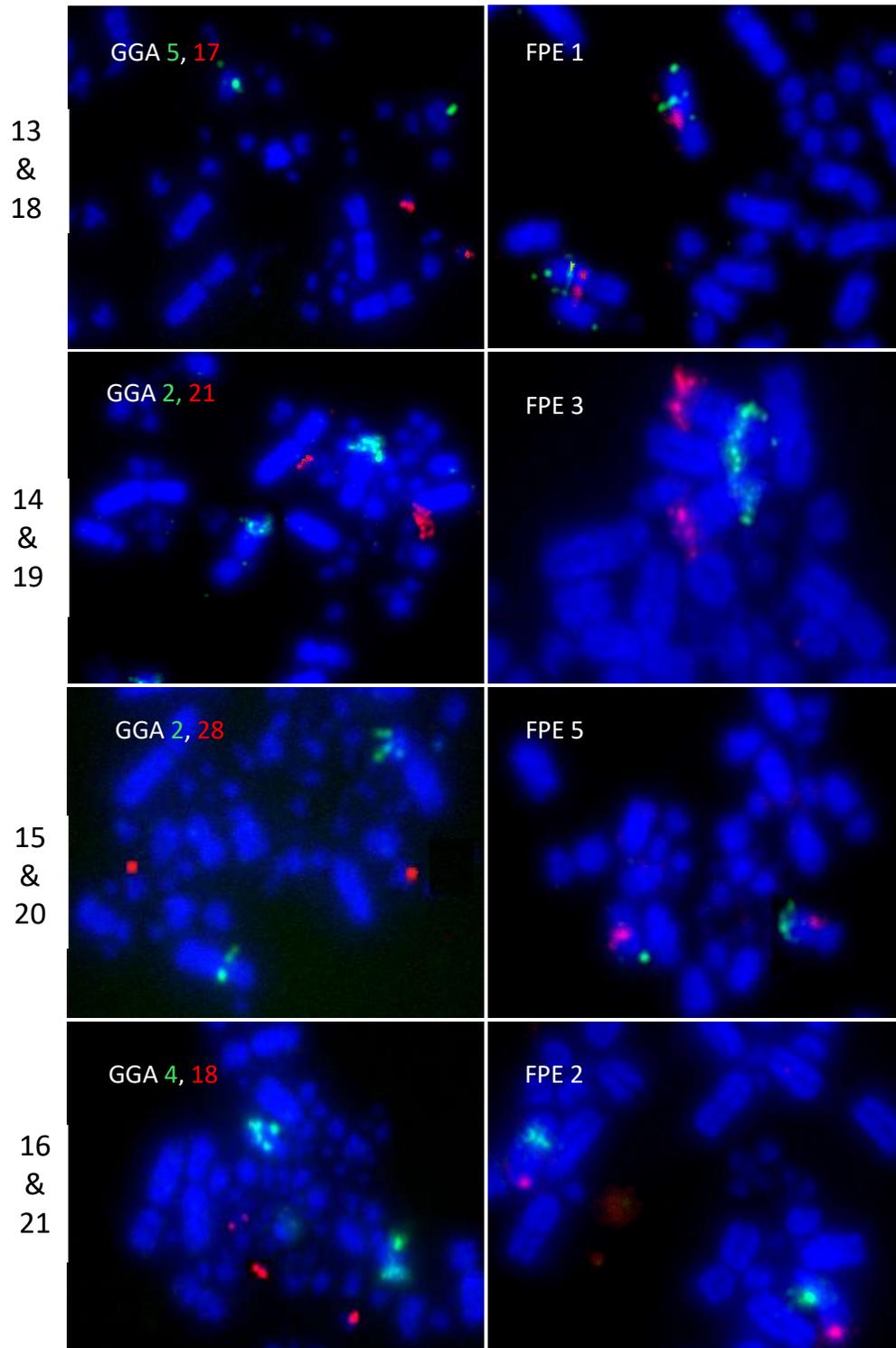


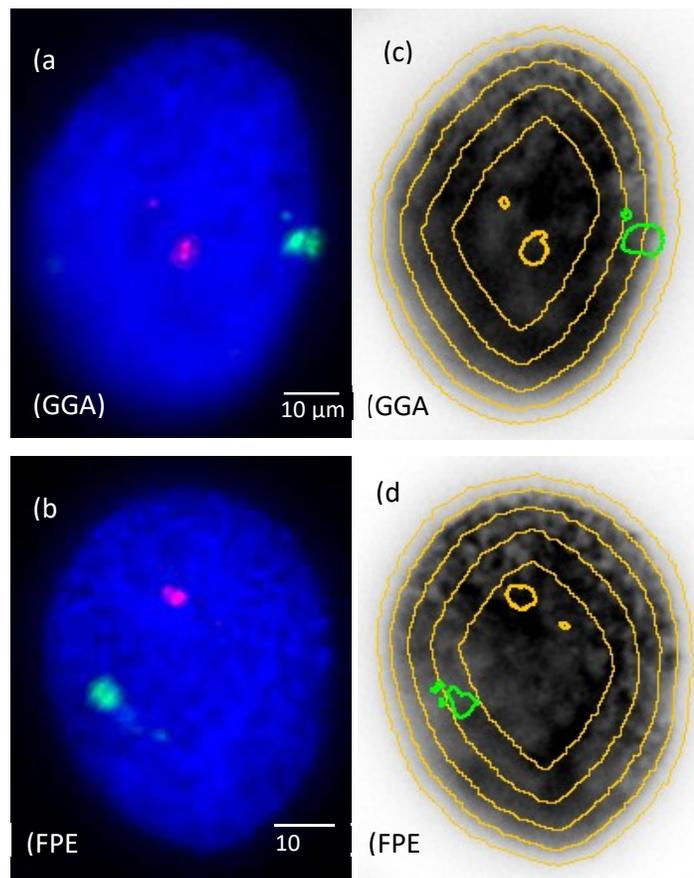
Figure 6-1 (Continued): FISH hybridization results of selected probe sets to GGA and FPE metaphase chromosomes.



6.3.2.1 Identification of chromosome territory position

Chromosome territory position analysis on FISH images was carried out in ImageJ using the plugin Nuclear Morphology Analysis version 1.13.7 (<https://bitbucket.org/bmskinner/nuclearmorphology/wiki/Home>). As outlined in section 2.7 of the materials and methods, each nucleus was divided into 5 shells of equal area and the distribution and the percentage of signal within the nucleus was measured. An image of a chicken and Peregrine falcon nucleus showing signals for 2 different probe sets along with the five-ring template overlaid is illustrated in (Figure 6-2). The overall position of chromosomes (from periphery, 0 to interior, 4) in chicken and Peregrine falcon nucleus are summarized in Figure 6-3.

Figure 6-2: DAPI stained chicken (a) and Peregrine falcon fibroblast interphase nucleus (b) showing signals with dual colour FISH probe set 3 and 4; (c) and (d) same images after shell analysis, dividing the nucleus with five concentric shells of equal area showing two different signals within the nucleus.



6.3.3 Specific Aim 3c: Testing the hypothesis that macrochromosomes are peripherally located, microchromosomes are centrally located but “former” microchromosomes still “behave” as though they are microchromosomes and thus occupy a central position

As illustrated in Figure 6-4 onwards, results suggest that the equivalent probes for macrochromosomes occupy similar nuclear positions in both chicken (GGA) and Peregrine falcon (FPE).

Figure 6-3: Chart showing the mean overall position of chromosomes (from periphery, 0, to interior, 4) in chicken and Peregrine falcon nucleus.

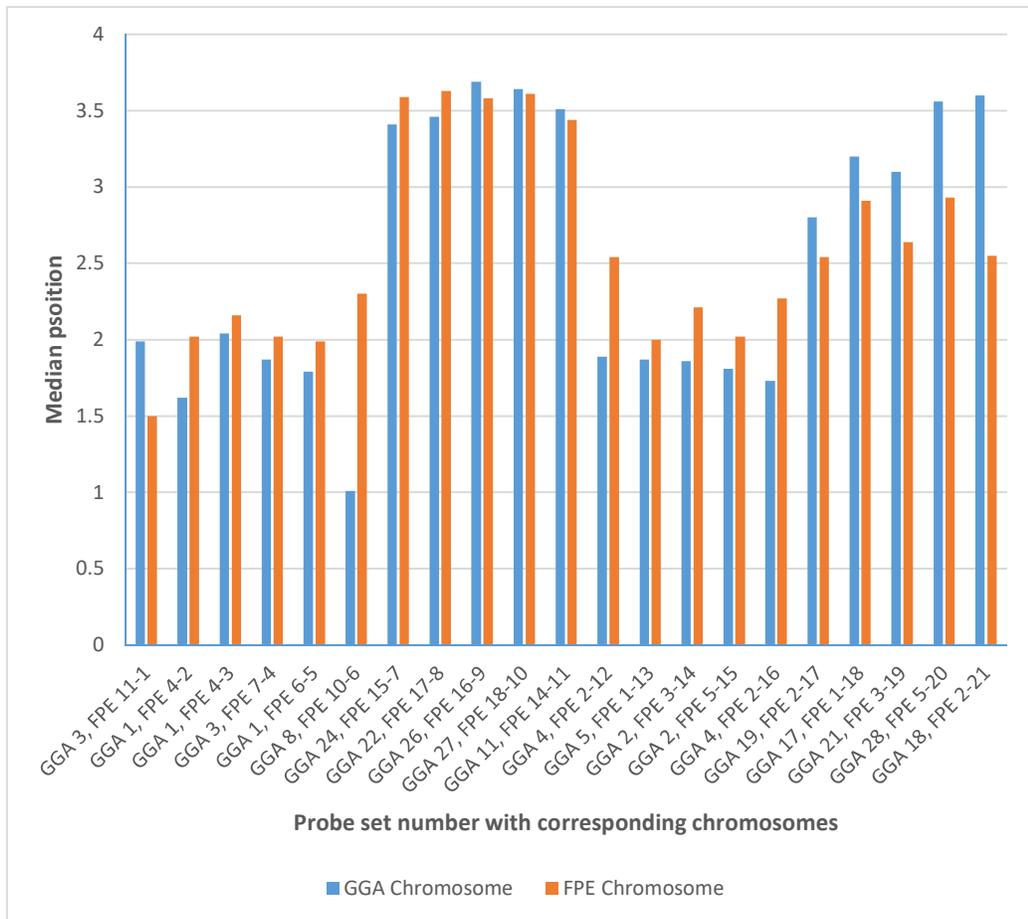


Figure 6-4: Signal distribution for macrochromosomes in chicken (GGA) and Peregrine falcon (FPE). Orange indicates GGA and blue indicates FPE.

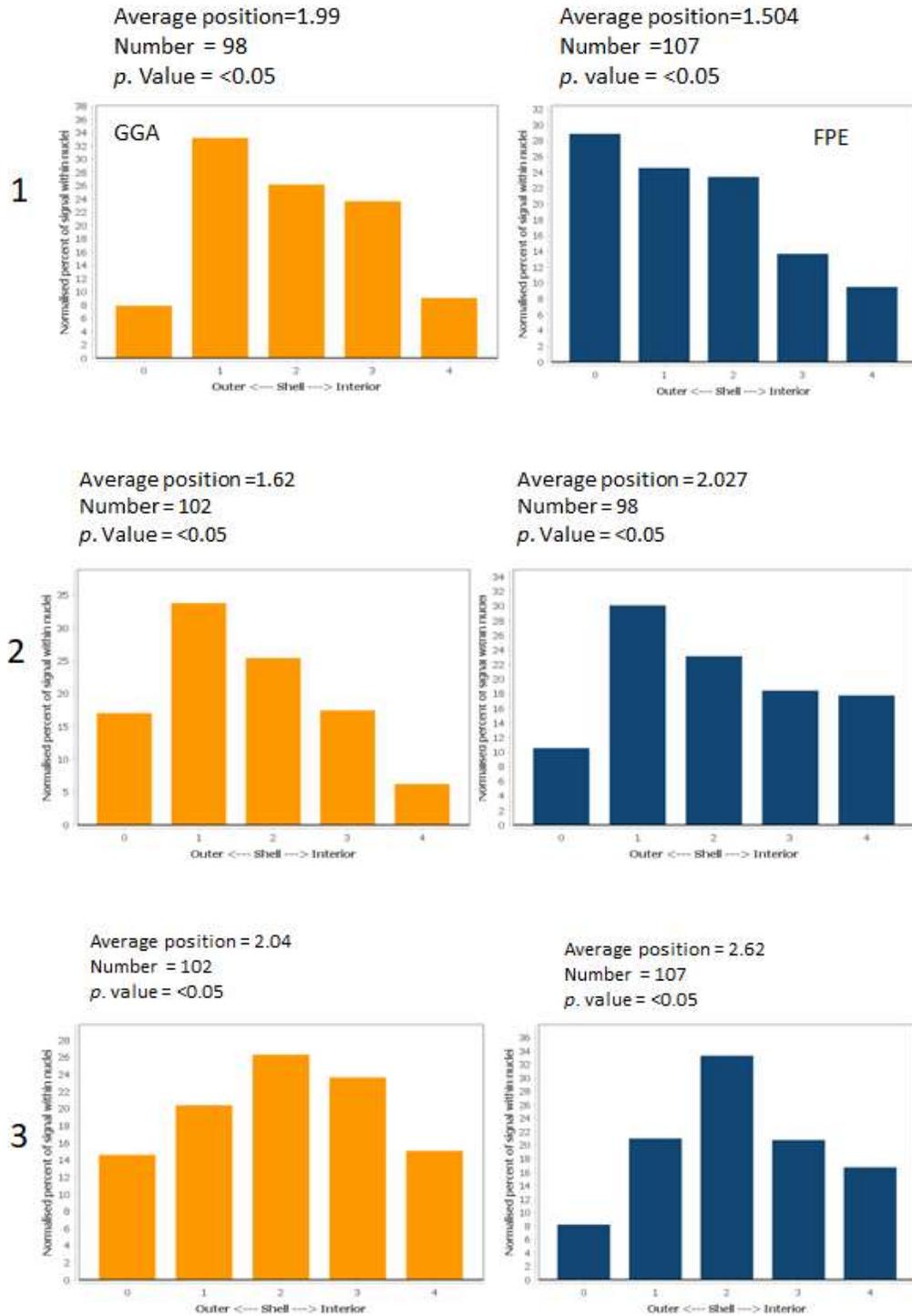
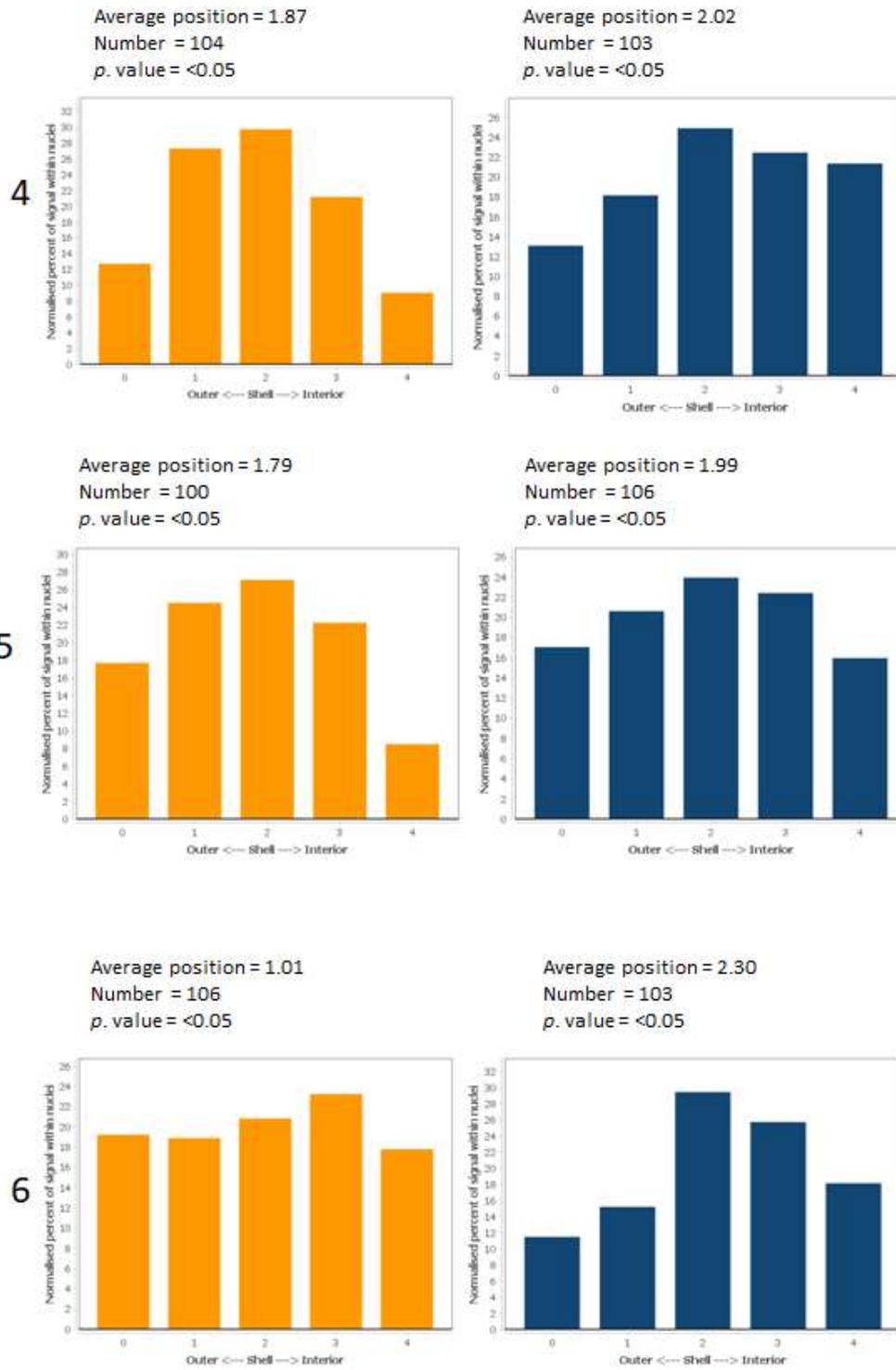


Figure 6-4 (Continued): Signal distribution for macrochromosomes in chicken (GGA) and Peregrine falcon (FPE). Orange indicates GGA and blue indicates FPE.



Results suggest that the microchromosomes tested are centrally located in both chicken and Peregrine falcon (Figure 6-5).

Figure 6-5: Signal distribution for microchromosomes in chicken (GGA) and Peregrine falcon (FPE). Green indicates GGA and brown indicates FPE.

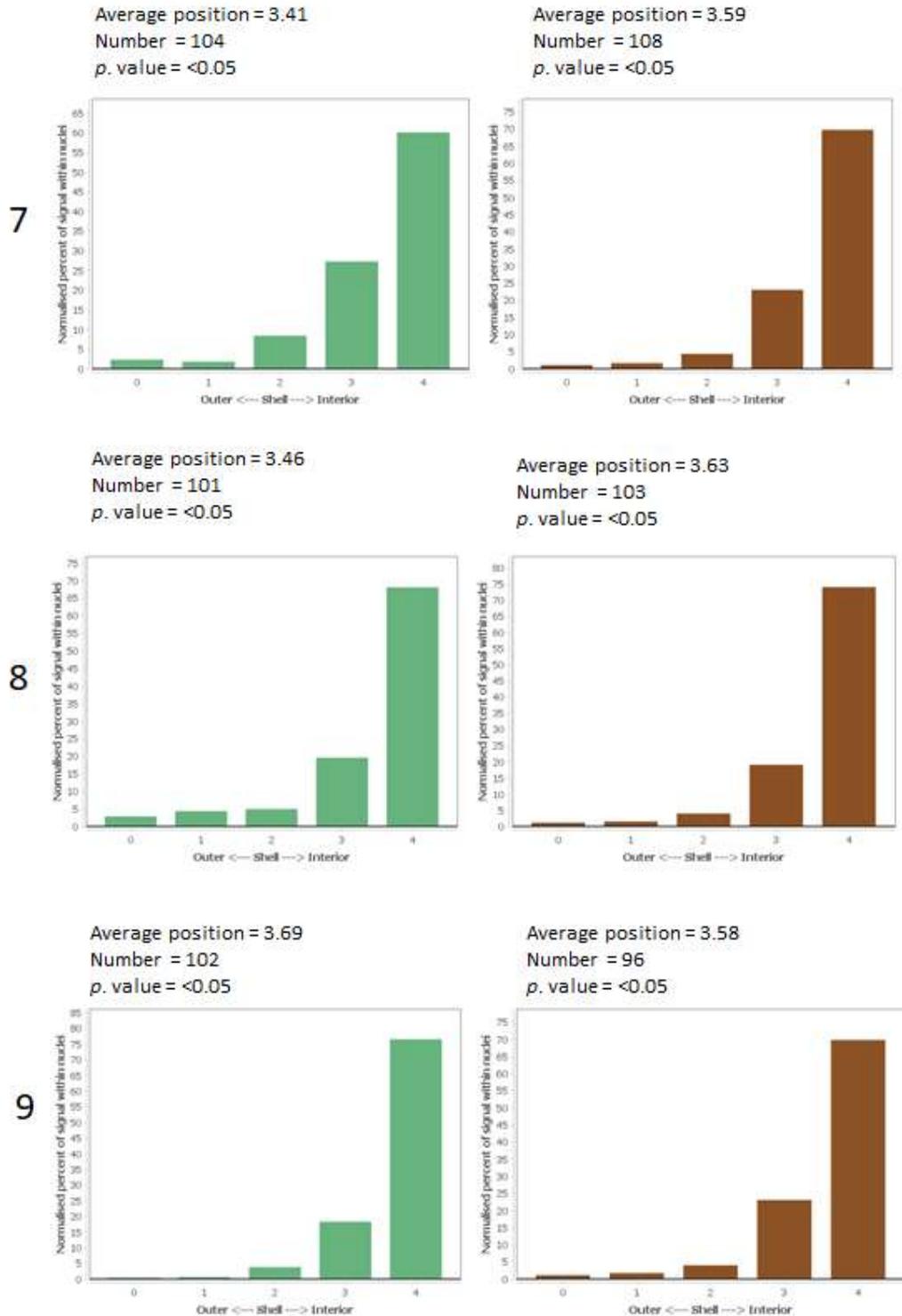
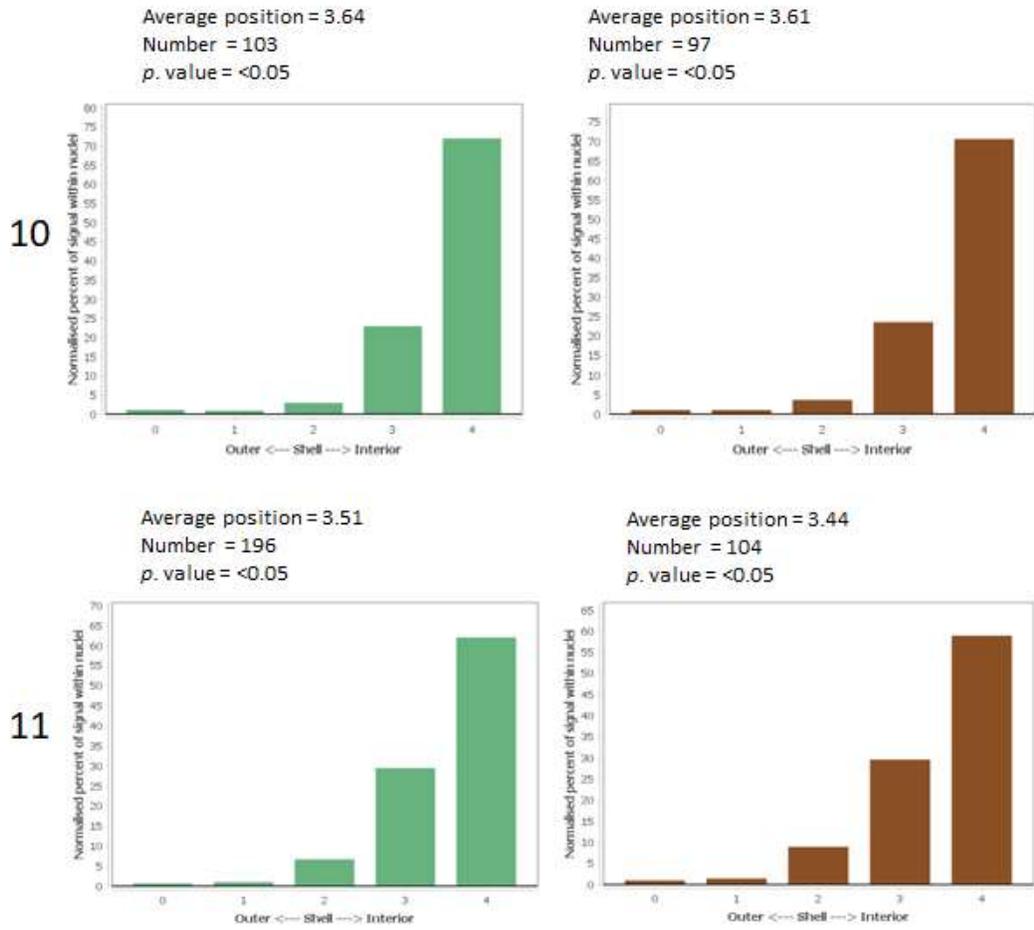


Figure 6-5 (Continued): Signal distribution for microchromosomes in chicken (GGA) and Peregrine falcon (FPE). Green indicates GGA and brown indicates FPE.



Results suggest that macrochromosomes that now have microchromosomes fused to them (in falcon) broadly occupy similar positions (Figure 6-6).

Figure 6-6: Signal distribution for macrochromosomes in chicken (GGA) and Peregrine falcon (FPE). In FPE, they were formerly macrochromosomes but now are fused with microchromosomes. Orange indicates GGA and violet indicates FPE.

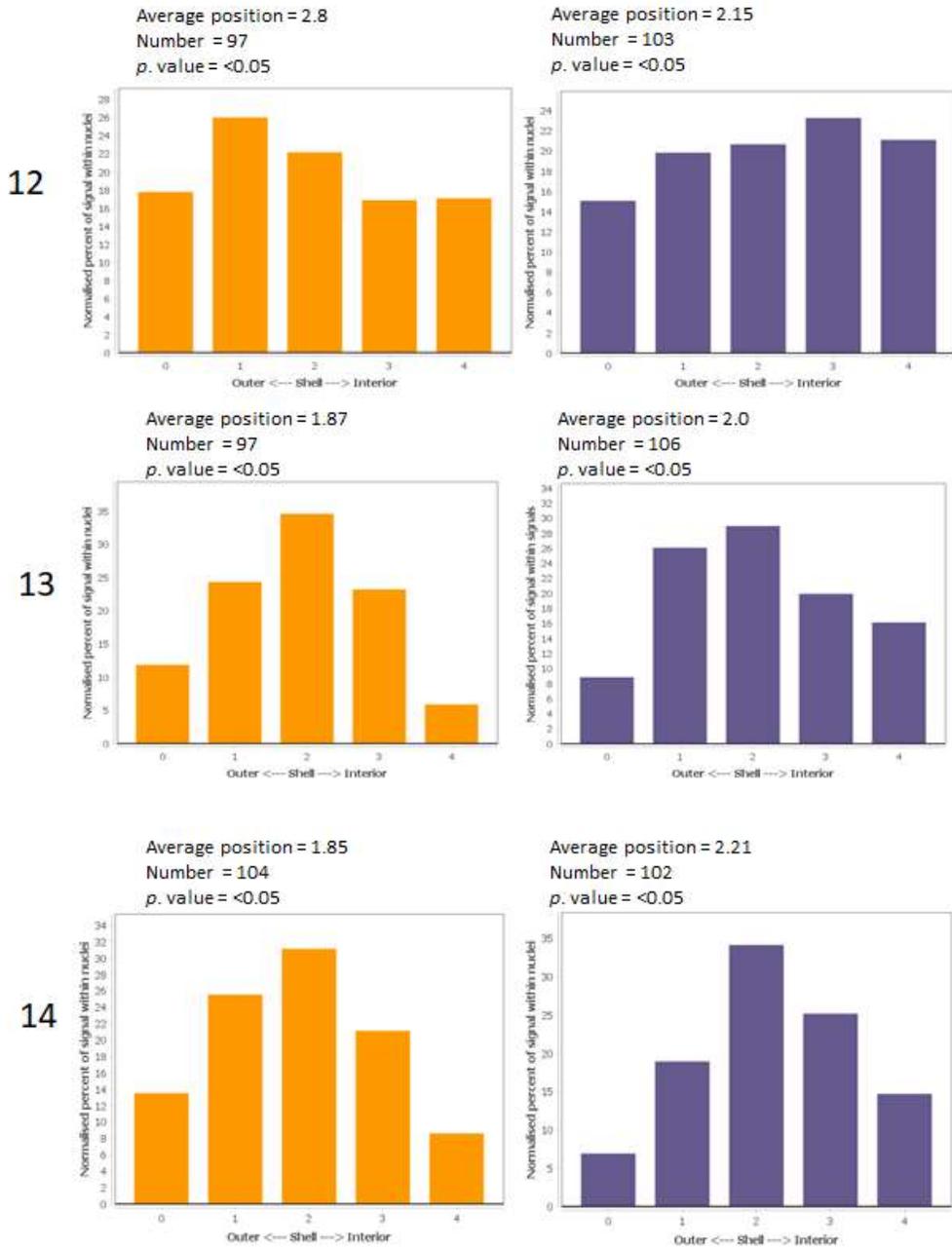
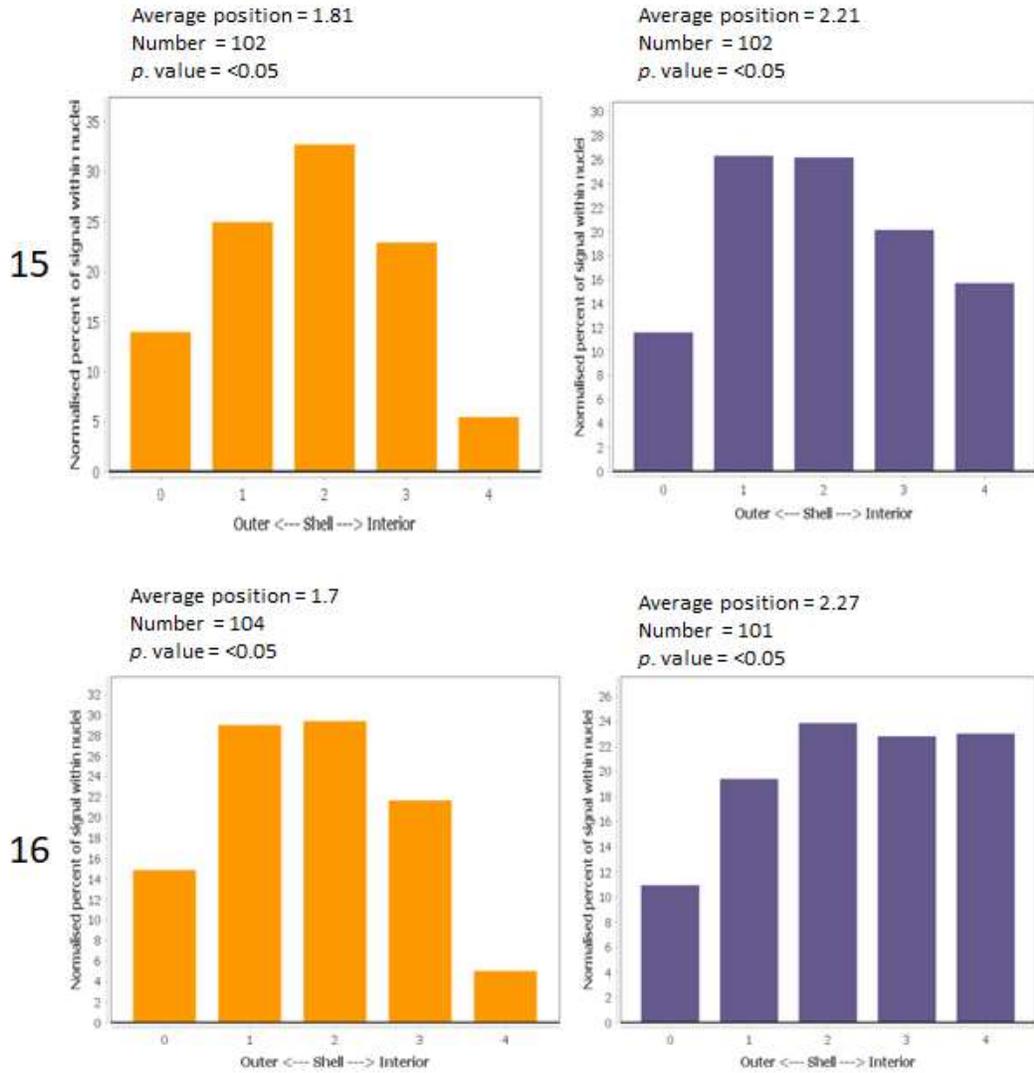


Figure 6-6 (Continued): Signal distribution for macrochromosomes in chicken (GGA) and Peregrine falcon (FPE). In FPE, they were formerly macrochromosomes but now are fused with microchromosomes. Orange indicates GGA and violet indicates FPE.



These results confirm the hypothesis that fused microchromosomes occupy a central position like their chicken counterparts (Figure 6-8). It is noteworthy however that the mean position in falcon is slightly more towards the periphery (Figure 6-7).

Figure 6-7: Chart showing the overall position of microchromosomes that are fused to macrochromosomes in Peregrine falcon and their corresponding chromosome position in chicken.

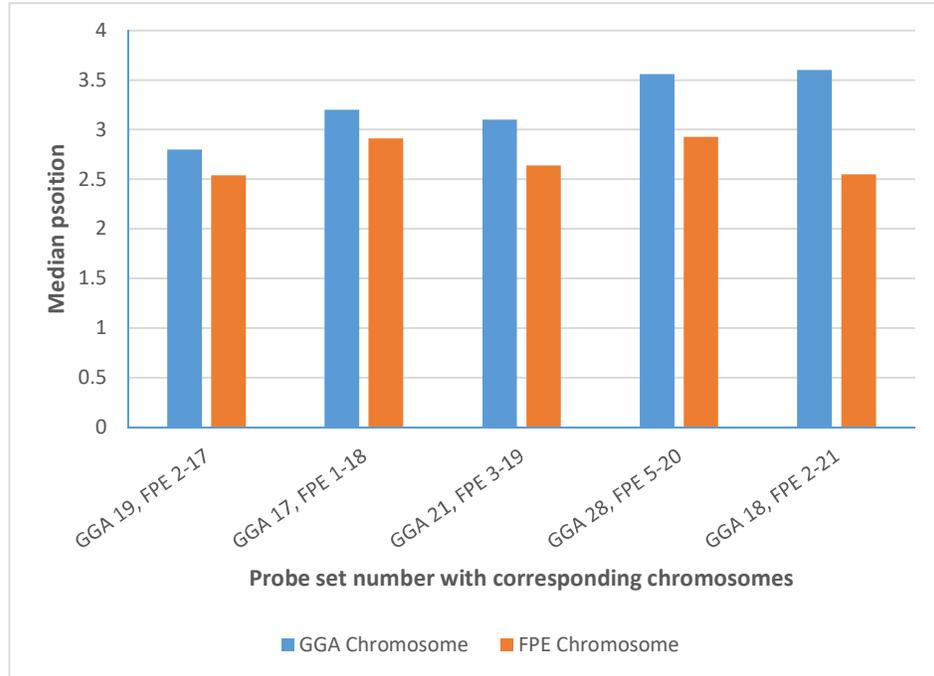


Figure 6-8: Signal distribution for microchromosomes in GGA and FPE. In FPE, they were formerly microchromosomes but now are fused with macrochromosomes. Green indicates GGA and purple indicates FPE.

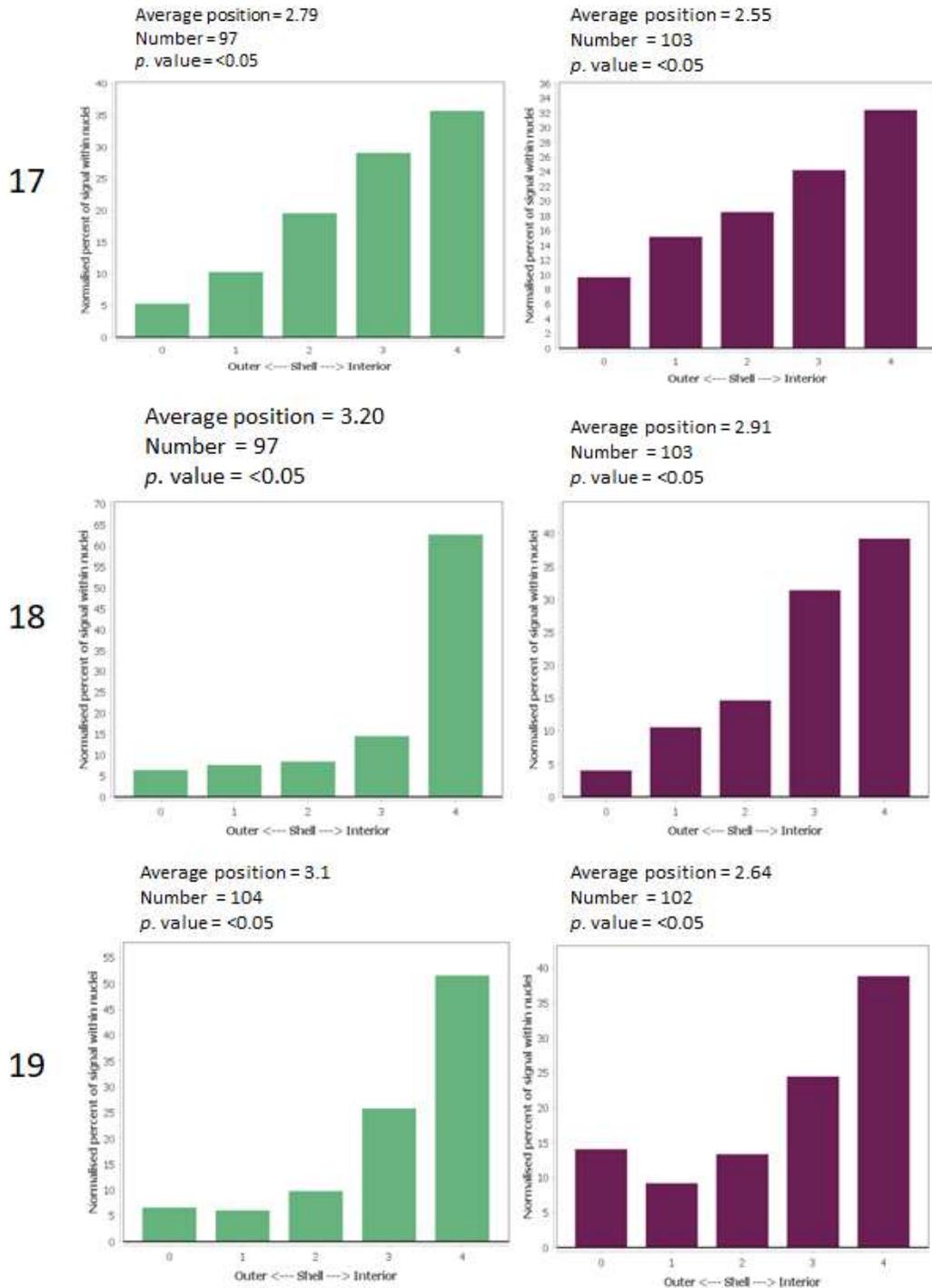
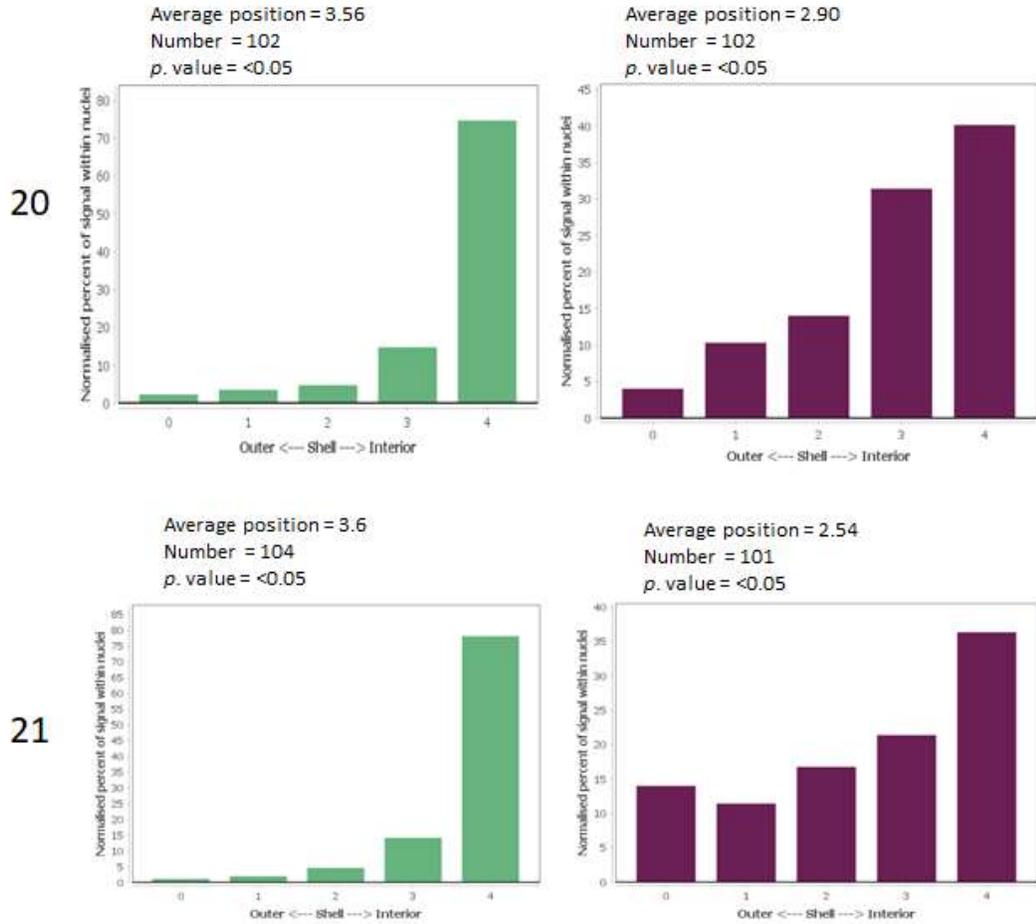


Figure 6-8 (continued): Signal distribution for microchromosomes in GGA and FPE. In FPE, they were formerly microchromosomes but now are fused with macrochromosomes. Green indicates GGA and purple indicates FPE.



In terms of the original hypothesis therefore the positions of the chromosome were largely similar, regardless to the chromosome on which there are attached. In other words, the hypothesis is accepted that “former microchromosome” occupy a central location despite being attached to larger chromosomes as a result of evolutionary events. Visual inspection of the graphs in Figure 6-8 suggests however that some “former microchromosomes” in FPE are not quite so centrally located as their counterparts in chicken.

7 Discussion

This thesis has made a significant contribution towards the understanding of genome organization and evolution in the genus *Falco*. From a 'molecular cytogenomic' point of view, this thesis was successful in fulfilling its stated aims as follows:

1. Conventional characterization of the Saker falcon karyotype ($2n=52$) was successful. This species is classified as 'endangered' according to the International Union for Conservation of Nature (IUCN) Red List of Threatened Species (IUCN 2017). This thesis has produced improved karyotypes and ideograms than those previously published on the Peregrine falcon and Gyrfalcon and made initial comparative genomic analyses among 3 falcon species using molecular and sequence-based approaches. This study has supported upgrading the fragmented Saker genome assembly to the chromosome level using a newer approach hitherto only published for the Peregrine falcon (and pigeon). This method combines computational algorithms to merge scaffolds into chromosomal fragments, with PCR-based scaffold verification and physical mapping to chromosomes. In particular, the present study performed physical mapping of PCFs directly to Saker chromosomes using a universal set of BACs applicable to all birds.
2. A comparison of genome-wide BAC-based comparative studies in three falcon species and bioinformatic analysis (MGRA2) revealed the chromosomal changes (inter- and intra-) that led to the falcon lineage. Also, the present study established that common mechanisms of chromosomal fusion do not recur in two different groups of species with rearranged karyotypes (falcons and parrots).
3. This work has provided an overview of the telomeric DNA profile in three different falcon species. It has established that the highly rearranged karyotypes studied (plus those of the budgerigar and crocodile) do not appear to possess interstitial telomeres at evolutionary fusion points. Also, the thesis demonstrated the existence of mega-telomeres in falcon species, their nature differing between Peregrine and the other two species studied.

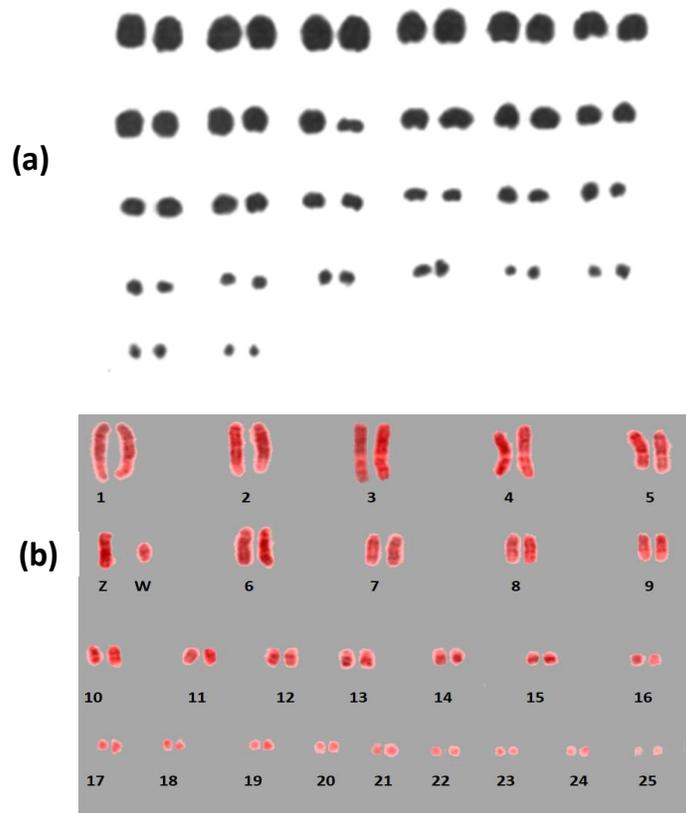
4. This study gives the first detailed description of nuclear organization in a bird species (Peregrine falcon) other than the *Galloanserae*. Non-fused macro and microchromosomes behave the same way in chickens and falcons. This implies that the same general nuclear organization mechanisms are present in falcons as well as in chickens, ducks and turkeys, whose last common ancestor existed around 89 million years ago. Most notably, fused microchromosomes in the Peregrine falcon retain the same nuclear organization pattern which has been evolutionarily conserved for more than 300 million years. The findings from this study give insight into the basic nature of chromosome territory patterns in bird species with highly rearranged karyotypes.

7.1 The overall structure (karyotype) of the Saker falcon genome and links to the genome assembly

Extensive karyotype studies in the genus *Falco* are not recorded even though falcons have significant importance in the Middle East for hunting and conservation, as well as for their characteristic chromosome features ('atypical' genome organization). The genus *Falco* consists of around 40 different species (<https://en.wikipedia.org/wiki/Falcon>) only 10 of which have been partially karyotyped (del Hoyo *et al.* 1994): *Falco columbarius* (Merlin falcon) (2n=40), *Falco mexicanus* (Prairie falcon) (2n=48), *Falco chicquera* (Red-necked falcon), *Falco jugger* (Laggar falcon), *Falco sparverius* (American kestrel), *Falco Subbuteo* (Eurasian hobby) and *Falco peregrinus* (Peregrine falcon) (2n=50), *Falco rusticolus* (Gyrfalcon) and *Falco tinnunculus* (Common kestrel) (2n=52) and *Falco biarmicus* (Lanner falcon) (2n= 52 or 54) (de Boer 1975; Belterman and de Boer 1984; Sasaki *et al.* 1984; Schmutz and Oliphant 1987; Longmire *et al.* 1988). Before the introduction of FISH techniques, karyotypic analysis gained intense attention for its ability to reveal cytological traits which provided a basis for an understanding of evolution (Liang and Chen 2015). Karyotypic studies also can provide information about diseases and tumorigenesis, and they represent a low resolution of the whole genome (Masabanda *et al.* 2004). Moreover, karyotypic variation involving chromosome number and structure plays an important role in evolution and speciation (Weiss-Schneeweiss *et al.* 2008).

This thesis, is the first to describe the karyotype for the Saker falcon ($2n=52$), a species which is classified as 'endangered' according to the International Union for Conservation of Nature (IUCN) Red List of Threatened Species (IUCN 2017). Also, it has produced improved karyotypes and ideograms than previously published for the Peregrine falcon and Gyrfalcon (Figure 7-1). Moreover, karyotypes of these birds have provided an important guideline for comparative cytogenomic studies as well as for the precise and easy mapping of each BAC on the chromosome.

Figure 7-1: Saker falcon karyotype (resolution) comparison of previous studies with present study (a) Amaral *et al.* 2003, (b) current study.



There are few reports of cytogenetic studies for falcons (O'Connor 2016; Lithgow *et al.* 2014; Al Mutery 2011) and only one comprehensive zoo-FISH study (Nishida *et al.* 2008) has characterized the chromosomes to provide a baseline understanding of the overall

genome structure. However, its usefulness is restricted due to the limitations of zoo-FISH.

With the genome sequencing of the Peregrine and Saker falcons, falcon genomic studies have gained tremendous momentum (Zhan *et al.* 2013). In addition to the Peregrine falcon (Damas *et al.* 2017), physical mapping of the Saker falcon (this thesis) has upgraded its genome status from a sub-chromosomal level genome assembly to that of a chromosomally assembled genome. The overall strategy used here for scaffold assembly by RACA, and physical mapping using a panel of universal BACs, provides an additional example for this approach which could be applied to any animal genome. Moreover, the chromosomally assembled Saker falcon genome provides an additional reference genome which has a different karyotype pattern from that of the Peregrine falcon. Furthermore, by uploading the chromosomally-assembled Saker falcon genome in Evolution Highway, users will be able to compare multiple species, including the Saker falcon, in order to identify evolutionary breakpoint regions and regions of synteny.

One of the prime roles of whole genome sequences is to provide a better understanding of evolutionary history of genome organization and chromosome structural variation caused by chromosome rearrangements (Kim *et al.* 2013). We live in an era where genome sequencing of new species is rapidly progressing with the advancement of various modern sequencing technologies (Groenen *et al.* 2012); this is summarized in 1.3.1. Though such modern DNA sequencing methods have advantages, especially being more cost effective, each time the genome of a new species is sequenced, it is often problematical and very expensive to assign large blocks of sequence to an overall genomic “map”.

Multiple projects, such as the Bird 10K project are working to generate draft genome sequences of thousands of extant bird species over the next couple of years using next generation sequencing (NGS) technologies, producing *de novo* assemblies (Kim *et al.* 2013; Zhang *et al.* 2015). *De novo* genome assembly efforts always aim to create a ‘chromosome level’ assembly, however, the limitation of NGS read length makes it very challenging to assemble the reads into chromosomes for large genomes (Damas *et al.* 2017; Kim *et al.* 2013). The number of avian genomes with physical or genetic maps for anchoring the assemblies to chromosomes is far less than the number of genomes

sequenced by NGS technology. Without assembling a whole genome sequence to the level of one “super scaffold” per chromosome, the resultant assembly can be studied for gene structure and function but cannot be utilized effectively to address biological questions related to critical aspects of avian genome evolution. It is essential to have an accurate mapping (Lewin *et al.* 2009) of each sequence attached to a specific locus on a chromosome for every *de novo* genome assembly of those species not previously studied. The core aim of a *de-novo* genome assembly is to represent all the sequences correctly mapped contiguously, but it does not always provide the information needed to map the new genomes to the chromosomes assemblies in scaffold format. When this method falls short of this “chromosome level assembly” in whole genome sequences, their use is significantly limited in the critical aspects of evolutionary and applied biology of species. For example, chromosome level assemblies have been essential for agricultural species where an established order of DNA markers is required to establish phenotype-to-genotype associations for gene-assisted selection and breeding (Andersson and Georges 2004). High-resolution SNP genotyping is very effective for association studies among different species which in turn facilitated the mapping of Mendelian disorders, accurate identification of (e.g. cryptic) chromosome translocations, discovery of quantitative trait nucleotides (QTNs) and expression quantitative trait loci (eQTLs), and studies of long-range regulatory interactions (Hu *et al.* 2009). This has resulted in significant economic improvement, more efficient food production and improved global food security (Hu *et al.* 2009).

Chromosome level assemblies are essential to establish genomic selection and genome-assisted breeding and/or conservation regimes and to study genome-phenome relationships, particularly of complex (e.g. quantitative) traits. Comparative genomics becomes possible *in-silico* (Farré *et al.* 2016) when such assemblies are built for multiple species. Moreover, they help to identify chromosome rearrangements by selection of (e.g. BAC) probes for FISH (Damas *et al.* 2017) which are not easily detected by basic karyotyping (e.g. cryptic translocations) (O’Connor *et al.* 2017). To generate chromosome level assemblies, extensive funding resources are needed which is an extremely pressing challenge.

Birds are among the most speciose of vertebrates with around 10,000 extant representatives. A large number are on the CITES threatened / endangered list; they are critical to agriculture (both meat and eggs), and many are models for human disease and development. With impending global warming, tools for the study of ecology and conservation of birds are essential. It is paramount to construct complete and reference-able genome assemblies for many of the numerous *de-novo* avian vertebrate genome projects that have recently been generated. Without this work, newly sequenced genomes will remain simply catalogues of genes (at best, collections of scaffolds) with limited reference to the overall genomic structure and organization.

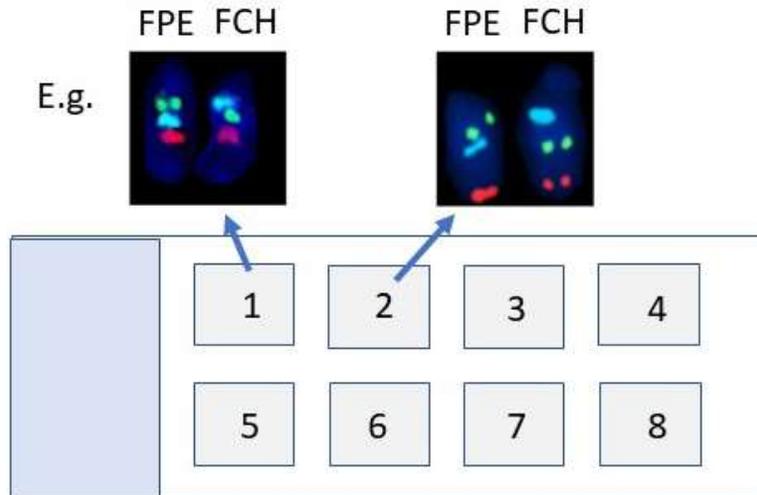
To resolve this problem, Damas *et al.* (2017) developed a novel, inexpensive and transferrable approach and tools for assigning the sequences to their proper position in chromosomes. In this method, different processes such as computational algorithms for ordering scaffolds into PCFs, PCR verification of scaffolds are combined with physical mapping directly to chromosomes with a universal set of avian bacterial artificial chromosome (BAC) probes.

This study is the first to establish, comparative genomic maps between three *Falco* species; the Peregrine falcon, Saker falcon and Gyrfalcon. It provides a direct, genome-wide overall chromosome homology between these birds and chromosome rearrangements that have occurred in each species since they diverged from a common ancestor approximately 2.1 million years ago (Zhan *et al.* 2013). Comparative genomic maps generated among these falcons are quite similar, with complete synteny between Saker falcon and Gyrfalcon. Lack of apparent chromosome rearrangements between Gyrfalcon and Saker falcon can indicate that they could be considered the same species. Helbig *et al.* (1994) conducted a phylogenetic relationship study among Falcon species based on cytochrome b gene variations. His study reported that the Saker falcon mtDNA haplotypes are almost identical to those of the Gyrfalcon. Moreover, Gyrfalcons and Saker falcons can produce fertile hybrids in the wild as well as in captivity with extended viability over indefinite generations (Eastham 2005; Heidenreich 1997). It is worth mentioning that no scaffold-based assemblies for Gyrfalcons on which genome mapping can rely.

The present study corroborates the findings of a cytogenomic analysis conducted by Zhan *et al.* (2013) who reported synteny between Peregrine falcons and Saker/Gyr falcons. The present results show, one inter-chromosomal and 9 intrachromosomal changes between them. The reduction of chromosome diploid number $2n=50$ found in the Peregrine falcon originated from the centric fusion of a Saker/Gyrfalcon chromosomes 7 and 9, which forms the metacentric chromosome 1 of the Peregrine falcon and thus forms the chromosome signature in this falcon species. This finding agrees with previous studies (Nishida *et al.* 2008). Inversions are the main cause for the intrachromosomal rearrangements, involving 7 chromosomes. Apart from the comparative Zoo-FISH data generated in 3 falcon species using chicken chromosome paints (Nishida *et al.* 2008), to the best of my knowledge, comparative genomic maps presented here represent the first, extensive study performed between these three-falcon species (Saker falcon, Gyrfalcon and Peregrine falcon).

The detection of hybrid falcons is becoming increasingly important in falcon racing. A total number of 9 intrachromosomal differences were identified between the Peregrine falcons and Saker falcons which can form the basis for establishing a testing device (FISH based) that could detect hybrids (Peregrine x Saker, Peregrine x Gyr). Such a device could have 8 spatially separated squares, each of which carries specific DNA FISH probes that are labelled and designed to identify Peregrine and Saker/Gyr chromosomes. Each square will have 3 probes which will be directly labelled with a different coloured fluorophore, green (FITC), red (Texas Red) and aqua (Alexa Fluor) (Figure 7-2).

Figure 7-2: Schematic representation of proposed device showing different hybridization pattern on Peregrine falcon chromosome (FPE) and Saker falcon chromosome (FCH).



For instance, in a hybrid falcon, 75% Saker and 25% Peregrine, at least one of the squares is expected to have the Saker pattern for one chromosome homologue and the Peregrine for the other chromosome. Commercially such devices can be designed based on directly-labeled BAC probes for simultaneous analysis of complex chromosomal rearrangements between Peregrine and Saker/Gyr. The arrangement of DNA FISH probes on such a device should be designed to facilitate the identification of the presence of Peregrine and Saker chromosomes in a sample. Developing such a testing tool to identify Gyr x Saker falcon hybrids is not possible as a result of there being no apparent intrachromosomal differences between them.

7.1.1 Application for falcon hybrids

A wide range of falcon hybrid species has been produced by artificial insemination in the *Falco* genus. A skilled falconer knows each falcon's good and bad qualities. To produce the best performing falcons, falconers usually choose cross breeding techniques. Due to hybrid vigor, falconers are mostly successful in producing the strong traits of parents in their offspring. In the U.A.E, Gyrfalcon, Saker and Peregrine falcons are usually crossed and frequently they are further crossed creating 1/2, 1/4, 3/4, 3/8, 5/8 hybrids and so

on. These numbers represent a simple way to indicate the proportion of genes the hybrid inherit from the parent species.

Falcon racing has become a most prestigious sport where captive falcons are competing for millions of Dirham's worth of prizes. It is often difficult to identify hybrids even for an expert using accurate morphometric characters for identification. This can be a risk to the integrity of falconry. To date, there are no available DNA markers to identify different falcon species, a concern which has recently drawn the attention of scientists around the world.

Future research should characterize interspecific hybrids between Peregrine and Saker, and Peregrine and Gyrfalcon species from F1, F2 and F3. Detailed comparative cytogenomic data generated within the present study show chromosomal rearrangements which have yet to be characterized in different falcon hybrids. Ultimately this would help to develop a testing method to distinguish the origin of Falco chromosomes in hybrid species.

7.1.2 Atypical karyotypes

Falconidae and *Accipitridae*, together with *Psittaciformes* members, are recognized as avian species with 'atypical' karyotypes and previous studies have shown that these avian species have the highest numbers of rearrangements occurring on their macrochromosomes (de Oliveira *et al.* 2015; Nie *et al.* 2015; Nishida *et al.* 2008; Seibold-Torres *et al.* 2015). Recent cytogenomic studies conducted on such 'atypical' karyotype species collectively highlight the substantial amount of rearrangements of macrochromosomes that have occurred throughout their evolutionary history. Until now, the only information available was that of their homology to chicken chromosomes GGA 1-9 and Z while knowledge of their microchromosomal rearrangements is still limited (de Oliveira *et al.* 2015; Nie *et al.* 2015; Nishida *et al.* 2008; Seibold-Torres *et al.* 2015). Therefore, by performing microchromosomes analysis and thus evaluating the inter-and intrachromosomal relationship between more falcon species, Gyps species and *Psittaciformes* our understanding of avian chromosome evolution will improve. Moreover, the recent availability of several sequenced bird' genomes including the White-tailed eagle (*Falconiformes*), Kea (*Psittaciformes*) and Turkey vulture

(*Cathartiformes*) (Zhang *et al.* 2014), giving the possibility to carryout gene synteny comparison which will give insight into the evolutionarily conserved synteny blocks and their role in avian evolution.

Comparative cytogenomic studies using successful hybridizing probes can be an excellent approach for karyotype evolution studies (Skinner and Griffin 2012, Damas *et al.* 2017). Comparative studies between *Falco* representatives (Saker falcon and Gyrfalcon), *Psittaciformes* representatives (budgerigar and cockatiel) and a phylogenetically distant species, the ostrich, shows that karyotypes of these avian species are not conserved. As expected (Nanda *et al.* 2007), but not predicted by Romanov *et al.* (2014) in the ostrich, no karyotype rearrangements are observed except that this species possesses 2 synteny segments for chicken (GGA) chromosome 4 (which is present in most birds studied). In contrast to this study, Romanov *et al.* (2014) suggested that chicken was most similar to the common ancestor when compared to ostrich, turkey, duck, budgerigar and zebra finch. That is, the chicken lineage underwent the least number of intrachromosomal rearrangements. It is worth mentioning that fragmented (scaffold-based) genome assemblies of the ostrich and budgerigar were used to reconstruct the common ancestor. They noticed a decrease in the number of reconstructed contiguous ancestral regions (CARs) when excluding fragmented genome assemblies. The importance of having 'chromosome level' assembled genomes is reiterated in section 7. 1

In falcon species, 2 homologous segments exist for the first 5 chicken chromosomes whereas only 3 GGA chromosomes (1, 4 and 5) are split in the budgerigar and cockatiel species. Following the fission of ancestral macrochromosomes, comparative data feature a tandem fusion of microchromosomes in the distal ends of macrochromosomes. Homologs of chicken chromosomes 10, 12, 13, 14, 15, 17, 18, 19, 20, 21, 23 and 28 are fused to different macrochromosomes in falcon species. The budgerigar shows a fusion of homologs of GGA 10, 11, 12, 13, 14 and 17 to macrochromosomes whereas homologues of GGA 13 and 20 are fused to form a single chromosome in the cockatiel. Also, GGA 11 and 14 are fused to separate macrochromosomes in cockatiels. The only common syntenic fusion combinations *Psittaciformes* species so far studied are 6/7 and 8/9 (de Oliveira *et al.* 2015; Nanda *et*

al. 2007; Nie *et al.* 2015; Seibold-Torres *et al.* 2015). But the budgerigar GGA 5 has fused to 6/7 and formed 5/6/7. Furthermore, GGA 4 has fused to GGA 8/9 to form 4/8/9. In the three falcon species rearrangements pattern are the same.

In other words, while chromosomal diversity is common in both *Psittaciformes* (de Oliveira *et al.* 2015) and *Falconiformes* the present study shows that in contrast, more extensive chromosome rearrangements have occurred in the lineage of the latter.

To date, cytogenetic studies point out that the majority of avian genomes are conserved in terms of chromosome number (in 60-70% of species $2n \sim 80$). Despite having a higher number of chromosomes, birds rarely exhibit interchromosomal changes except in representatives of *Psittaciformes* (parrots), *Sphenisciformes* (penguins) and *Falconiformes* (falcons) (Griffin *et al.* 2007; Schmid *et al.* 2015).

Chromosome evolution is most importantly concerned with the maintenance of syntenies and changes in the order of DNA fragments (Lemaitre *et al.* 2009) that are fixed throughout evolution. It has been established that the primary evidence of chromosomal change in birds are homologous synteny blocks (HSBs) that are demarked by evolutionary breakpoint regions (EBRs) (Griffin *et al.* 2015). Karyotype differences between species are associated with repetitive sequences used for non-allelic homologous recombination (NAHR) that underwent at least one large chromosomal structural change in the descendant specie's genome (Murphy *et al.* 2005; Damas *et al.* 2017). Precise detection and comparison of rearrangement breakpoints in multi-species chromosomes led to the fact that during bird and mammal evolution, breakpoints appeared to have occurred in the same region over independent lineages more often than expected and such breakpoint re-use is extensively noticed in mammals compared to birds. Those breakpoint hotspots correspond to evolutionarily-stable fragile regions (Lemaitre *et al.* 2009).

Romanov *et al.* (2014) performed comparative mapping of 21 avian genome sequences and used six best-assembled genomes (chicken, turkey, duck, zebra finch, ostrich, budgerigar) to assemble a putative karyotype of the dinosaur ancestor for each chromosome. Evolutionary events were reconstructed in these species which led to their present genome organization. Intra- and inter- chromosomal changes occurred in these

six species' genomes, detailed by a series of inversions and translocations with common break point re-use. The Damas *et al.* (2017); avian karyotype study (Peregrine falcon) revealed that the most common mechanism of interchromosomal change in this species is fusion with some exhibiting as many as four fused ancestral chromosomes without any trace of reciprocal translocations and leaving all microchromosomes intact. Compared to other animals, avian genomes have a smaller number of transposable elements which indicate that there are fewer opportunities for chromosome fusion using NAHR, suggesting that this may underlie the presence of multiple chromosomes (Damas *et al.* 2017).

Avian genomes remain evolutionarily stable interchromosomally with microchromosomes representing blocks of conserved synteny, with relatively rare exceptions (Romanov *et al.* 2014; Farré *et al.* 2016). Interchromosomal rearrangements are absent in most birds suggesting an evolutionary advantage to retain their genome configuration or having less opportunity for change. Analysis of several interchromosomal EBRs in the falcon genome shows that rare interchromosomal rearrangements that are fixed in the avian lineage-specific evolution are seen in areas of a low density of conserved non-coding elements (CNE's) which applies to both fission and fusion events. In order to consider the sites of interchromosomal EBRs as suitable for chromosomal fission, they need to be significantly distant from the areas with high CNE density which explains why falcon-specific fission breakpoints that are reused in other avian lineages as intrachromosomal EBRs (Damas *et al.* 2017). Intrachromosomal EBRs are prominent in areas of significantly higher CNE density than interchromosomal EBRs (Skinner and Griffin 2012; Romanov *et al.* 2014).

The chromosome rearrangement and subsequent speciation in the avian genome mainly depends on the mutation rate and fixation rate (Burt *et al.* 1999). Chromosomal rearrangements can be enhanced by repeat structures in general (e.g. CNVs) and transposable elements. Since the avian genome is constrained by size, the chances for mutation is limited for chromosomal rearrangements. Thereby only fission or intrachromosomal change, like inversions, are possible, which explains the reason for fewer interchromosomal rearrangements occurring in avian karyotypes.

7.2 Chromosome evolution in falcons

Romanov *et al.* 2014 suggest that highly rearranged genomes like *Psittaciformes* can undergo rapid intra- and interchromosomal rearrangements which is common in mammals, reptiles and amphibians with larger, repeat rich genomes. However, there is no evidence to support that highly rearranged avian genomes are especially large, or significantly more repeat-rich than other avian genomes. Intrachromosomal changes observed in pigeons, falcons (Damas *et al.* 2017) and Passeriform species (Skinner and Griffin 2012; Romanov *et al.* 2014) suggests that intrachromosomal events have a less effect on *cis* gene regulation than interchromosomal events. Avian species, such as falcons and parrots, undergo wholesale interchromosomal rearrangements, but with fission restricted to a few events and fusion more common (Damas *et al.* 2017), such a process has remained undiscovered to date.

7.3 Mapping of telomeric sequences in falcons and other animals

Telomeres are highly conserved, non-coding repetitive DNA sequences that are organized in such a way that along with several associated proteins, they cap the ends of eukaryotic chromosomes (Armanios and Blackburn 2012); however, the role and significance of telomeres that are not at the chromosomal ends (interstitial) are not well understood. Telomere length decreases with each incomplete DNA replication, when its length reaches a critical-level, cellular proliferation stops, this is mediated by either senescence or by apoptosis. However, it is not established whether this is the case for the interstitial telomeres. Their fundamental role in maintaining the genome stability and chromosome replication is of paramount importance (Nussey *et al.* 2014). The cellular function of telomeres is well understood mainly because of *in vitro* studies but their *in vivo* significance in terms of their distribution, the organism's health, longevity and reproductive ability is poorly understood, as is their role within biology, epidemiology and medicine (Aubert *et al.* 2012).

In vertebrates' telomere size varies from 10-20Kb (Davis and Kipling 2005); in birds they are approximately 100Kb (Nanda *et al.* 2002) although observation from FISH signals alone suggests that this is highly variable from chromosome to chromosome. In the

chicken, three different classes of telomere arrays have been identified: Class I, interstitial; Class II and Class III, terminal (Delany *et al.* 2000; Delany *et al.* 2007; Nanda *et al.* 2002). Early FISH studies concluded that, in general, primitive birds (Ratites and *Galloanserae*) have only one telomeric site, while Neoaves may have more than one telomeric site (Meyne *et al.* 1990). However, this idea is not supported by the observation of telomeric sites in several extant birds which do not display non-interstitial telomeric sites (Nanda *et al.* 2002). The chicken has 3 interstitial sites, the turkey has one (Griffin *et al.* 2008); however, ostrich (this study and Nanda *et al.* 2002) has 2. Their absence in Neoavian birds may suggest interstitial telomeres were lost gradually after the divergence of the Neoaves from the rest of the Neognathae (Nanda *et al.* 2002).

So far, telomeric (TTAGGG)_n sequences had previously been observed by FISH only in 3 falcon species: the common kestrel (*Falco tinnunculus*), Peregrine falcon (*Falco peregrinus*) and merlin (*Falco columbarius*) (Nishida *et al.* 2008). Hence, the present study extends the telomeric (TTAGGG)_n sequence information to two further falcon species and re-analyzes the previously published telomeric signal pattern in Peregrine falcons. Moreover, the use of a peptide nucleic acid (PNA) probe, rather than the DNA probe used by Nanda and colleagues, may, a) increase the chance of observing any interstitial telomeres and b) differentiate between smaller and larger motifs. The apparent absence of interstitial telomeric sequences in three different falcon species, crocodiles and budgerigars suggests that the interstitial telomeric sequences are not exclusive to chromosomes where fusions have occurred. Moreover, this study supports previous studies which suggested that, during or after the chromosomal fusion activity, either telomeric DNA sequences were lost or that degeneration of telomeric DNA sequences occurred that are undetectable by FISH. This finding agrees well with previous studies in that interstitial telomeric DNA arrays were not seen in birds which inherit highly rearranged karyotypes e.g. the scarlet macaw (*Ara macao*) (Seabury *et al.* 2013), budgerigar (*Melopsittacus undulatus*) (Nanda *et al.* 2002), California condor (*Gymnogyps californianus*) (Raudsepp *et al.* 2002), Griffon vulture (*Gyps fulvus*) and bearded vulture (*Gypaetus barbatus*) (Nanda *et al.* 2006).

Microchromosomes are gene-rich, stable and highly conserved (Burt *et al.* 1999) and linear microchromosomes are held to have telomeric sequences at both ends, a concept

that has been established by Nanda *et al.* (2002) in 16 different avian species. Moreover, microchromosomes exhibit greater numbers of telomeric repeats than macrochromosomes (Nanda *et al.* 2002); this may be an important factor in accounting for the high recombination rate in avian microchromosomes. Similar observations are demonstrated in this study, showing abundant telomeric signals on microchromosomes compared with macrochromosomes in 3 different falcon species; a fact that was established more markedly with the use of a PNA probe.

In addition, the FISH results using the PNA probe showed that some falcon microchromosomes possess mega-telomeres (see section 1.2.4) similar to chicken and Japanese quails (Delany *et al.* 2007; McPherson *et al.* 2014). In Peregrine falcons, 6 chromosome pairs exhibited mega-telomeres whereas only one such pair showed these features in the Saker and Gyrfalcon. Nishida *et al.* (2008) demonstrated a similar finding in Peregrine falcon microchromosomes. Different chicken genotypes exhibit different telomeric DNA profiles (O'Hare and Delany 2009), hence such variation exhibited among different falcon species is less surprising. Interestingly, mega-telomeres (identified by their strong overall fluorescence) were observed only on microchromosomes. These findings, however, require further confirmation and validation in other falcon species, and in different cell types as well as with different test methods. Telomere profiling could, in the future, be extended to a number of other falcon species to evaluate whether mega-telomeres are a characteristic feature of other species in the genus *Falco*. Mega-telomere numbers vary among these falcons, hence further telomere studies on other falcon species can delineate the distribution pattern of mega-telomeres in them. Furthermore, a detailed study of telomere array length in different falcon cell types, using different test methods such as Southern hybridization, quantitative PCR and terminal restriction fragment analysis (TRF) is suggested to confirm and validate the nature of mega-telomeres.

PNA probes are ideal tools for FISH because of their high affinity and specificity to target telomere DNA sequences on chromosomes (Kawano *et al.* 2014; Slijepcevic 2001). The genetically determined variation in telomerase activity between individual birds makes telomere length measurements relevant for the determination of aging which can be measured using suitable software. The quantitative PNA-FISH tests have allowed for the

detection of telomere repeat lengths in individual chromosomes of birds. Telomere length measurements in Falcons using PNA probe Q-FISH can act as a biomarker to determine the onset of cellular senescence. Telomere length in Falcons can be considered as a marker for ageing and somatic cell fitness which is important in the Middle East where falconry and falcon breeding centres are popular. The relationship between telomere length and age has been well studied in birds (Pauliny *et al.* 2012). Long lived birds and mammals have a very effective maintenance system in that they lose telomere repeats more slowly than short lived species (Hausmann *et al.* 2003).

7.4 Nuclear organization of “former microchromosome”

Territorial organization of chromosomes has mostly been studied in mammals with convincing experimental evidence for the existence of chromosome territory (CT) features in the interphase nuclei. ‘X’ shaped chromosomes are easily visualized during mitosis whilst during interphase, chromosomes tend to relax and their positions are considered as chromosome territory (Cremer and Cremer 2001; Cremer *et al.* 2006). Chromosome organization patterns in CTs during interphase are now generally accepted in the scientific community (Fritz *et al.* 2016) and it is well established that birds have highly organized chromosome territories with microchromosomes taking a central nuclear position and macrochromosomes occupying a peripheral position.

Chromosomes in different species and in different types of cells, tend to arrange radially. Radial arrangements can be according to their gene density, their size or both, with birds fitting both models (Bolzer *et al.* 2005; Boyle *et al.* 2001; Mehta *et al.* 2010). One of the dominating principles of nuclear organization is that the chromosome size is the major influencing factor which determines the nonrandom position of CTs with shorter chromosomes localized internally (Fritz *et al.* 2016). It is possible that cell shape and cell cycle can affect their nuclear position (Sun *et al.* 2000). CT size-radial position correlations are more common in ellipsoid fibroblast nuclei (Bolzer *et al.* 2005; Mehta *et al.* 2010).

A number of studies have demonstrated that G- light G- dark bands on mitotic chromosomes exist in distinct higher-order chromatin compartments. Major parts of the

gene-dense, early replicating chromatin forms an interior compartment in the nucleus. In contrast, gene-poor, late replicating chromatin locates at the periphery (Sadoni *et al.* 1999; Visser *et al.* 1998). This reflects their complex organization. Cremer and Cremer (2001) explained the correlation between gene density and chromosome position of human chromosomes 18 and 19 in the nuclei of proliferating lymphocytes and lymphoblastoid cells. Though human chromosomes 18 and 19 have similar DNA content, their nuclear distributions are determined by their gene density. Gene-rich human chromosome 19 territories were found to be located in the interior while the gene-poor chromosome 18 was found at the periphery (Croft *et al.* 1999; Tanabe *et al.* 2002). Habermann *et al.* (2001) demonstrated that larger chromosomes in chicken embryonic fibroblasts and neurons tend to locate towards the periphery and microchromosomes tend to locate at the center. Skinner *et al.* (2009b) reported similar findings in the turkey and duck, whose smaller chromosomes localize towards the center and larger chromosomes towards the periphery. Microchromosomes are known to be gene-dense, enriched with CpG and early replicating, whereas, macrochromosomes are generally considered as later replicating and comparatively gene-poor. It is important to note that microchromosomes represent 50% of the whole chicken genome, even though their overall size accounts for only 23% of its genome (McQueen *et al.* 1996; Tanabe *et al.* 2002). The results of this study clearly point to a “gene density” based level of genome organization in falcons. The presence of a gene-rich “microchromosome-like” portion and a regular relatively gene “macrochromosome-like” portion of a single chromosome, the former occupying a central nuclear position provides the evidence of this. Indeed, the present study is the first to describe the nuclear organization in a bird species other than the *Galloanserae*. Non-fused macro- and microchromosomes behave the same way in chickens and falcons. Specifically seven homologous genes of gene-poor human chromosome (HAS) 18 were mapped on chicken chromosome (GGA) 2 and Z, which are gene-poor and 17 homologous genes on human chromosome 19 were identified on chicken chromosome 11 and 28 which are gene-rich (Tanabe *et al.* 2002). The positions of these chromosomes were previously examined in human fibroblast nuclei by Tanabe *et al.* (2002). Chromosome territory positions of both human chromosome 18 and its homologous chicken macrochromosomes (GGA 2 and Z) were located towards the nuclear periphery, while, chromosome territory positions of both human 19 and chicken

11 and 28 were located towards the nuclear center. Strikingly, the present study shows GGA 11 and its homologous falcon microchromosome (FPE) 14 have a chromosome territory positioned towards the nuclear center with median positions of 3.51 and 3.44 respectively. Moreover, GGA 28 which is fused to macrochromosome 5 in the Peregrine falcon, retains its ancestral character, being located towards the center of the nucleus (median position 2.93). GGA 2 is split across FPE 3 and FPE 5. Both are fused with GGA microchromosomes in the Peregrine falcon and retain the same nuclear organization pattern which has been conserved for more than 300 million years. The findings from this study can provide information on the basic nature of chromosome territory patterns in bird species with rearranged karyotypes. This implies that the same general nuclear organization mechanisms are present in falcons as well as in chickens, ducks and turkeys whose last common ancestor existed around 89 million years ago.

Although many models have been suggested for chromosome positioning in interphase and how they fold within chromosome territories, these models are unable to explain the relationship between higher order chromatin structure and genome function (Dixon *et al.* 2012). Microscopic approaches such as 3D chromatin arrangements, coupled with chromosome conformation capture strategies such as 3C, 4C, 5C, Hi-C can detect intra- and inter- physical chromosomal interactions (Cremer *et al.* 2015; Fritz *et al.* 2016). The nuclear organization pattern of chromatin into individual chromosome territories that has been proposed by the FISH studies is further validated by newer technique such as Hi-C (Wang *et al.* 2015). This technique described the three-dimensional organization of the human and mouse genomes in embryonic stem cells and terminally differentiated cell types with extraordinary resolution. It also identified 'topological domains' (chromatin interacting domains) and such special organization appears to be the general feature of the genome, and stable across different cell types as well as being conserved in humans and mice (Dixon *et al.* 2012; Bonev and Cavalli 2016; Wang *et al.* 2015).

Chromatin architecture influences genomic rearrangements during evolution. 3D nuclear organization accurately reproduced the genomic rearrangements as predicted by the ancestral genome reconstruction and statistical modelling (Bonev and Cavalli 2016). Studies into 3D chromatin organization in Falcons should give insight into genomic rearrangements that have occurred during evolutionary events. It has been

demonstrated that 3D genome architecture is directly linked to regulating gene expression during development of physiological processes (Bonev and Cavalli 2016); and can help in future research on functional expression of genes related to water conservation and sodium secretion in Falcons. This may explain the genetic basis by which certain falcon species cope with desert and steppe habitats and heat stress (Zhan *et al.* 2013). Recent genome sequencing efforts in falcons have provided opportunities for studying chromatin structure with 3C based data for understanding the genetic basis of their wide distribution across the globe. This would create a vital tool to support the long-term conservation of falcons especially with Saker falcons being globally classified as vulnerable.

Further studies on different falcon cell types, in different species and with newer techniques may delineate questions of whether nuclear organization has a major role in avian embryology, disease phenotypes and genome evolution. Nuclear organization information generated through this study for falcons can be considered as a baseline for species with atypical karyotypes. An analysis of nuclear organization patterns in different falcon cell types will determine whether these patterns vary with cell type. Parrots are an ideal group to explore this area, as they represent birds that have undergone significant rearrangements compared to the “norm” of $2n \sim 80$. Such studies could provide insight into the evolution of chromosome positioning. Defining higher-order structures in nuclear organization (3C, 4C and 5C approaches) can demonstrate regulatory interactions within and between other chromosomes in these birds of interest. Such studies would reveal whether a common nuclear organizing pattern exists in *Falconiformes* and *Psittaciformes*.

7.5 Future Work

It is essential to have an accurate map (Lewin *et al.* 2009) with each sequence assigned to a specific locus on a chromosome for every *de-novo* genome assembly of a species that has not previously been studied. The prime importance of a *de-novo* genome assembly is a representation of all the sequences, correctly ordered contiguously. Now there are many “next generation” technologies that can generate sufficient read lengths

and compositions to assemble *de-novo* genomes with long scaffolds of contiguous sequence. This includes Dovetail (<http://dovetailgenomics.com/technology>), optical mapping (Davydov *et al.* 2010), BioNano (www.bionanogenomics.com) and “next-next generation” long read sequencing.

To achieve chromosome level assembly, these technologies require substantial funding. A good example is the recently published PacBio/HiC goat genome in which a series of technical hurdles such as contigs and scaffolds may not extend across multiple DNA nick site fragile regions, centromeres or large heterochromatin blocks (Bickhart *et al.* 2017). Rectifying this problem requires additional cost of reagents, manpower and time (e.g. estimated current cost of the PacBio/HiC goat genome is \$100,000 as reported by authors) (Kong *et al.* 2011). Recently more financially affordable efforts have been tested, one example, the dog genome which is assembled using Illumina *Discover* and Dovetail (cost of Dovetail was ~\$10,000). These generated large sized super scaffolds which were sub-chromosomal in size, this applies to most new genomes generated by these technologies. Numerous individual projects or consortia such as Bird 10K are progressing in assembling thousands of genomes to scaffold level. Bird 10K is conducting the project in four phases, based on the avian classification hierarchy. The first, ordinal phase (for 34 orders of birds) has been accomplished (Zhang *et al.* 2014). Collection of genomic data for the second, familial phase (about 240 families) is ongoing. Specimen and trait-data collection for the third phase (2,250 genera) and the fourth phase (the remaining 8,000 or so species) is under way (<http://b10k.genomics.cn>).

Recently a relatively low-cost solution to this problem has been developed in birds by taking pre-existing scaffold-based assemblies (e.g. generated by Dovetail or improved bioinformatically by RACA) and “upgrading” them to full chromosome level by mapping super scaffolds directly to chromosomes (Damas *et al.* 2017). Of the >10000 extant avian species, to date 61 species have had their genomes sequenced of which only 7 have genomes assembled to chromosomes. Among the 7 species, Peregrines and Pigeons have been recently added using this novel approach which combines computational algorithms to merge scaffolds into chromosomal fragments, scaffold verification by PCR and physical mapping to chromosomes (Damas *et al.* 2017).

To date, in the genus *Falco*, genome sequencing of the Peregrine and Saker falcons has been accomplished (Zhan *et al.* 2013) at sub-chromosomal level. The present study has upgraded the Saker falcon genome from sub-chromosomal sized “scaffolds” to chromosome level assembly, thus adding its status to that of the Peregrine falcon (Damas *et al.* 2017). This approach is theoretically applicable to any avian genome in the future. Furthermore, in birds, since only few reference genomes are available such as the chicken, turkey and zebra finch, complete chromosomal assembly needs to be achieved by other means till a significant number of assembled avian genomes become available (e.g. at least one per phylogenetic order).

The cross-species avian BACs developed by Damas *et al.* (2017) were successfully hybridized on the Saker falcon genome and can be applied to all avian species. This will provide understanding not only of the individual comparative genomic question but also will be a method to upgrade the scaffold based genomes of sequenced avian genomes. Similar universal BAC set with the same level of mapping resolution as in birds would be possible in mammals. Since mammalian genomes have more repetitive sequences with triple the size of birds, building a mammalian universal BAC set would be a greater challenge. The fact that this avian BAC set has been successfully hybridized on lizard and turtle chromosomes (O'Connor 2016) suggests that the approach is applicable for building chromosomal assemblies for all vertebrate and eventually for entire animal groups in the near future.

7.6 Concluding Remarks

I am very glad to complete this thesis journey successfully, a journey which started in 2012. The road to completion was never smooth. There were times I have fought to see the way forward, but my curiosity always got the better side of me and boosted my quest to explore new arenas in the yet hidden scientific treasures of cytogenetic studies. The period of initial sample sourcing proved to be extremely challenging, as falcons are an integral part of the Arabian culture and heritage: not only can they be hugely expensive, but falcons are also cherished family members of their proud and very protective owners. Keeping that in mind I still experienced overwhelming support from falcon breeders, as far afield as Spain and Scotland. Since they all understood how ground

S. Joseph

breaking scientific research will help to shed more light on the genetic complexity in the evolution of the genus *Falco*, this magnificent bird of prey. At the various falcon breeding centers in the United Arab Emirates I was intrigued to come across breeders with a deep-rooted knowledge and insight into falconry going back many generations. I am so grateful to also have learned a lot during these visits. This research project allowed me to submerge in a completely new part of a fascinating line of science. I am convinced it is beneficial for the falcon breeders and owners of the United Arab Emirates and worldwide. Not only will it be a tremendous joy for me to receive the academic appreciation of 'Doctor of Philosophy', but it makes me equally proud to be able to repay with the submission of my thesis the trust and support the authorities of the United Arab Emirates extended to me.

8 References:

Al Mutery AF (2011) *Cytogenetic and genomic characterization of the Gyr Falcon and related species: individual genomic signature and genome evolution* (Doctoral thesis). University of Kent, UK.

Alekseyev MA and Pevzner PA (2009) Breakpoint graphs and ancestral genome reconstructions. *Genome Res.* **19**, 943-957.

Amadon D and Bull J (1988) Hawks and owls of the world: A distributional and taxonomic list. *Proc. West. Found. Vert. Zool.* **3**(4), 295-357.

Amaral KF and Jorge W (2003) The chromosome of the order falconiformes: A review. *Ararajuba* **11**[1], 65-73.

Amirsadeghi H (2008) *Sky Hunters: The Passion of Falconry*. Thames & Hudson, U.K.

Andersson L and Georges M (2004) Domestic-animal genomics: deciphering the genetics of complex traits. *Nat. Rev. Genet.* **5**, 202-212.

Armanios M and Blackburn EH (2012) The telomere syndromes. *Nat. Rev. Genet.* **13**, 693-704.

Aubert G, Hills M, Lansdorp PM (2012) Telomere length measurement-caveats and a critical assessment of the available technologies and tools. *Mutat. Res.* **730**, 59-67.

Avdeyev P, Jiang S, Aganezov S, Hu F, Alekseyev MA (2016) Reconstruction of Ancestral Genomes in Presence of Gene Gain and Loss. *J. Comput. Biol.* **23**, 150-164.

Axelsson E, Webster MT, Smith NG, Burt DW, Ellegren H (2005) Comparison of the chicken and turkey genomes reveals a higher rate of nucleotide divergence on microchromosomes than macrochromosomes. *Genome Res.* **15**, 120-125.

Barton NWH (2000) Trapping estimates for saker and peregrine falcons used for falconry in the United Arab Emirates. *J. Raptor Res.* **34**, 53-55.

Bauch C, Becker PH, Verhulst S (2012) Telomere length reflects phenotypic quality and costs of reproduction in a long-lived seabird. *Proc. Biol. Sci.* **280**, 2012.2540.

Becker JJ (1987) Revision of *Falco ramentata* wetmore and the neogene evolution of the falconidae. *The Auk* **104**, 270-276.

Beltermann RHR and de Boer LEM (1984) A karyological study of 55 species of birds, including karyotypes of 39 species new to cytology. *Genetica* **65**, 39-82.

S. Joseph

Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S *et al.* (2017) Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nat Genet.* **49**, 643-650.

Bickmore WA (2001) Karyotype Analysis and Chromosome Banding. In: *eLS. John Wiley & Sons Ltd, Chichester*. <http://www.els.net> doi: 10.1038/npg.els.000116.

Binothman AM (2016) *Current Status of Falcon Populations in Saudi Arabia* (Doctoral thesis). South Dakota State University, U.S.A.

Bishop R (2010) Applications of fluorescence in situ hybridization (FISH) in detecting genetic aberrations of medical significance. *Bioscience Horizons: The International Journal of Student Research* **3**, 85-95.

Bolzan AD and Bianchi MS (2006) Telomeres, interstitial telomeric repeat sequences, and chromosomal aberrations. *Mutat. Res.* **612**, 189-214.

Bolzer A, Kreth G, Solovei I, Koehler D, Saracoglu K, Fauth C *et al.* (2005) Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol.* **3**, e157.

Bonev B and Cavalli G (2016) Organization and function of the 3D genome. *Nature Reviews Genetics* **17**, 661-678.

Bourque G and Pevzner PA (2002) Genome-scale evolution: reconstructing gene orders in the ancestral species. *Genome Res.* **12**, 26-36.

Boveri T (1909) Die Blastomerenkerne von *Ascaris megalocephala* und die Theorie der Chromosomenindividualität. *Arch. Zellforsch* **3**, 181-268.

Boyce DA (1989) *A systematic study of the family Falconidae: protein electrophoretic analysis of genera* (Doctoral thesis). Brigham Young University, Provo, U.S.A.

Boyd LL and Boyd N (1975) Hybrid falcons. *Hawk Chalk* **14**, 54.

Boyle S, Gilchrist S, Bridger JM, Mahy NL, Ellis JA, Bickmore WA (2001) The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum. Mol. Genet.* **10**, 211-219.

Bumstead N and Palyga J (1992) A preliminary linkage map of the chicken genome. *Genomics* **13**, 690-697.

Burt DW, Bruley C, Dunn IC, Jones CT, Ramage A, Law AS *et al.* (1999) The dynamics of chromosome evolution in birds and mammals. *Nature* **402**, 411-413.

S. Joseph

Burt DW (2002) Origin and evolution of avian microchromosomes. *Cytogenet. Genome Res.* **96**, 97-112.

Burt DW (2004) The chicken genome and the developmental biologist. *Mech. Dev.* **121**, 1129-1135.

Cade TJ (1982) *Falcons of the world*. William Collins Sons and Co. Ltd., London.

Cai WW, Mao JH, Chow CW, Damani S, Balmain A, Bradley A (2002) Genome-wide detection of chromosomal imbalances in tumors using BAC microarrays. *Nat. Biotech.* **20**, 393-396.

Caspersson T, Zech L, Johansson J (1970) Differential banding of alkylating fluorochromes in human chromosomes. *Experimental Cell Research* **60**, 315-319.

Chaves LD, Harry DE, Reed KM (2009) Genome-wide genetic diversity of 'Nici', the DNA source for the CHORI-260 turkey BAC library and candidate for whole genome sequencing. *Anim. Genet.* **40**, 348-352.

Chen H and Wang X (2013) CrusView: a Java-based visualization platform for comparative genomics analyses in Brassicaceae species. *Plant Physiol.* **163**, 354-362.

Cheung VG, Nowak N, Jang W, Kirsch IR, Zhao S, Chen XN *et al.* (2001) Integration of cytogenetic landmarks into the draft sequence of the human genome. *Nature* **409**, 953-958.

Chiappe LM and Witmer LM (2002) *Mesozoic birds: Above the Heads of Dinosaurs*. University of California Press, Berkeley, California, USA.

Chiari Y, Cahais V, Galtier N, Delsuc F (2012) Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (*Archosauria*). *BMC Biol.* **10**, 65-10.

Chowdhary BP, Raudsepp T, Fronicke L, Scherthan H (1998) Emerging patterns of comparative genome organization in some mammalian species as revealed by Zoo-FISH. *Genome Res.* **8**, 577-589.

Christidis L (1990) Chordata 3B. Aves. In *Animal Cytogenetics*, Bernard J, Kayano H, Levan A (eds). Berlin: Gebrüder Borntraeger; 4.

Clancy S and Shaw K (2008) DNA deletion and duplication and the associated genetic disorders. *Nature Education* **1**, 23.

Clayton DF, Balakrishnan CN, London SE (2009) Integrating genomes, brain and behavior in the study of songbirds. *Curr. Biol.* **19**, R865-R873.

S. Joseph

Cremer T and Cremer C (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* **2**, 292-301.

Cremer T and Cremer C (2006) Rise, fall and resurrection of chromosome territories: a historical perspective. Part II. Fall and resurrection of chromosome territories during the 1950s to 1980s. Part III. Chromosome territories and the functional nuclear architecture: experiments and models from the 1990s to the present. *Eur J Histochem.* **50**, 223-272.

Cremer T and Cremer M (2010) Chromosome Territories. *Cold Spring Harb. Perspect. Biol.* **2**(3), doi: 10.1101/cshperspect.a003889.

Cremer T, Cremer M, Hubner B, Strickfaden H, Smeets D, Popken J *et al.* (2015) The 4D nucleome: Evidence for a dynamic nuclear landscape based on co-aligned active and inactive nuclear compartments. *FEBS Lett.* **589**, 2931-2943.

Crittenden LB, Provencher L, Santangelo L, Levin I, Abplanalp H, Briles RW *et al.* (1993) Characterization of a Red Jungle Fowl by White Leghorn Backcross Reference Population for Molecular Mapping of the Chicken Genome. *Poultry Science* **72**, 334-348.

Croft JA, Bridger JM, Boyle S, Perry P, Teague P, Bickmore WA. (1999) Differences in the localization and morphology of chromosomes in the human nucleus. *J. Cell Biol.* **145**, 1119-1131.

Crooijmans RP, Vrebalov J, Dijkhof RJ, van der Poel JJ, Groenen MA (2000) Two-dimensional screening of the Wageningen chicken BAC library. *Mamm. Genome.* **11**, 360-363.

Dalloul RA, Long JA, Zimin AV, Aslam L, Beal K, Ann L *et al.* (2010) Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. *PLoS Biol.* **8**, e1000475.

Dalloul RA, Zimin AV, Settlege RE, Kim S, Reed KM (2014) Next-generation sequencing strategies for characterizing the turkey genome. *Poult. Sci.* **93**, 479-484.

Damas J, O'Connor R, Farre M, Lenis VP, Martell HJ, Mandawala A *et al.* (2017) Upgrading short read animal genome assemblies to chromosome level using comparative genomics and a universal probe set. *Genome Res.* **27**, 875-884.

Davis T and Kipling D (2005) Telomeres and telomerase biology in vertebrates: progress towards a non-human model for replicative senescence and ageing. *Biogerontology.* **6**, 371-385.

Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S (2010) Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput. Biol.* **6**, e1001025.

S. Joseph

de Boer LEM (1990) The somatic chromosome complements of 16 species of Falconiformes (Aves) and the karyological relationships of the order. *Genetica* **46**, 77-113.

de Boer LEM (1975) Karyological heterogeneity in the Falconiformes (Aves). *Experientia* **31**(10), 1138-1139.

de Oliveira EH, de Moura SP, dos Anjos LJ, Nagamachi CY, Pieczarka JC, O'Brien PC *et al.* (2008) Comparative chromosome painting between chicken and spectacled owl (*Pulsatrix perspicillata*): implications for chromosomal evolution in the Strigidae (Aves, Strigiformes). *Cytogenet Genome Res.* **122**, 157-162.

de Oliveira FI, Kretschmer R, O'Brien PC, Ferguson-Smith MA, de Oliveira EH (2015) Chromosomal Diversity and Karyotype Evolution in South American Macaws (*Psittaciformes*, *Psittacidae*). *PLoS One.* **10**, e0130157.

del Hoyo J, Elliott A, Sargatal J (eds.) (1994). '*Handbook of the Birds of the World*'. Vol. 2: New World Vultures to Guinea-fowl. Lynx Edicions, Barcelona, Spain.

del Hoyo J, Elliott A, Sargatal J (eds.) (2004) '*Handbook of the Birds of the World*.' Vol. 9: Contingas to Pipits and Wagtails Lynx Edicions, Barcelona, Spain.

Delany ME, Daniels LM, Swanberg SE, Taylor HA (2003) Telomeres in the chicken: genome stability and chromosome ends. *Poult. Sci.* **82**, 917-926.

Delany ME, Gessaro TM, Rodrigue KL, Daniels LM (2007) Chromosomal mapping of chicken mega-telomere arrays to GGA9, 16, 28 and W using a cytogenomic approach. *Cytogenet. Genome Res.* **117**, 54-63.

Delany ME, Krupkin AB, Miller MM (2000) Organization of telomere sequences in birds: evidence for arrays of extreme length and for in vivo shortening. *Cytogenet. Cell Genet.* **90**, 139-145.

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M *et al.* (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature.* **485**, 376-380.

Dodson P (2000) Origin of birds: The final solution? *Amer. Zoo.* **40**, 504-512.

Doupe AJ and Kuhl PK (1999) Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* **22**, 567-631.

Ducos A, Revay T, Kovacs A, Hidas A, Pinton A, Bonnet-Garnier A *et al.* (2008) Cytogenetic screening of livestock populations in Europe: an overview. *Cytogenetic and Genome Research* **120**, 26-41.

S. Joseph

Dyke G and Kaiser GW (2011) *Living Dinosaurs the Evolutionary History of Modern Birds*. Wiley-Blackwell, Hoboken, New Jersey, USA.

Eastham CP and Nicholls MK (2005) Morphometric analysis of large falco species and their hybrids with implications for conservation. *J. Raptor Res.* **39**, 384-393.

Ellegren H (2010) Evolutionary stasis: the stable chromosomes of birds. *Trends Ecol. Evol.* **25**, 283-291.

Ellegren H, Smeds L, Burri R, Backstroem N, Kawakami T, Kuenstner, Makjnen H *et al.* (2012) The genomic landscape of species divergence in *Ficedula flycatchers*. *Nature* **491**, 756-760.

Ericson PG, Anderson CL, Britton T, Elzanowski A, Johansson US, Kallersjo M *et al.* (2006) Diversification of Neoaves: integration of molecular sequence data and fossils. *Biol. Lett.* **2**, 543-547.

Farré M, Bosch M, Lopez-Giraldez F, Ponsa M, Ruiz-Herrera A (2011) Assessing the role of tandem repeats in shaping the genomic architecture of great apes. *PLoS One.* **6**, e27239.

Farré M, Narayan J, Slavov GT, Damas J, Auvil L, Li C *et al.* (2016) Novel Insights into Chromosome Evolution in Birds, Archosaurs, and Reptiles. *Genome Biol. Evol.* **8**, 2442-2451.

Fawcett JJ, Longmire JL, Martin JC, Deaven LL, Cram LS (1994) Large-scale chromosome sorting. *Methods Cell Biol.* **42**, 319-330.

Federico C, Cantarella CD, Scavo C, Saccone S, Bed'Hom B, Bernardi G (2005) Avian genomes: different karyotypes but a similar distribution of the GC-richest chromosome regions at interphase. *Chromosome Res.* **13**, 785-793.

Feduccia A (2003) 'Big bang' for tertiary birds? *Trends Ecol. Evol.* **18**, 172-176.

Feschotte C and Pritham EJ (2007) DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **41**, 331-368.

Fleming LV, Douse AF, Williams NP (2011) Captive breeding of peregrine and other falcons in Great Britain and implications for conservation of wild populations. *Endang. species. Res.* **14**, 243-257.

Forstmeier W and Ellegren H (2010) Trisomy and triploidy are sources of embryo mortality in the zebra finch. *Proc. Biol. Sci.* **277**, 2655-2660.

Foster HA and Bridger JM (2005) The genome and the nucleus: a marriage made by evolution. Genome organization and nuclear architecture. *Chromosoma* **114**, 212-229.

S. Joseph

Franca LTC, Carrilho E, Kist TBL (2002) A review of DNA sequencing techniques. *Quarterly Reviews of Biophysics* **35**, 169-200.

Fritz AJ, Barutcu AR, Martin-Buley L, van Wijnen AJ, Zaidi SK, Imbalzano AN *et al.* (2016) Chromosomes at Work: Organization of Chromosome Territories in the Interphase Nucleus. *J. Cell Biochem.* **117**, 9-19.

Gill F and Donsker D (eds.) (2017) *IOC World Bird List (v 7.2)*. doi: 10.14344/IOC.ML.7.2.

Gordon JL, Byrne KP, Wolfe KH (2009) Additions, losses, and rearrangements on the evolutionary route from a reconstructed ancestor to the modern *Saccharomyces cerevisiae* genome. *PLoS Genet.* **5**, e1000485.

Grasser F, Neusser M, Fiegler H, Thormeyer T, Cremer M, Carter NP, Cremer T, Muller S (2008) Replication-timing-correlated spatial chromatin arrangements in cancer and in primate interphase nuclei. *J. Cell Sci.* **121**, 1876-1886.

Green RE, Braun EL, Armstrong J, Earl D, Nguyen N, Hickey G *et al.* (2014) Three crocodylian genomes reveal ancestral patterns of evolution among archosaurs. *Science* **346**(6215), 1254449.

Gregory TR (2017) The Animal Genome Size Database. <http://www.genomesize.com>

Griffin DK, Romanov, Michael N, O'Connor R, Fowler KE, Larkin DM (2015) *Avian cytogenetics goes functional, in: Third Report on Chicken Genes and Chromosomes 2015. Cytogenetic and Genome Research*, 145 (2).

Griffin D and Burt DW (2014) All chromosomes great and small: 10 years on. *Chromosome Res.* **22**, 1-6.

Griffin DK, Haberman F, Masabanda J, O'Brien P, Bagga M, Sazanov A *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 LB, 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, Vignal A, Fillon V, Crooijmans RP *et al.* (2008) Whole genome comparative studies between chicken and turkey and their implications for avian genome evolution. *BMC Genomics* **9**, 168.

Griffiths CS (1994) Syringeal Morphology and the Phylogeny of the Falconidae. *The Auk* **96**, 127-140.

S. Joseph

Griffiths CS (1999) Phylogeny of the Falconidae inferred from molecular and morphological data. *The Auk* **116**, 116-130.

Griffiths CS, Barrowclough GF, Groth JG, Mertz L (2004) Phylogeny of the Falconidae (Aves): a comparison of the efficacy of morphological, mitochondrial, and nuclear data. *Mol. Phylogenet. Evol.* **32**, 101-109.

Groenen MA, Archibald AL, Uenishi H, Tuggie CK, Takeuchi Y, Rothschild MF *et al.* (2012) Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* **491**, 393-398.

Groenen MA, Crooijmans RP, Veenendaal A, Cheng HH, Siwek M, van der Poel JJ (1998) A comprehensive microsatellite linkage map of the chicken genome. *Genomics* **49**, 265-274.

Gulsuner S and McClellan JM (2015) Copy Number Variation in Schizophrenia. *Neuropsychopharmacology* **40**, 252-254.

Guttenbach M, Nanda I, Feichtinger W, Masabanda JS, Griffin DK, Schmid M (2003) Comparative chromosome painting of chicken autosomal paints 1-9 in nine different bird species. *Cytogenet Genome Res.* **103**, 173-184.

Habermann FA, Cremer M, Walter J, Kreth G, von HJ, Bauer K *et al.* (2001) Arrangements of macro- and microchromosomes in chicken cells. *Chromosome Res.* **9**, 569-584.

Hackett SJ, Kimball RT, Reddy S, Bowie RC, Braun EL, Braun MJ *et al.* (2008) A phylogenomic study of birds reveals their evolutionary history. *Science* **320**, 1763-1768.

Hansmann T, Nanda I, Volobouev V, Yang F, Scharl M, Haaf T *et al.* (2009) Cross-species chromosome painting corroborates microchromosome fusion during karyotype evolution of birds. *Cytogenet. Genome Res.* **126**, 281-304.

Hassold T and Hunt P (2001) To err (meiotically) is human: the genesis of human aneuploidy. *Nat. Rev. Genet.* **2**, 280-291.

Hausmann MF, Winkler DW, O'Reilly KM, Huntington CE, Nisbet IC, Vleck CM (2003) Telomeres shorten more slowly in long-lived birds and mammals than in short-lived ones. *Proc. Biol. Sci.* **270**, 1387-1392.

Hausmann MF, Winkler DW, Vleck CM (2005) Longer telomeres associated with higher survival in birds. *Biol. Lett.* **1**, 212-214.

Heidenreich M (1997) *Birds of Prey - Medicine and Management*. Blackwell Science Ltd., Oxford, London.

S. Joseph

Heidenreich M and Küspert H (1992) Metodos de la reproduccion en cautividad de las diferentes especies de halcones. el problema de los hibridos y la cooperacion entre cradores y autoridades. *Proc. Congr. Nac. Cria. Aves Cetreia* **1**, 15-17.

Helbig AJ, Seibold I, Bednarek W, Brüning H, Gaucher P, Ristow D *et al.* (1994) Phylogenetic relationships among falcon species (genus *Falco*) according to DNA sequence variation of the cytochrome b gene. *In: Meyburg BU and Chancellor RD (eds.) Raptor conservation today: 593-599.*

Hepperger C, Mannes A, Merz J, Peters J, Dietzel S (2008) Three-dimensional positioning of genes in mouse cell nuclei. *Chromosoma* **117**, 535-551.

Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP *et al.* (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695-716.

Hooper DM and Price TD (2015) Rates of karyotypic evolution in Estrildid finches differ between island and continental clades. *Evolution* **69**, 890-903.

Hu, F, Zhou, L., and Tang, J (2013) Reconstructing Ancestral Genomic Orders Using Binary Encoding and Probabilistic Models. In Cai, Z, Eulenstein, O, Janies, D, and Schwartz, D (eds.), *Proceedings of the 9th International Symposium on Bioinformatics Research and Applications (ISBRA)*, volume 7875 of Lecture Notes in Computer Science, pages 17–27. Springer Berlin Heidelberg.

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.

Huang Y, Li Y, Burt DW, Chen H, Zhang Y, Qian W *et al.* (2013) The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. *Nat. Genet.* **45**, 776-783.

Huang Y, Zhao Y, Haley CS, Hu S, Hao J, Wu C, Li N (2006) A genetic and cytogenetic map for the duck (*Anas platyrhynchos*). *Genetics* **173**, 287-296.

Hughes AL and Hughes MK (1995) Small genomes for better flyers. *Nature* **377**, 391.

Hughes AL and Friedman R (2008) Genome size reduction in the chicken has involved massive loss of ancestral protein-coding genes. *Mol. Biol. Evol.* **25**, 2681-2688.

Hutt FB (1936) Genetics of the fowl. *Journal of Genetics* **32**, 277

International Chicken Genome Sequencing Consortium (ICGSC) (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695-716.

S. Joseph

Islam FB, Uno Y, Nunome M, Nishimura O, Tauri H, Agata K *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 by Comparative Chromosome Mapping. *J. of Poult. Sci.* **51**, 1-13.

Itoh Y and Arnold AP (2005) Chromosomal polymorphism and comparative painting analysis in the zebra finch. *Chromosome Res.* **13**, 47-56.

IUCN (2017) The IUCN Red List of Threatened species. Version 2017-2.
<http://www.iucnredlist.org>.

Jarvis ED (2004) Learned birdsong and the neurobiology of human language. *Ann. N. Y. Acad. Sci.* **1016**, 749-777.

Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C *et al.* (2014) Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320-1331.

Johnson JA and Burnham KK (2011) Population differentiation and adaptive selection on plumage color distributions in Gyrfalcons. In *Gyrfalcons and Ptarmigan in a Changing World*. Watson RT, Cade TJ, Fuller M, Hunt G, Potapov E (eds.) pp. 71-91. The Peregrine Fund, Boise, Idaho, USA.

Johnson JA, Ambers AD, Burnham KK (2012) Genetics of plumage color in the Gyrfalcon (*Falco rusticolus*): analysis of the melanocortin-1 receptor gene. *J. Hered.* **103**, 315-321.

Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F *et al.* (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* **258**, 818-821.

Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **38**, D355-D360.

Kasai F, O'Brien PC, Martin S, Ferguson-Smith MA (2012) 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.

Kawano Y, Ishikawa N, Aida J, Sanada Y, Izumiyama-Shimomura N, Nakamura K *et al.* (2014) Q-FISH measurement of hepatocyte telomere lengths in donor liver and graft after pediatric living-donor liver transplantation: donor age affects telomere length sustainability. *PLoS One.* **9**, e93749.

S. Joseph

Kim EB, Fang X, Fushan AA, Huang Z, Lobanov AV, Han L *et al.* (2011) Genome sequencing reveals insights into physiology and longevity of the naked mole rat. *Nature* **479**, 223-227.

Kim J, Larkin DM, Cai Q, Asan, Zhang Y, Ge RL *et al.* (2013) Reference-assisted chromosome assembly. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1785-1790.

Kimball RT, Wang N, Heimer-McGinn V, Ferguson C, Braun EL (2013) Identifying localized biases in large datasets: a case study using the avian tree of life. *Mol. Phylogenet. Evol.* **69**, 1021-1032.

Kircher M and Kelso J (2010) High-throughput DNA sequencing-concepts and limitations. *Bioessays*. **32**, 524-536.

Kirkpatrick M (2010) How and why chromosome inversions evolve. *PLoS. Biol.* **8**, e1000501.

Kirkpatrick M and Barton N (2006) Chromosome Inversions, Local Adaptation and Speciation. *Genetics* **173**, 419.

Koepfli KP, Paten B, O'Brien SJ (2015) The Genome 10K Project: a way forward. *Annu. Rev. Anim Biosci.* **3**, 57-111.

Kong F, Zhu J, Wu J, Peng J, Wang Y, Wang Q *et al.* (2011) dbCRID: a database of chromosomal rearrangements in human diseases. *Nucleic Acids Res.* **39**, D895-D900.

Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G *et al.* (2012) Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat. Biotechnol.* **30**, 693-700.

Kozubek S, Lukasova E, Jirsova P, Koutna I, Kozubek M, Ganova A *et al.* (2002) 3D Structure of the human genome: order in randomness. *Chromosoma* **111**, 321-331.

Kupper C, Augustin J, Edwards S, Szekely T, Kosztolanyi A, Burke T *et al.* (2012) Triploid plover female provides support for a role of the W chromosome in avian sex determination. *Biol. Lett.* **8**, 787-789.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921.

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.

Lejeune J, Gautier M, Turpin R (1959) Study of somatic chromosomes from 9 mongoloid children. *C R Hebd. Seances Acad. Sci.* **248**, 1721-1722.

S. Joseph

Lemaitre C, Zaghloul L, Sagot MF, Gautier C, Arneodo A, Tannier E *et al.* (2009) Analysis of fine-scale mammalian evolutionary breakpoints provides new insight into their relation to genome organization. *BMC Genomics* **10**:335.

Lewin HA, Larkin DM, Pontius J, O'Brien SJ (2009) Every genome sequence needs a good map. *Genome Res.* **19**, 1925-1928.

Li H and Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493-496.

Li H, Coghlan A, Ruan J, Coin LJ, Heriche JK, Osmotherly L *et al.* (2006) TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res.* **34**, D572-D580.

Liang G and Chen H (2015) Scaling Chromosomes for an Evolutionary Karyotype: A Chromosomal Tradeoff between Size and Number across Woody Species. *PLoS One.* **10**, e0144669.

Lin M, Thorne MH, Martin IC, Sheldon BL, Jones RC (1995) Development of the gonads in the triploid (ZZW and ZZZ) fowl, *Gallus domesticus*, and comparison with normal diploid males (ZZ) and females (ZW). *Reprod. Fertil. Dev.* **7**, 1185-1197.

Lithgow PE, O'Connor R, Smith D, Fonseka G, Al Mutery A, Rathje C *et al.* (2014) Novel tools for characterizing inter and intra chromosomal rearrangements in avian microchromosomes. *Chromosome Res.* **22**, 85-97.

Livezey BC, Zusi RL (2007) Higher-order phylogeny of modern birds (Theropoda, Aves: Neornithes) based on comparative anatomy. II. Analysis and discussion. *Zool. J. Linn. Soc.* **149**, 1-95.

Longmire JL, Lewis AK, Brown NC, Buckingham JM, Clark LM, Jones MD *et al.* (1988) Isolation and molecular characterization of a highly polymorphic centromeric tandem repeat in the family falconidae. *Genomics* **2**, 14-24.

Longrich NR, Tokaryk T, Field DJ (2011) Mass extinction of birds at the Cretaceous-Paleogene (K-Pg) boundary. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 15253-15257.

Louis A, Muffato M, Roest CH (2013) Genomicus: five genome browsers for comparative genomics in eukaryota. *Nucleic Acids Res.* **41**, D700-D705.

Lu H, Giordano F, Ning Z (2016) Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics Proteomics Bioinformatics* **14**, 265-279.

Ma J, Zhang L, Suh BB, Raney BJ, Burhans RC, Kent WJ *et al.* (2006) Reconstructing contiguous regions of an ancestral genome. *Genome Res.* **16**, 1557-1565.

S. Joseph

Magadum S, Banerjee U, Murugan P, Gangapur D, Ravikesavan R (2013) Gene duplication as a major force in evolution. *J. Genet.* **92**, 155-161.

Mardis ER (2008) Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.* **9**, 387-402.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-380.

Masabanda JS, Burt DW, O'Brien PC, Vignal A, Fillon V, Walsh PS *et al.* (2004) Molecular cytogenetic definition of the chicken genome: the first complete avian karyotype. *Genetics* **166**, 1367-1373.

McCormack JE, Harvey MG, Faircloth BC, Crawford NG, Glenn TC, Brumfield RT (2013) A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. *PLoS. One.* **8**, e54848.

McPherson MC, Robinson CM, Gehlen LP, Delany ME (2014) Comparative cytogenomics of poultry: mapping of single gene and repeat loci in the Japanese quail (*Coturnix japonica*). *Chromosome Res.* **22**, 71-83.

McQueen HA, Fantès J, Cross SH, Clark VH, Archibald AL, Bird AP (1996) CpG islands of chicken are concentrated on microchromosomes. *Nat. Genet.* **12**, 321-324.

Mehta IS, Amira M, Harvey AJ, Bridger JM (2010) Rapid chromosome territory relocation by nuclear motor activity in response to serum removal in primary human fibroblasts. *Genome Biol.* **11**, R5-11.

Meyne J, Baker RJ, Hobart HH, Hsu TC, Ryder OA, Ward OG *et al.* (1990) Distribution of non-telomeric sites of the (TTAGGG)_n telomeric sequence in vertebrate chromosomes. *Chromosoma.* **99**, 3-10.

Miller JR, Delcher AL, Koren S, Venter E, Walenz BP, Brownley A *et al.* (2008) Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics* **24**, 2818-2824.

Mirzaghaderi G and Marzangi K (2015) IdeoKar: an ideogram constructing and karyotype analyzing software. *Caryologia* **68**.

Moore CM and Best RG (2001) Chromosome Preparation and Banding. In 'eLS'. (John Wiley & Sons, Ltd) DOI: 10.1038/npg.els.0001444

Moret BM, Wyman S, Bader DA, Warnow T, Yan M (2001) A new implementation and detailed study of breakpoint analysis. *Pac. Symp. Biocomput.* 583-594.

S. Joseph

Muffato M, Louis A, Poisnel CE, Roest CH (2010) Genomicus: a database and a browser to study gene synteny in modern and ancestral genomes. *Bioinformatics*. **26**, 1119-1121.

Muffato M and Roest CH (2008) Paleogenomics in vertebrates, or the recovery of lost genomes from the mist of time. *Bioessays*. **30**, 122-134.

Munster VJ, Baas C, Lexmond P, Waldenstrom J, Wallensten A, Fransson T *et al.* (2007) Spatial, temporal, and species variation in prevalence of influenza A viruses in wild migratory birds. *PLoS. Pathog.* **3**, e61.

Murphy WJ, Larkin DM, Everts-van der Wind A, Bourque G, Tesler G, Auvil L *et al.* (2005) Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science* **309**, 613-617.

Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ *et al.* (2000) A whole-genome assembly of *Drosophila*. *Science* **287**, 2196-2204.

Nam K, Mugal C, Nabholz B, Schielzeth H, Wolf JB, Backstrom N *et al.* (2010) Molecular evolution of genes in avian genomes. *Genome Biol.* **11**, R68-11.

Nanda I, Benisch P, Fetting D, Haaf T, Schmid M (2011) Synteny conservation of chicken macrochromosomes 1-10 in different avian lineages revealed by cross-species chromosome painting. *Cytogenet. Genome Res.* **132**, 165-181.

Nanda I, Karl E, Griffin DK, Scharl M, Schmid M (2007) Chromosome repatterning in three representative parrots (Psittaciformes) inferred from comparative chromosome painting. *Cytogenet. Genome Res.* **117**, 43-53.

Nanda I, Karl E, Volobouev V, Griffin DK, Scharl M, Schmid M (2006) Extensive gross genomic rearrangements between chicken and Old World vultures (Falconiformes: Accipitridae). *Cytogenet. Genome Res.* **112**, 286-295.

Nanda I, Schrama D, Feichtinger W, Haaf T, Scharl M, Schmid M (2002) Distribution of telomeric (TTAGGG)(n) sequences in avian chromosomes. *Chromosoma*. **111**, 215-227.

Nie W, O'Brien PC, Fu B, Wang J, Su W, He K *et al.* (2015) Multidirectional chromosome painting substantiates the occurrence of extensive genomic reshuffling within Accipitriformes. *BMC Evol. Biol.* **15**, 205-0484.

Nie W, O'Brien PC, Ng BL, Fu B, Volobouev V, Carter NP *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 Res.* **17**, 99-113.

S. Joseph

Nishida C, Ishijima J, Kosaka A, Tanabe H, Habermann F, Griffin D *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.

Nittinger F, Haring E, Inskler W, Wink M, Gamauf A (2005) Out of Africa? Phylogenetic relationships between *Falco biarmicus* and the other hierofalcons (Aves: Falconidae). *J. Zoological Syst. Evol. Res.* **43**, 321-331.

Nittinger F, Gamauf A, Inskler W, Wink M, Haring E (2007) Phylogeography and population structure of the saker falcon (*Falco cherrug*) and the influence of hybridization: mitochondrial and microsatellite data. *Mol. Ecol.* **16**, 1497-1517.

Nussey DH, Baird D, Barrett E, Boner W, Fairlie J, Gemmell N, Hartmann N, Horn T, Haussmann M, Olsson M, Turbill C, Verhulst S, Zahn S, Monaghan P (2014) Measuring telomere length and telomere dynamics in evolutionary biology and ecology. *Methods Ecol. Evol.* **5**, 299-310.

O'Brien SJ, Haussler D, Ryder O (2014) The birds of Genome10K. *GigaScience* **3**, 32-33.

Ocalewicz K (2012) Genomic Distribution of Telomeric DNA Sequences - What Do We Learn from Fish About Telomere Evolution? *In Reviews on Selected Topics of Telomere Biology*. Li, B (ed.) pp. 271-294 DOI: 10.5772/38397:

O'Connor C (2008a) Karyotyping for chromosomal abnormalities. *Nature Education* **1**, 27.

O'Connor C (2008b) Fluorescence in situ hybridization (FISH). *Nature Education* **1**, 171.

O'Connor RE (2016) *Reproductive isolation, in individuals and during evolution, as result of gross genomic rearrangement in pigs, birds and dinosaurs*. (Doctoral dissertation) The University of Kent, U.K.

O'Connor RE, Fonseka G, Frodsham R, Archibald AL, Lawrie M, Walling GA, Griffin DK (2017) Isolation of subtelomeric sequences of porcine chromosomes for translocation screening reveals errors in the pig genome assembly. *Anim Genet.* **48**, 395-403.

O'Hare TH and Delany ME (2009) Genetic variation exists for telomeric array organization within and among the genomes of normal, immortalized, and transformed chicken systems. *Chromosome Res.* **17**, 947-964.

Oleksyk T, Pombert JF, Siu D, Mazo-Vargas A, Ramos B, Guiblet W *et al.* (2012) A locally funded Puerto Rican parrot (*Amazona vittata*) genome sequencing project increases avian data and advances young researcher education. *GigaScience* **1**, 14.

S. Joseph

Olsen B, Munster VJ, Wallensten A, Waldenstrom J, Osterhaus AD, Fouchier RA (2006) Global patterns of influenza A virus in wild birds. *Science* **312**, 384-388.

Otto SP (2007) The evolutionary consequences of polyploidy. *Cell* **131**, 452-462.

Pauliny A, Larsson K, Blomqvist D (2012) Telomere dynamics in a long-lived bird, the barnacle goose. *BMC Evol. Biol.* **12**, 257.

Pennacchio LA and Rubin EM (2003) Comparative genomic tools and databases: providing insights into the human genome. *J. Clin. Invest.* **111**, 1099-1106.

Ponitz B, Schmitz A, Fischer D, Bleckmann H, Brucker C (2014) Diving-flight aerodynamics of a peregrine falcon (*Falco peregrinus*). *PLoS. One.* **9**, e86506.

Porter SL (1993) Pesticide poisoning in birds of prey. In *Raptor biomedicine* Redig PT, Cooper JE, Remple JD, Remple D, Hunter DB (eds.) pp. 239-245. University of Minnesota Press, Minneapolis, U.S.A.

Pourjafari B, Pour-Jafari H, Farimani M, Ghahramani S, Saleh EK (2012) Genetic counseling in carriers of reciprocal translocations involving two autosomes. *Indian J. Hum. Genet.* **18**, 250-253.

Pratt RC, Gibb GC, Morgan-Richards M, Phillips MJ, Hendy MD, Penny D (2009) Toward resolving deep neaves phylogeny: data, signal enhancement, and priors. *Mol. Biol. Evol.* **26**, 313-326.

Prum RO, Berv JS, Dornburg A, Field DJ, Townsend JP, Lemmon EM *et al.* (2015) A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* **526**, 569-573.

Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR *et al.* (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**, 341-13.

Rabl C (1885) Über Zelltheilung. *Morph. Jb.* **10**, 214-330.

Rajaraman A and Ma J (2016) Reconstructing ancestral gene orders with duplications guided by synteny level genome reconstruction. *BMC Bioinformatics.* **17**, 414-1262.

Rands CM, Darling A, Fujita M, Kong L, Webster MT, Clabaut C *et al.* (2013) Insights into the evolution of Darwin's finches from comparative analysis of the *Geospiza magnirostris* genome sequence. *BMC Genomics.* **14**, 95-14.

Rao M, Morisson M, Faraut T, Bardes S, Feve K, Labarthe E *et al.* (2012) A duck RH panel and its potential for assisting NGS genome assembly. *BMC Genomics* **13**, 513.

S. Joseph

Raudsepp T, Houck ML, O'Brien PC, Ferguson-Smith MA, Ryder OA, Chowdhary BP (2002) Cytogenetic analysis of California condor (*Gymnogyps californianus*) chromosomes: comparison with chicken (*Gallus gallus*) macrochromosomes. *Cytogenet. Genome Res.* **98**, 54-60.

Remple D and Gross C (1993) *Falconry and birds of prey in the Gulf*. Motivate Publishing, Dubai, UAE.

Ren C, Lee MK, Yan B, Ding K, Cox B, Romanov MN *et al.* (2003) A BAC-based physical map of the chicken genome. *Genome Res.* **13**, 2754-2758.

Rhoads A and Au KF (2015) PacBio sequencing and its applications. *Genomics, Proteomics & Bioinformatics* **13**, 278-289.

Riegel M (2014) Human molecular cytogenetics: From cells to nucleotides. *Genetics and Molecular Biology* **37**, 194-209.

Romanov MN, Farre M, Lithgow PE, Fowler KE, Skinner BM, O'Connor R *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**, 1060.

Roper TJ (1999) Olfaction in birds. *Adv. Stud. Behav.* **28**, 247-332.

Ruiz-Herrera A, Castresana J, Robinson TJ (2006) Is mammalian chromosomal evolution driven by regions of genome fragility? *Genome Biol.* **7**, R115.

Sadoni N, Langer S, Fauth C, Bernardi G, Cremer T, Turner BM, Zink D (1999) Nuclear organization of mammalian genomes. Polar chromosome territories build up functionally distinct higher order compartments. *J Cell Biol.* **146**, 1211-1226.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463-5467.

Sankoff D (2009) The where and wherefore of evolutionary breakpoints. *J. Biol.* **8**, 66.

Sasaki M, Takagi N, Nishida C (1984) Current profiles of avian cytogenetics, with notes on chromosomal diagnosis of sex in birds. *The Nucleus* **27**, 63-73.

Schardin M, Cremer T, Hager HD, Lang M (1985) Specific staining of human chromosomes in Chinese hamster x man hybrid cell lines demonstrates interphase chromosome territories. *Hum. Genet.* **71**, 281-287.

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467-470.

S. Joseph

Schimid M, Smith J, Burt DW, Aken BL, Antin BL, Archibald AL *et al.* (2015) Third Report on Chicken Genes and Chromosomes 2015. *Cytogenet. Genome Res.* **145**, 78-179.

Schmid M, Nanda I, Hoehn H, Schartl M, Haaf T, Buerstedde JM *et al.* (2005) Second report on chicken genes and chromosomes 2005. *Cytogenet. Genome Res.* **109**, 415-479.

Schmutz S and Oliphant, L W (1987) Chromosome study of peregrine, prairie, and gyrfalcons with implications for hybrids. *J. Hered.* **78**(6), 388-390.

Schrock E, du MS, Veldman T, Schoell B, Wienberg J, Ferguson-Smith MA *et al.* (1996) Multicolor spectral karyotyping of human chromosomes. *Science* **273**, 494-497.

Seabury CM, Dowd SE, Seabury PM, Raudsepp T, Brightsmith DJ, Liboriussen P *et al.* (2013) A Multi-Platform draft de novo genome assembly and comparative analysis for the Scarlet Macaw (*Ara macao*). *Plos One* **8**(5) e62415.

Seibold I, Helbig AJ, Wink M (1993) Molecular systematics of falcons (family Falconidae). *Naturwissenschaften* **80**, 87-90.

Seibold-Torres C, Owens E, Chowdhary R, Ferguson-Smith MA, Tizard I, Raudsepp T (2015) Comparative Cytogenetics of the Congo African Grey Parrot (*Psittacus erithacus*). *Cytogenet. Genome Res.* **147**, 144-153.

Sessarego M, Panarello C, Coviello DA, Boccaccio P, Ajmar F (1987) Karyotype evolution in CML: high frequency of translocations other than the Ph. *Cancer Genet. Cytogenet.* **25**, 73-80.

Shapiro B and Hofreiter M (2014) A paleogenomic perspective on evolution and gene function: new insights from ancient DNA. *Science.* **343**, 1236573.

Shapiro MD, Kronenberg Z, Li C, Domyan, ET, Pan H, Campbell M *et al.* (2013) Genomic diversity and evolution of the head crest in the rock pigeon. *Science* **339**, 1063-1067.

Shedlock AA and Edwards SV (2009) Amniotes. In the *Timetree of Life*. Hedges SB and Kumar S (eds.) pp. 375-379, Oxford University Press.

Shetty S, Griffin DK, Graves JA (1999) Comparative painting reveals strong chromosome homology over 80 million years of bird evolution. *Chromosome. Res.* **7**, 289-295.

Shibusawa M, Nishibori M, Nishida-Umehara C, Tsudzuki M, Masabanda J, Griffin DK, Matsuda Y (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. *Cytogenet. Genome Res.* **106**, 111-119.

S. Joseph

Shibusawa M, Nishida-Umehara C, Masabanda J, Griffin DK, Isobe T, Matsuda Y (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.

Sibley CG and Ahlquist JE (1990) *Phylogeny and Classification of Birds: A Study in Molecular Evolution*. Yale Univ. Press, New haven, U.S.A.

Slijepcevic P (2001) Telomere length measurement by Q-FISH. *Methods Cell Sci.* **23**, 17-22.

Skinner BM and Griffin DK (2012) Intrachromosomal rearrangements in avian genome evolution: evidence for regions prone to breakpoints. *Heredity* (Edinb.) **108**, 37-41.

Skinner BM, Al Mutery A, Smith D, Volker M, Hojjat N, Raja S et al. (2014) Global patterns of apparent copy number variation in birds revealed by cross-species comparative genomic hybridization. *Chromosome Res.* **22**, 59-70.

Skinner BM, Robertson LB, Tempest HG, Langley EJ, Ioannou D, Fowler KE et al. (2009a) Comparative genomics in chicken and Pekin duck using FISH mapping and microarray analysis. *BMC Genomics.* **10**, 357-10.

Skinner BM, Volker M, Ellis M, Griffin DK (2009b) An appraisal of nuclear organization in interphase embryonic fibroblasts of chicken, turkey and duck. *Cytogenet. Genome Res.* **126**, 156-164.

Smith J, Bruley CK, Paton IR, Dunn I, Jones CT, Windsor D et al. (2000) Differences in gene density on chicken macrochromosomes and microchromosomes. *Anim. Genet.* **31**, 96-103.

Soorae PS, Suwaidi KA, Shamsi AA (2007) Implementation of the cites convention in the UAE. *Wildlife Middle East* **1**, 3.

Speicher MR, Carter NP (2005) The new cytogenetics: blurring the boundaries with molecular biology. *Nat. Rev. Genet.* **6**, 782-792.

Speicher MR and Gwyn BS, Ward DC (1996) Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nat. Genet.* **12**, 368-375.

Stack SM, Brown DB, Dewey WC (1977) Visualization of interphase chromosomes. *J. Cell Sci.* **26**, 281-299.

Stapley J, Birkhead TR, Burke T, Slate J (2008) A linkage map of the zebra finch (*Taeniopygia guttata*) provides new insights into avian genome evolution. *Genetics* **179**, 651-667.

S. Joseph

Stefansson H, Meyer-Lindenberg A, Steinberg S, Magnusdottir B, Morgen K, Arnarsdottir S (2014) CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature* **505**, 361-366.

Steidl RJ (1991) Reproductive success and eggshell thinning of a reestablished peregrine falcon population. *J. Wildl. Manage.* **55**, 294-299.

Steiger SS, Fidler AE, Valcu M, Kempnaers B (2008) Avian olfactory receptor gene repertoires: evidence for a well-developed sense of smell in birds? *Proc. Biol. Sci.* **275**, 2309-2317.

Steiger SS, Kuryshev VY, Stensmyr MC, Kempnaers B, Mueller JC (2009) A comparison of reptilian and avian olfactory receptor gene repertoires: species-specific expansion of group gamma genes in birds. *BMC Genomics* **10**, 446-10.

Sun HB, Shen J, Yokota H (2000) Size-dependent positioning of human chromosomes in interphase nuclei. *Biophys. J.* **79**, 184-190.

Suschkin P (1905) Zur Morphologie des Vogelskelets. Vergleichende Osteologie der normalen Tagrauhvögel (Accipitres) und die Fragen der Classification. *Nouv. Mem. Soc. Imp. Nat. Moscou* **16**, 1-247.

Tajika Y, Matsuzaki T, Suzuki T, Ablimit A, Aoki T, Hagiwara H et al. (2005) Differential regulation of AQP2 trafficking in endosomes by microtubules and actin filaments. *Histochem. Cell Biol.* **124**, 1-12.

Tamazian G, Dobrynin P, Krashennnikova K, Komissarov A, Koepfli KP, O'Brien SJ (2016) Chromosomer: a reference-based genome arrangement tool for producing draft chromosome sequences. *Gigascience* **5**, 38-0141.

Tanabe H, Muller S, Neusser M, von HJ, Calcagno E, Cremer M, Solovei I, Cremer C, Cremer T (2002) Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4424-4429.

Theisen A (2008) Microarray-based Comparative Genomic Hybridization (aCGH). *Nature Education* **1**, 45.

Touchman J (2010) Comparative genomics. *Nature Education Knowledge* **3**, 13.

Tucker VA (1998) Gliding flight: speed and acceleration of ideal falcons during diving and pull out. *The Journal of Experimental Biology* **201**, 403-414.

Vetro A, Bouman K, Hastings R, McMullan DJ, Vermeesch JR, Miller K et al. (2012) The introduction of arrays in prenatal diagnosis: a special challenge. *Hum. Mutat.* **33**, 923-929.

S. Joseph

Visser AE, Eils R, Jauch A, Little G, Bakker PJ, Cremer T, Aten JA (1998) Spatial distributions of early and late replicating chromatin in interphase chromosome territories. *Exp. Cell Res.* **243**, 398-407.

Volker M, Backstrom N, Skinner BM, Langley EJ, Bunzey SK, Ellegren H *et al.* (2010) Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution. *Genome Res.* **20**, 503-511.

Voss SR, Kump DK, Putta S, Pauly N, Reynolds A, Henry RJ *et al.* (2011) Origin of amphibian and avian chromosomes by fission, fusion, and retention of ancestral chromosomes. *Genome Res.* **21**, 1306-1312.

Wallis JW, Aerts J, Groenen MA, Crooijmans RP, Layman D, Graves TA *et al.* (2004) A physical map of the chicken genome. *Nature* **432**, 761-764.

Wang B, Ekblom R, Bunikis I, Siitari H, Hoglund J (2014a) Whole genome sequencing of the black grouse (*Tetrao tetrix*): reference guided assembly suggests faster-Z and MHC evolution. *BMC Genomics.* **15**, 180-15.

Wang Y, Li Y, Gao J, Zhang MQ (2015) A novel method to identify topological domains using Hi-C data. *Quantitative Biology* **3**, 81-89.

Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Kunstner A *et al.* (2010) The genome of a songbird. *Nature* **464**, 757-762.

Weiss MM, Hermsen MA, Meijer GA, van Grieken NC, Baak JP, Kuipers EJ *et al.* (1999) Comparative genomic hybridisation. *Molecular Pathology* **52**, 243-251.

Weiss-Schneeweiss H, Tremetsberger K, Schneeweiss GM, Parker JS, Stuessy TF (2008) Karyotype diversification and evolution in diploid and polyploid South American *Hypochaeris* (Asteraceae) inferred from rDNA localization and genetic fingerprint data. *Ann. Bot.* **101**, 909-918.

Wernery U (2017) Falconry becomes a cultural asset to mankind. In *Asil Arabians VII, The Noble Arabian Horses*. Olms G (ed.), Georg Olms, Heidesheim, Zürich, Newyork, Asil Club.

Wienberg J (2004) The evolution of eutherian chromosomes. *Curr Opin Genet Dev.* **14**, 657-666.

Wilcox BR, Knutsen GA, Berdeen J, Goekjian V, Poulson R, Goyal S *et al.* (2011) Influenza-A viruses in ducks in northwestern Minnesota: fine scale spatial and temporal variation in prevalence and subtype diversity. *PLoS. One.* **6**, e24010.

Zann RA (1996) *The Zebra Finch: A synthesis of Field and Laboratory Studies*. Oxford Ornithology Series, UK.

S. Joseph

Zhan XJ, Dixon A, Fox NC, Bruford MW. (2012). Missense SNP of the MC1R gene is associated with plumage variation in the Gyrfalcon (*Falco rusticolus*). *Anim. Genet.* 43: 460–462.

Zhan X, Pan S, Wang J, Dixon A, He J, Muller MG *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 C, Li Q, Li B, Larkin DM, Lee C *et al.* (2014) Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* 346, 1311-1320.

Zhang G, Rahbek C, Graves GR, Lei F, Jarvis ED, Gilbert MT (2015) Genomics: Bird sequencing project takes off. *Nature* **522**, 34.

Zhao, H. and Bourque, G (2007) Recovering True Rearrangement Events on Phylogenetic Trees. In Tesler, G and Durand, D, (eds.) *Proceedings of the 5th Annual RECOMB Satellite Workshop on Comparative Genomics (RECOMB-CG)*, volume 4751 of Lecture Notes in Computer Science, pages 149–161. Springer Berlin Heidelberg.

9 Appendix

Table S1: List of BACs and their assigned position in the Saker falcon (FCH - *Falco cherrug*) PCFs.

| FCH Chrom | BAC Name | FCH Start Position | FCH End Position | GGA Homolog |
|-----------|---------------|--------------------|------------------|-------------|
| 1 | CH261-18C6 | 3,334,762 | 3,587,608 | 4 |
| 1 | TGMCBA-104H6 | 7,209,109 | 7,420,615 | 4 |
| 1 | CH261-89P6 | 14,860,757 | 15,002,448 | 4 |
| 1 | CH261-185L11 | 18,879,513 | 19,145,312 | 4 |
| 1 | CH261-93H1 | 2,643,794 | 2,859,725 | 4 |
| 1 | CH261-85H10 | 3,663,901 | 3,909,243 | 4 |
| 1 | TGMCBA-216A16 | 5,182,604 | 5,336,447 | 4 |
| 1 | TGMCBA-231D20 | 14,602,901 | 14,807,084 | 15 |
| 1 | CH261-48M1 | 14,481,079 | 14,664,964 | 15 |
| 1 | CH261-90P23 | 13,494,236 | 13,679,624 | 15 |
| 1 | CH261-40D6 | 8,482,011 | 8,675,153 | 15 |
| 1 | TGMCBA-266G23 | 7,713,316 | 7,868,177 | 15 |
| 1 | CH261-10F1 | 8,390,823 | 8,531,352 | 19 |
| 1 | TGMCBA-307H9 | 6,637,566 | 6,754,282 | 19 |
| 1 | TGMCBA-356O18 | 6,053,669 | 6,201,255 | 19 |
| 1 | CH261-50H12 | 2,352,774 | 2,504,186 | 19 |
| 1 | TGMCBA-84A3 | 1,856,318 | 1,980,932 | 19 |
| 1 | CH261-67N15 | 65,829 | 249,847 | 18 |
| 1 | TGMCBA-263I20 | 4,424,202 | 4,569,506 | 18 |
| 1 | CH261-72B18 | 5,746,324 | 5,918,647 | 18 |
| 1 | CH261-137B21 | 8,046,910 | 8,198,673 | 18 |
| 1 | CH261-118D24 | 8,801,284 | 8,971,729 | 18 |
| 2 | CH261-17J16 | 4,755,817 | 4,936,970 | 2 |
| 2 | CH261-44H14 | 19,718,549 | 19,924,265 | 2 |
| 2 | CH261-44D16 | 28,735,657 | 28,940,732 | 2 |
| 2 | CH261-1J20 | 35,059,326 | 35,294,717 | 2 |
| 2 | CH261-169E4 | 44,364,589 | 44,593,848 | 2 |
| 2 | TGMCBA-78C11 | 49,063,209 | 49,189,450 | 2 |
| 2 | TGMCBA-340P4 | 11,304,634 | 11,459,796 | 2 |
| 2 | CH261-10A18 | 44,683 | 245,368 | 21 |
| 2 | CH261-83I20 | 2,056,292 | 2,250,140 | 21 |
| 2 | CH261-122K8 | 1,191,880 | 1,372,993 | 21 |
| 2 | TGMCBA-134A8 | 2,419,865 | 2,575,824 | 21 |
| 2 | TGMCBA-48O8 | 2,797,079 | 2,922,593 | 23 |
| 2 | CH261-90K11 | 2,518,573 | 2,680,838 | 23 |
| 2 | CH261-191G17 | 1,620,868 | 1,839,145 | 23 |
| 2 | CH261-49G9 | 1,419,736 | 1,641,357 | 23 |
| 2 | TGMCBA-272G9 | 1,385,064 | 1,584,975 | 23 |
| 2 | TGMCBA-173N15 | 1,060,484 | 1,195,888 | 23 |

Table S1(Continued): List of BACs and their assigned position in the Saker falcon (FCH - *Falco cherrug*) PCFs.

| FCH Chrom | BAC Name | FCH Start Position | FCH End Position | GGA Homolog |
|-----------|---------------|--------------------|------------------|-------------|
| 3 | CH261-98G4 | 10,179,570 | 10,362,438 | 1 |
| 3 | CH261-184E5 | 32,723,924 | 32,957,933 | 1 |
| 3 | CH261-58K12 | 26,622,038 | 26,824,691 | 1 |
| 3 | CH261-107E2 | 15,646,509 | 15,875,568 | 1 |
| 3 | CH261-83O13 | 6,062,360 | 6,302,017 | 1 |
| 3 | CH261-168O17 | 769,258 | 1,015,280 | 1 |
| 3 | CH261-9B17 | 31,891,430 | 32,096,024 | 1 |
| 3 | TGMCBA-204O4 | 24,042,794 | 24,217,513 | 1 |
| 3 | CH261-29N14 | 9,745,037 | 9,941,716 | 1 |
| 3 | CH261-118M1 | 9,573,731 | 9,803,041 | 1 |
| 3 | TGMCBA-146O14 | 2,708,651 | 2,873,156 | 1 |
| 4 | CH261-123O22 | 894,898 | 1,072,944 | 2 |
| 4 | CH261-50C15 | 4,654,610 | 4,889,746 | 2 |
| 4 | CH261-40G6 | 11,026,420 | 11,214,283 | 2 |
| 4 | CH261-186J5 | 19,653,684 | 19,881,513 | 2 |
| 4 | CH261-177K1 | 25,028,646 | 25,272,224 | 2 |
| 4 | CH261-172N3 | 35,353,391 | 35,501,661 | 2 |
| 4 | CH261-169N6 | 43,333,754 | 43,529,335 | 2 |
| 4 | CH261-64A15 | 395,201 | 565,281 | 28 |
| 4 | TGMCBA-37M13 | 16,494 | 271,636 | 28 |
| 4 | CH261-72A10 | 126,006 | 336,162 | 28 |
| 4 | TGMCBA-205N19 | 14,531,687 | 14,676,648 | 14 |
| 4 | CH261-122H14 | 14,267,138 | 14,465,254 | 14 |
| 4 | CH261-49P24 | 12,702,049 | 12,866,007 | 14 |
| 4 | CH261-69D20 | 4,817,822 | 5,005,181 | 14 |
| 4 | CH261-4M5 | 1,346,351 | 1,538,673 | 12 |
| 4 | TGMCBA-342P15 | 2,250,485 | 2,391,544 | 12 |
| 4 | CH261-95H20 | 2,853,308 | 3,074,263 | 12 |
| 4 | CH261-60P3 | 7,654,537 | 7,796,531 | 12 |
| 4 | TGMCBA-305E19 | 4,437,620 | 4,615,575 | 12 |
| 5 | CH261-89G23 | 6,346,445 | 6,583,395 | 1 |
| 5 | TGMCBA-206D5 | 30,991,530 | 31,139,651 | 1 |
| 5 | CH261-120J2 | 16,059,377 | 16,292,990 | 1 |
| 5 | CH261-119K2 | 11,659,521 | 11,895,837 | 1 |
| 5 | CH261-25P18 | 3,582,503 | 3,825,147 | 1 |
| 5 | CH261-36B5 | 3,586,415 | 3,794,018 | 1 |

Table S1(Continued): List of BACs and their assigned position in the Saker falcon (FCH - *Falco cherrug*) PCFs.

| FCH Chrom | BAC Name | FCH Start Position | FCH End Position | GGA Homolog |
|-----------|---------------|--------------------|------------------|-------------|
| 6 | CH261-97P20 | 6,934,589 | 7,171,053 | 3 |
| 6 | CH261-120H23 | 11,022,299 | 11,203,000 | 3 |
| 6 | CH261-17B14 | 61,659,292 | 61,883,310 | 3 |
| 6 | TGMCBA-64D9 | 32,963,930 | 33,085,015 | 3 |
| 6 | CH261-169K18 | 24,553,218 | 24,693,790 | 3 |
| 6 | TGMCBA-250J17 | 5,660,867 | 5,832,201 | 3 |
| 7 | CH261-78F13 | 309,189 | 496,208 | 5 |
| 7 | CH261-2I23 | 19,421,864 | 19,589,758 | 5 |
| 7 | TGMCBA-24C1 | 2,160,637 | 2,312,160 | 5 |
| 7 | CH261-161B22 | 1,782,348 | 1,941,077 | 5 |
| 7 | TGMCBA-310P11 | 643,861 | 817,199 | 10 |
| 7 | CH261-118E15 | 2,618,017 | 2,802,450 | 10 |
| 7 | CH261-71G18 | 3,488,835 | 3,707,222 | 10 |
| 7 | CH261-115G24 | 8,218,803 | 8,430,968 | 10 |
| 7 | TGMCBA-48L18 | 21,995,517 | 22,129,373 | 10 |
| 8 | TGMCBA-224O13 | 577,580 | 724,131 | 7 |
| 8 | TGMCBA-34L13 | 4,194,398 | 4,298,655 | 7 |
| 8 | CH261-112D24 | 4,260,178 | 4,427,569 | 7 |
| 8 | CH261-56K7 | 33,157,653 | 33,319,307 | 7 |
| 8 | CH261-180H18 | 15,990,861 | 16,224,103 | 7 |
| 8 | CH261-186K14 | 10,442,986 | 10,619,170 | 7 |
| 8 | TGMCBA-356H21 | 8,486,277 | 8,607,193 | 7 |
| 8 | CH261-38E18 | 107,506 | 298,616 | 7 |
| 8 | TGMCBA-266O5 | 11,604,295 | 11,775,255 | 13 |
| 8 | CH261-115I12 | 2,834,792 | 3,029,013 | 13 |
| 8 | CH261-59M8 | 2,770,199 | 2,970,480 | 13 |
| 8 | TGMCBA-321B13 | 1,332,561 | 1,482,294 | 13 |
| 9 | CH261-179F2 | 440,896 | 610,105 | 6 |
| 9 | TGMCBA-382J4 | 12,515,558 | 12,653,112 | 6 |
| 9 | CH261-94G14 | 14,533,386 | 14,717,200 | 6 |
| 9 | CH261-49F3 | 6,811,223 | 6,923,943 | 6 |
| 9 | CH261-165L8 | 6,524,327 | 6,720,054 | 6 |
| 9 | CH261-67H5 | 6,274,734 | 6,524,207 | 6 |
| 9 | TGMCBA-375I5 | 2,006,136 | 2,135,424 | 17 |
| 9 | TGMCBA-67H23 | 3,826,985 | 3,947,417 | 17 |
| 9 | TGMCBA-185B22 | 4,945,113 | 5,046,612 | 17 |
| 9 | CH261-113A7 | 10,217,493 | 10,364,542 | 17 |
| 9 | TGMCBA-197G19 | 10,616,773 | 10,798,076 | 17 |
| 9 | CH261-42P16 | 10,932,885 | 11,104,907 | 17 |
| 9 | CH261-69M11 | 11,201,650 | 11,370,663 | 17 |

Table S1(Continued): List of BACs and their assigned position in the Saker falcon (*FCH - Falco cherrug*) PCFs.

| FCH Chrom | BAC Name | FCH Start Position | FCH End Position | GGA Homolog |
|-----------|---------------|--------------------|------------------|-------------|
| 10 | CH261-122F8 | 4,359,445 | 4,529,874 | 5 |
| 10 | CH261-49B22 | 3,303,066 | 3,496,466 | 5 |
| 10 | TGMCBA-145C6 | 146,887 | 295,380 | 5 |
| 10 | TGMCBA-341F20 | 8,962,854 | 9,090,534 | 20 |
| 10 | TGMCBA-225I12 | 6,624,006 | 6,781,353 | 20 |
| 10 | TGMCBA-250E3 | 718,666 | 867,524 | 20 |
| 11 | TGMCBA-346F6 | 4,521,291 | 4,693,655 | 8 |
| 11 | CH261-34H16 | 8,436,325 | 8,631,209 | 8 |
| 11 | TGMCBA-208D17 | 21,312,872 | 21,520,187 | 8 |
| 11 | CH261-96D24 | 22,351,449 | 22,544,739 | 8 |
| 12 | CH261-115J5 | 1,938,971 | 2,121,607 | 3 |
| 12 | CH261-18I9 | 4,558,851 | 4,781,284 | 3 |
| 12 | CH261-160I6 | 17,022,071 | 17,250,153 | 3 |
| 12 | CH261-130M12 | 21,723,515 | 21,888,968 | 3 |
| 13 | CH261-183N19 | 1,725,161 | 1,899,763 | 9 |
| 13 | TGMCBA-150E19 | 5,608,321 | 5,767,194 | 9 |
| 13 | CH261-95N3 | 11,961,758 | 12,155,423 | 9 |
| 13 | CH261-187M16 | 15,232,762 | 15,421,162 | 9 |
| 13 | TGMCBA-321L6 | 17,799,573 | 17,970,696 | 9 |
| 13 | TGMCBA-217A3 | 23,042,433 | 23,308,239 | 9 |
| 13 | CH261-68O18 | 23,447,830 | 23,639,377 | 9 |
| 14 | TGMCBA-220A5 | 2,191,038 | 2,339,395 | 4 |
| 14 | CH261-111A15 | 1,783,127 | 2,005,434 | 4 |
| 14 | TGMCBA-330J11 | 7,959,580 | 8,091,736 | 4 |
| 14 | CH261-83E1 | 7,339,415 | 7,511,502 | 4 |
| 14 | TGMCBA-280M7 | 3,248,848 | 3,390,398 | 4 |
| 15 | CH261-154H1 | 780,307 | 995,477 | 11 |
| 15 | CH261-138H13 | 7,365,815 | 7,546,392 | 11 |
| 15 | CH261-121N21 | 8,785,273 | 9,027,905 | 11 |
| 15 | TGMCBA-192A10 | 18,604,140 | 18,721,785 | 11 |
| 16 | CH261-65O4 | 35,053 | 186,611 | 24 |
| 16 | TGMCBA-111K1 | 2,862,067 | 3,002,831 | 24 |
| 16 | CH261-103F4 | 3,568,230 | 3,720,181 | 24 |
| 16 | CH261-154H17 | 6,300,513 | 6,498,093 | 24 |
| 17 | TGMCBA-97D20 | 716,499 | 801,912 | 26 |
| 17 | TGMCBA-297G21 | 2,146,297 | 2,255,284 | 26 |
| 17 | CH261-186M13 | 2,125,752 | 2,302,412 | 26 |
| 17 | TGMCBA-332G15 | 2,406,453 | 2,482,143 | 26 |
| 17 | CH261-170L23 | 2,071,246 | 2,265,819 | 26 |

Table S1(Continued): List of BACs and their assigned position in the Saker falcon (FCH - *Falco cherrug*) PCFs.

| FCH Chrom | BAC Name | FCH Start Position | FCH End Position | GGA Homolog |
|-----------|---------------|--------------------|------------------|-------------|
| 18 | CH261-18G17 | 4,487,150 | 4,701,972 | 22 |
| 18 | TGMCBA-113N13 | 4,496,512 | 4,656,382 | 22 |
| 18 | CH261-40J9 | 4,069,354 | 4,247,664 | 22 |
| 18 | TGMCBA-151I22 | 1,852,797 | 2,024,213 | 22 |
| 19 | TGMCBA-324P4 | 1,319,515 | 1,416,633 | 27 |
| 19 | TGMCBA-23C5 | 409,144 | 567,696 | 27 |
| 19 | CH261-66M16 | 384,826 | 560,963 | 27 |
| 19 | CH261-28L10 | 762,690 | 978,807 | 27 |
| 19 | CH261-100E5 | 978,902 | 1,143,628 | 27 |
| Z | TGMCBA-200J22 | 3,439,653 | 3,619,858 | Z |
| Z | CH261-129A16 | 4,791,598 | 5,017,472 | Z |
| Z | CH261-133M4 | 2,128,091 | 2,320,310 | Z |
| Z | TGMCBA-270I9 | 1,065,757 | 1,222,505 | Z |