Exploring Chromosomal Rearrangements in Avian and Non-Avian Reptiles to Study Amniote Genome Evolution

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

DOCTOR OF PHILOSOPHY

in the Faculty of Science

Lucas G. Kiazim

February 2020

The School of Biosciences

Declaration

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

Lucas G. Kiazim

February 2020

Incorporation of Published Work

Original Research Manuscripts:

O'Connor, R.E., Farré, M., Joseph, S., Damas, J., **Kiazim, L.G,** *et al.* (2018). Chromosome-level assembly reveals extensive rearrangement in saker falcon and budgerigar, but not ostrich, genomes. *Genome Biology*

O'Connor, R.E., **Kiazim, L.G**., *et al.* (2018). Patterns of microchromosome organisation remain highly conserved throughout avian evolution. *Chromosoma*

O'Connor R.E., Romanov M.N, **Kiazim, L.G,** *et al.* (2018). Reconstruction of the diapsid ancestral genome permits chromosome evolution tracing in avian and non-avian dinosaurs. *Nature Communications*

Original Research Manuscripts in Preparation:

Kiazim, L.G., Griffin, D.K, Romanov, M.N., Larkin, D., and O'Connor, R.E. Comparative mapping of the macrochromosomes of 7 avian species to investigate phylogenetic relationships and lineage-specific traits arising from chromosomal rearrangements

Ribas, T.F.A, Griffin, D.K., **Kiazim, L.G.,** Pieczarka, J.C., Nagamachi, C.Y., O'Brien, P.C.M., Ferguson-Smith, M.A., Yang, F., Aleixo, A., and O'Connor, R.E. Analysis of Chromosomal Break Points in the Genome of *Willisornis vidua* Reveals Multiple Rearrangements on a Supposedly Conserved Karyotype

Griffin, D.K., O'Connor, R.E., Silvestri, G., and **Kiazim, L.G**. Improving cattle production by counting chromosomes. Open Access Government article.

Poppleton, D., **Kiazim, L.G.** *et al.,* Genome Assembly of the Houbara Bustard

Published Abstracts:

O'Connor, R.E., …, **Kiazim, L.G.**, *et al.* (2016). Upgrading molecular cytogenetics to study reproduction and reproductive isolation in mammals, birds, and dinosaurs. *Cytogenetic and Genome Research*, 148(2–3), pp.151-152.

O'Connor, R.E., Damas, J., Fonseka, G., Jennings, R., **Kiazim, L.G.**, *et al.* (2016). Exceptional Degree of Avian Microchromosome Conservation Revealed by Cross Species BAC Mapping. *Chromosome Research*, 24: 1, p.23

Poster Presentations:

Kiazim, L.G., *et al.* International Chromosome Conference, Prague, August 2018. *Comparative genomics through the development of universal cross-species BAC sets*

Kiazim, L.G., *et al.* Genome 10K and Genome Science Conference, Norwich, September 2017. *Comparative BAC Mapping of Macrochromosomes from Nine Avian Species Reveals Strong Chromosome Homology Over 98 Million Years of Evolution.*

Kiazim, L.G., *et al.* European Cytogeneticists Association, Florence, July 2017. *Development of a chromosomal translocation screening device reveals errors in the pig genome assembly*

Grants Awarded:

Genetics Society Junior Scientist Conference Grant, non-GS Meeting (2018): £734

Acknowledgements

Before all else, I would like to thank my supervisor Professor Darren Griffin, who not only provided the opportunity to complete my PhD, but always offered support and guidance. The opportunities for both personal growth and career progression have been endless, and I cannot thank him enough for this. Likewise, I would like to thank Dr. Becky O'Connor (my unofficial co-supervisor) for patiently answering my unrelenting wave of questions. Your support, trust, and incredible organisation allowed for my research to be carried out seamlessly, and you instilled a confidence in me that has made me the scientist I am today.

My time in the lab has forged many friendships, but there is one person who I need to give special thanks to: Becca, you are one of the most caring people I know and you provided comic relief in some of the hardest moments of my life. It has been an honour to complete my PhD alongside you and I could not think of a better person to have done so with. To Claudia, Giuseppe, Jacob, Nikki, Lucy, and Ane, our lab shenanigans have been hilarious and incredible, and I will never forget them. To all other members of the Griffin, Ellis, and Farré labs, both past and present, each and everyone one of you have been influential in your own way, so thank you for your support.

Finally, I would like to thank my family. Even when you had no idea what I was rambling about, you let me express myself without questioning every unknown/obscure detail. You have kept me grounded and I never could have done this without each of you. To the newest member of the family, María, even though our journey has just begun you have supported me endlessly and have helped to transform my life into something that felt impossible.

Table of Contents

List of Figures

List of Tables

Table 3.3: Summary of rearrangements in the budgerigar genome using the chicken genome as a reference**.** The ancestral avian chromosome is represented in the left-hand column, with the subsequent columns indicating the number of inter- and intrachromosomal changes detected that have led to the evolution of the budgerigar. GGA = Gallus gallus domesticus.

Table 4.1: Summary of diploid number in the 7 avian species studied. [1]) Fillon *et al.,* 2007. [2]) O'Connor *et al.,* 2018b. [3]) Damas *et al.,* 2017. [4]) Shibusawa *et al.,* 2002. [5]) Hammar, 1970. [6]) dos Santos *et al.,* 2017.

Table 4.2: Percentage of successful BAC hybridisation in all 7 avian species tested.

Table 4.3: Results for chicken FLpter values, with all BACs corresponding to chicken chromosome 1. GGA = *Gallus gallus.* S.D = Standard deviation. n = number of chromosomes measured,

Table 4.4: Summary of inter- and intrachromosomal rearrangements.

Table 4.5: Unpaired T-tests for the Atlantic canary, with all BACs corresponding to chicken chromosome 1. GGA = *Gallus gallus.* S.D = Standard deviation. S.E.M = Standard error of mean.

Table 5.1: Chicken microchromosome BACs used for verification of microchromosome paints R1 to R9.

Table 5.2: Chicken macrochromosome paints for chromosomes 1-9 tested on the spiny softshell turtle (*Apalone spinifera*). The location in which the paints hybridised were determined, and how many signals were present.

Table 5.3: Chicken macrochromosome paints for chromosomes 1-9 tested on the yellow spotted river turtle (*Podocnemis unifilis*). The location in which the paints hybridised were determined (i.e. macrochromosome or microchromosome), and how many signals were present.

Table 5.4: The number of primary and secondary signals present in microchromosome paints R1 to R9. N = No signal present. Y = signal present (green). ++ = secondary signal present (yellow).

Table 5.5: Chicken microchromosome paints R1 to R9 tested on the spiny softshell turtle (*Apalone spinifera*). The location in which the paints hybridised were determined, and how many signals were present.

Table 5.6: Chicken microchromosome paints R1 to R9 tested on the yellow spotted river turtle (*Podocnemis unifilis*). The location in which the paints hybridised were determined, and how many signals were present.

Table 5.7: BAC clones successfully hybridised to the yellow spotted river turtle (*Podocnemis unifilis*, PUN). Orthologues of the chicken chromosomes were determined based on karyotype and chromosome size.

Table 5.8: Hybridisation success of chicken and zebra finch BACs on the yellow spotted river turtle (*Podocnemis unifilis*) and spiny softshell turtle (*Apalone spinifera*). N = No signal present. Y = signal present. N/A = not tested.

Table 6.1: BAC clones spanning genes newly found or with reassigned placements in the *Gallus_gallus*-5.0 genome assembly, and the corresponding chromosome position revealed by FISH testing.

Table 6.2: BAC clones spanning genes newly found or with reassigned placements, and the corresponding locations in the Gallus_gallus-4.0 and *Gallus_gallus*-5.0 genome assemblies.

Table 6.3: Purified PCR products for all 7 genes tested, with the PCR additive required for each primer pair.

Abbreviations

aCGH – Array Comparative Genomic Hybridisation
AKAP8L – A-Kinase Anchor Protein 8-like
APL – *Anas platyrhynchos*
AT – Adenine-Thymine
BAC – Bacterial Artificial Chromosome
BAZ2A – Bromodomain Adjacent to Zinc Finger Domain 2A
bp – Base Pair
CCD – Charge-coupled Device
CEFs – Chicken Embryonic Fibroblasts
CGH – Comparative Genomic Hybridisation
CLI – *Columba livia*
CNE – Conserved Non-coding Element
CNV – Copy Number Variation
CUN – *Chlamydotis undulata*
DAPI – 4',6-diamidino-2-phenylindole
DIPI – Diimidazolinophenylindole
ddH_2O – Double Distilled Water
DMSO – Dimethyl Sulphoxide
DNA – Deoxyribonucleic Acid
dNTP – Dideoxyribose Nucleotide Triphosphate
dNTP – Deoxyribose Nucleotide Triphosphate
DOP – Degenerate Oligonucleotide Primed
DSBs – Double-stranded DNA Breaks
DTT – Dithiothreitol
EBR – Evolutionary Breakpoint Region
E. coli – *Escherichia coli*
FISH – Fluorescence *in situ* Hybridisation
FITC – Fluorescein Isothiocyanate
FUS_1 –Fused in Sarcoma RNA-binding Protein
Gb – Gigabase
GC – Guanine-Cytosine
GGA – *Gallus gallus domesticus*
HBSS – Hank's Balanced Salt Solution

HSB – Homologous Synteny Block

HS – Hot Start

ISH – *in situ* Hybridisation

IKZF4 – IKAROS Family Zinc Finger 4

kbp – Kilobase Pairs

KCl – Potassium Chloride

KOD – *Thermococcus kodakarensis*

LB – Luria Bertani

Mb – Megabase

MBG – Molecular Biology Grade

$MgCl_2$ – Magnesium Chloride

$MgSO_4$ – Magnesium Sulphate

mya – Million Years Ago

NAHR – Non-allelic Homologous Recombination

NGS – Next Generation Sequencing

NHEJ – Non-homologous End Joining

NME – *Numida meleagris*

PacBio – Pacific Biosciences

PCF – Predicted Chromosome Fragment

PCR – Polymerase Chain Reaction

PBS – Phosphate Buffered Saline

RACA – Reference Assisted Chromosome Assembly

RVC – Royal Veterinary College, London

SNP – Single Nucleotide Polymorphism

SCD – *Serinus canaria*

SKIV2L – Ski2 like RNA helicase

SMARCC2 – SWI/SNF-related, Matrix Associated, Actin-dependent Regulator of Chromatin Subfamily C Member 2

SMRT – Single Molecule Real Time

SSC – Saline Sodium Citrate

SRU – *Scolopax rusticola*

TE – Transposable Element

TBE – Tris Borate EDTA

TGS – Third Generation Sequencing

TGU – *Taeniopygia guttata*

TME – *Turdus merula*

TYK2 – Tyrosine Kinase 2

UKC – University of Kent, Canterbury

VCFS – Velocardiofacial Syndrome

WGS – Whole Genome Sequencing

WGSS – Whole Genome Shotgun Sequencing

YAC – Yeast Artificial Chromosome

Abstract

Karyotyping is, in some ways, the original whole genome analysis that explores genome structure, with an emphasis on how chromosomal rearrangements result in disease and speciation. Whole genome sequencing is a more recent technology, however only the highest quality genome assemblies can convincingly link the karyotype to the genome sequence. The assembly of whole genomes to a so-called chromosome-level assembly, facilitated by fluorescence *in situ* hybridisation (FISH), provides an enhanced resolution to study chromosomal rearrangements to explore the diversity of chromosomal rearrangements within and between taxa. Studying genomes at this resolution allows us to trace chromosomal rearrangements between species, providing insights into phylogenomics, genome organisation, trait linkage, and genotype-phenotype associations

Chromosomal rearrangements fixed during evolution cause reproductive isolation and subsequent speciation, and thus the main purpose of this thesis was to identify chromosomal rearrangements between phylogenetically distant species. Despite millions of years of evolution, the genomes of avian and non-avian reptiles remain remarkably alike, with similarities in gross genomic structure and a high degree of synteny observed. The majority of avian and non-avian reptiles do not have chromosome-level assemblies, and for most of those that do, their *de novo* sequenced genomes remain heavily fragmented. This limits their use in comparative analyses and hinders our understanding of the chromosomal rearrangements that ultimately underpin genome evolution. Moreover, the very smallest of avian microchromosomes remain largely under-

characterised, both cytogenetically and genomically. With this in mind, this thesis has four specific aims:

The first was to upgrade the scaffold-based genome assembly of the budgerigar (a key species as a model for vocal learning and as a companion animal) to that of a chromosome-level using *in silico* and molecular cytogenetic techniques. This was successfully achieved and a number of interchromosomal rearrangements (rare in birds) were characterised. The second aim was to identify chromosomal rearrangements in 7 species that lack whole genome sequence data, but nonetheless their genome structure could be defined by mapping bacterial artificial chromosomes (BACs) to the macrochromosomes. This allowed for the identification of fissions, fusions, duplications, and inversions, all of which contribute to the chromosomal changes that influence speciation in birds. The third was to study genome conservation between avian and non-avian reptiles, exploiting evolutionary conserved sequences in BACs and chromosome paints to study the stability of genome organisation and the conservation of microchromosomes. Remarkable similarities between birds and turtles were identified and this shed light on the likely karyotypes of extinct dinosaurs. The final aim was to develop a panel of probes for the smallest of avian microchromosomes that would aid sequencing efforts in completing the chicken genome, furthering our understanding of avian karyotype evolution. This aim was partially successful, identifying two of the smallest microchromosomes. Taken together, these results have furthered the study of avian genomics and demonstrates that they evolved from their reptilian ancestors.

1 Introduction

1.1 Eukaryotic Chromosome Structure

The normal state of DNA and its associated histone proteins (chromatin) is a 30 nm fibre folded in an interphase nucleus. Chromatin exists in two states: heterochromatin and euchromatin, which differ as a result of epigenetic histone modification. Heterochromatin is a compact and often transcriptionally inactive structure, with DNA being more tightly condensed (Okada and Comings, 1974; Sullivan and Karpen, 2004). Euchromatin on the other hand is relatively de-condensed and facilitates the process of transcription, allowing DNA to be accessible for repair, replication, and gene expression (Grewal and Moazed, 2003). During the process of cell division, a radical restructuring occurs, and the chromatin forms supercoiled structures known as chromosomes (as seen in Figure 1.1). These tightly wound structures temporarily suspend transcription, while they facilitate the segregation of DNA into separate daughter cells (Fong, 1967; Raccaud and Suter, 2018).

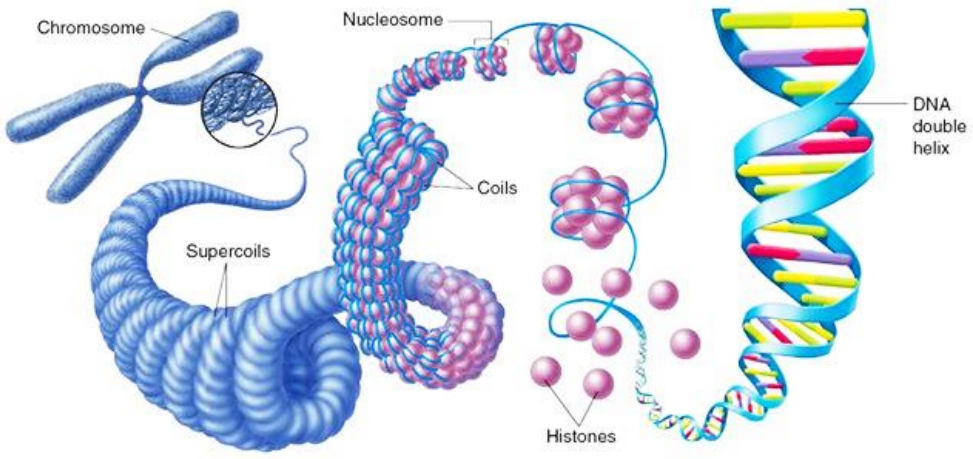


Figure 1.1: Condensation of DNA into mitotic chromosomes (© Pearson Prentice Hall).

The key morphological characteristics of any chromosome is the presence of constriction points (centromeres), arms, and sister chromatids. The length of the

chromosome arms on either side of the centromere determines how they are classified, with the shortest being referred to as the p-arm and the longest being referred to as the q-arm. The sister chromatids are genetically identical mirror images of one another, each segregating to one daughter cell. Moreover, the location of the centromere, and thus the length of the arms, allows for the classification of chromosome morphology into four key groups: metacentric, submetacentric, acrocentric, and telocentric (Levan, Fredga, and Sandberg, 1964), as illustrated in Figure 1.2. In metacentric chromosomes, the chromosome arms are typically equal in length, with the centromere positioned in the middle of the chromosome. With submetacentric chromosomes, the centromere is off-centre and exhibits chromosome arms that differ in length, with the p-arm being relatively shorter. Acrocentric chromosomes have the centromere located close to the end of the chromosome, with the p-arms significantly shorter in length but still visible. In telocentric chromosomes, the centromere located at the end of the chromosome, with only one pair of arms visible.

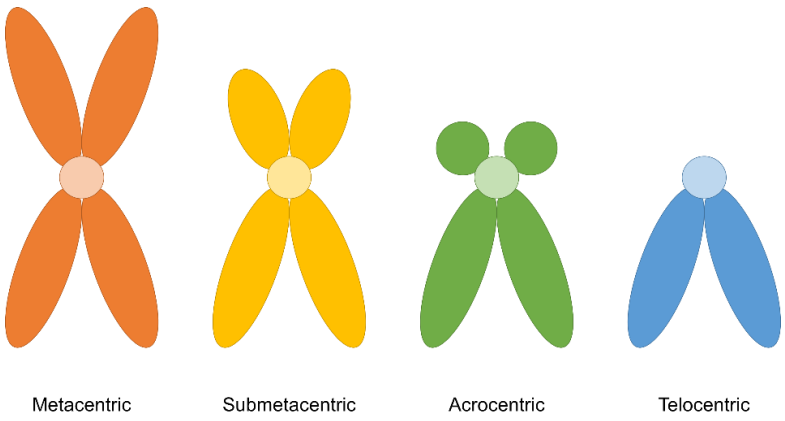


Figure 1.2: Classification of chromosome morphology based on centromere location, demonstrating metacentric, submetacentric, acrocentric, and telocentric morphologies.

1.2 Chromosomal Abnormalities

The integrity of the genome is dependent on the accurate replication and segregation of chromosomes during cell division. Chromosomal abnormalities arise when errors occur during mitosis, meiosis, and fertilisation, with these abnormalities responsible for disease traits or spontaneous abortion. Similar changes are also responsible for speciation, creating reproductive barriers as species diverge. A key aspect of this thesis is the examination of these chromosomal changes that lead to speciation, in which there are two broadly categorised groups of chromosomal abnormalities: numerical and structural.

1.2.1 Numerical Abnormalities

1.2.1.1 Aneuploidy

Aneuploidy occurs as a result of chromosome mis-segregation during meiosis or mitosis and can arise via non-disjunction or anaphase lag, resulting in the loss or gain of chromosomes (Fitzgerald, 1975; Pfau and Amon, 2012). Classical non-disjunction can occur during meiosis I or mitosis, whereby homologous chromosomes are unable to separate and both migrate to one spindle pole (and thus only one cell), leaving the other cell nullisomic for the chromosome. However, non-disjunction secondary to chromatid pre-division may also occur during meiosis II, whereby sister chromatids are unable to separate, forming one cell with an extra chromosome copy and the other with none (Hassold and Hunt, 2001). With anaphase lag, the failure of chromosomal binding to the spindle fibres or failure to migrate to the spindle pole results in a lagging chromosome. Consequently, this lagging chromosome is absent in one of the daughter cells, and can either become enveloped in micronuclei or become caught in the

cleavage furrow during telophase (Kato and Sandberg, 1968; Janssen *et al.,* 2011; Crasta *et al.,* 2012).

Mis-segregation events such as non-disjunction can occur from several complications with cellular division, such as weakened centromere cohesion, incorrect spindle fibre attachment and incorrect separation of chromosomes, failure in the spindle assembly checkpoints, and the presence of extra centrosomes (Minhas *et al.,* 2003; Ganem, Godinho, and Pellman, 2009; Chiang *et al.,* 2010; Janssen *et al.,* 2011). Most aneuploidies are incompatible with life, with only five known abnormalities known to result in live births in humans. For trisomies, these are chromosomes 13, 18, and 21, X and Y, causing disease phenotypes of Patau syndrome, Edwards syndrome, Down syndrome, triple X syndrome, Klinefelter Syndrome, and XYY syndrome respectively (Ji *et al*., 2015; Mandrioli *et al.,* 2016). For monosomies, only chromosome X monosomy is compatible with life, causing the disease phenotype of Turner syndrome (Tuke *et al.,* 2019). Unlike other chromosomal changes, aneuploidy is rarely a cause of speciation, however different sex chromosome systems (e.g. XX/XO; XX/XY; XX/XX) can be seen amongst similar, related species.

1.2.1.2 **Polyploidy**

Polyploidy involves the presence of an additional set or multiple sets of chromosomes in the genome, and typically arises from errors in fertilisation. Triploidy is the most common form of polyploidy in human embryos, with an additional set arising as a result of polyspermy or from parthenogenic oocytes (Brancati, Mingarelli, and Dallapiccola, 2003; Demyda-Peyrás *et al.,* 2015). In humans, polyploidy usually results in spontaneous abortion and survival beyond birth is rare, yet polyploidy is tolerated in plants, insects, fish, amphibians, and

reptiles (Otto and Whitton, 2000; Gregory and Mable, 2005). For example, polyploidy in plants is a common form of adaptation and speciation, with wheat being the most cited example; of the 31 wheat species (*Triticum* and *Aegilops*), 18 have polyploid genomes (Mirzaghaderi and Mason, 2017), with polyploidy shown to improve cultivars and tolerance to biotic/abiotic stress (Sattler, Carvalho, and Clarindo, 2016; Fang and Morrell, 2016).

Polyploidy in fish has also been well-documented, occurring frequently, spontaneously, and often in many groups. This polyploidy has been associated with phenotypic diversity and adaptation, such as electrogenic and electrosensory organs (Schultz, 1980; Moriyama and Koshiba-Takeuchi, 2018), but there is no definitive advantage due to the difficulties in studying and understanding complex fish genomes. Furthermore, there are groups of fish species where polyploidy remains an ongoing process, an example being the Rutilus alburnoides complex (*Cyprinidae, Leuciscinae*), in which multiple ploidy levels of diploid, triploid, and tetraploid animals exist (Leggatt and Iwama, 2003).

1.2.2 Structural Abnormalities

Structural abnormalities are generated by double-stranded DNA breaks (DSBs) that are aberrantly repaired when homologous recombination fails, and includes chromosomal deletions, duplications, fusions, insertions, inversions, and translocations (Figure 1.3). The erroneous repair of gross genomic rearrangements typically occurs when non-allelic homologous sequences recombine (Inoue and Lupski, 2002; Parks, Lawrence, and Raphael, 2015). Furthermore, there are more complex abnormalities that can occur, such as chromothripsis, in which there is exchange of genetic material between two or more chromosomes from a minimum of three chromosomal breakpoints

(Rosenberg et al., 2005). These structural abnormalities can be categorised as either balanced or unbalanced: In balanced abnormalities, the whole genome remains complete, with differences only in the orientation or location of genetic material. Occasionally balanced abnormalities may remain undetected if there is disease phenotype displayed. Whereas in unbalanced abnormalities, there is a loss or gain of genetic material and often disease phenotypes.

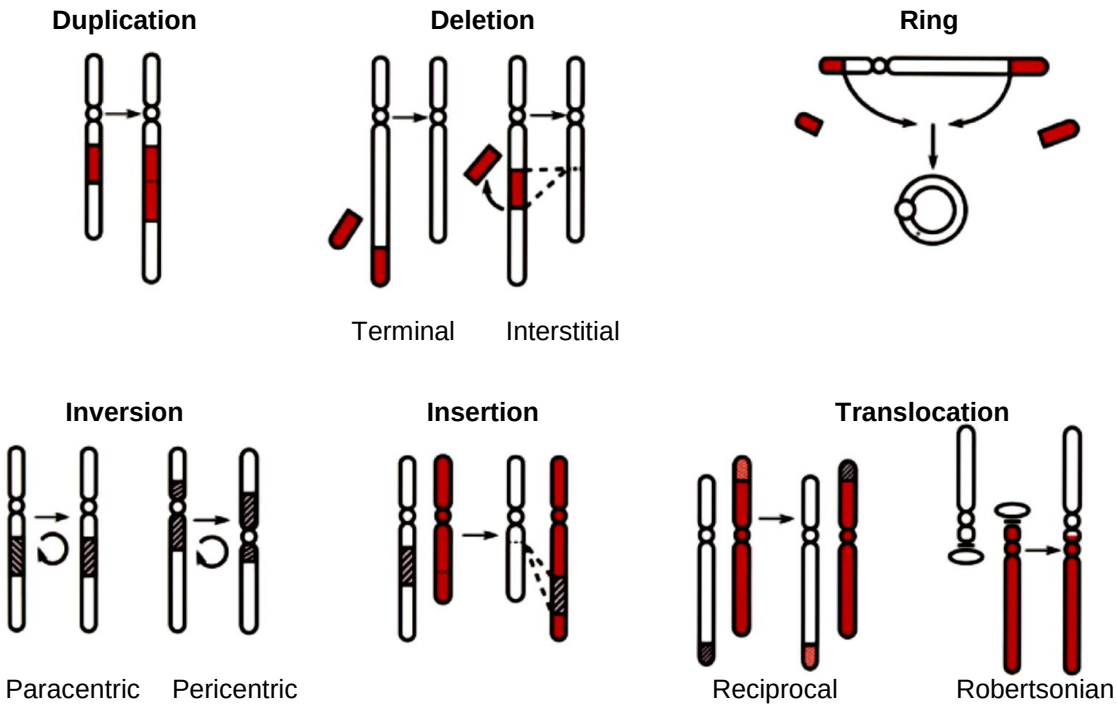


Figure 1.3: Common structural chromosomal abnormalities (Arsham, Barch, and Lawce, 2017).

In addition to the disease traits arising from structural abnormalities, changes in chromosome structure are also known to contribute to speciation for numerous reasons, which will be discussed further in the sections below. Nevertheless, variations in chromosome structure may reduce hybrid fitness as a result of mis-segregation during meiosis, and may suppress recombination. Subsequently,

these may reduce gene flow and introduce genomic incompatibilities, which are often associated with reproductive isolation and speciation.

1.2.2.1 Insertions

Insertions include chromosomal rearrangements in which there is a translocation of genomic DNA from one chromosome into a non-homologous chromosome or to non-adjacent locus on a homologous chromosome. Chromosomal insertions can have various phenotypic effects (or none at all) depending on the size and gene content of the chromosomal fragment, including other factors such as the orientation of the insertion, disruption of genes, effects on dosage-sensitive genes, and the site of insertion (Baptista *et al*., 2008). Numerous case studies have observed problems such as infertility, recurrent miscarriage, and multiple congenital anomalies in offspring due to unbalanced insertions (Doheny *et al.,* 1997; Xanthopoulou *et al.,* 2010).

The molecular mechanisms that underlie the presence of chromosomal insertions are not fully understood, with most studies elucidating to the mis-repair of DSBs (Croll, Zala, and McDonald, 2013; Kato *et al.,* 2017). However, insertions have been implicated as a cause of speciation and are particularly known to arise from the presence of transposable elements (TEs). TEs are commonly found in eukaryotic genomes, and are sequences of DNA that copy and subsequently insert themselves throughout the genome. The insertion of these TEs can be highly disruptive, interrupting regulatory elements or genes, or inducing ectopic homologous or non-homologous recombination (Deininger and Batzer, 1999; Raskina *et al.,* 2008).

1.2.2.2 Deletions and Duplications

Deletions are the loss of genomic DNA, and can range from small changes of 1 base pair (bp) to several megabases in length. As is the case with chromosomal insertions, the effects of chromosomal deletions can have vary depending on the size and gene content of the chromosomal fragment lost, the disruption and reduction of genes, and any effects on dosage-sensitive genes. There are also instances of complex insertions and deletions, typically in cancerous cells, whereby nucleotides are simultaneously inserted and deleted (Ye *et al.,* 2016).

Duplications refer to the presence of an additional copy of a chromosomal region, in which the genetic material can be an intrachromosomal tandem duplication, or located at different loci on homologous or non-homologous chromosomes. Duplications can be direct or inverted depending on whether the orientation of the duplicated genetic material remains the same or is inverted (with respect to centromere position). Irrespective of the orientation or loci within the genome, duplications in the absence of any other chromosomal imbalances will generate problems during chromosome segregation and could yield partial trisomies. Small-scale deletions arise from errors during DNA replication and repair processes, such as DNA slippage during replication with DNA polymerases (Manjari, Pata, and Banavali, 2014), whereas large-scale deletions arise from DSBs that are not repaired or from non-allelic homologous recombination (NAHR) during meiosis (Jelesko *et al.,* 2004).

The deleterious effects of deletions and duplications depend on the size of the regions involved. The presence of low-copy repeats has been known to facilitate NAHR during meiosis for both deletions and duplications, and thus the disease phenotypes associated with both deletions or duplications are similar (Lupski

and Stankiewicz, 2005). For example, deletions on human chromosome 22q11.2 results in DiGeorge syndrome and velocardiofacial syndrome (VCFS), where VCFS is a related chromosomal disorder considered to have a broader range of symptoms. Yet the clinical effect of duplications is dependent on whether the maternal or paternal chromosomes inherit the duplications. In Prader-Willi/Angelman syndrome, a phenotype that includes intellectual disability and autism, maternally inherited proximal 15q11-q13 duplications are involved. Inherited duplications on the paternal proximal 15q11-q13 typically exhibits no disease phenotype (Cook *et al.,* 1997). Although, females that inherit the paternal 15q duplication becomes carriers, with a 50% chance of transmitting the duplication to offspring.

Both deletions and duplications are known to contribute to speciation due to changes in gross genomic structure, sequence variation, and copy number variation. For example, the presence of segmental duplications has been associated with speciation in primates, with duplications increasing significantly in the ancestral branch leading to African great apes and humans (Marques-Bonet *et al.,* 2009). As for deletions, the sequence variation may interrupt coding sequences and splice sites, leading to the inactivation of genes, emergence of new genes/proteins, the suppression of recombination, or the promotion of NAHR (Shaw and Lupski, 2004; Yao and Schnable, 2005; Volfovsky *et al.,* 2009).

1.2.2.3 Fusions and Translocations

Chromosome fusion and translocation can occur in both normal and cancerous cells. Fusions involve the joining of two chromosomes (or more in the case of tandem fusions), and can fuse via their telomeres to produce a chromosome

with one active and one latent telomere, or via centric fusion (Robertsonian translocation). Chromosomal fusion is known to be a driver of karyotype evolution and speciation, as observed by the telomeric fusion (Robertsonian translocation) of two separate acrocentric chromosomes from the genomes of great apes to produce human chromosome 2 (Idjo *et al.,* 1991; Chiantante *et al.*, 2017).

Translocations may be balanced or unbalanced depending on whether the translocation is reciprocal or Robertsonian, and can exhibit both normal and disease phenotypes. Balanced translocations result in the equal exchange of genetic material between non-homologous chromosomes, with no gain or loss of DNA. On the other hand, unbalanced translocations, including Robertsonian translocations, result in the unequal exchange of genetic material between chromosomes, with some genetic loss. Chromosomal fusions and translocations typically cause problems associated with chromosome segregation, resulting in genomic instability and chromosomal abnormalities, but can also produce normal phenotypes in carriers if there has been no loss or gain of genetic material and if genes are not truncated (Robinson *et al.,* 1994; Baptista *et al.,* 2005).

1.2.2.3.1 Reciprocal Translocations

Reciprocal translocations occur when two non-homologous chromosomes exchange genetic material. In humans, most reciprocal translocations result in normal phenotypes, with approximately 6% of carriers exhibiting symptoms such as congenital abnormalities, intellectual disability, or autism (Fryns *et al.,* 1986; Hu *et al.,* 2016). However, male carriers with reciprocal translations on the X chromosome may experience spermatogenic arrest as a result of incomplete X

inactivation (Braekeleer and Dao, 1990). The cause of reciprocal translocations has been associated with non-homologous end joining (NHEJ), both inherited and *de novo* in origin, exposure to chemical and radiation, and AT-rich palindromic sequences (Weinstock *et al.,* 2006; Tucker, 2008).

The role of reciprocal translocations in local adaptation and speciation has been well established for several species, including fish and mammals (Ryu, Murooka, and Kaneko, 1998; Berggren *et al.,* 2016; Dobigny, Britton-Davidian, and Robinson, 2017). These translocations effect meiotic segregation (Figure 1.4) producing both balanced and unbalanced gametes. In turn, this reduces hybrid fertility and contributes to reproductive isolation, particularly if chromosomes contain large deletions or duplications, altering gross genomic structure and contributing to speciation as mentioned in section 1.2.2.2.

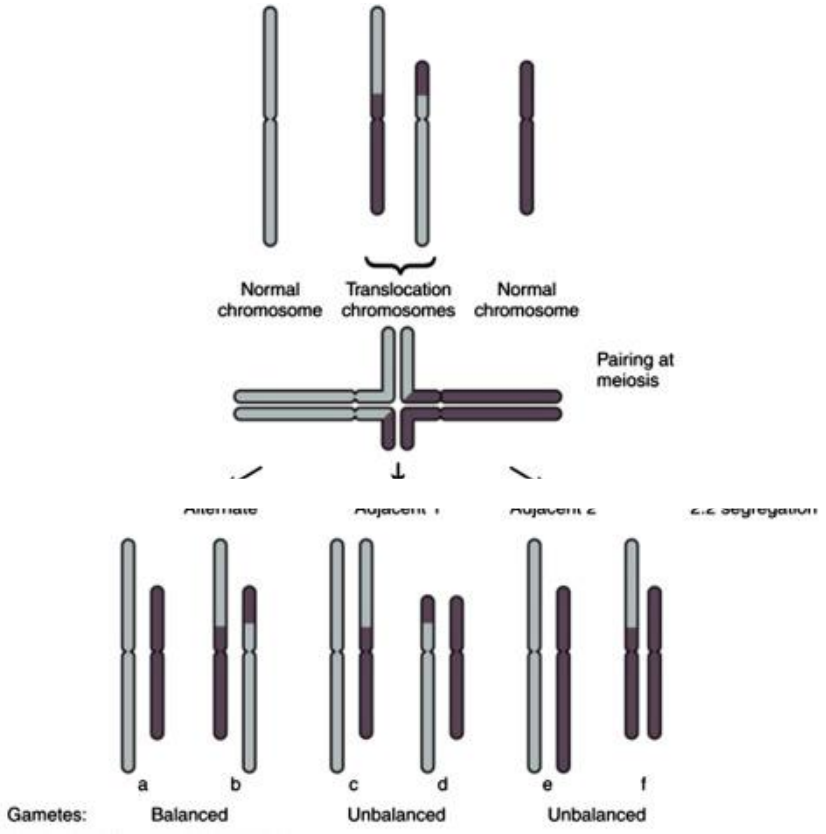


Figure 1.4: Meiotic segregation with the presence of a reciprocal translocation (modified image from Strauss *et al.,* 2009).

1.2.2.3.2 Robertsonian Translocations

Robertsonian translocations occur when two chromosomes, typically acrocentric, fuse at the centromere. The most common form of Robertsonian translocation is between non-homologous chromosomes but has also been observed between homologous chromosomes. The outcome of Robertsonian translocations is the production of a metacentric or submetacentric chromosome, and a small, gene poor, chromosome derivative which is subsequently lost in meiosis (Herschler and Fechheimer, 1966; Morin *et al.,* 2017). The loss of genetic material bears little impact on cellular function as the genetic material tends to lack essential genes, however issues arise at meiosis due to the formation of trivalents (Figure 1.5).

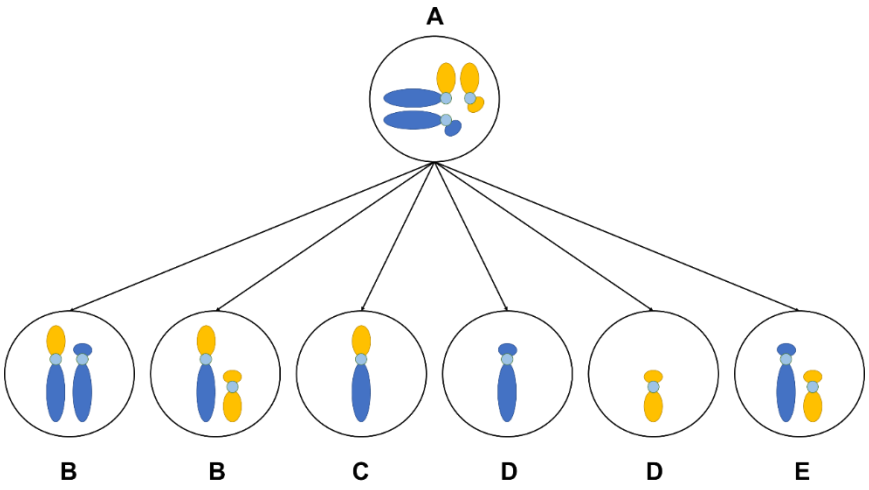


Figure 1.5: Generation of gametes after Robertsonian translocation. A) Trivalent formation at synapsis. B) Disomic gametes. C) Carrier gametes. D) Nullisomic gametes. E) Normal gametes.

The segregation of trivalents produces gametes that are disomic or nullisomic for one chromosome involved in the translocation, resulting in a monosomic or trisomic zygote. The pregnancy risks are dependent on the chromosomes involved and the parental origin of the translocation; translocations with a

maternal origin show a considerably increased chance of testing positive at a second trimester screen than translocations of paternal origin (Boue and Gallano, 1984).

Robertsonian translocations are also known to be a driver in karyotype evolution and speciation. For example, the Indian muntjac genome exhibits the fusion of multiple ancestral acrocentric chromosomes (Hartmann and Scherthan, 2004). Associations between karyotype evolution and speciation has also been demonstrated in spiny lizards (genus *Sceloporus*), in which phylogenomic analyses correlated extensive Robertsonian fusions with higher speciation rates (Leaché *et al.,* 2016).

1.2.2.4 Inversions

Chromosomal inversions involve the presence of two DSBs within the same chromosome, followed by the 180-degree rotation of the segment and subsequent repair of the DSBs. These inversions can be inherited or *de novo,* and can be classified as pericentric or paracentric, whereby pericentric inversions involve the centromere within the inverted segment and paracentric inversions do not. Inversions result in a disruption to the gene order with respect to the rest of the chromosome, yet most inversions have no/minimal phenotypic abnormalities as there is no gain or loss of genetic material.

However, if there are disruptions to regulatory elements and reading frames, this may ultimately impact gene expression and the production of fully functional proteins (Kok *et al.,* 1995; Phippard *et al.,* 2000). Furthermore, carriers may face issues with fertility depending on the size and type of inversion. Individuals that are homozygous for an inversion will not exhibit a reduction in chromosome

recombination or a reduction in fertility as inverted chromosomes are capable of recombining and segregating normally during meiosis. On the other hand, heterozygous individuals exhibit a reduction in chromosome recombination and errors in chromosome segregation, resulting in the formation of inversion loops (Figure 1.6). These inversion loops result in reduced fertility as some gametes will be inviable.

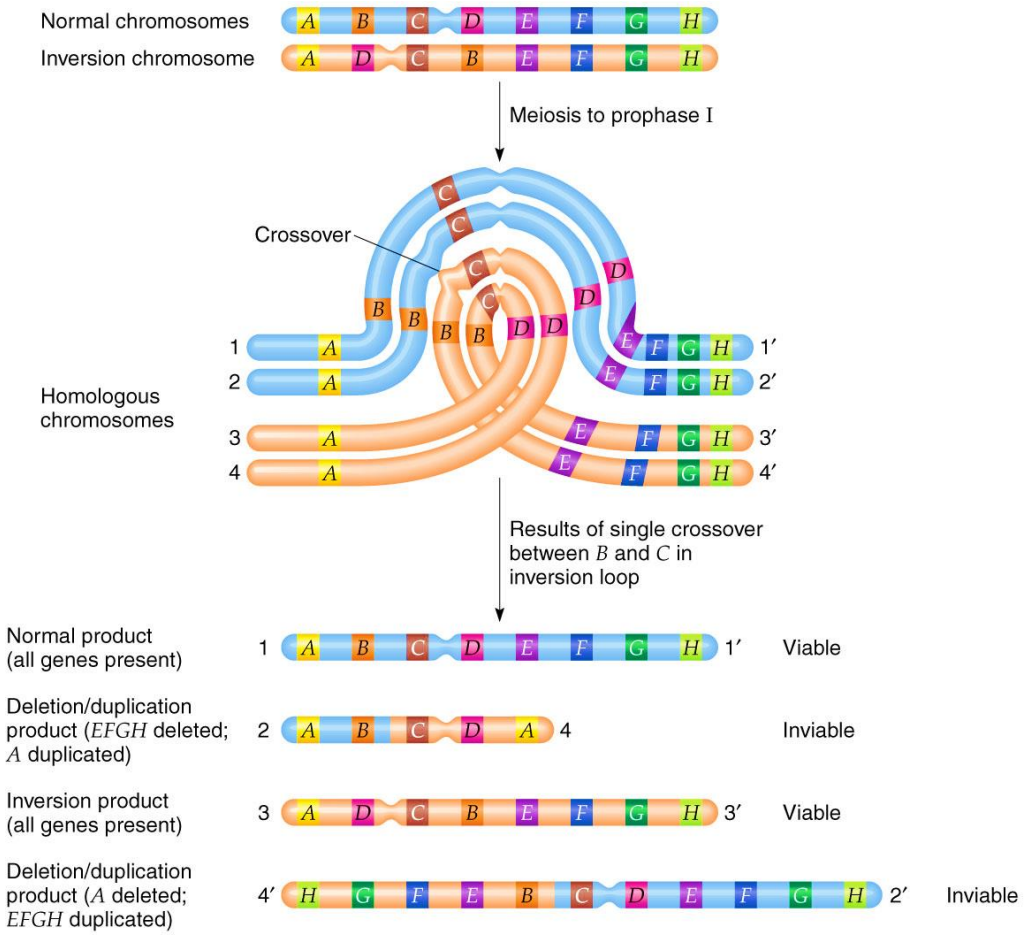


Figure 1.6: Inversion loop formed during meiosis in heterozygous individuals (© 2010 Pearson Education).

Inversions are also known to be a key factor in speciation and reproductive isolation, causing a reduction in recombination across regions of the chromosome, which can be significant if there are beneficial/essential genes

located within the inversion. Co-adapted alleles that contribute to reproductive isolation may be fixed within the inversion, reducing gene flow and promoting speciation in the early stages of divergence (Bush *et al.,* 1977; Lohse *et al.,* 2015). Moreover, the presence these co-adapted alleles that are involved in isolating or adaptation barriers have been shown to correlate with phenotypic evolution (Fishman *et al.,* 2013).

1.2.2.5 Isochromosomes and Ring Chromosomes

Ring chromosomes and isochromosomes are considered unusual relative to other structural abnormalities. Isochromosomes are considered to be whole-arm translocations, whereby both chromosome arms are identical on either side of the centromere, essentially creating a mirror image. These chromosomes are commonly associated with tumours and aneuploidy (Santana, Gardner, and Neu, 1977; Berend *et al.,* 2000; Jin *et al.,* 2000), and phenotypic abnormalities tend to arise due to dose variation.

Ring chromosomes form when two DSBs occur in both chromosomal arms, and the subsequent repair fuses the two ends of the chromosome to produce a ring. This can involve one or multiple chromosomes, and thus the ring structure may have one or more centromeres. The formation of ring chromosomes is typically *de novo* in origin but can be inherited, and result in phenotypic abnormalities specific to the chromosomes involved in the ring (Kosztolányi, Méhes, and Hook, 1991).

1.3 Classical Cytogenetics

In eukaryotic cells, the genome is distributed across multiple chromosomes, with the structure of each chromosome defined by the unique pattern of supercoiling. The formation of supercoils is sequence dependent (Kim *et al.,* 2018), and consequently, this pattern typically remains identical in every cell division and between individuals of the same species. Subsequently, this unique pattern of supercoiling determines the gross genomic structure for each species, which is known as a karyotype.

Classical cytogenetics provides an inexpensive way to observe gross genomic structure and diagnose genetic diseases through the studying karyotypes. Initial observations of karyotypes can determine chromosome number, size, and morphology, but can remain ineffective when attempting to correlate chromosome morphology to phenotypic or disease traits. Therefore, the invention of a staining techniques (examples shown in Table 1.1) enhances the contrast between different chromosomal components and produces characteristic bands as a comprehensive method for analysing and/or characterising chromosomes. The bands produced by stains can be classified into two groups: those that generate bands along the whole length of the chromosome, and those that generate a limited number of bands and stain specific chromosomal structures.

Staining Technique	Stain	Banding Pattern
C-Banding	Giemsa	Non-coding constitutive heterochromatin (active and latent centromeres)
Cd-Banding	Giemsa	Positive: active centromeres Negative: latent centromeres
G-Banding	Giemsa; Wright's	AT-rich regions = dark bands GC-rich regions = light bands
NOR-Banding	Silver nitrate; Chromomycin A3; Mythramycin; DAPI (4',6'-diamidino-2-phenylindole)	Active nucleolar organising regions (5.8S, 18S and 28S ribosomal RNA)
Q-Banding	Quinacrine; Hoescht 33258; DAPI (4',6'-diamidino-2-phenylindole); DIPI (diimidazolinophenylindole)	AT-rich regions = bright fluorescence GC-rich regions = weak fluorescence
R-Banding	Giemsa; Acridine orange; Hoescht 33258; Methyl green; Chromomycin A3/distamycin A	AT-rich regions = light bands GC-rich regions = dark bands
T-Banding	Giemsa	Telomeres
DA-DAPI	Distamycin A/DAPI (4',6'-diamidino-2-phenylindole); Distamycin A/Hoechst	AT-rich regions = dark bands GC-rich regions = light bands

Table 1.1: An overview of commonly used stains for each chromosome staining technique, and a summary of the banding patterns produced.

Chromosomes are typically treated via enzymatic digestion or heat denaturation and then stained, exhibiting the presence of light and dark bands when visualised. This digestion/degradation of chromosomal proteins and subsequent decondensing of the chromatin structure allows the dyes to more readily access the DNA. With these techniques, amplifications of chromosome segments, deletions, duplications, insertions, inversions, fragile sites, primary constrictions, satellites, stalks, and translocations are readily recognisable (Lubs *et al.,* 1973; Moore and Best, 2001; Huang and Chen, 2017). Chromosome staining techniques, such as G-, Q-, and R-banding, are used for karyotype preparation since each species produces a unique pattern of dark and light bands for each chromosome (illustrated in Figure 1.7). The production of consistent landmarks is useful for clinical diagnostics and identifying any polymorphisms within a population.

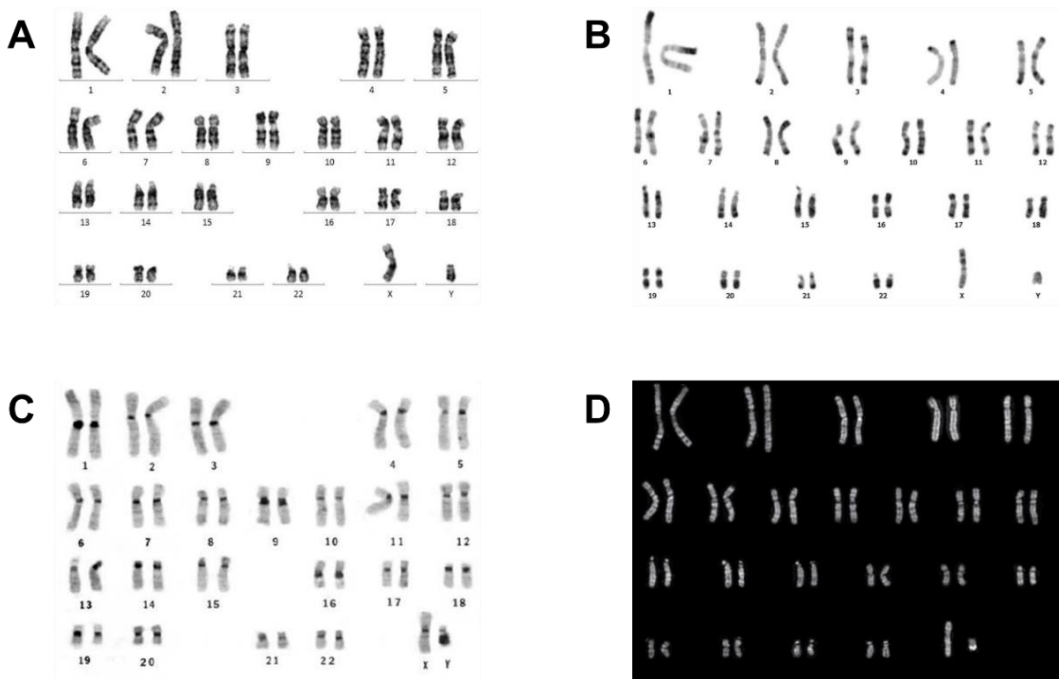


Figure 1.7: Chromosome banding by commonly used staining techniques. A) Human G-banded karyotype (Huang and Chen, 2017). B) Human R-banded karyotype (Huang and Chen, 2017). C) Human C-banded karyotype (Di Tomaso *et al.,* 2013). D) Human Q-banded karyotype (Arsham, Barch, and Lawce, 2017).

Furthermore, studying chromosomal banding patterns can elucidate evolutionary relationships by revealing changes in gross genomic structure. Consequently, genotype to phenotype associations can be determined, which may also be a contributing factor to speciation. However, the resolution to which these associations can be made is relatively restricted due to limitations in the number of bands produced on chromosomes. For example, gross genomic changes should be easily identifiable by a trained cytogeneticist, such as large insertions/deletions and extra chromosomes. A skilled cytogeneticist should be able to detect a deletion of 5-10 Mb (Huang and Chen, 2017), but it becomes increasingly difficult to detect smaller abnormalities, and correlating these genotypes to phenotypes may be exponentially harder.

There are essentially two reasons to analyse chromosomes: The first is to detect deviations from the norm (chromosome abnormalities), which can be in live-born individuals (including new-borns), prenatal diagnoses (including spontaneous abortions) or preimplantation embryos. The second is to map genes, which can be within species or between species (comparative genomics). In this context, a karyotype can be considered a low-resolution genome map of any particular species. The mechanism through which chromosome changes occur have parallels, both when studying disease-related chromosome abnormalities and changes that impact on speciation. Most of these investigations came about as a result of banding studies.

1.4 Molecular Cytogenetics with Fluorescence *in situ* Hybridisation

The ability to band chromosomes laid the foundation for the basis of all cytogenetics. A revolution occurred however with the advent of molecular cytogenetics. Suddenly it became possible to light up individual chromosomes, to map genes to chromosomes, to attain a far better resolution for clinical cytogenetics, and to trace evolutionary relationships more accurately.

All of this was made possible through the isolation and labelling of DNA probes. The invention of *in situ* hybridisation (ISH) relied on hybridising these labelled nucleic acid probes to their complementary sequences, either in the form of chromosome paints or individual locus probes. In hybridising the labelled DNA or RNA to its complementary sequence, ISH allows for the localisation of specific nucleic acid sequences. The detection of these nucleic acid sequences via ISH was first described in 1969, in which radioactive isotopes were used for labelling ribosomal RNA in *Xenopus* oocytes (Gall and Pardue, 1969). The progression to the use of non-radioactive variants for labelling occurred in 1981, whereby fluorophore-tagged streptavidin was used to detect biotin (Langer, Waldrop, and Ward, 1981). Additionally, the use of labelling variants with differing excitation and emission spectra allows for the simultaneous detection of one or more probes, all of which require the denaturation and the subsequently reannealing of both the labelled probe and complementary sequence (as seen in Figure 1.8). Once hybridised, the unbound probe is removed via washes and the hybridised probe can be visualised microscopically. It is necessary for the labelled probes to be digested into fragments smaller than 500 bp, with an average 150-250 bp, to facilitate entry into the nucleus of fixed cells (Lichter *et al.,* 1988), and

unlabelled repetitive sequences are essential to counteract repetitive sequences found within the genome.

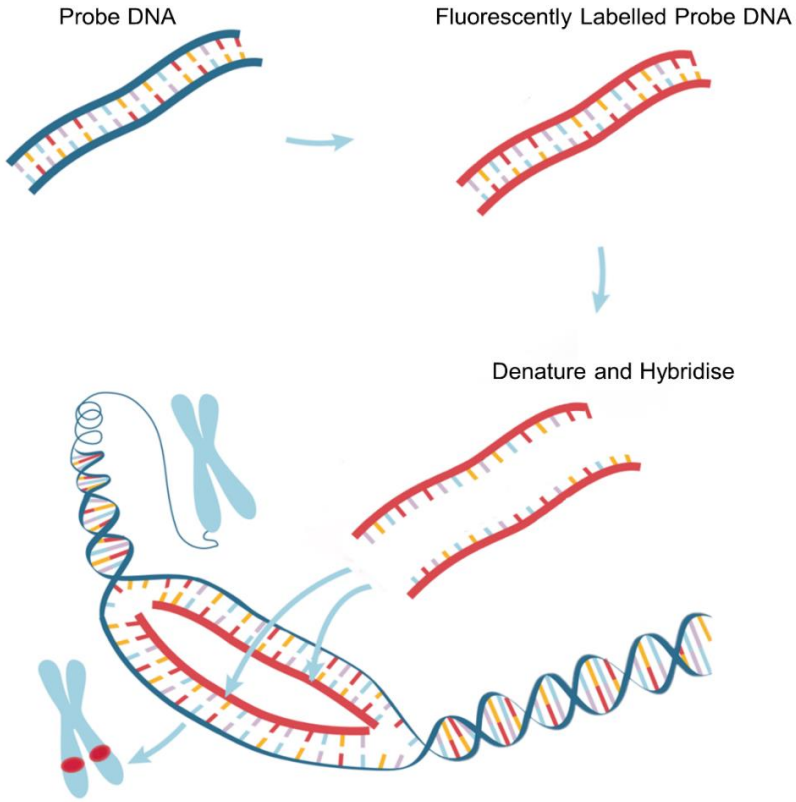


Figure 1.8: An overview of *in situ* hybridisation (ISH) using sequence-specific DNA probes (modified from Abnova promotional material).

To date, fluorescence *in situ* hybridisation (FISH) is the most commonly used form of ISH as fluorophores provide superior detection and stability, and delivers fewer safety risks (Speicher and Carter, 2005). The sensitivity and versality of FISH were recognised during the Human Genome Project, in which it supported the mapping and sequencing efforts by providing physical maps. Consequently, there is now a diverse range of FISH-based applications, including diagnostic assays for fields such as cancer, prenatal diagnosis, reproductive medicine, neuroscience, evolutionary biology, and comparative genomics.

1.4.1 FISH Using Chromosome Paints

Chromosome painting is the process in which labelled chromosome-specific sequences are hybridised to individual metaphase chromosomes or interphase cells. Chromosome paints were initially generated by obtaining and combining plasmid clones from DNA libraries and were then labelled via nick translation (Pinkel, Straume, and Gray, 1986). This labour-intensive process was soon simplified by the generation of paints via degenerate oligonucleotide primed (DOP) polymerase chain reaction (PCR) using flow-sorted or micro-dissected chromosomes as template DNA (Carter *et al.,* 1992; Meltzer *et al.,* 1992). The application of chromosome paints is useful for phylogenetic and comparative genomic studies, in which homologous chromosomes or homologous blocks can be readily identified in cross-species FISH experiments, and large interchromosomal rearrangements can also be detected. Moreover, chromosome paints can also be used to determine chromosomal abnormalities in metaphase chromosomes within a population.

Despite the frequent use of chromosome painting for clinical diagnoses and comparative studies, the resolution of chromosome painting is limited due to the nature of the paint and the method in which they are generated. Chromosome painting results in the fluorescence of large chromosomal regions or entire chromosomes. Therefore, intrachromosomal rearrangements such as inversions, deletions, and duplications cannot be identified, and chromosome paints are incapable of defining what is encoded within homologous regions or why they remain conserved. Moreover, cryptic rearrangements such as insertions or translocations can often be missed (Lee *et al.,* 2001). In addition to the limited resolution, using flow-sorted or micro-dissected DNA can also be

problematic as there may be co-amplification of DNA from non-homologous chromosomes, which is further restrictive for comparative chromosome painting studies.

1.4.1.1 Comparative Chromosome Painting in Mammals

Using human chromosome paints, the Japanese macaque (*Macaca fuscata*) was the first species in which comparative chromosome painting was used to establish chromosome homology for its entire karyotype (Wienberg *et al.,* 1992). The application of chromosome paints for cross-species use provides the means to visualise chromosome homology across multiple mammalian orders (Scherthan *et al.,* 1994; Wienberg *et al.,* 2000), whilst allowing for the elucidation of phylogenetic relationships and the identification of interchromosomal rearrangements (Dumas and Mazzoleni, 2017).

Chromosome paints derived from the human genome are the most commonly used due to the high-quality genome assembly, in addition to the highly conserved syntenic chromosome organisation that is similar to the ancestral organisation of all *Placentalia* (Graphodatsky, Trifonov, and Stanyon, 2011). Consequently, comparative chromosome painting using human chromosome paints has been widely applied to several mammalian species (as demonstrated in Figure 1.9) and has also been extensively applied to primate species. Nevertheless, chromosome paints derived from other mammalian species are available, particularly those from derived from domestic species, such as cattle, horse, pig, and sheep (Rubes *et al.,* 2009). To date, species from almost all mammalian Orders have chromosome painting data associated with their genomes, which has led to large-scale studies on modes of chromosome

evolution and identifying ancestral syntenies (for example Ferguson-Smith *et al.,* 2005; Romanenko *et al.,* 2007; Martinez *et al.,* 2017).

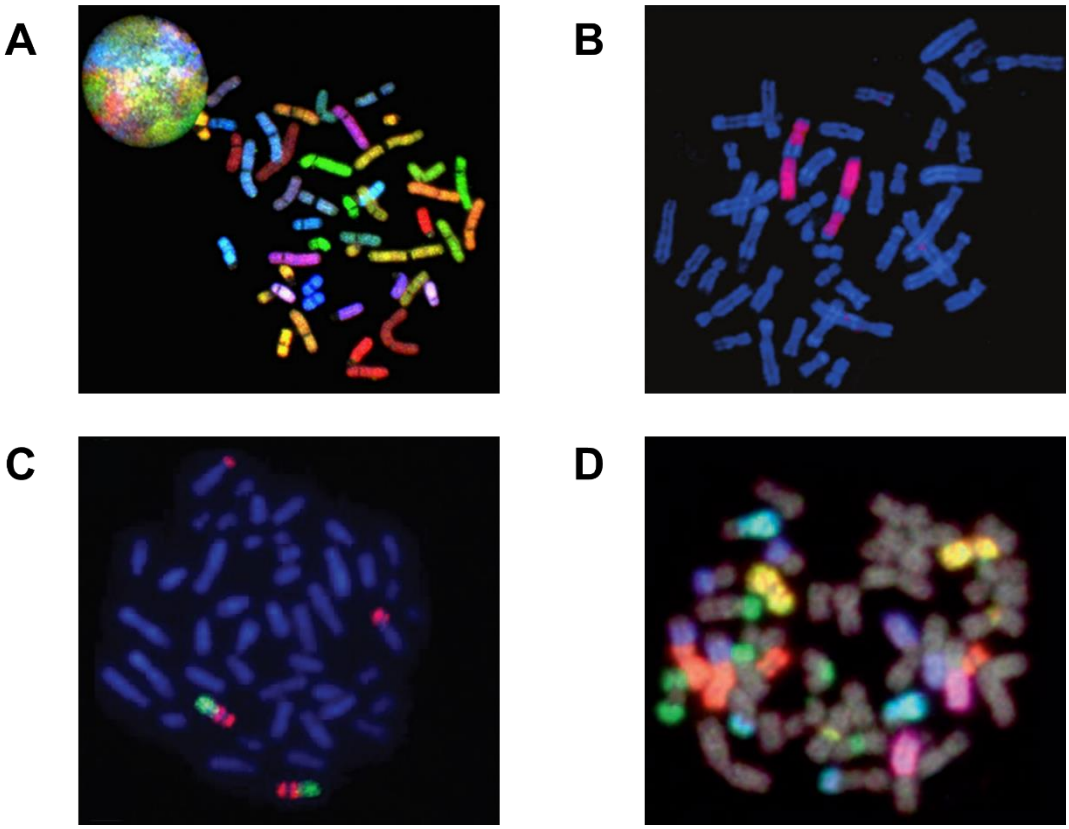


Figure 1.9: Chromosome painting in mammals using human chromosome paints A) Human chromosomes visualised via spectral karyotyping (Reid, 2015). B) Human chromosome paint 1 applied to Sei whale (*Balaenoptera borealis*) metaphase chromosomes (Murphy *et al.,* 2003). C) Human chromosome paints 14 (green) and 15 (red) applied to golden-backed uakari (*Cacajao melanocephalus)* metaphase chromosomes (Gifalli-Iughetti and Koiffmann, 2009). D) Human chromosome paints applied to lar gibbon (*Hylobates lar)* to metaphase chromosomes (Wienberg, 2005).

1.4.1.2 Comparative Chromosome Painting in Birds

The generation of chromosome paints provides a comprehensive and rapid method for tracing chromosomal evolution in avian species, and have been applied to approximately 80 avian species to date (Kretschmer, Ferguson-Smith, and De Oliveira, 2018). There are four commonly used sets of chromosome paints generated from the chicken (*Gallus gallus*), stone curlew (*Burhinus oedicnemus*), white hawk (*Leucopternis albicollis*), and griffon vulture (*Gyps*

fulvus), all of which are applied with aim of furthering our understanding of avian genome organisation (Griffin *et al.,* 1999; 2007; Nie *et al.,* 2009; De Oliveira *et al.,* 2010; Nie *et al.,* 2015).

Of all the chromosome paints, the most commonly used in literature are those generated from the chicken (an example seen in Figure 1.10), which include paints for both the macrochromosomes and microchromosomes. Whilst the macrochromosome paints have demonstrated incredible success across multiple avian species and have provided valuable insights into the avian genome conservation (Shetty, Griffin, and Graves, 1999; Griffin *et al.,* 1999; 2007), the microchromosomal paints have had limited success. This is in part due to the pooling of microchromosomes, rather than paints being assigned to individual chromosomes (Lithgow *et al.,* 2014), and thus most comparative painting data refers only to macrochromosomes. The use of chromosome paints to study karyotypic variation and chromosome evolution in birds is picked up further (in the context of avian genomics) in section 1.7.1.2.1.

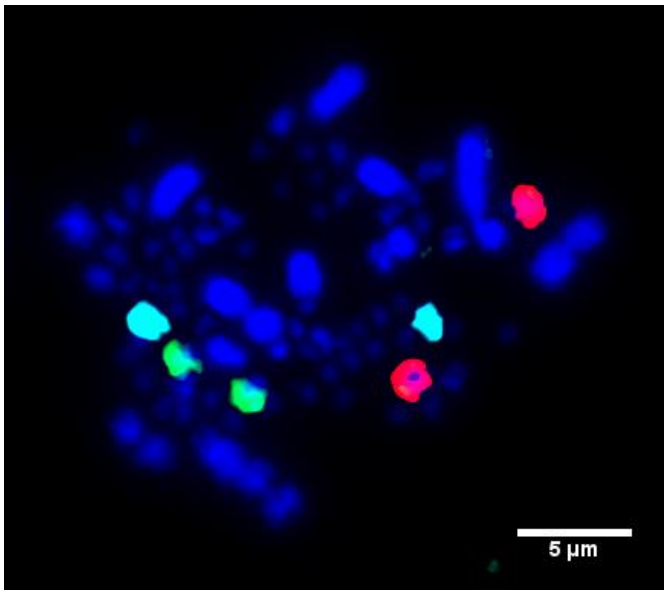


Figure 1.10: Chicken chromosome paints hybridised to Japanese quail (*Coturnix japonica*) chromosomes 6 (green), 7 (red), and 9 (aqua) (prepared personally).

As previously mentioned, the biggest limitation of chromosome paints is the resolution. For species that are highly diverged or with poor sample preparations, suboptimal binding can result in ambiguous results, particularly for microchromosome pools. Given that analysing FISH images is subjective and criteria for defining signals varies between users, there is the possibility that cryptic rearrangements could be overlooked or non-specific binding is classed as positive signals.

1.4.1.3 Comparative Chromosome Painting in Non-avian Reptiles

The Class Reptilia consists of over 10,600 species, exhibiting the same diversity as the Class Aves, yet non-avian reptile chromosomes have not been as thoroughly studied in comparison to avian chromosomes (Deakin and Ezaz, 2019). Nevertheless, comparative chromosome painting in non-avian reptiles demonstrates a varied degree of success, and as is the case with birds,

comparative chromosome painting has largely been limited to the macrochromosomes (Kichigin *et al.,* 2016; Deakin and Ezaz, 2019).

Comparative chromosome painting in non-avian reptiles typically relies on paints generated from the chicken, with multiple studies demonstrating the conservation of genomic sequences between avian and non-avian reptiles despite 275 million years of independent evolution. Many studies have focused on the high degree of conservation of the sex chromosomes (Pokorná *et al.,* 2011), but extensive homology is also demonstrated between autosomes (Pokorná *et al.,* 2012; O'Connor *et al.,* 2018c). Other studies have also developed paints from the common sandfish (*Scincus scincus*), Japanese gecko (*Gekko japonicus*), Chinese soft-shelled turtle (*Pelodiscus sinensis*), and the Caucusus emerald lizard (*Lacerta strigata*) for studying genome diversity within/between Orders and Families (Giovannotti *et al.,* 2009; Trifonov *et al.,* 2011; Kawagoshi *et al.,* 2014; Lisachov *et al.,* 2019).

1.4.2 FISH Using Individual locus Probes

FISH probes can be characterised as repetitive sequence probes or locus-specific probes (an example seen in Figure 1.11), and differ to chromosome paints in that they are relatively smaller in size and provide a more targeted approach to identifying nucleic acid sequences. Repetitive sequence probes consist of short monomers that hybridise to chromosomal regions consisting of tandem repeats, such as telomeric, centromeric, and ribosomal DNA. Locus-specific probes are often used as positional markers along the length of the chromosome, allowing for the orientation to be determined for each probe and detecting intrachromosomal rearrangements.

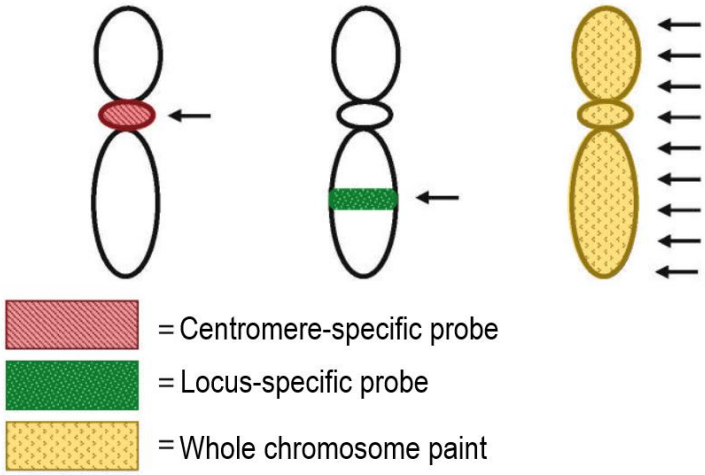


Figure 1.11: Example of differences between chromosome probes (modified from Bridge, 2008).

Locus-specific probes are much larger in size, ranging from 10 kilobase pairs (kbp) for plasmid vectors to 350 kbp for genomic clones (Bishop, 2010), and the generation of clone libraries has been vital in comparative and functional genomic studies. A single genomic construct can span numerous genes/loci and is capable of detecting smaller intrachromosomal rearrangements that would remain undetected by chromosome paints. Of the genomic clones, bacterial artificial chromosome (BAC) vectors are preferentially used over yeast artificial chromosomes (YAC) for generating probes as there is increased stability of insert propagation over numerous generations (Hall *et al.,* 2012)

The creation of probes has made it possible to overcome the limitations imposed by chromosome paints, allowing for the finer mapping of chromosomal rearrangements, such as inversions, translocations, deletions, and duplications, in addition to the identification of specific loci. Yet, as is the case with chromosome paints, there are limitations in using FISH probes. This includes suboptimal binding when using locus-specific probes for cross-species experiments, as the degree of sequence divergence will affect the hybridisation

of the probe. Subsequently, this suboptimal binding may result in background fluorescence and autofluorescence, and may generate ambiguous results. Nonetheless, the advantages of FISH probes are significant for physical mapping studies (see section 1.5.6) and are informative in both metaphase chromosomes and interphase nuclei.

1.4.2.1 Comparative BAC Mapping

The limited resolution of chromosome paints, and particularly in the case of microchromosomes, spurred the development for more refined cytogenetic tools that provided an improved resolution of the genome. Comparative BAC mapping utilises a similar approach to that of chromosome paints, but allows for the detection of a wider range chromosomal rearrangements, including both inter- and intrachromosomal and cryptic rearrangements. The use of BACs can identify specific homologous regions of the genome, providing insights into genome evolution and linkage relationships. Furthermore, BAC libraries typically have several-fold coverage of the genome (Ariyadasa and Stein, 2012), which is important for the assembly of genomes and improving the order and orientation of genetic and physical maps. However, the generation of BAC libraries is heavily dependent on the availability of genomic sequences and a well-defined clone library, and comparative BAC mapping suffers from the reduced probability of sufficient sequence homology for successful binding. Consequently, comparative BAC mapping had been relatively underrepresented in the literature due to lower success rates compared to chromosome paints.

A key example of the resolution provided by comparative BAC mapping is demonstrated in the studies of muntjac genomes. The Chinese muntjac (*Muntiacus reevesi*) and the Indian muntjac (*Muntiacus muntjak vaginalis*) are

phenotypically similar but differ substantially in diploid number, with 2*n*= 46 and 2*n*= 6/7 respectively, despite a short evolutionary divergence of approximately 4.7 million years. Regardless of these differences in diploid number, the two muntjac species are capable of interbreeding to produce viable offspring, albeit sterile (Tsipouri *et al.,* 2008). Comparative BAC mapping of the Indian muntjac genome determined that the reduction in diploid number arose exclusively through multiple centromere-telomere tandem fusions, with the orientation of each fused chromosome determined (Chi *et al.,* 2005).

1.4.2.2 Exploiting Evolutionary Conserved Sequences

Comparative BAC mapping has been more successful in birds, with the development of a universal BAC probe set by Damas *et al.* (2017) demonstrating the ability to map to any avian genome with high efficiency (>90%). This was pivotal in highlighting the significance of BAC mapping for comparative genomics and genome assembly efforts, and revived the use and reporting of BACs in avian species. The subsequent application of these BACs to numerous avian species allowed for comparisons of genomic structure, revealing the high degree of conserved synteny and gene organisation through detectable changes in BAC order/location (Griffin *et al.*, 2008). The use of conserved BACs to study karyotypic variation and chromosome evolution in birds is largely under-explored however, and picked up further in chapters 3 and 4 of this thesis.

The application of comparative BAC mapping in non-avian reptiles has been relatively limited in comparison to birds and mammals for numerous reasons, such as the difficulties in obtaining metaphase cell preparations, the lack of sequence data available for preparation of BAC libraries, and the lack of representative genomes for comparison. Nevertheless, the use of avian BACs

in non-avian reptiles for comparative mapping has been possible as the BACs are able to effectively hybridise across long evolutionary distances. This is due to the high degree of genome conservation in and between avian and non-avian reptiles, especially as both macrochromosomes and microchromosomes are truly homologous to that of the chicken (O'Connor *et al.,* 2018c). Therefore, comparative mapping of non-avian reptiles using avian BACs offers insight into the evolutionary history of reptilian chromosomes. One of the chapters in this thesis (chapter 5) deals with the use of BACs on reptilian chromosomes, a hitherto under-explored area.

1.4.3 Comparative Genomic Hybridisation

Comparative genomic hybridisation (CGH) is a technique in which chromosomal losses or gains are detected through comparisons between a reference genome and that of the subject genome being studied (Kallioniemi *et al.,* 1992). The DNA samples of the reference and subject DNA are labelled using two different fluorophores (typically red or green), mixed together, and then hybridised to metaphase chromosome preparations. The simultaneous hybridisation of both DNA samples results in competitive hybridisation at each locus of origin, creating differences in signal intensity between the red and green fluorophores. This ratio difference can be measured along the length of chromosomes, with a ratio value less than 1 on a linear scale representing a loss of DNA (deletion) and a value greater than 1 representing a gain (duplication). Whilst CGH provides a high resolution comparison between genomes, it is only capable of detecting unbalanced rearrangements and is unable to detect rearrangements such as inversions, translocations or mosaicism (Weiss *et al*., 1999). Additionally, the resolution to detect rearrangements is limited by the metaphase

chromosomes, and for most clinical applications has been limited to 5-10 Mbp (Theisen, 2008).

As with CGH, array CGH (aCGH) utilises competitive hybridisation between DNA samples from a reference and subject genome, but utilises microchips that contain numerous DNA probes of known sequence and location in the genome (Figure 1.12). In using a microchip, the evolution of array CGH (aCGH) has mostly overcome the limitations of resolution with CGH by removing the requirement of metaphase chromosomes and instead relying on oligonucleotides, BACs, or plasmids. In doing so, aCGH can simultaneously detect thousands of genetic loci and provides a higher resolution of 100 kbp (Ahn *et al.,* 2015). Yet, as with CGH, aCGH is still unable to detect balanced rearrangements, but can detect mosaicism when combined with single nucleotide polymorphism (SNP) data.

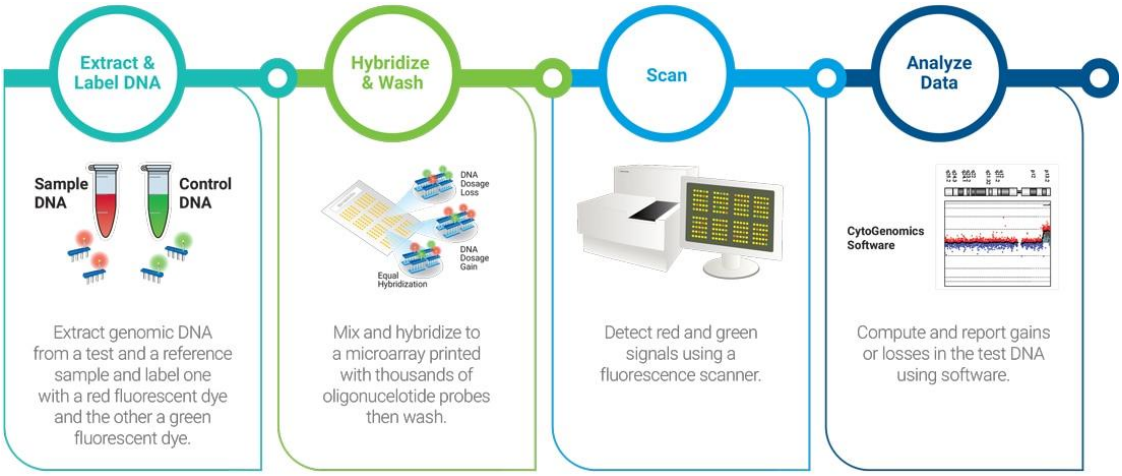


Figure 1.12: Overview of array-CGH technology (Agilent promotional material).

Nevertheless, CGH/aCGH studies allow for the detection of chromosomal imbalances without the need to culture cells. This was particularly useful for the

studies of tumour suppressor genes and oncogenes, in which some tumour samples proved to be problematic in generating satisfactory chromosome preparations (Kallioniemi *et al.,* 1994). Furthermore, CGH analysis of tumours was pivotal in studying the genetic heterogeneity present in some tissues, the localisation of tumour suppressor genes and oncogenes, and inferring the possible mechanisms of tumour development (Ostroverkhova, Nazarenko, and Cheremnykh, 2002). The use of CGH in human samples progressed to applications in prenatal diagnostics of foetal chromosomal imbalances, replacing prenatal karyotyping studies from amniocytes and chorionic villi. Moreover, the resolution of the aCGH studies in prenatal diagnostics can reliably detect duplications, deletions, aneuploidy, and mosaicism, overcoming problems associated with karyotyping producing false-positive and false-negative results (Lichtenbelt, Knoers, and Schuring-Blom, 2011). To date, aCGH is one of the most commonly used diagnostic tool for genetic diagnosis in prenatal and clinical disorders, and has been able to detect chromosomal abnormalities with relatively mild phenotypes (Ballif *et al.,* 2008).

The use of aCGH has not been limited to the identification of chromosomal abnormalities and polymorphisms within a species, but has also been used for cross-species comparisons of multiple species, such as birds and primates. The application of CGH techniques in these cross-species studies has been successful in identifying lineage-specific chromosomal variations that arise throughout evolution, helping to elucidate the genetic events which may have contributed to speciation. Such studies have focused on copy number variations, in addition to the analysis gene expression patterns, between species. Furthermore, an additional benefit of cross-species aCGH compared to other

genomic studies, such as BAC mapping, is that it does not rely on the availability of a physical map or genome assembly for a reference, nor does the species being studied require its genome to be sequenced, assembled, or mapped.

1.4.3.1 Copy Number Variation

Copy number variations (CNVs) are polymorphisms that include amplifications, deletions, or duplications of DNA, and play an important role in genetic variation. The extent to which CNVs are tolerated is dependent on the genes involved, and specifically how well that gene tolerates duplications or deletions. Consequently, CNVs are can either be responsible for causing disease, speciation, and phenotypic variation. For example, within the human genome, CNVs are known risk factors in many diseases, such as autism and cancer (Lu *et al.,* 1988; Volik *et al.,* 2006; Poultney *et al.,* 2013). Yet despite being typically deleterious, CNVs have been shown to be essential for the development of lineage-specific traits between humans and other primates (Dumas *et al.,* 2007), and to promote adaptation under stress conditions and in natural environments (Gresham *et al.,* 2008; DeBolt, 2010).

Previously, the role of CNVs in influencing speciation was underestimated, largely due to assumptions that CNVs represented minor regions of genomic variation and that smaller CNVs were underrepresented in genomic databases (Gökçümen and Lee, 2009; Conrad *et al*., 2010). However, the use of inter-specific arrays for cross-species analysis spurred deeper investigation into the role of CNVs on genome evolution in a wide range of species, such as plants, birds and primates. Studies on the evolution of primate genomes identified genomic regions in gorillas, chimpanzees, macaques, orangutans, gibbons, and bonobos that exhibit large, high-frequency CNVs in comparison to the human,

which exhibits no or very low-frequency CNVs. It has been suggested that the differences in these lineages and the fixation of human-specific non-polymorphic regions was an early event in the history of modern humans (Gazave *et al.,* 2011; Hellen and Kern, 2015).

CNVs within avian genomes have also been extensively studied, representing the importance of CNVs on phenotypic diversity. Studies by Skinner *et al.* (2014) demonstrated that CNVs occur at a higher frequency per megabase in microchromosomes in comparison to macrochromosomes, and that the majority of CNVs detected (70%) contained genes, indicating functional importance. Examples can be seen with the silkie (hyperpigmentation of the connective tissue and skin) and pea comb phenotype (reduced in the size of the comb and the wattles) in chickens; the silkie phenotype exhibits an inverted duplication and junction of two genomic regions more than 400 kbp apart (Dorshorst *et al*., 2011), and the pea comb phenotype exhibits a 20 to 40-fold increase in copy number as opposed to the 2 copies found in wild-type (Vignal and Eory, 2019). These changes in copy number are thought to arise through increases in sequence-related meiotic recombination due the large number of CNVs found within microchromosomes (Skinner *et al.,* 2014; Weissensteiner and Suh, 2019).

1.5 Genome Sequencing Technologies

The ability to determine the order, conservation, and variation of nucleotides in genomic sequences allows us to make genotype to phenotype associations, in addition to tracing the evolutionary history of a genome. This has been made possible through unprecedented genome sequencing efforts, with the invention of first generation, next generation, and third generation sequencing revolutionising the field of genomics.

1.5.1 First Generation Sequencing

The early endeavours to obtain DNA sequences were cumbersome and time consuming, taking several years to obtain short fragments. In 1968, primer extension methods were used to identify RNA bacteriophage lambda and only yielded 12 base pairs of sequence (Wu and Kaiser, 1968). This was followed by the identification of 24 base pairs of the lactose-repressor binding site by Gilbert and Maxam in 1973, which took two years in total to complete (Gilbert and Maxam, 1973). Nevertheless, it was the discovery of two methods in 1976 that revolutionised sequencing methods; these were developed by Sanger and Coulson using chain termination, and by Maxam and Gilbert using chemical cleavage methods, but it was the method developed by Sanger and Coulson that was revolutionary in launching the genome sequencing era.

Now referred to as Sanger sequencing, this method relies on DNA polymerase to generate four extensions of a primer, with each of the four reactions containing trace amounts of a dideoxynucleotide (ddNTP) to produce DNA fragments of various lengths (Sanger *et al.,* 1977). As ddNTPs do not contain a 3’ hydroxyl groups necessary to form 5’ phosphate bonds with deoxynucleotides (dNTPs), the DNA chain cannot be extended beyond the incorporation of the ddNTP and thus hinders further progression of the polymerase extension. Through incorporating ddNTPs in four parallel base-specific reactions and measuring the size of the DNA fragments on a gel, it provides a single base resolution in which the DNA sequence can be determined (as seen in Figure 1.13).

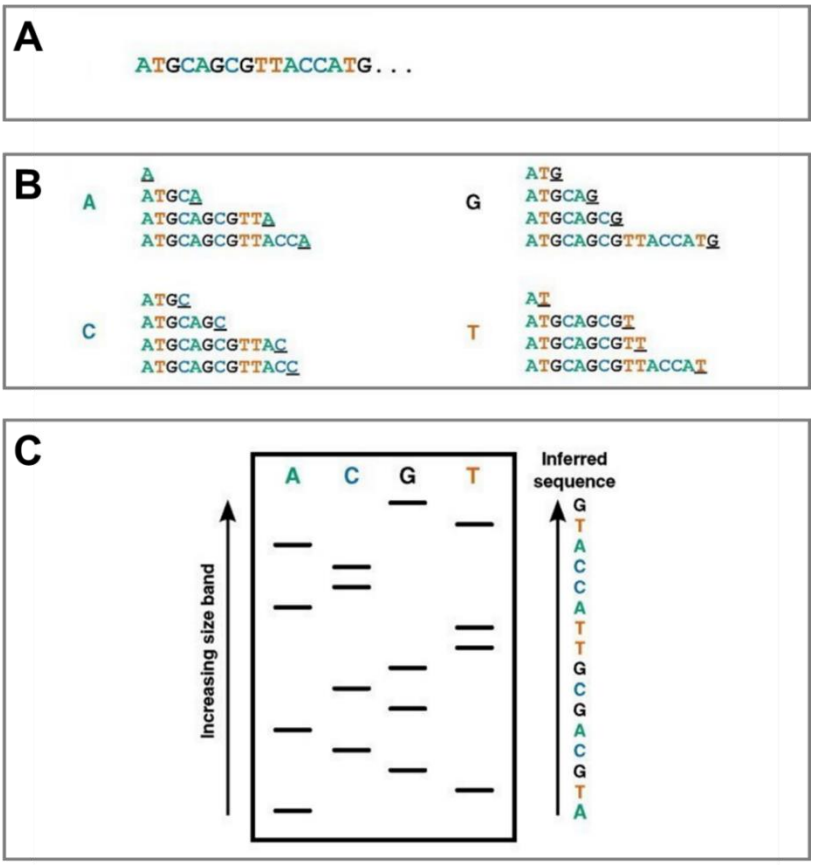


Figure 1.13: Overview of Sanger sequencing. A): DNA sample to be sequenced. B): Four reactions of sample DNA extended by DNA polymerase, with base-specific labelled ddNTPs (underlined terminal character) randomly incorporated. C): Schematic of a gel demonstrating that the labelled ddNTP corresponds to the base at that position, and thus the sequence can be determined by locating the lane in which the band is present. (Modified from Heather and Chain, 2016).

A series of improvements were applied to the Sanger sequencing method, including the substitution of radiolabelled ddNTPs with fluorometric-based detection, and the use of capillary-based electrophoresis for enhanced detection (Heather and Chain, 2016). These improvements led to the automation of the Sanger sequencing method, ensuing the development of commercial DNA sequencing (Hunkapiller *et al.,* 1991) and the sequencing of complex genomes. The creation of these automated first generation sequencing machines was significant for the field of genomics but had limitations in that the DNA sequences generated were less than 1 kbp in length, meaning other methodologies were required for assembling DNA fragments from species with larger and/or more

complex genomes. These methodologies included shotgun sequencing, whereby DNA fragments with overlapping regions were assembled for long contiguous sequences (contigs).

1.5.2 Next Generation Sequencing

Despite the revolutionary introduction of Sanger sequencing and its implications in the field of both genetics and bioinformatics, whole genome sequencing (WGS) using this technology was laborious, time consuming, and incredibly expensive. The demand for high-throughput sequencing was exemplified by the Human Genome Project, which took 13 years to complete and cost $3.8 billion. This first draft of the human genome produced an assembly consisting of 250,000 gaps and represented 90% of the euchromatic genome (International Human Genome Sequencing Consortium, 2001; Lander, 2011).

Next generation sequencing (NGS) technology, also known as short-read technology, superseded that of Sanger sequencing, providing a relatively easier and cheaper method to sequence DNA. NGS technology utilises massively parallel sequencing, sequencing the genome numerous times in small and random fragments to produce thousands to millions of sequences in parallel. To date, there are numerous NGS systems which differ in biochemistry and read length, yet the workflows typically include similar steps (Figure 1.14): DNA extraction, library preparation, the addition of adaptors and barcodes/indexes, amplification, template preparation, and automated sequencing.

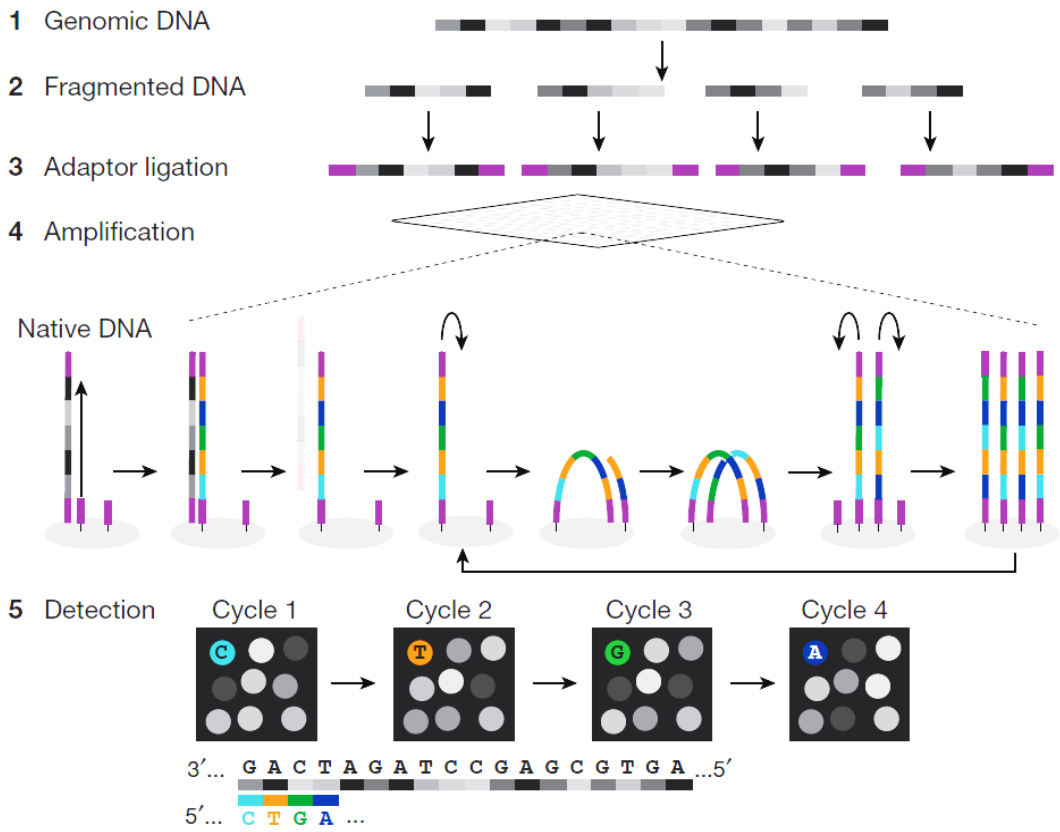


Figure 1.14: Schematic of next generation sequencing (modified from Shendure *et al.*, 2017).

Alongside the significant advantages of NGS technologies came numerous limitations, with the biggest being the relatively short reads in comparison to Sanger sequencing. Genomes typically contain repeat sequences, many of which are longer than NGS read sequences, resulting in a reduction in the length of contigs, the presence of gaps, and mis-assemblies in the genome (Goodwin, McPherson, and McCombie, 2016). As a result, genomes sequenced by NGS are heavily fragmented and limit in-depth studies for disease traits, genotype to phenotype associations, and comparative genomics. Nonetheless, these limitations can be overcome with higher coverage of the sequenced genome, allowing for the reads to be assembled into scaffolds, but incurs additional costs.

1.5.3 Third Generation Sequencing

Eukaryotic genomes are generally complex, consisting of copy number variations, repetitive elements, and structural variations, many of which are associated with disease traits, evolution, and adaptation (McCarroll and Altshuler, 2007; Völker *et al.,* 2010; Hull *et al.,* 2017). As mentioned above, NGS reads are inadequate in resolving these complex elements owing to their short length, and hence long-read sequencing becomes a necessity.

Third generation sequencing (TGS) technology, also known as long-read technology, superseded that of Sanger sequencing, generating reads that are several kilobases in length and providing a higher resolution of complex genomic elements. By spanning complex regions with single continuous reads, it reduces gaps and ambiguity in the location and size of these genomic elements. Furthermore, TGS allows for single molecule real time (SMRT) sequencing without the need for clonal or amplified DNA (Schadt, Turner, and Kasarskis, 2010). Pacific Biosciences (PacBio) is currently the most established TGS technology, generating reads of several kilobases long by combining molecular biology and nanotechnology with fluorometric detection (Eid *et al.*, 2009). Other TGS platforms include the MinION device developed by Oxford Nanopore Technologies, which is the smallest sequencing device currently available. Whilst the MinION is capable of generating read lengths similar to that of PacBio, there are documented difficulties in sequencing GC-rich regions.

However, the cost of using TGS technology is significantly higher and the throughput is relatively lower compared to NGS (illustrated in Table 1.2), with large amounts of starting material required for library preparation. Nevertheless, the generation of long reads using TGS is ideal for the assembly of *de novo*

genomes as it produces less fragments and gaps, yet remains unable to produce a contiguous sequence from the p-terminus to the q-terminus. Therefore, as is the case with other sequencing technologies, *in silico* methods are required to generate full chromosome-level assemblies. In addition to the requirement of *in silico* assembly methods, TGS technologies have inherently high error rates, often attributed to homopolymeric regions, and therefore also require assembly algorithms that rely on error correction (Besser *et al.,* 2018).

Method/Instrument	Read length (bp)	Strength	Weakness
Sanger			
ABI 3500/3730	Up to 1000	Read accuracy; length	Cost; throughput
454 Pyrosequencing			
GS 20/FLX	600 to 800	Read accuracy; throughput; speed	High initial investment; homopolymers
Applied Biosystems			
SOLiD	50 to 75	Low cost per base; accuracy	Slow; palindromic sequences.
Illumina			
MiniSeq	1×75 to 2×150	Low initial investment	Run and read length
MiSeq	1×36 to 2×300	Read length; scalability	Run length
NextSeq	1×75 to 2×150	Throughput	Run and read length
HiSeq (2500)	1×50 to 2×250	Read accuracy; throughput; low per sample cost	High initial investment; run length
NovaSeq 5000/6000	2×50 to 2×150	Read accuracy; throughput; low per sample cost	High initial investment; run and read length
IonTorrent			
PGM	Up to 400	Read length; speed	Throughput; homopolymers
S5	Up to 400	Read length; speed; scalability	Homopolymers
Proton	Up to 200	Speed; throughput	Homopolymers
Pacific BioSciences			
PacBio RSII	Up to 60,000	Read length; speed	High error rate; high initial investment; low throughput
Sequel	Up to 60,000	Read length; speed	High error rate
Oxford Nanopore Technologies			
MInION	Up to 100,000	Read length; portability	High error rate; run length; low throughput

Table 1.2: Characteristics of first, next, and third generation sequencing technologies (modified from Besser *et al.,* 2018).

1.5.4 Sequence Assembly

Regardless of improvements in sequencing technologies, all sequencers have limitations in read length and error rates, and consequently the sequences produced must first be assembled to reconstruct a genome. These sequences can either be assembled by whole genome shotgun sequencing (WGSS) or map-based approaches (illustrated in Figure 1.15), or a combination of the two.

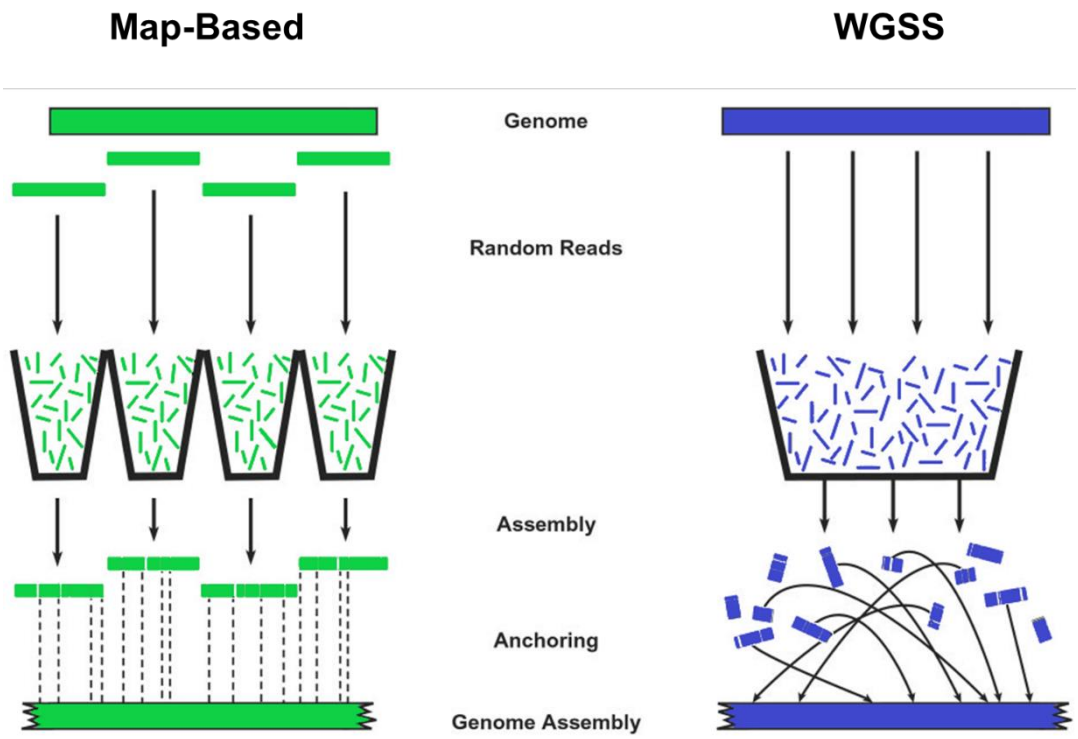


Figure 1.15: Overview of map-based sequencing versus whole genome shotgun sequencing (WGSS) (modified from Waterson, Lander, and Sulston, 2002).

Initial sequence assembly relied on map-based approaches, in which the whole genome is fragmented and inserted into BAC vectors to produce an overlapping BAC library. Using restriction fingerprinting (shared restriction bands) and BAC-end sequences, contigs are assembled which remain true to the order and orientation of the BACs (Pan *et al.,* 2016), producing a panel of BACs from a

minimal tiling path. These BACs are fragmented, and shotgun sequencing is used to determine the sequence of each BAC, generating the sequence of the genome being studied. Using map-based approaches, high quality genomes are assembled and with little error as the chromosomal location for each BAC is known, though this procedure is cumbersome and expensive.

WGSS bypasses the generation of physical maps and instead only uses overlapping clones to define the map, providing a relatively faster and cheaper approach to map-based methods. This involves fragmenting the DNA and cloning the fragments to generate plasmids libraries, which in turn are directly sequenced from both ends of the fragment, heavily relying on pair reads and computer algorithms for the assembly of contigs (Green, 2001). Despite being faster than map-based methods, the nature of WGSS renders this approach prone to errors and produces lower quality assemblies. Nevertheless, these limitations can be overcome with combined approaches. WGSS and map-based approaches are complementary to each other, in that map-based approaches produce higher quality assemblies with minimal mistakes but WGSS provides speed and cheaper costs. In combining WGSS with physical mapping data, complex genomes can be assembled to a high quality with relatively cheaper costs and at a faster rate (Chen *et al.,* 2002).

1.5.5 Complexities of Genome Sequencing

As previously mentioned, understanding genomes and elucidating the finer details of composition and complexity allows us to correlate sequence data and genomic structure with disease traits, evolution, and adaptation. Still, producing accurate and high-quality representations of genomes can become particularly problematic when sequencing complex genomes.

The quality of genome assemblies is determined by metrics such as the proportion of reads that are assembled, the number of gaps, the length and number of contigs and scaffolds, and the total length of the contigs and scaffolds relative to the size of the genome. N50 values, which are defined as the minimum contig length required to represent 50% of the genome, are the most commonly used metric, but does not consider other aspects which affect the quality of genome assemblies. For example, the artificial inflation of contigs generated by over aggressive joining of reads can produce scaffolds containing large numbers of repeats, producing mis-assemblies and misleading metrics (Salzberg and Yorke, 2005). Moreover, the quality of the genome assembly ultimately relies on the sequencing technology, the use of reference genomes, physical mapping data, and the type of assemblers/algorithms. To date, countless genomes have been assembled using of NGS technology, which produces vast numbers of short reads. The short reads result in the inability to reliably order and orientate scaffolds on chromosomes for complex genomes that consist of elements such as large gene families, duplications, and an abundance of repetitive sequences, ultimately generating incomplete assemblies (Pan *et al.,* 2016).

In addition to the abovementioned issues, NGS technology mostly relies on PCR for amplification of targeted genomic regions (Casbon *et al.,* 2011), which becomes problematic with GC-rich content due to inefficient polymerase amplification. For example, GC bias has been known to result in uneven coverage or no coverage of reads across the genome (Kozarewa *et al.,* 2009), with the issue of GC bias being well documented when generating genome assemblies using NGS data. Whilst it is not definitive as to how GC bias impairs

genome assembly, key findings have pointed to decreases in complete assemblies, assembly fragmentation due to low coverage of reads at GC-poor or GC-rich regions, and increased assembly errors when encountering tandem repeats (Chen *et al.,* 2012).

1.5.6 The Importance of Physical Genome Mapping

With the development of genome sequencing technologies, our understanding of genome composition and complexity is continuously evolving. Nonetheless, as aforementioned, sequencing technologies remain incapable of sequencing entire lengths of DNA that span a chromosome. Complex genomes with high repeat content, high GC content, and gene duplications still introduce limitations (Kajitani *et al.,* 2014). Furthermore, the reliance on sequencing technology and algorithms alone has resulted in a decline in quality of published genomes due to mis-assemblies and gaps (Kelley and Salzberg, 2010; Alkan, Sajjadian, and Eichler, 2011), and the high error rates associated with TGS technology hinder accurate assembly (Au *et al.,* 2012). Hence physical maps are incredibly important in helping to resolve these limitations and improve the quality of genome assemblies.

In the progression from genetics to genomics, the introduction of molecular cloning, BACs, microarrays, linkage maps, and radiation hybrids have been crucial in generating genome assemblies and addressing problems with the quality. Yet, molecular cytogenetic tools have not been the only solution to problems posed by NGS technology and complex genomes. For example, the development of computational algorithms, such as Bambus, OSLay, ABACAS, and Reference Assisted Chromosome Assembly (RACA) can be used to join contigs, and generate, order, and orientate scaffolds. In particular, RACA uses

paired-end reads and comparative genome information (Kim *et al.,* 2013), consisting of data from a closely related reference genome and one or more outgroup genomes. Predicted chromosome fragments (PCFs) can then be generated from scaffolds (as illustrated in Figure 1.16), and subsequently validated using PCR (a full methodology given in section 2.4). The combination of this algorithm with FISH has been successful in upgrading 5 avian genomes to that of a chromosome-level (Damas *et al.,* 2017; O'Connor *et al.,* 2018a), and was implemented in chapter 3 of this thesis.

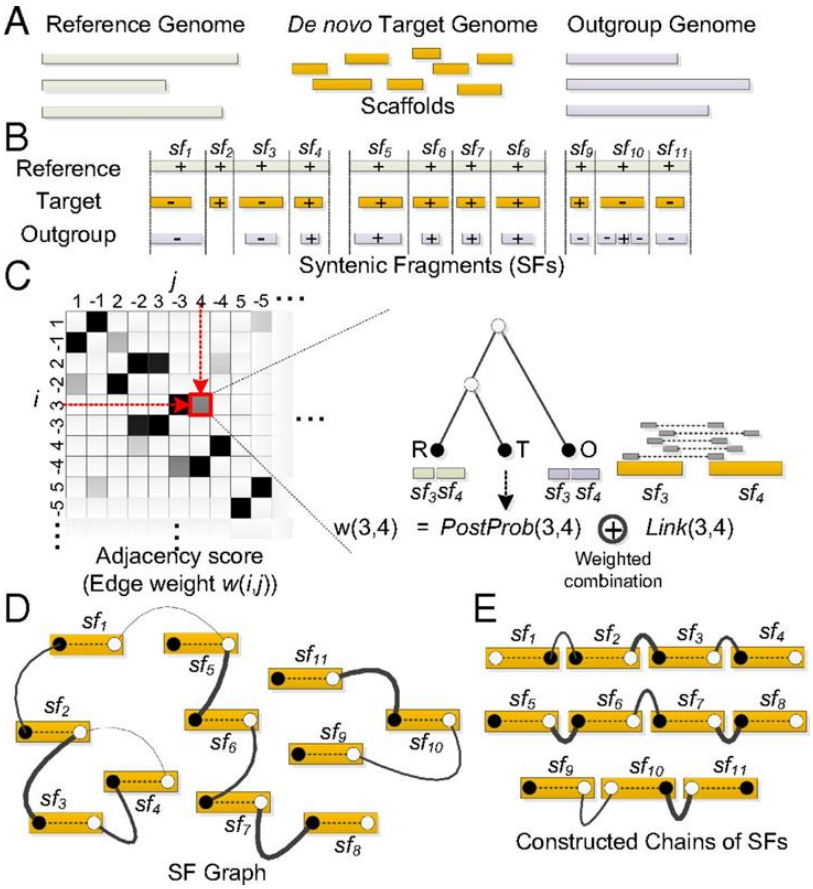


Figure 1.16: Overview of the RACA algorithm. A) Genomic data inputted from reference, de novo target, ad outgroup genome. B) Alignment of genomes to generate syntenic fragments (SF), with the orientations denoted with '+' and '-'. C) Syntenic fragments are scored, representing the adjacency. D) The generation of a syntenic fragment graph. E) Constructed syntenic fragment chains extracted by the RACA algorithm (Kim *et al.,* 2013).

1.6 Chromosomal Change and Speciation

The advent of genome sequencing technology (particularly assembling to chromosome-level) made it possible to study chromosome evolution at a much deeper level. In molecular terms, evolution of chromosomes occurs through rearrangements of genetic material, whether that be via deletion, acquisition, and/or modification of DNA, or structural rearrangements. These non-random chromosome rearrangements have been known to contribute to genome evolution and adaptation (Pevzner and Tesler, 2003), with fixation of chromosomal rearrangements arising in new species (Livingstone and Rieseberg, 2004). Nevertheless, the underlying mechanisms of chromosomal change in vertebrates can be explained by theories such as DNA proximity in chromatin (Branco and Pombo, 2006) and the role of repetitive sequences for NAHR in evolutionary breakpoint regions (EBRs) (Murphy *et al.,* 2005). In studying genomes at this resolution, it can provide insight into the role of chromosomal rearrangements in evolution and speciation.

1.6.1 Homologous Synteny Blocks and Evolutionary Breakpoint Regions

Homologous synteny blocks (HSBs) are chromosomal regions that share a common order of homologous genes between two or more genomes, and can remain undisturbed for millions of years of evolution. The continued presence of HSBs in multiple genomes suggests that the preservation of genes and gene order offers a selective advantage, especially considering that HSBs are enriched for distinct DNA features such as evolutionary conserved sequences and genes involving organismal development (Larkin *et al.,* 2009; Farré *et al.,* 2016). The disruption of HSBs can lead to alterations in expression of adjacent genes by the separation of existing regulatory elements and/or addition of new

regulatory elements, or through rearrangements of topologically associating domains which modifies chromatin interactions (Farré *et al*., 2019). Therefore, interchromosomal or intrachromosomal rearrangements that disrupt HSBs can generate phenotype diversity and speciation.

The main mechanism in which HSBs are rearranged arises from NAHR during meiosis (Lupski and Stankiewicz, 2005), in which conserved genomic regions are separated (typically along EBRs) and subsequently repaired in a different order. EBRs disrupt the conservation of synteny and are capable of disturbing HSBs, particularly in instances where unstable EBRs delineate stable HSBs (O'Connor *et al.,* 2018a). Rather than demonstrating high degrees of conservation, EBRs are enriched for genes relating to lineage-specific traits (Farré *et al*., 2016) and have been associated with chromosome fragile sites (Ruiz-Herrera, Castresana, and Robinson, 2006). These fragile sites are enriched for repetitive elements, segmental duplications, and are amongst members of a gene family which are located on distinct chromosomes (Kehrer-Sawatzki and Cooper, 2007; Lupski and Stankiewicz, 2005). The presence of these fragile sites and the subsequent chromosomal rearrangements can introduce post-zygotic barriers by suppressing recombination, which in turn leads to a reduction in fertility. Despite this reduction in fertility, these chromosomal changes may still be fixed due to mechanisms such as the facilitation of adaptive evolution via recombination suppression, meiotic drive, and the new rearrangement of genes that promotes positive selection (Potter *et al.,* 2017). With this reduction of recombination in chromosomes, populations will begin to diverge and eventually new species will arise.

1.7 Variations in Amniote Genome Structure

The structure of any genome can vary both across and within taxonomic groups. These variations include karyotypic diversity, degrees of synteny, genome size, chromosomal rearrangements, repeat content, and CNVs.

1.7.1 Avian Genomes

1.7.1.1 Avian Genome Size

Avian genomes are typically the smallest amongst amniotes, ranging from 0.9 Gb in the black-chinned hummingbird to 2.1 Gb in the ostrich, with an average of 1.35 Gb (Janes *et al.,* 2010; Suh, Smeds, and Ellegren, 2018). Compared to mammalian genomes, the chicken genome is approximately 60% smaller than the human genome, with 1.05 Gb and 3.1 Gb respectively, and approximately 50% smaller than the mouse genome (2.5 Gb).

These compact genomes demonstrate a low frequency of repetitive elements, which are known to contribute to the diversity in vertebrate genome size and structure. In comparison to mammalian genomes, avian genomes contain roughly 10% of interspersed repeats as opposed to the 40-50% observed in mammalian genomes. In addition, it is thought that the shortening of introns and reduction in intergenic distance, further reducing genome size, correlates with high metabolic rates and rapid gene regulation required for flight (Zhang *et al.,* 2014). However, it has also been suggested that the reduction in genome size is non-adaptive and neutral, preceding the ability of flight (Nam and Ellegren, 2012).

1.7.1.2 Avian Karyotype Diversity

Despite variations in genome structure and millions of years of evolution, avian karyotypes demonstrate a remarkable degree of karyotypic stability. Most avian lineages exhibit diploid numbers of 2*n*=~80, with the key morphological feature of these “standard” karyotypes being the presence 10 pairs of macrochromosomes and approximately 30 pairs of microchromosomes (Christidis, 1990; Masabanda *et al.,* 2004), an example of which can be seen in Figure 1.17. The size of macrochromosomes ranges from 30 Mb to 250 Mb, and the average length of microchromosomes is 12 Mb, with the smallest being 3 Mb (Pichugin *et al.,* 2001). The size, number, and morphology of microchromosomes is problematic when analysing avian chromosomes, and so most karyotypes produced by classical cytogenetic techniques are only partially characterised.

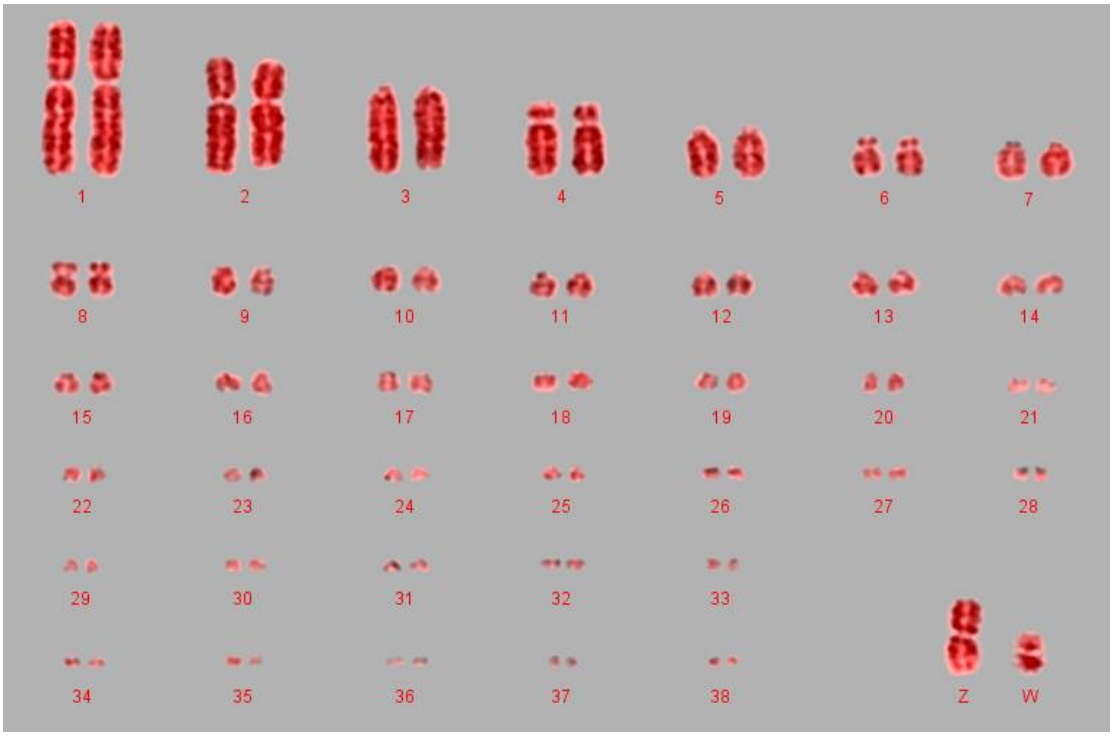


Figure 1.17: Karyotype of the chicken (*Gallus gallus*). Chromosomes are stained with DAPI and propidium iodide (prepared personally).

Variations in karyotype have been observed in avian species (as seen in Figure 1.18), with diploid numbers ranging from 2*n*=42 in the stone curlew (Nie *et al.,* 2009) to 2*n*=142 in the grey lourie (Christidis, 1990). Species which exhibit unusually small diploid numbers tend to demonstrate the fusion of microchromosomes to macrochromosomes and/or other microchromosomes, and those that exhibit large diploid numbers demonstrate the fission of macrochromosomes.

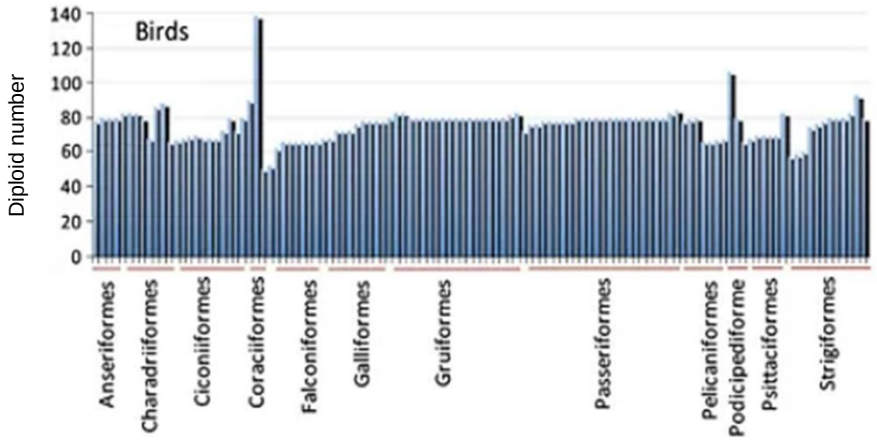


Figure 1.18: Variation in avian diploid number (modified from Ruiz-Herrera, Farré, and Robinson, 2012).

It is thought that process of chromosomal fusions and fissions have arisen as a result of natural evolution (Burt, 2002; Giannuzzi *et al.,* 2013), and these processes may have accelerated evolution (Wang *et al.,* 2019). Additionally, there are some well characterised fusions and fissions of avian chromosomes reported in literature that would affect diploid number. For example, in the lineage leading to the chicken, there was a fusion of a microchromosome to the p-arm of chicken chromosome 4. This p-arm exhibits particularly high GC content and recombination rates, features that are consistent with its origins as a microchromosome (Griffin *et al.,* 2008).

1.7.1.2.1 Identification of Rearrangements with Chromosome Painting

As mentioned in section 1.4.1.2, chicken chromosome paints have been extensively applied to numerous avian species, with examples given in Table 1.3. Many species, including those that are phylogenetically distant, exhibit strong chromosome homology. For example, the macrochromosomes of the elegant crested tinamou (*Eudromia elegans*) and ostrich (*Struthio camelus*) demonstrate identical chromosome homology to the chicken, with almost every chromosome paint hybridising to a single pair of macrochromosomes despite approximately 111.4 million years of divergence (http://www.timetree.org; Hedges, Dudley, and Kumar, 2006). However, there are some notable exceptions, such as chromosome 4 hybridising to one pair of macrochromosomes and one pair of microchromosomes, and paints for chromosome Z also hybridising to chromosome W (Nishida-Umehara *et al.,* 2007).

The hybridisation of chicken chromosome paint 4 to both macrochromosomes and microchromosomes is a recurrent rearrangement observed in several bird species, as the chicken exhibits an ancestral fusion not commonly observed in other birds (Guttenbach *et al.,* 2003). Another example of recurrent rearrangements is a synapomorphic trait observed in all *Passeriformes*, in which a fission in chromosome 1 produces chromosomes 1 and 1A, with chromosome 1A being the fifth largest chromosome (Nanda *et al.,* 2011).

Order	Species	Chr1	Chr2	Chr3	Chr4	Chr5	Chr6	Chr7	Chr8	Chr9
Struthioniformes	*Struthio camelus*	GGA1	GGA2	GGA3	GGA4	GGA5	GGA6	GGA7	GGA8	GGA9
Rheiformes	*Rhea americana*	GGA1	GGA2	GGA3	GGA4	GGA5	GGA6	GGA7	GGA8	GGA9
	Rhea pennata	GGA1	GGA2	GGA3	GGA4	GGA5	GGA6	GGA7	GGA8	GGA9
Casuariiformes	*Casuarius casuarius*	GGA1	GGA2	GGA3	GGA4	GGA5	GGA6	GGA7	GGA8	GGA9
	Dromaius novaehollandiae	GGA1	GGA2	GGA3	GGA4	GGA5	GGA6	GGA7	GGA8	GGA9
Tinamiformes	*Eudromia elegans*	GGA1	GGA2	GGA3	GGA4	GGA5	GGA6	GGA7	GGA8	GGA9
Anseriformes	*Coscoroba coscoroba*	GGA1	GGA2	GGA3	GGA4	GGA5	GGA6	GGA7	GGA8	GGA9
	Anser anser	GGA1	GGA2	GGA3	GGA4	GGA5	GGA6	GGA7	GGA8	GGA9
Galliformes	*Phasianus colchicus*	GGA1	GGA3	GGA2	GGA4	GGA5	GGA2	GGA6	GGA7	GGA8
	Coturnix coturnix	GGA1	GGA2	GGA3	GGA4	GGA5	GGA6	GGA7	GGA8	GGA9
Accipitriformes	*Nisaetus nipalensis orientalis*	GGA4	GGA2/Mic	GGA2	GGA6/Mic	GGA7	GGA1	Mic	GGA8	GGA9/Mic
	Cathartes aura	GGA1	GGA2	GGA3	GGA4	GGA5	GGA6	GGA7	GGA8	GGA4
Eurypygiformes	*Eurypyga helias*	GGA2/GGA5	GGA1	GGA3	GGA4	GGA1	GGA6	GGA7	GGA2	GGA8
Gruiformes	*Fulica atra*	GGA1	GGA2	GGA3	GGA4/GGA5	GGA6/GGA7	GGA8	GGA4	GGA9	GGA10
	Gallinula chloropus	GGA1	GGA2	GGA3	GGA4/GGA5	GGA6/GGA7	GGA8	GGA4	GGA9	GGA10
Charadriiformes	*Burhinus oedicnemus*	GGA1	GGA2	GGA3	GGA4q	GGA7/GGA8	GGA5	GGA9/2/Mic	GGA4p/Mic	GGA6/Mic
	Vanellus chilensis	GGA1	GGA2	GGA3	GGA7/GGA8	GGA4q	GGA5	GGA6	GGA9	GGA10

Opisthocomiformes	*Opisthocomus hoazin*	GGA1	GGA3	GGA2	GGA2/GGA9	GGA 4q	GGA6/ GGA8	GGA5	GGA1	GGA7
Columbiformes	*Leptotila verreauxi*	GGA1	GGA2	GGA3	GGA6/ GGA7	GGA4	GGA5	GGA8	GGA9	GGA10
	Zenaida auriculata	GGA1	GGA2	GGA3	GGA4	GGA5	GGA6	GGA7	GGA8	GGA9
Strigiformes	*Strix nebulosa*	GGA2	GGA4q/ GGA5	GGA1	GGA3	GGA1	GGA7	GGA6	GGA8	GGA9
	Pulsatrix perspicillata	GGA1/ GGA2	GGA3	GGA4/GGA5	GGA1	GGA6 /GGA 7	GGA9/ GGA4	GGA5/ GGA8	GGA10	
Trogoniformes	*Trogon surrucura surrucura*	GGA1	GGA3	GGA2	GGA6/GGA7	GGA 4q	GGA5	GGA8	GGA2	GGA9
Falconiformes	*Falco columbarius*	GGA3/ GGA4/ Mic	GGA1	GGA2/GGA5 /Mic	GGA2/GGA3/Mic	GGA5 /GGA 7/Mic	GGA6/ GGA8/ Mic	GGA9	GGA4/ Mic	Mic
	Pandion haliaetus	GGA1/ GGA9	GGA1/Mic	GGA1/GGA6	GGA1/GGA4	GGA1	GGA3		GGA4	GGA2
Psittaciformes	*Nymphicus hollandicus*	GGA2	GGA3	GGA1	GGA4/GGA8/GGA9	GGA6 /GGA 7	GGA1	GGA5		GGA10
	Ara macaw	GGA1/ GGA4	GGA2	GGA3	GGA1	GGA5 /Mic	GGA6/ GGA7	GGA8/ GGA9/ Mic	Mic	GGA1
Passeriformes	*Sitta europaea*	GGA2	GGA1	GGA3	GGA4	GGA1	GGA5/ GGA10	GGA8/ GGA9	GGA6	GGA7
	Satrapa icterophrys	GGA3	GGA1q	GGA2q	GGA4q	GGA1 p	GGA5	GGA2p	GGA6	GGA7

Table 1.3: Chromosome painting data of macrochromosomes 1 to 9 using chicken chromosome paints. Chr= Chromosome. GGA = *Gallus gallus.* Mic= Microchromosome. (Modified from Garnero *et al.,* 2019).

1.7.1.3 Evolution of the Chicken Genome Assembly

The chicken plays an important role in society, constituting an excellent model organism and agricultural animal. The chicken genome was first sequenced in 2004, becoming the first avian and non-mammalian amniote genome to be sequenced (International Chicken Genome Sequencing Consortium, 2004) and marking the start of a revolution in avian genomics.

The initial draft of the chicken genome was assembled using a WGSS approach that mainly consisted of paired-end plasmid reads produced using Sanger sequencing (Burt, 2005). Combining sequence data with genetic and physical maps produced a high-quality assembly, representing 85% of the genome and providing coverage for only 26 of the 38 autosomes (Warren *et al.,* 2017; Cheng and Burt, 2018). Yet, this initial draft contained gaps and low sequence coverage in specific genomic regions, resulting in numerous microchromosomes that were partially assembled or not assembled at all. Based on the original data from the initial draft, the second version of the chicken genome was released in 2006 (*Gallus_gallus*-2.1; GCA_000002315.1). This version provided a further 198,000 reads and incorporated data from radiation hybrid maps (Morisson *et al.,* 2007). Furthermore, revisions to linkage and physical maps provided improved resolution, in which 95% of the genome was represented (Groenen *et al.,* 2009).

Released in 2011, the third version of the chicken genome (*Gallus_gallus*-4.0; GCA_000002315.2) utilised NGS technology, combining that with existing sequence and mapping data. This version identified approximately 10 Mb of erroneous duplications from the *Gallus_gallus*-2.1 assembly and helped to improve the data associated with the Z chromosome (Bellott *et al.,* 2010; Rubin

et al., 2010). Nevertheless, there was only a small increase in genomic representation, increasing from 95% to 96% in *Gallus_gallus*-4.0. The fourth and current version of the chicken genome (*Gallus_gallus*-5.0, GCA_000002315.3) implemented TGS technology, using PacBio SMRT to generate long-read data. This data was error-corrected (refer to section 1.2.3.3), with the new *de novo* assembled contigs being merged with existing data from the *Gallus_gallus*-4.0 assembly. Despite the increase in read length and incorporation of all available data, this newest assembly demonstrates a reduction in genomic representation, decreasing from 96% to 87%. This reduction is thought to arise from the sequencing of previously unsequenced regions, which now remain partially assembled or unplaced (Peona, Weissensteiner, and Suh, 2018).

After 15 years of ongoing efforts to further improve the chicken genome assembly, the presence of gaps continues to hinder our insight into the true size and composition of avian genomes. Avian species exhibit one of the highest numbers of sequenced genomes, yet sequence data for microchromosomes 29-31 and 33-38 is near absent. Furthermore, the lack of physical mapping data for microchromosomes 29-38 reduces the degree of certainty in anchoring sequence data, restricting evolutionary and comparative genomic studies. Therefore, the generation of reliable tools for the detection of these smallest microchromosomes, whether they be sequencers or molecular cytogenetic tools, is essential for generating a full chromosome-level assembly to represent all avian chromosomes. This is an issue that is explored further in chapter 6 of this thesis.

1.7.1.4 Other Avian Genome Assemblies

To date, there are numerous genome sequences available for avian species, yet very few are fully assembled. Of those with assembled genomes, the most studied belong to the zebra finch (*Taeniopygia guttata*), mallard duck (*Anas platyrhynchos*), and domestic turkey (*Meleagris gallopavo*). However, of all assembled avian genomes, including that of the chicken, microchromosomes beyond that of chromosome 28 have little or no sequence data associated with them. Thus, despite the other chromosomes typically containing substantial structural and genetic information, all avian genomes are yet to be fully assembled (discussed further in chapter 6).

1.7.1.4.1 The Zebra Finch Genome

After the chicken genome, the zebra finch became the second avian species and the first *Passeriforme* to have its genome sequenced. The sequencing of the zebra finch was incredibly significant for avian genomics considering that Passerines represent the largest of all avian orders, constituting over half of all avian species (Ricklefs, 2012). Furthermore, as a representative of Passerine birds, the zebra finch exhibits the lineage-specific trait of vocal learning, a trait shared only with humans, parrots, and hummingbirds (Balakrishnan *et al.,* 2010; Nowicki and Searcy, 2014).

The development of a BAC library, expressed sequence tag databases, and linkage maps occurred before the genome sequencing of the zebra finch, which was then assembled using a shotgun-based Sanger sequencing approach (Balakrishnan *et al.,* 2010). The draft assembly of the zebra finch genome (*Taeniopygia_guttata*-3.2.4) spanned a total of 1.2 Gb, in which 1.0 Gb was assigned to 33 chromosomes and 3 linkage groups (Warren *et al.,* 2010). As of

December 2019, there are 7 assemblies of the zebra finch genome according to Assembly (https://www.ncbi.nlm.nih.gov/assembly), of which 5 belong to the Vertebrate Genomes project and are pseudo-haplotypes. Previously, most avian studies on relied on the chicken genome assembly considering the first genome sequence of the zebra finch was only made available in 2010, 6 years after the initial sequencing of the chicken genome and 2 chicken genome assemblies later (Stapley *et al.,* 2008; dos Santos *et al.,* 2017). Nevertheless, the zebra finch is now an important reference genome for many sequencing and assembly efforts.

1.7.1.4.2 Falcon Genomes

The sequencing of two falcon species, the peregrine falcon (*Falco peregrinus*) and saker falcon (*Falco cherrug*), was also important in avian genomics as these genomes represent species which deviate from the “standard” avian karyotype. In the *Falco* genus, there is a notable reduction in the number of microchromosomes, with karyotypes consisting of 7-11 macrochromosomes and 13-16 microchromosomes (De Boer, 1975). Thus, diploid numbers have been shown to range from 2*n*= 40 in the Merlin (*Falco columbarius*) to 2*n*= 54 in lanner falcon (Joseph *et al.,* 2018). It has been inferred that the highly rearranged karyotypes observed in falcons is associated with the behavioural, morphological, and physiological adaptations observed, such as the visual acuity and flying speeds. Both genomes were sequenced using Illumina HiSeq 2000 deep sequencing and assembled using SOAPdenovo, with estimated genome sizes of 1.21 Gb in the peregrine falcon and 1.18 Gb in the saker falcon (Zhan *et al.,* 2013).

1.7.2 Non-avian Reptile Genomes

1.7.2.1 Non-avian Reptiles Genome Structure

As is the case with karyotypic diversity, non-avian reptiles demonstrate large variation in genome size and composition in numerous Orders and Suborders. Genome size ranges from 1.03 Gb in the mionecton skink (*Chalcides mionecton*) to 5.3 Gb in the Mediterranean spur-thighed tortoise (*Testudo graeca*), with the range of genome sizes given in Table 1.4. The variation in genome size is dependent on non-long terminal repeat (LTR) retrotransposons, and typically accounts for 30% of the entire genome (Canapa *et al.,* 2015).

Taxon	Genome Size (Gb)
Testudines	1.4 to 5.3
Crocodilia	1.3 to 3.9
Sphenodontia	4.9
Squamata	1.03 to 3.8

Table 1.4: Variation in genome size in non-avian reptiles (Alam *et al.,* 2018).

Unfortunately, non-avian reptiles are vastly underrepresented in genomic studies and thus the depth of information available is relatively limited. To date, there are only 66 genomes that have sequenced data associated with them, of which 34 belong to lepidosaurs, 26 to testudines, and 6 to crocodilians (https://www.ncbi.nlm.nih.gov/assembly, last accessed on December 19th 2019).

1.7.2.2 Non-avian Reptile Karyotype Diversity

There are three distinct lineages of extant non-avian reptiles, consisting of lepidosaurs (squamate reptiles and tuataras), archosaurs (crocodiles), and testudines (turtles) (Olmo, 2008). The chromosomes of non-avian reptiles are interesting to study as they demonstrate a high degree of diversity in both chromosome number and morphology (Olmo, 2008), with karyotypic patterns similar to that of birds (as seen in Figure 1.19). As is the case with birds, most karyotypes produced by classical cytogenetic techniques are only partially characterised due to the size, number, and morphology of microchromosomes.

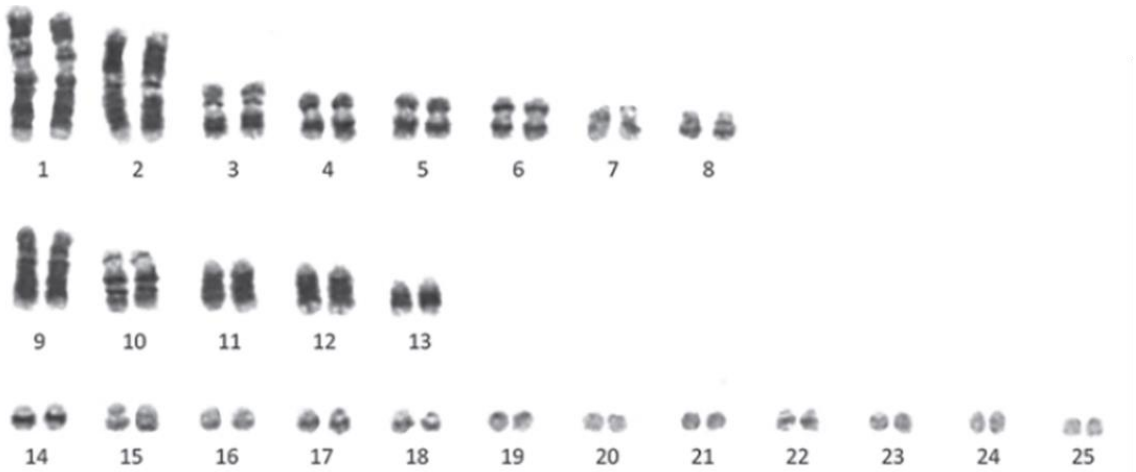


Figure 1.19: Karyotype of the European pond turtle (E*mys orbicularis*) after GTG-banding (Iannucci *et al.,* 2019).

Furthermore, non-avian reptiles are karyologically heterogenous, with diploid numbers ranging from 2*n*= 20 in the spectral pygmy chameleon (*Rampholeon spectrum*) to 2*n*= 68 in the twist-necked turtle (*Platemys platycephala*). These variations in karyotype are attributed to the presence or absence of microchromosomes, with squamate reptiles and turtles demonstrating the highest diversity in diploid number (as summarised in Table 1.5).

Order	Diploid Range	Macrochromosome Range	Microchromosome Range
Testudines	26 to 68	10 to 36	0 to 56
Crocodilia	30 to 42	30 to 42	0
Sphenodontia	36	28	8
Squamata	20 to 62	10 to 38	0 to 36

Table 1.5: Diversity of diploid numbers in non-avian reptiles (Deakin and Ezaz, 2019).

1.8 Specific Aims of this Thesis

Due to numerous constraints, whether it be the limitations of technology, unoptimised experimental procedures, or simply human error/bias, there remains a vast number of under-explored areas in the fields of genomics. The absence in general terms of chromosome-level assemblies for most species impedes our understanding of chromosome evolution and the mechanisms in which this occurs. In birds, exceptions to the usual pattern of 2*n*=~80 (e.g. *Psittaciformes and Falconiformes*) provide much needed insight into this question and thus chromosome-level assemblies of such animals are urgently needed.

Once chromosome-level assemblies are established for a small handful of species then comparative genomics at a chromosome-level becomes much easier, even in species that are not sequenced. As pointed out in section 1.4.2.2 there now exist tools (selected BAC clones) that can be used for this purpose however, although inter-chromosomal rearrangement is well studied, intrachromosomal rearrangement between birds is less. Similarly, in non-avian reptiles, karyotype evolution remains largely understudied because of the lack of sufficient tools (see section 1.4.2.2). Finally, the “black box” that is the smallest of microchromosomes remains frustratingly under discovered in both

molecular genomic and cytogenetic terms (despite the efforts of Masabanda *et al.,* 2004). Taking all these factors into consideration, the purpose of this thesis is to address the shortfalls in our knowledge through the pursuit of the following specific aims:

Specific aim 1: To upgrade the scaffold-based budgerigar (*Psittaciformes*) genome assembly, which is known to have an atypical avian karyotype, to chromosome-level using bioinformatics (in collaboration with the Royal Veterinary College, London) and molecular cytogenetics (FISH).

Specific aim 2: To produce comparative cytogenomic maps for 7 avian species to investigate phylogenetic relationships and lineage-specific patterns arising from chromosomal rearrangements.

Specific aim 3: To investigate genome structure and conservation between avian and non-avian reptiles, comparing the chicken and two karyotypically dissimilar turtle species (yellow spotted river turtle and spiny softshell turtle) using chromosome paints and sequence conserved BACs.

Specific aim 4: Identify genes within the newest chicken genome assembly (*Gallus_gallus*-5.0; GCA _ 000002315.3) to generate a panel of fluorescent markers for the hitherto undiscovered microchromosomes 29 to 38.

2 Materials and Methods

2.1 Chromosome Preparation

2.1.1 Culturing Fibroblasts

Chicken embryonic fibroblasts (CEFs) were obtained from local suppliers (Pirbright Institute). Cells were incubated at 40^{o}C with 5% CO_2.

2.1.1.1 Culture Media Preparation

CEFs were cultured in medium consisting of MEM Alpha Modification (Gibco), supplemented with 10% fetal bovine serum (Gibco) and 1% Penicillin-Streptomycin-L-Glutamine (Sigma-Aldrich). The medium was prepared in a class II hood and stored at 4^{o}C until needed.

2.1.1.2 Refreshing Medium

The medium within the flasks was refreshed every 2 days, replacing spent medium with complete medium. Depending on growth within the flask, 3.5-4.5 mL of spent media was removed and replaced to a total of 5 mL in T25 flasks, 10 mL in T75 flasks, and 20 mL in T175 flasks.

2.1.1.3 Passaging Cells

Cells were passaged when the growth within the flask had reached 80-90% confluency. Spent medium was aspirated, and the flask rinsed with 1 mL (2 mL for T75, 4 mL for T175) Hank's Balanced Salt Solution (HBSS) (Gibco) which was removed after sufficiently coating the cells. 1 mL (2 mL for T75, 4 mL for T175) of 0.05% Trypsin-1x ethylenediaminetetraacetic acid (EDTA) (Gibco) was added to the flask and then warmed on a 37^{o}C hotplate. Once the fibroblasts had rounded up and detached from the flask (observed under a microscope), the cells were resuspended in 9.5 mL of complete medium. 10 mL of the cell

suspension was transferred to a T75 flask and the T25 flask was refreshed with complete medium to a final volume of 5 mL.

Cells passaged between T75 flasks had the cells resuspended in 9.5 mL complete medium, of which 6 mL was transferred to the new T75 flask and both replenished with complete medium to a final volume of 10 mL. Cells passaged from T75 flasks to T175 flasks had the cells resuspended in 20 mL of complete medium, of which 16 mL was transferred to the new T175 flask and both replenished with complete medium to a final volume of 10 mL (T75) or 20 mL (T175).

2.1.2 Harvesting Chromosomes

Cultures were harvested when growth was optimal, with a confluency of 80-90% and evidence of many mitotic doublets. Colcemid solution (Gibco) was added to the flasks 1 hour prior to harvesting, with a final concentration of 100 µg/mL in each flask. After 1 hour, the medium was discarded and the flasks rinsed with 1 mL (2 mL for T75, 4 mL for T175) HBSS (Gibco), which was removed after sufficiently coating the cells. 1 mL (2 mL for T75, 4 mL for T175) of 0.05% Trypsin-1xEDTA (Gibco) was added to the flask and then warmed on a 37^{o}C hotplate. Once the fibroblasts had rounded up and detached from the flask (observed under a microscope), the cells were resuspended in 5 mL HBSS (Gibco) and transferred to a 15 mL falcon tube. Flasks were further rinsed with HBSS (Gibco) to remove all cells from within the flask, with the cell suspension being transferred to the 15 mL falcon tube.

The cell suspensions were centrifuged for 10 minutes at 1,000 RPM, with the supernatant being discarded and the cell pellet resuspended. 75 mM pre-

warmed (37ºC) potassium chloride (KCl) was added in a dropwise manner with gentle agitation, and the samples incubated at 37ºC for 20 minutes.

Three droplets of chilled fixative (3:1 absolute methanol:absolute acetic acid) were added with gentle agitation, and the samples centrifuged for 10 minutes at 1,000 RPM. After discarding the supernatant, pellets were resuspended gently using a Pasteur pipette. The sample was drawn into the pipette, and 5 mL chilled fixative was added to the falcon tube, followed by the sample being slowly released into the fixative. The samples were centrifuged for 10 minutes at 1,000 RPM, and the fixing process repeated an additional 2 times. Samples were labelled and suspended in 5 mL fixative at -20ºC until needed.

2.2 Preparation of BAC Probes

2.2.1 BAC Selection

Bacterial artificial chromosomes (BACs) were obtained from the chicken CHORI-261 BAC library (BACPAC) and the zebra finch TGMCBA library (Wageningen). BACs were received in Luria-Bertani (LB) agar stab format.

2.2.1.1 LB Broth

LB broth was created with 10 g LB brother powder (Invitrogen) and 500 mL of ddH_2O. The agar solution was autoclaved, and chloramphenicol (60 µg/mL) was later added when the agar solution had sufficiently cooled. Sterile pipette tips were inserted into the agar stab and then transferred to 50 mL falcon tubes containing 20 mL LB broth. Cultures were left overnight at 37ºC in a shaking incubator at 140 RPM.

2.2.1.2 LB Agar Plates

LB agar plates were created with 16 g LB agar powder (Invitrogen) and 500 mL of deionised water (ddH_2O). The agar solution was autoclaved, and chloramphenicol (60 µg/mL) was later added when the agar solution had sufficiently cooled.

2.2.2 Isolation and Purification of BAC Clone DNA

A sterile pipette tip was used to streak glycerol stocks of *Escherichia coli* (*E. coli*) transformed with BAC clones, onto LB agar plates. Plates were cultured overnight at 37^{o}C. Plates were washed with 2 mL 1x phosphate buffer solution (PBS) to resuspend the colonies. DNA from the colonies was purified using a Miniprep kit (Qiagen) following standard protocol.

2.2.3 Amplification of BAC Clone DNA

Each DNA sample isolated was analysed for a suitable DNA concentration and 260/280 purity ratio using a NanoDrop™ 1000 spectrophotometer (Thermo-Fisher Scientific). The DNA was amplified using GenomiPhi V2 DNA Amplification Kit (GE Healthcare). A modified protocol was followed, with all reagents obtained from the kit.

3 µL purified BAC DNA was transferred to 0.5 mL eppendorf tubes and mixed with 27 µL sample buffer. Samples were incubated at 95^{o}C in a thermocycler for 3 minutes and then immediately placed on ice. 30 µL enzyme/reaction buffer solution was added to each sample. The volume of Phi29 DNA polymerase enzyme was calculated at a ratio of 3 µL x the number of tubes x 1.2, with the reaction buffer volume calculated at 9x the volume of enzyme. The samples

were placed in a thermocycler at 30ºC for 1.5 hours, followed by heat inactivation at 65ºC for 10 minutes.

60 µL ddH_2O and 12 µL sodium acetate/EDTA buffer (3 M sodium acetate pH 8.00 and 0.5 M EDTA pH 8.00) were added to the sample, followed by 300 µL 100% ethanol prior to centrifugation for 15 minutes at 11,000 RPM. The supernatant was discarded, 500 µL 70% ethanol was added, and the sample was centrifuged again at 11,000 RPM for 2 minutes. The supernatant was discarded and the residual ethanol was left to evaporate. The pellet was resuspended in 60 µL 10 mM Tris-HCl buffer (pH 8.00) and samples were stored at 4ºC.

2.2.4 Nick Translation

Each amplified DNA sample was analysed for a suitable DNA concentration and 260/280 purity ratio using a NanoDrop™ 1000 spectrophotometer (Thermo-Fisher Scientific). Samples were diluted with 10 mM Tris-HCl buffer (pH 8.00) to a final concentration of 166.5 µg/µL

Probe mixes were prepared to a total volume of 100 µL, consisting of 12 µL BAC DNA, 0.01 M Dithiothreitol (DTT), 8 µL Nucleotide Mix A (Cytocell Ltd), 0.01 M nick translation buffer (Cytocell Ltd), 49.5 µL MBG H_2O, 4 µL DNA polymerase I (10 U/mL), and 5 µL DNase I (0.01 U/mL). To these mixes, either 1.5 µM FITC-Fluorescein-12-UTP (Roche), 1.5 µM Texas Red-12-dUTP (Invitrogen) or 1.5 µM Aqua PromoFluor-532-aadUTP (Promokine) was added. The probe mixes were pulsed and incubated for 2 hours at 15ºC in a thermocycler, followed by heat inactivation at 65ºC.

2.2.5 Gel Preparation

To ensure the BAC DNA was digested below 500 bp, a 1.4% agarose gel was prepared. Agarose (Bio-Rad) was dissolved in 1xTris-borate-EDTA (TBE) (Sigma-Aldrich), with 2 µL SYBR Safe (Invitrogen) added when the gel solution had sufficiently cooled. 2 µL BAC DNA and 2 µL 100 bp ladder (Promega) were mixed with 2 µL 6x loading buffer (Promega) before loading. Samples were run at 90 V/58 mA.

2.2.6 Probe Purification

Once the BAC DNA was digested to the correct size, the samples were purified using the QIAquick Nucleotide Removal kit (Qiagen) following standard protocol.

2.3 Preparation of Chromosome Paints

2.3.1 Amplification of DNA

Flow-sorted chromosomal DNA was amplified via degenerate oligonucleotide (DOP) polymerase chain reaction (PCR) using a DOP primer designed by Cytocell Ltd (OGT, Cambridge). Primers were produced by Eurofins Genomics.

PCR reactions were prepared, consisting of 0.2 mM dNTPs, 0.01 U/µL High Fidelity platinum Taq polymerase (Invitrogen), 1x platinum Taq buffer (Invitrogen), 4 µM DOP primer, chromosomal DNA (10 ng), and MBG H_2O to a total volume of 100 µL. The following cycle was run for primary and secondary DNA amplification (Table 2.1):

Step	Temperature (ºC)	Duration	Cycles
Denaturation	94	30 seconds	20
Annealing	56	30 seconds	
Extension	72	2 minutes	
Final Extension	72	5 minutes	1

Table 2.1: Summary of the PCR cycling conditions used for DOP-PCR DNA amplification.

2.3.2 Purification of DNA

After amplification, PCR products were purified using the QIAquick PCR Purification kit (Qiagen) following standard protocols.

2.3.3 Labelling DNA

Each amplified DNA sample was analysed for a suitable DNA concentration and 260/280 purity ratio using a NanoDrop™ 1000 spectrophotometer (Thermo-Fisher Scientific). For FITC labelled chromosome paints, the PCR reactions were prepared, consisting of 0.2 U/µL KOD XL Polymerase (Novagen), 1x KOD XL buffer (Novagen), 0.2 mM dATP, dCTP, and dGTP, 0.075 mM dTTP, 22.5 µM FITC-Fluorescein-12-UTP (Roche), 4 µM DOP primer, amplified chromosomal DNA (10 ng), and MBG H_2O to a total volume of 100 µL. The following cycle was run for primary and secondary DNA amplification (Table 2.2):

Step	Temperature (ºC)	Duration	Cycles
Denaturation	94	30 seconds	35
Annealing	55	5 seconds	
Extension	72	1 minute	
Final Extension	74	10 minutes	1

Table 2.2: Summary of the PCR cycling conditions for the generation of FITC chromosome paints using DOP-PCR.

For Texas Red labelled chromosome paints, the PCR reactions were prepared, consisting of 0.02 U/µL KOD HS Polymerase (Novagen), 1x KOD HS buffer (Novagen), 1 mM $MgSO_4$, 0.2 mM dATP, dCTP, and dGTP, 0.05 mM dTTP, 7.5 µM Texas Red-12-dUTP (Invitrogen), 4 µM DOP primer, 6 µL amplified chromosomal DNA (10 ng), and MBG H_2O to a total volume of 100 µL. The following cycle was run for primary and secondary DNA amplification (Table 2.3):

Step	Temperature (ºC)	Duration	Cycles
Activation	94	2 minutes	1
Denaturation	94	15 seconds	35
Annealing	55	30 seconds	
Extension	72	1 minute	

Table 2.3: Summary of the PCR cycling conditions for the generation of Texas Red chromosome paints using DOP-PCR.

2.4 Genome Mapping

The following methodologies were performed at the Royal Veterinary College, London (RVC) by the Larkin lab. These approaches were used to generate predicted chromosome fragments (PCFs) which were then mapped via FISH in our lab. This joint methodology devised at the University of Kent and the RVC London consists of 4 main steps: (1) constructing PCFs from scaffold-based assemblies using the RACA algorithm to align raw sequencing read data (RVC); (2) computational and PCR verification of the generated PCFs (RVC); (3) using the previous verification set to develop a refined set of PCFs; (4) anchoring PCFs to chromosomes with a set of BACs designed to hybridise in phylogenetically divergent species using fluorescence *in situ* hybridisation (zoo-FISH). Outlined below are detailed steps performed at the RVC London.

2.4.1 Generation of PCFs using the RACA Algorithm (RVC)

The RACA (Kim *et al*., 2013) assembly was generated for the budgerigar from fragmented Illumina assemblies previously published (Zhang *et al.,* 2014b). The zebra finch genome was used as the reference genome as it was more closely related (divergence of 81 million years), and the chicken genome as the outgroup (divergence of 98 million years). The initial assembly generated 84 PCFs with an N50 of 46.54 Mbp using default RACA parameters, representing 96.29% of the original assembly and using 254 scaffolds. Due to insufficient read or comparative evidence to support their structures, 31.5% of the scaffolds were split by RACA.

2.4.2 Verification of PCFs (RVC)

As aforementioned, the default RACA parameters resulted in split regions of the target genome scaffolds and thus verification of these regions was required. Colleagues at the RVC London used PCR across the split regions less than 6 kbp in the target genome, representing 28% of all split scaffolds in the budgerigar assembly. Of these, 20 regions (46%) yielded positive PCR results with amplicons of the expected length. For the split regions that yielded negative PCR results, an alternative RACA-suggested order of the flanking syntenic fragments (SFs) was tested. Of these, 11 (25.58%) regions yielded positive PCR results with amplicons of the expected length, confirming the original scaffolds were chimeric in nature as indicated by RACA. To determine which of the remaining 111 split regions greater than 6 kbp were chimeric, they empirically determined the genome-wide minimum physical coverage level in the SF joining regions for which the PCR results were most consistent with the RACA

predictions; a physical coverage of 216x was estimated to produce the highest agreement between scaffolds and PCR results.

2.4.3 Creation of an Improved Set of Budgerigar PCFs (RVC)

An improved set of PCFs were constructed by adjusting the physical coverage thresholds, including scaffolds with the structures confirmed by PCR as additional inputs. Consequently, the number of PCFs increased to 95, the N50 decreased to 37.96 Mbp, and the number of chimeric fractions decreased to 21%.

2.5 Fluorescence *in situ* Hybridisation (FISH)

2.5.1 BAC Probe Mixture

For dual colour FISH, combinations of Texas red and FITC labelled probes, FITC and Aqua labelled probes or Texas red and Aqua labelled probes were used. The probe mixtures were made to a final volume of 10 µL, consisting of 1.5 µg chicken Hybloc (Applied Genetics Laboratories), 1.5 µL each of Texas red/FITC/Aqua labelled probes, and 5.5 µL Hybridisation solution I (Cytocell Ltd).

2.5.2 Chromosome Paint Mixture

For single colour FISH, chromosome paint mixture consisted of 1.5 µg chicken Hybloc (Applied Genetics Laboratories), 1.5 µL Texas red/FITC labelled paints, and 7 µL Hybridisation solution E (Cytocell Ltd).

2.5.3 Standard Slide Preparation

Chromosome suspensions derived from fibroblasts were centrifuged at 1,000 RPM for 10 minutes. The supernatant was discarded and the pellet resuspended in 0.5 mL fixative. 10 µL chromosome suspension was pipetted onto slides and

fixed with 10 µL fixative. Once dry, slides were washed in a dehydration series as follows: 2xSSC (saline-sodium citrate) (Thermofisher Scientific), 70% ethanol, 85% ethanol, and 100% ethanol for two minutes per solution at room temperature.

10 µL probe mixture was pipetted onto 22x22 mm coverslips, which were inverted onto the slides and sealed with rubber cement (Fixogum). Slides were heated for 2 minutes on a Hybrite hotplate at 37ºC, and then both probe and template DNA denatured at 75ºC for 2 minutes. The slides were incubated for 24 hours (same species) or 72 hours (cross-species) in a humidified chamber at 37ºC.

2.5.4 Octochrome Slide Preparation

Octochrome slides differ from standard microscope slides in that they have 8 labelled boxes per slide. These slides require a template slide for the chromosome suspension and an Octochrome device which has 8 unique probe mixtures on each square (Figure 2.1).

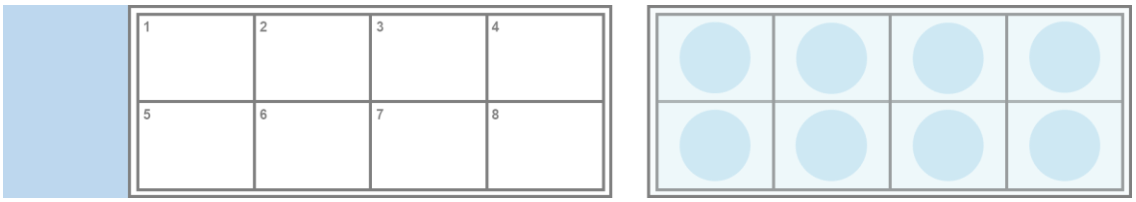


Figure 2.1: Schematic representation of an Octochrome template slide (left) and Octochrome device (right).

Chromosome suspensions derived from fibroblasts were centrifuged at 1,000 RPM for 10 minutes. The supernatant was discarded and the pellet resuspended

in 0.5 mL fixative. 4 µL chromosome suspension was pipetted on each square and fixed with 4 µL fixative. Once dry, slides were washed in a dehydration series as follows: 2xSSC (Thermofisher Scientific), 70% ethanol, 85% ethanol, and 100% ethanol for two minutes per solution at room temperature.

4 µL probe mixture was pipetted onto each square of the device, and the template slide then inverted onto the device. Slides were heated for 10 minutes on a Hybrite hotplate at 37ºC, and then both probe and template DNA denatured at 75ºC for 5 minutes. The slides were incubated for 24 hours (same species) or 72 hours (cross-species) in a humidified chamber at 37ºC.

2.5.5 Multiprobe Slide Preparation

Multiprobe slides differ from standard microscope slides in that they have 24 labelled boxes per slide. These slides require a template slide for the chromosome suspension and a Multiprobe device which has 24 unique probe mixtures on each square (Figure 2.2).

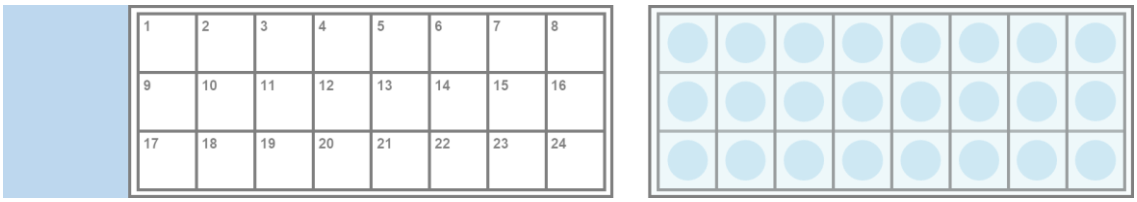


Figure 2.2: **Schematic representation of a Multiprobe template slide (left) and Multiprobe device (right).**

Chromosome suspensions derived from fibroblasts were centrifuged at 1,000 RPM for 10 minutes. The supernatant was discarded and the pellet resuspended in 0.5 mL fixative. 2 µL chromosome suspension was pipetted on each individual square and fixed with 2 µL fixative. Once dry, slides were washed in a

dehydration series as follows: 2xSSC (Thermofisher Scientific), 70% ethanol, 85% ethanol, and 100% ethanol for two minutes per solution at room temperature.

2 µL probe mixture was pipetted onto each square of the device, and the template slide then inverted onto the device. Slides were heated for 10 minutes on a Hybrite hotplate at 37ºC, and then both probe and template DNA denatured at 75ºC for 5 minutes. The slides were incubated for 24 hours (same species) or 72 hours (cross-species) in a humidified chamber at 37ºC.

2.5.6 Second Day FISH

After incubation, the devices or rubber cement and coverslips were removed. Slides were washed in 0.4xSSC at 72ºC for 2 minutes without agitation, followed 2xSSC with 0.05% Tween-20 at room temperature for 30 seconds. For cross-species hybridisations, the 0.4xSSC wash was omitted. Once dry, slides were counterstained using VECTASHIELD anti-fade medium with 4′,6-diamidino-2-phenylindole (DAPI) (Vectorlab) and 22x50 mm or 24x60 mm coverslips were placed on the slide, which were then developed in the dark for 10 minutes.

2.5.7 Microscopy

Slides were visualised under an Olympus BX61 epifluorescence microscope. A cooled charge-coupled device (CCD) camera captured images with DAPI, Texas red, FITC, and Aqua filters. Metaphase chromosomes visualised with DAPI were artificially coloured, with chromosomes either being coloured as black with a white background or white with a black background. The choice of chromosome colour was dependent on the visibility of the probe, with black chromosomes being used for small or difficult to visualise probes The Texas red,

FITC, and Aqua filters were used to visualise the probe DNA, none of which were artificially coloured. Images were captured at x1000 magnification using SmartCapture3 software (Digital Scientific UK).

2.6 Chromosome Morphology

2.6.1 Karyotype Analysis

Chromosome suspensions derived from fibroblasts were centrifuged at 1,000 RPM for 10 minutes. The supernatant was discarded and the pellet resuspended in 0.5 mL fixative. 10 µL chromosome suspension was pipetted onto a microscope slide and fixed with 10 µL fixative. Once dry, DAPI (Vectorlab) mixed with propidium iodide (0.6 µg/mL) was added. Slides were developed in the dark for 10 minutes before being visualised.

Taking into consideration the nomenclature described by the International System for Standardized Avian Karyotypes (ISSAK) regarding chromosome size, 5 karyotype images were produced per species using SmartType3 software (Digital Scientific UK). In the case of the songbirds, the nomenclature describing chicken chromosome homology was used.

2.6.2 Ideogram Generation

Ideograms were created based on the karyotype images produced from section 2.6.1 using Microsoft Powerpoint. Banding patterns were replicated by visual interpretation, with measurements (where possible) being made for a degree of accuracy. The results were verified by comparing multiple karyotype images to account for any variance in banding between metaphase spreads.

2.6.3 FLpter Measurements

Fractional length relative to the p-terminus (FLpter) measurements were made on 15 to 40 chromosomes for each probe using ImageJ (version 1.51r, Rasband W., National Institutes of Health, USA) and an ImageJ FLpter plugin developed by Dr. Benjamin Skinner (Department of Pathology, University of Cambridge, UK). The FLpter value of the probe is the mean of all measurements taken and the order of the probes was determined by the values obtained.

2.6.4 FLpter Measurements on Ideograms

The position of the BACs on the ideogram was calculated by applying the FLpter value to the ideogrammatic chromosome axis relative to the p-arm of the chromosome. For visual clarity, the BACs are numbered in ascending order based on their position on the chicken chromosome, with number 1 being at the topmost position of the p-arm.

2.7 Generating Microchromosome Markers

2.7.1 Gene Selection

240 genes were identified in the newest chicken genome assembly (*Gallus_gallus*-5.0; GCA_000002315.3) that were not found in the previous build (*Gallus_gallus*-4.0;GCA_000002315.2), and a further 111 genes have newly assigned chromosomal placements. These genes were refined based on those already mapped, size (<5 kb), poor predictions, poor matches, and predicted chromosome position, resulting in 35 selected genes for testing.

2.7.2 Primer Design

Of the 35 genes, primers were designed for 7 using ThermoAlign software (Table 2.4) by Dr. Benjamin Skinner (Department of Pathology, University of

Cambridge, UK), and primers were produced by Eurofins Genomics. Full details of each primer are given in the appendix, Supplementary Tables S1 and S2.

Gene	Contig	Primer Pairs	Genomic Length (bp)	GC Content Range (%)	T_m Range
FUS_1	NT_464478.1	5	20847	47.85 – 54.65	61.73 – 62.97
TYK2	NT_469030.1	6	10277	52.83 – 59.09	60.01 – 62.60
IKZF4	NC_008465.3	6	12358	58.74 – 66.46	60.18 – 62.85
SKIV2L	NC_006113.4	6	12317	61.81 – 67.70	60.54 – 62.81
AKAP8L	NC_028739.1	5	10987	62.46 – 68.66	61.95 – 63.00
SMARCC2	NC_008465.3	5	8863	63.82 – 68.39	57.85 – 62.84
BAZ2A	NC_008465.3	5	12152	66.17 – 70.73	60.52 – 62.90

Table 2.4: Summary of the 7 selected avian genes for testing.

2.7.3 PCR Optimisation

Following the manufacturer's guidelines, PCR was carried out for *KOD* Hot Start Polymerase (Novagen) at a final concentration of 0.02 U/µL, using chicken genomic DNA (Zyagen) as a template. Amplification of genomic DNA was run for 40 cycles, and all PCR experiments had unchanged final concentrations of genomic DNA (100 ng), $MgSO_4$ (1.5 mM), *KOD* Hot Start Polymerase buffer (1x), primer (0.5 µM) and dNTPs (0.2 mM). Modified parameters included betaine (Sigma-Aldrich) concentration, dimethyl sulfoxide (DMSO) (Sigma-Aldrich), formamide (Sigma-Aldrich), and KOD HS polymerase (Novagen) (0.02-0.04 U/µL). Temperature gradients were applied for annealing steps (58-70^{o}C) in native conditions, and with varying concentrations of betaine (1-1.5 M), formamide (5-10%), DMSO (1-5%), and 1 x Q5 High GC Enhancer (New England Biolabs).

After optimal annealing temperatures and additive concentrations were deduced the following programs were run (Table 2.5):

Step	Temperature (ºC)	Duration	Cycles
Activation	95	2 minutes	1
Denaturation	95	20 seconds	40
Annealing	61.9 – 63	10 seconds	
Extension	70	1 minute	
Final Extension	75	10 minutes	1

Table 2.5: Summary of the optimal PCR cycling conditions using KOD HS Polymerase.

The optimal annealing temperatures for each gene varied, with a summary shown in Table 2.6:

Gene	Annealing Temperature (ºC)
FUS_1	61.9
TYK2	61.5
IKZF4	61.1
SKIV2L	62.8
AKAP8L	62.0
SMARCC2	61.9
BAZ2A	62.9

Table 2.6: Summary of the optimal annealing temperatures for each gene.

PCR products were run on a 1.5% agarose gel (see 2.2.5 Gel Preparation) with a 1 kb (Promega) ladder and 100 bp ladder (Promega) to screen for size.

2.7.4 PCR Purification

After amplification, PCR products were purified using the QIAquick PCR Purification kit (Qiagen) following standard protocols.

2.7.5 Gel Extraction

For products that could not be optimised to yield a single band, gel extraction was performed using the QIAquick Gel Extraction kit (Qiagen) following standard protocols.

2.7.6 Product (A)-Tailing

Poly (A)-tailing was undertaken for all PCR products using GoTaq Flexi DNA Polymerase (Promega). A final reaction of 10 µL was made, consisting of 3.5 µL purified PCR product, 1X GoTaq reaction buffer (Promega), 0.2 mM dATP, 0.5U/µL GoTaq Flexi DNA polymerase (Promega), 1.5 mM MgCl2, and 0.9 µL MBG H_2O. The reaction mixture was incubated at 70°C for 30 minutes in the thermocycler.

2.7.7 T/A Ligation

T/A ligation was performed using T4 DNA ligase (New England Biolabs) and pGEM T-Easy vector (New England Biolabs) using a 3:1 (insert:vector) ratio. Ligation mixtures consisted of 1X T4 DNA ligase buffer (New England Biolabs), 50 ng vector DNA, appropriate mass of insert DNA, 0.5 µL T4 DNA ligase (New England Biolabs), made to a final volume of 10 µL with MBG H_2O. Ligation mixtures were incubated overnight at 4^{o}C before heat inactivation at 70^{o}C for 10 minutes.

2.7.8 Restriction Enzyme Digestion

Amplicon sequences were checked using Restriction Mapper (version 3) to ensure there were no digestion sites in the sequence in order to select a suitable restriction enzyme. Digestion using SphI (New England Biolabs) was undertaken to determine the correct insert size for FUS_1 ligated products, and

EcoRI for IKZF4. For SphI, the reaction mixture consisted of 0.2 µg ligated DNA, 1X NEBuffer 2.1 (New England Biolabs), 0.2 µL SphI (1U) (New England Biolabs), and made to a final volume of 10 µL with MBG H_2O. The reaction mixture was incubated at 37ºC for 1 hour. For EcoRI, the reaction mixture consisted of 0.2 µg ligated DNA, 1X NEBuffer EcoRI (New England Biolabs), 0.2 µL EcoRI (4U) (New England Biolabs), and made to a final volume of 10 µL with MBG H_2O. The reaction mixture was incubated at 37ºC for 1 hour. Digested products were run on a 1.5% agarose gel.

2.7.9 Bacterial T7 Transformation

High efficiency T7 express competent *E. coli* cells (New England Biolabs) were used for bacterial transformation of ligated products following standard protocols (C2566). Untransformed cells were spread on LB agar plates (ampicillin$^-$) to determine the viability beforehand. Transformed cells were spread on LB agar plates (ampicillin$^-$) to determine viability of the cells after transformation and on ampicillin$^+$ LB agar plates (100 µg/mL) at 10^0, 10^{-1}, and 10^{-2} dilutions to determine the success of the transformation.

2.7.10 Screening Colonies

Individual colonies were picked and plated on LB agar plates (ampicillin$^+$), and then screened following the protocol outlined in section 2.7.3, replacing genomic DNA with plasmid DNA and run for either 25 or 40 cycles.

3 Specific Aim 1: To upgrade the scaffold-based budgerigar (*Psittaciformes*) genome assembly, which is known to have an atypical avian karyotype, to chromosome-level using bioinformatics (in collaboration with the Royal Veterinary College, London) and molecular cytogenetics (FISH).

3.1 Background

DNA sequencing has been applied to many fields, such as medicine, agriculture, and genetic engineering, and has been revolutionary in advancing the field of science. With the rise of NGS, there has been a significant reduction in the cost and time required for sequencing large genomes, and thus the number of available *de novo* genome sequences are increasing exponentially. This increase has spurred many genome sequencing projects, from ambitious projects (Bird 10,000 Genomes) to idealistic projects (Earth Biogenome Project).

The ultimate aim of any genome sequencing project is to produce a single contig for each chromosome from the p- to q- terminus, referred to as a 'chromosome-level' assembly. For genome sequences that have relied on short read NGS technology, these can prove to be problematic when assembling due to the length of the reads, polymerase errors, large data sets, and repetitive structures. Ultimately, this fails to provide a contiguous *de novo* assembly which limits the ability to fully characterise and compare genomes. With improved assembly methods and sequencing technologies, such as optical mapping (Neely, Deen, and Hofkens, 2011), BioNano (Mak *et al.,* 2016), Dovetail (Putnam *et al*., 2016), PacBio (Rhoads and Au, 2015), and Nanopore (Venkatesan and Bashir, 2011), these problems are partially resolved with longer DNA reads, greater read depth, and assemblies with both fewer and longer contigs per genome (Koepfli *et al.,*

2015; Gordon *et al.,* 2016). Yet, these long read assemblies provide problems of their own: BioNano contigs fail to map across large heterochromatin blocks, centromeres, and multiple DNA nick site regions; and PacBio requires high molecular weight DNA in the region of hundreds of micrograms, restricting its use to species where this is obtainable. Moreover, PacBio and Nanopore sequencing have an error rate of 5-20% (Weirather *et al.,* 2017; Jain *et al.*, 2018), requiring further work for correction, which results in the need for a higher coverage compared to short read sequencing.

This demonstrates that it is not sufficient to construct accurate and contiguous *de novo* genome assemblies through the use of either short or long read sequencing technology alone. A combination of technologies is required to achieve a chromosome-level assembly in addition to the sequencing method, such as linkage mapping, Hi-C (Lieberman-Aiden *et al.,* 2009), pre-existing chromosome-level reference assemblies, and/or molecular cytogenetics (Larkin *et al.,* 2012; Damas *et al.,* 2017). Through the application of a synteny based bioinformatics approach, such as RACA (Kim *et al*., 2013), sub-chromosome sized predicted chromosome fragments (PCFs) can be generated for a *de novo* NGS genome. RACA has its limitations in that a closely related reference species is needed for comparison, and size limitations also apply, with FISH mapping of PCFs to chromosomes being essential. This integrated approach allows for the mapping *de novo* assembled genomes onto chromosomes and the subsequent visualisation on interactive browsers (e.g. Evolution Highway; UCSC) for chromosome-level comparison.

Chromosome-level assemblies are necessary for addressing biological questions, especially those regarding overall karyotype evolution. An assembly

at this level allows for genotype to phenotype associations to be identified through the means of an established order of DNA markers, which can then be applied for marker-assisted selection for species regularly bred for companionship or conservation purposes (Andersson and Georges, 2004). For agricultural animals (cattle, chicken, pig, sheep), chromosome-level assemblies have been rapidly established (Hillier *et al.,* 2004; Elsik, Tellam, and Worley 2009; Groenen *et al.,* 2012; Jiang *et al.,* 2014; Warren *et al.,* 2017) with early versions sequenced and assembled by Sanger sequencing, and to date are very well annotated. For many other species, the genomes are poorly represented as they were initially assembled using short read NGS data alone. By generating a chromosome-level assembly, we can further elucidate how chromosomal changes fixed during evolution give rise to speciation (White, 1969; Rieseberg, 2001; de Villena, 2001; Lewin, Larkin, and O'Brien, 2009; Dobigny, Britton-Davidian and Robinson, 2017).

The aim of this chapter was to use a previously established, inexpensive and integrated approach to upgrade an existing scaffold-based genome assembly to that of a chromosome-level for the common budgerigar (*Melopsittacus undulatus* - MUN). The budgerigar was selected for many reasons, the predominant one being that it has a highly rearranged karyotype with multiple fusions (2*n*=62). Previous studies have identified the degree of chromosome rearrangements through fusions using chicken macro- and microchromosome paints (Nanda *et al.,* 2007; Lithgow *et al.,* 2014). Moreover, the budgerigar, a member of the *Psittaciformes* (parrots), is a popular companion animal globally and is a key model for vocal learning studies (Striedter, 1994; Gahr, 2000; Brauth *et al.,* 2002; Webb and Zhang, 2004)

3.2 Specific Aims

The purpose of this chapter was to confirm the placement of PCFs to chromosomes, defining the order and orientation using a previously established panel of BAC clones designed to hybridise in avian species that are phylogenetically distant. The specific aims of this chapter were:

- Specific aim 1a: Apply a conserved panel of BACs to budgerigar (*Melopsittacus undulates*) chromosomes to upgrade an existing scaffold-based genome assembly to that of a chromosome-level
- Specific aim 1b: To map the chromosomal rearrangements between the chicken and budgerigar genome that gave rise to lineage-specific traits
- Specific aim 1c: To provide the raw FISH data from mapping PCFs to the RVC for uploading to the Evolution Highway genome browser

3.3 Materials and Methods

3.3.1 Upgrading Genome Assemblies with the RACA Algorithm

All methodologies described in section 2.4 were conducted at the RVC London, with the subsequent data used for the FISH mapping by our lab. The Larkin lab generated PCFs for the budgerigar genome (Ganapathy *et al.,* 2014) through multispecies RACA alignments, as described in section 2.4.1. The physical mapping of the budgerigar genome using FISH was guided by the PCFs produced. To determine the validity of the RACA algorithm, BACs within the same PCF were mapped with those predicted to be within the same PCF. After all PCFs were verified and confirmed, each BAC was systematically tested with those from other PCFs for the assembly of a preliminary genomic structure, and to also identify the orientation of the PCFs.

3.3.2 Culturing Budgerigar Fibroblasts

Fibroblast cell lines were established by Dr. Rebecca O'Connor (University of Kent) from collagenase treatment of tracheal dissections to generate metaphase chromosomes. The fibroblasts were cultured following protocols outlined in section 2.1. Sampling of avian tissues was reviewed and approved by the Animal Welfare and Ethics Review Board (AWERB) at the University of Kent.

3.3.3 BAC Generation and FISH

The labelling of FISH probes was performed as described in section 2.2. FISH was performed as described in section 2.5.

3.4 Results

3.4.1 Specific aim 1a: Apply a conserved panel of BACs to budgerigar (*Melopsittacus undulates*) chromosomes in order to upgrade an existing scaffold-based genome assembly to that of a chromosome-level

3.4.1.1 Assigning Predicted Chromosome Fragments to Chromosomes

PCFs were generated for budgerigar whole-genome sequences using RACA, using the zebra finch chromosome assembly as a reference and the chicken chromosome assembly as the outgroup. 84 PCFs were generated, with ~31% of the initial PCF sets contained chimeric scaffolds due to insufficient comparative evidence and/or reads to support their structures (Table 3.1).

Assembly Statistics	
Pair-end Read Physical Coverage Within Tested Scaffolds	0-631
Split SF Adjacencies by RACA (default parameters)	154
Tested Scaffold Split Regions	43 (100%)
Amplified Split Regions (confirmed SF joints)	20 (46%)
Non-amplified Split Regions	23 (54%)
Tested RACA-suggested Adjacencies	18
Amplified Adjacencies (chimeric SF joints)	11
Final No. Ambiguous SF Joints from Tested Split Regions	12
Selected Pair-end Read Spanning Threshold	216

Table 3.1: Statistics for the scaffold split regions tested by PCR.

Using adjusted physical coverage thresholds, colleagues at the RVC London generated a new set of PCFs, with the PCF structures confirmed via PCR. As a result, there was an increase in PCFs from 84 to 95, a lower portion of chimeric scaffolds, and a reduction in the N50.

3.4.1.2 Upgrading a scaffold-based genome assembly to a chromosome-level assembly

For the PCFs to be assigned to the correct place within the genome, 119 conserved BAC clones identified within PCFs were successfully hybridised to budgerigar chromosomes. The complete list of BACs and their coordinates in the budgerigar genome is given in the appendix, Supplementary Table S3. Chromosome homology between the budgerigar and chicken was established for all sequenced chromosomes, with the exception of chromosome 16; chromosome 16 contains clusters of major histocompatibility complex (MHC) genes (Miller *et al.,* 2013) which varies immensely between species and especially in Passerine birds (Ekblom *et al.,* 2011; Chen *et al.,* 2015).

The budgerigar (2*n*=62) deviated from the “standard” avian karyotype (2*n*=~80), with FISH mapping (as seen in Figure 3.1) improving the assembly of 21 pairs of autosomes and the Z chromosome with a four-fold improvement on the scaffold N50 from 11 Mbp to 38 Mbp.

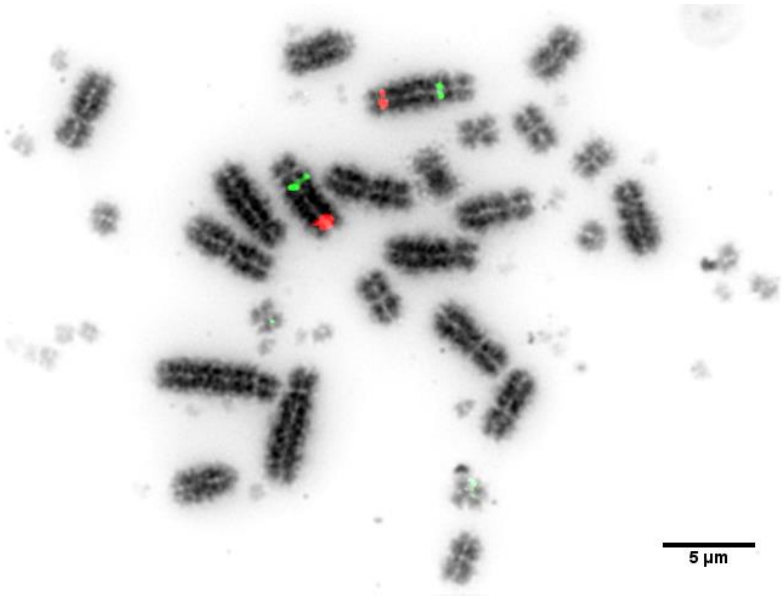


Figure 3.1: BAC clones hybridised to budgerigar chromosome 2. The FITC (green) labelled signal represents TGMCBA-375I5 (chicken 17 homolog), and maps to PCF 17. The Texas red labelled signal represents CH261-169K18 (chicken 3 homolog), and maps to PCF 3c_5a.

Of the 95 scaffolds, 46 were placed to represent 1.01 Gb of the genome (93.56% of the combined scaffold length), and 844.43 Mb (77.93%) were fully oriented on the chromosome (Table 3.2). The scaffolds that remained unoriented were due a number of factors, such as the PCF length being too small, an insufficient number of BACs to map the full length of the PCF, or the distance between BACs was too small and thus the hybridisations failed or the result was ambiguous.

Original Assembly	
Genome Size (Mbp)	1.12
Scaffolds longer 10 Kbp	1,138
Total Length (Gbp)	1.08
N50 (Mbp)	11.41
Default RACA Assembly	
PCFs Generated	84
Total length (Gbp)	1.04
N50 (Mbp)	46.54
Chimeric scaffolds	80 (31%)
Number of Scaffolds Used	254
Percentage of Original Assembly	96.29
RACA and PCR Assembly	
PCFs Generated	95
Total Length (Gbp)	1.04
N50 (Mbp)	37.96
Chimeric scaffolds	55 (21%)
Number of Scaffolds Used	254
Percentage of Original Assembly	96.29
RACA and FISH Assembly	
Successfully Hybridised BAC Clones	119
PCFs placed	46
PCFs oriented	28
Disagreements between RACA and FISH	4
Length Placed (Gbp)	1.01
Length Oriented (Mbp)	844.43
Original Assembly Placed (%)	93.56
Original Assembly Oriented (%)	77.93

Table 3.2: Assembly statistics from the original NGS genome, the default RACA assembly, the RACA and PCR assembly, and the combined RACA and FISH assembly.

3.4.1.3 Determining RACA algorithm accuracy for predicting PCFs

In testing the validity of the PCFs generated by RACA, another purpose of these experiments was to determine the accuracy of PCF predictions in terms of chromosome location, scaffold orientation, and the correct joining of scaffolds. Some of the PCFs generated for the budgerigar contained errors which were only detected via FISH. This was due to chimeric scaffolds in the original assembly, resulting in 4 occasions in which PCF predictions were incorrect. These errors can be seen in Figure 3.2, where PCFs had to be broken and rejoined in different orientations or to different scaffolds.

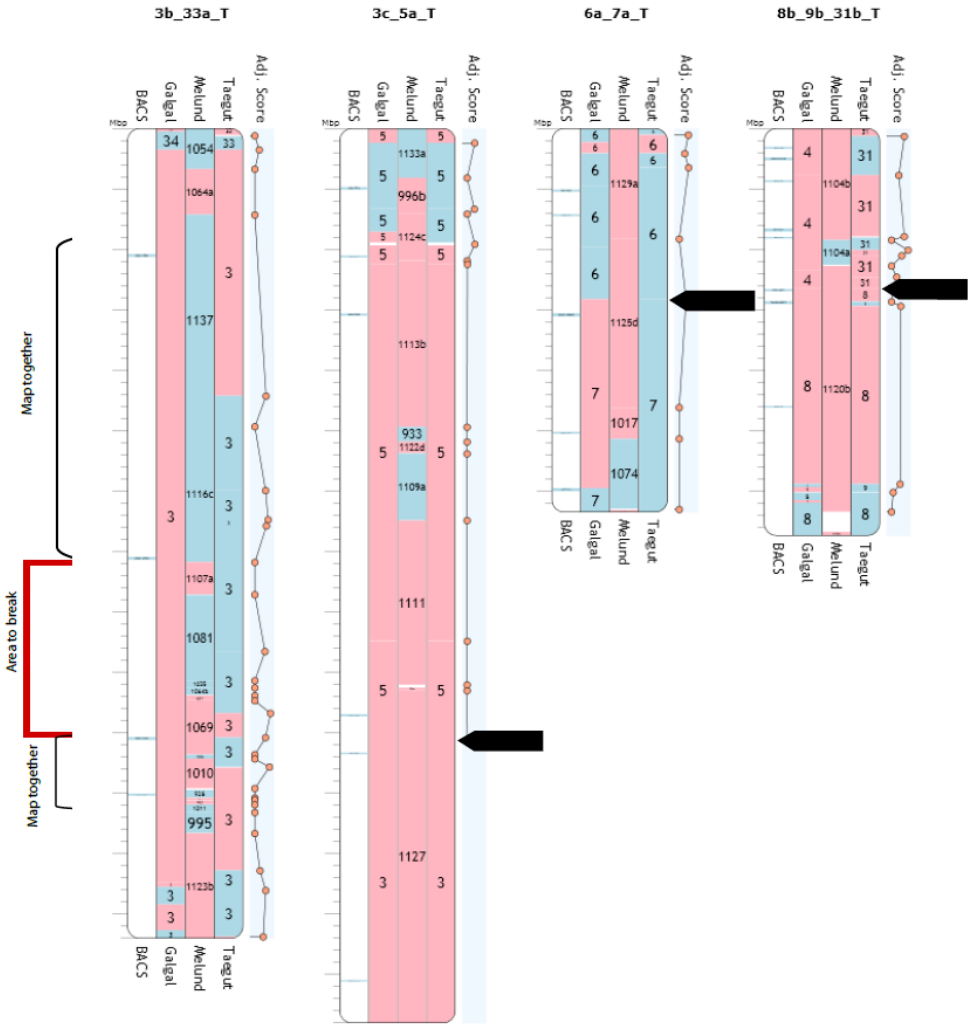


Figure 3.2: Errors in PCF predictions for budgerigar chromosomes 3, 6, and 8. PCF 3b_33a_T is annotated to show the regions where the PCF had to be split and reassembled due to incorrect scaffold joining. Black arrows denote where the other PCFs were split. BACs, PCFs, scaffolds, and both the chicken and zebra finch homologies shown alongside budgerigar chromosomes.

These errors were as a result of the over aggressive joining of scaffolds from the original assembly, generating PCFs that required splitting, which demonstrates the importance of having a physical map to verify any *in silico* genome assembly. Most assemblers and algorithms typically rely on reference genomes and multiple alignments. This can result in some syntenic regions remaining fragmented or unplaced, and the use of a reference genome introduces bias to the target genome (Ghurye and Pop, 2019). Moreover, the stages of generating scaffolds (e.g. Hi-C, paired end reads) introduces assembly errors, with aggressive scaffolding approaches being a common cause of contig and/or scaffold misjoin in several assemblers (Salzberg *et al.,* 2012). The combination of the target genome bias and mis-assembly errors can be significant, and these errors will often remain undetected until other studies or physical maps dispute the assembly.

3.4.2 Specific aim 1b: To map the chromosomal rearrangements between the chicken and budgerigar genome that gave rise to lineage-specific traits

A total 19 interchromosomal rearrangements were identified between budgerigar and chicken (Table 3.3). Three budgerigar chromosomes (MUN 4, 5, and 8) were identified as a fusion of 10 chicken homologs, and three budgerigar chromosomes (MUN 2, 9, and 10) revealed a fusion of two chicken homologs. Chicken homolog 1 demonstrated three fissions to form budgerigar chromosome 3 and 6, with no evidence of further fusion. Chicken homologs 5 and 7 split and fused to each other, with other chicken homologs, to form budgerigar chromosomes 4 and 8. Chicken homolog 4 exhibited a fission pattern

seen in most avian species, in which the p-arm of chicken chromosome 4 is a fused ancestral microchromosome.

Ancestral Chromosome (numbered according to chicken)	Budgerigar	
	Inter-	Intra-
1	Fission	1
2	-	2
3	Fusion to GGA 17	3
4a	-	0
4b	Fusion to GGA 9	0
5	Fission and Fusion to GGA 6	1
6	Fusion to GGA 5	0
7	Fission and fusion to GGA 6 and 5	1
8	Fusion to GGA 9	0
9	Fusion to GGA 8	1
10	Fusion to GGA 12	0
11	Fusion to GGA 4q	0
12	Fusion to GGA 10	1
13	Fusion to GGA 20	0
14	Fusion to GGA 5	1
15	-	2
16	No data	No data
17	Fusion to GGA 3	0
18	-	1
19	-	0
20	Fusion to GGA 13	0
21	-	0
22	-	0
23	-	2
24	-	0
25	No data	No data
26	-	0
27	-	0
28	-	0
Z	-	0

Table 3.3: Summary of rearrangements in the budgerigar genome using the chicken genome as a reference. The ancestral avian chromosome is represented in the left-hand column, with the subsequent columns indicating the number of inter- and intrachromosomal changes detected that have led to the evolution of the budgerigar. GGA = *Gallus gallus domesticus.*

An ideogram illustrating the overall genomic structure of the budgerigar, with chicken homologies, is shown in Figure 3.3.

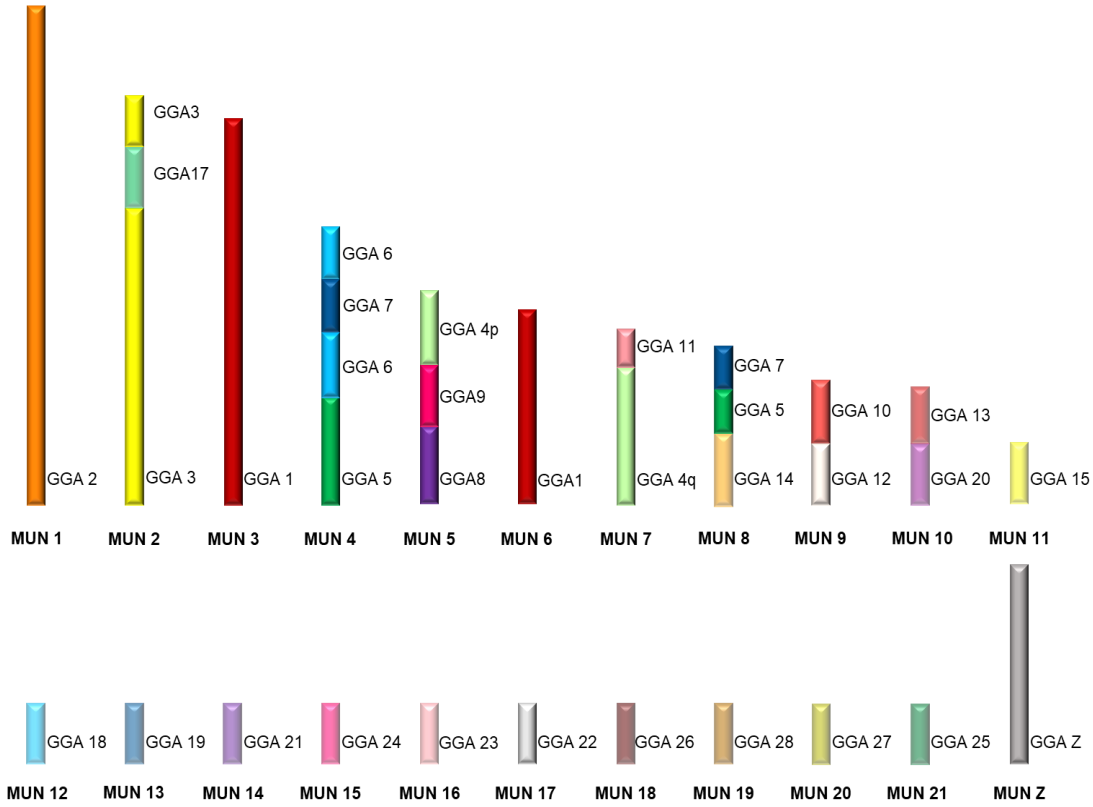


Figure 3.3: Ideogram of the gross genomic structure of the budgerigar (*Melopsittacus undulates*, MUN) with chicken homologies per chromosome. Each chicken (GGA) homolog is represented as a different colour (randomly assigned). Intrachromosomal differences are not shown and are given in the appendix, Supplementary Table S3.

Of the chicken microchromosome homologs, 11 remained as distinct microchomosomes and 7 fused to other chromosomes. In total, 16 intrachromosomal rearrangements were identified between the budgerigar genome compared to chicken, with 13 chicken homologs showing no evidence of fusions or fissions.

3.4.3 Specific aim 1c: To provide the raw FISH data from mapping PCFs to the RVC for uploading to the Evolution Highway genome browser

The newly assembled budgerigar genome has been uploaded as a reference genome to Evolution Highway, a comparative genome browser (http://eh-demo.ncsa.uiuc.edu/birds/) thanks to the work of the Larkin lab at the RVC London. Budgerigar chromosomes were numbered based on previous assignments (Nanda *et al.,* 2003), and where no previous assignment was documented, the chromosomes were assigned numbers based on decreasing PCF size. For the mapped chicken chromosomes (1-28, excluding 16, including Z), homologies were identified between the chicken and the budgerigar.

A screenshot representing Evolution highway is shown in Figure 3.4, showing chromosomes homologous to chicken chromosomes 1, 2, 3, and 4 (including scaffolds, BAC positions, and PCFs). The whole dataset is available on http://eh-demo.ncsa.uiuc.edu/birds/

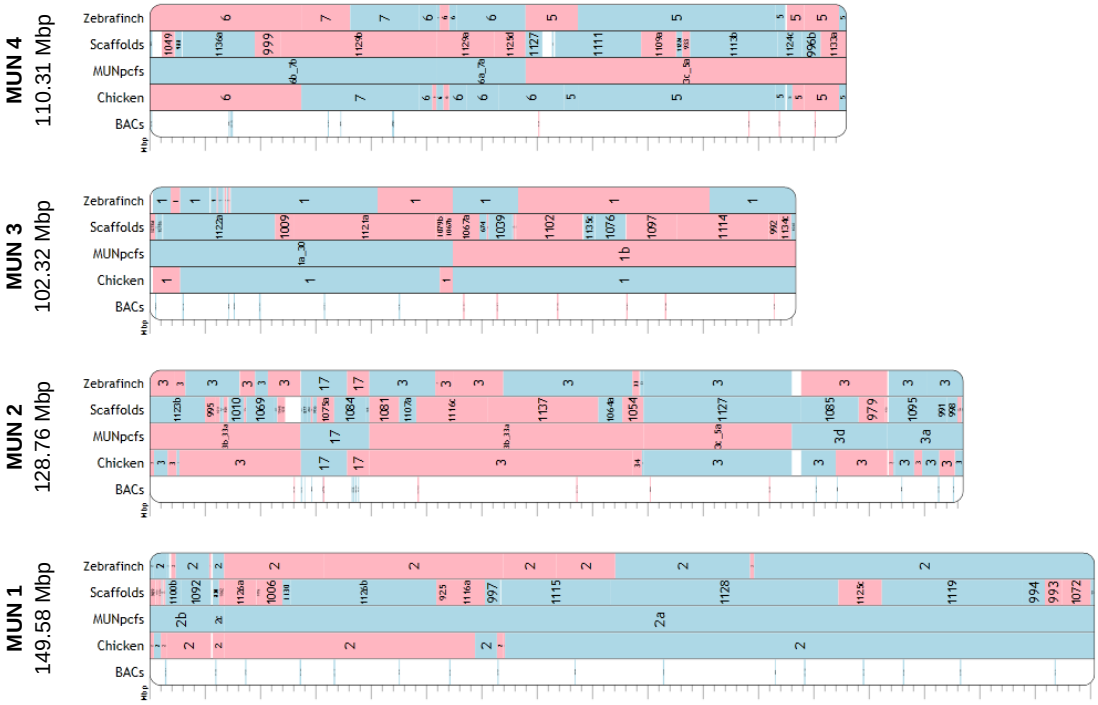


Figure 3.4: Chromosomes homologous to chicken chromosome 1,2, 3, and 4, with mapped BACs, PCFs, scaffolds, and zebra finch homologies shown. MUN (*Melopsittacus undulates*) 1, 2, 3, 4 = budgerigar chromosome 1, 2, 3 and 4.

3.5 Discussion

3.5.1 Generating a Chromosome-Level Genome Assembly

With the increasing number of newly sequenced genomes, tools are required to enable efficient and inexpensive chromosome-level assembly for the reasons previously mentioned. The integrated methodology established by the University of Kent and the RVC London has now generated 5 avian chromosome-level assemblies (pigeon, peregrine falcon, budgerigar, ostrich, saker falcon) for published but highly fragmented sequenced genomes (Damas *et al.,* 2017; O'Connor *et al.,* 2018a). Using this method has now generated assemblies with more than 80% of their genomes placed on chromosomes, upgrading them to an assembly comparable to genomes assembled using high-density physical or genetic mapping and Sanger sequencing (Lewin, Larkin, and O'Brien, 2009).

By identifying homologous synteny blocks and aligning the target genome to both a reference genome and an outgroup, RACA provides a means to generate larger scaffolds that can be physically mapped with FISH probes. The combination of RACA and FISH allows for any sequenced genome to be assembled regardless of the read length, providing a cheaper alternative to other approaches due to the ability to generate PCFs from a reference genome and existing read pair information only. However, RACA has its limitations in that it relies on a single reference genome which introduces bias to the assembly of the target genome. Moreover, the scaffolds generated are of a sub-chromosomal size and may have gaps of unresolved sequence between the original sequenced contigs.

3.5.2 Chromosomal Rearrangements

There have been very few studies of the karyotype structure among the *Psittaciformes*, and only one zoo-FISH study to characterise the overall genome structure (Nanda *et al.,* 2007). However, the inability to detect microchromosome rearrangements limited the results to only those involving the macrochromosomes. This chapter presents results that reveal undetected microchromosome rearrangements, demonstrating that the most common mechanism of interchromosomal rearrangement is fusion i.e. there was no evidence of reciprocal translocation. In some instances, predominantly in falcon genomes, many microchromosomes have fused together but have remained as discrete regions of conserved synteny, although now fused to larger chromosomes (Damas *et al.,* 2017; O'Connor *et al.,* 2018a). Another rearrangement revealed is a breakpoint in the chicken chromosome 1 homolog. This breakpoint occurs in the same region as the saker falcon (O'Connor *et al.,* 2018a) and the zebra finch genome (Itoh and Arnold, 2005), suggesting that this evolutionary breakpoint occurred in the Australavian ancestor and thus previously fixed in the three descendant lineages.

EBR reuse has been shown to play a role in regions of the genome being prone to chromosomal breakage (Stankiewicz and Lupski, 2002), with 'EBR genes' having a degree of correlation to biological features specific to individual lineages (Larkin *et al.,* 2009). Correlations have been identified between adaptive traits and EBRs in individual species, including forebrain development in the budgerigar. This is consistent with this species being a vocal-learning bird, but also defines it as a species with distinct neuronal connections compared to other vocal-learning species (Farré *et al.,* 2016). Moreover, in avian species,

such as the zebra finch, chicken, and turkey, EBR reuse has been suggested as a mechanism for the generation of recombination-based chromosome rearrangements (Völker *et al.,* 2010). A recent study by O'Connor and colleagues (O'Connor *et al.,* 2018a) studied conserved non-coding element (CNE) depletion, noting that CNEs are generally depleted in EBRs and are particularly depleted in interchromosomal rearrangements, especially fission. This corresponds with previous studies (Damas *et al.,* 2017), suggesting that CNEs in avian genomes play a role in defining where rearrangements (specifically interchromosomal ones) can be fixed in evolution without resulting in deleterious effects. Moreover, both studies demonstrated that chromosomal fissions are associated with genome intervals being fully depleted of CNEs.

Some avian lineages, such as the *Psittaciformes* and *Falconiformes*, demonstrate a high degree of interchromosomal rearrangement (Nanda *et al.,* 2006; Nanda *et al.,* 2007; Nishida *et al.,* 2008; Furo *et al*., 2018). This could suggest that the degree of rearrangement may be due to an exploitation of evolutionary niches, resulting in fixed interchromosomal rearrangements (Zhang *et al.,* 2014a). However, this fixation appears to be prevented in the majority of other avian species, allowing for the maintenance of a stable avian karyotype. In avian chromosomes, a large number of CNES (roughly twice as high compared to mammalian genomes) could be responsible for the formation of regulatory networks (Zhang *et al.,* 2014a; Farré *et al.,* 2016, O'Connor *et al.,* 2018a) that cannot be modified, which would explain the stability of the chromosomes.

3.5.3 The Budgerigar Genome

Results presented here between the budgerigar and the chicken genome demonstrated a total of 19 interchromosomal rearrangements and 16 intrachromosomal rearrangements. This suggests that despite the highly rearranged genome, the overall pattern of change is interchromosomal, and once fixed, changed relatively little intrachromosomally. Previous studies (O'Connor *et al.,* 2018b) have detected interchromosomal rearrangements in three *Psittaciformes,* the red-crowned parakeet (*Cyanoramphus novaezelandiae*), the cockatiel (*Nymphicus hollandicus*), and the budgerigar. In this study, homologs for chicken chromosomes 10, 11, and 14 exhibit a fusion. The budgerigar also demonstrated fusions of homologs for chicken chromosomes 12, 13, and 17, which were not seen in the other *Psittaciformes*. Given the budgerigar karyotype deviates from that of the standard avian pattern, it could indicate that the interchromosomal changes identified in this chapter are unique to the budgerigar lineage.

3.6 Conclusion

Through this integrated methodology, which combines comparative sequence analysis, targeted PCR, and FISH, this chapter provides further evidence that fragmented scaffold-level genomes can be upgraded to a chromosome-level assembly in a cost-effective and efficient manner. There was a four-fold improvement in the budgerigar N50, as well as the identification of previously undetected inter- and intrachromosomal rearrangements. 93.56% of the original scaffold assembly was assigned to chromosomes, which exceeds that of genomes previously assembled by traditional means. By assembling genomes to this level and as more genomes become available, it enables research into

chromosomal rearrangements and avian karyotype evolution as a much higher resolution, which could allow us to identify adaptive phenotypic traits in individual orders and families.

4 Specific Aim 2: To produce comparative cytogenomic maps for 7 avian species to investigate phylogenetic relationships and lineage-specific patterns arising from chromosomal rearrangements.

4.1 Background

Among terrestrial vertebrates, birds demonstrate incredible diversity and are the most species-rich Class, with approximately 10,900 extant representatives (Gill and Donsker, 2019). This diversity often means that birds are used as model organisms for phylogenetic and biological studies, such as virology (Alexander, 2000) and developmental biology (Nowicki, Searcy, and Peters, 2002). Additionally, birds are important both economically and environmentally (Maas *et al.,* 2015), and have shown to influence human behaviour (Bezerra *et al.,* 2013). Studying overall genome structure is an essential element to understanding avian biology, however, most avian species have no structural (karyotypic) data associated with their genome sequences (Kretschmer, Ferguson-Smith, and de Oliveira, 2018), despite ~460 avian genomes having been sequenced as of May 2019 (Stiller and Zhang, 2019; Genomes (B10K) project (https://b10k.genomics.cn/species.html)). In other words ~4% of avian species have sequence data associated with their genomes, but, nonetheless only 16 genomes (0.05%) are assembled to a chromosome-level (i.e. a single contig for each chromosome from the p- to q- terminus (https://www.ncbi.nlm.nih.gov/assembly)).

To address this problem, and with minimal cost, classical cytogenetics can provide a cheap and quick method for initial characterisation of chromosomes, such as the diploid number and chromosome morphology. This can signify the degree of chromosomal rearrangement alone, especially for orders that deviate

from the “standard” of 2*n*=80, such the *Charadriiformes, Falconiformes,* and *Bucerotiformes* (Griffin *et al.,* 2007; Nishida *et al.,* 2008). These studies are however typically limited to the macrochromosomes due to the indistinguishable morphology and lack of banding patterns of the microchromosomes. Nonetheless, classical cytogenetics provides a base upon which all mapping data relies on for a chromosome number. However, following the guidelines from the International System for Standardized Avian Karyotypes (ISSAK) (Ladjali-Mohammedi *et al.,* 1999) for generating karyotypes, this can be misleading with regards to chromosome homology i.e. chromosome 1A commonly found in passerine birds can be classified as chromosome 5 (dos Santos *et al.,* 2017). Additionally, successfully identifying the smallest of the microchromosomes can be difficult, resulting in inaccurate and/or inconsistent diploid numbers.

Coupling classical cytogenetics with molecular cytogenetics (e.g. zoo-FISH) provides a finer resolution of genomic structure and can be used to determine chromosome homology in addition to chromosomal rearrangements. Chromosome painting allows for the identification of homologous chromosomes (or homologous blocks) between species. However, due to the nature of chromosome paints, intrachromosomal rearrangements such as inversions and duplications, cannot be identified. These limitations can be circumvented through the use of BACs, providing a finer resolution to detect small rearrangements, but also reducing the success of hybridisation due a reduction in length of homologous sequences (Damas *et al.,* 2017). Through the use of a universal BAC probe set developed by Damas *et al. (*2017), these rearrangements can be mapped quantitatively by the fractional length relative to the p terminus (FLpter) value (Lichter *et al.,* 1990; Morris *et al.,* 2007; Mota-Velasco *et al.,* 2010). Using a reference genome with known BAC order for

comparison, the mapping of BACs can be used to track chromosomal rearrangements within different species, which provides an inexpensive way to partially characterise avian genomes without sequencing data. This data can also be used to generate comparative maps which can lay the foundations for other studies, such as upgrading assemblies to that of a chromosome-level (Damas *et al.,* 2017; O'Connor *et al.,* 2018a), as well as providing a means for easy visual comparison.

In this study, 7 avian genomes were studied, representing 6 of the 32 neognath orders, to further develop our understanding of avian genomes and the variation that leads to phenotypic and behavioural diversity.

The first species selected were the common blackbird (*Turdus merula*, TME) and Atlantic canary (*Serinus canaria*, SCD) as representatives of the Australave clade and the order *Passeriformes*, which constitutes over half of all avian species (Ricklefs, 2012). These are song learning birds that are used for studies pertaining to brain development (reviewed by Nowicki, Searcy, and Peters, 2002; Olson and Mello, 2010). To date, an assembled and annotated canary genome is available (Frankl-Vilches *et al.,* 2015). However, this genome is not assembled to a chromosome-level and there is no physical map to detect any mis-assemblies.

The next selected species, are the Eurasian woodcock (*Scolopax rusticola*, SRU), helmeted guinea fowl (*Numida Meleagris*, NME), houbara bustard (*Chlamydotis undulata*, CUN), and the rock dove or pigeon (*Columba livia*, CLI). The Eurasian woodcock is a wading bird known for its 360-degree vision and recognised among game hunters due to its erratic flight patterns, speed, and

size (Hoodless, 1995; Duriez *et al.,* 2005; Braña, Prieto, and González-Quirós, 2010), in addition to having an atypical diploid number of 2*n*=96 (O'Connor *et al.,* 2018b); the helmeted guinea fowl provides a reference point for the *Galliformes* other than that of the chicken, as the chicken has ancestral rearrangements not commonly seen in other birds (Guttenbach *et al.,* 2003; O'Connor *et al.,* 2018b); the houbara bustard is significant for heritage and culture in Arabian countries, in addition to being listed as an IUCN vulnerable specie (BirdLife International, 2016); and the rock dove (pigeon) exhibits extreme phenotypic diversity not seen within other avian species (Vickrey *et al.,* 2018). Moreover, its genome has recently been upgraded to that of a chromosome-level (Damas *et al.,* 2017), and thus mapping this species will help in determining the validity of this method for identifying chromosomal rearrangements without sequencing data.

Finally, the mallard duck (*Anas platyrhynchos*) was selected from the Galloanserae clade. This is a well-studied bird, with much interest in the field of immunology (Jourdain *et al*. 2010) and also having a whole genome radiation hybrid panel (Rao *et al.,* 2012). Moreover, the mallard has a divergence time from the chicken genome of ~80 million years (http://www.timetree.org; Hedges, Dudley, and Kumar, 2006), which is shorter than most of the other species selected in this study, and may help to pinpoint when any chromosomal changes became fixed in the evolution of the Neoaves that deviate from that of the chicken or the avian ancestor.

The focus of this study was the macrochromosomes and tracing chromosome evolution through the mapping of genomic rearrangements between species. Using the chicken (*Gallus gallus*, GGA) genome as a reference, the mapping of

individual BAC clones to the macrochromosomes allows for the identification of fissions, fusions, duplications, and inversions, all of which contribute to the chromosomal changes that influence speciation. The chicken was selected as the reference genome as it is currently the best assembled and annotated genome of any avian species (Cheng and Burt, 2018), with an extensive panel of BACs available to map gross genomic rearrangement. The purpose of this study was to produce comparative cytogenetic maps for the macrochromosomes of 7 avian species to investigate phylogenetic relationships and lineage-specific traits arising from chromosomal rearrangements. Specifically, we generated new karyotypes and ideograms, applied a panel of 74 selected chicken BACs for the fine mapping of macrochromosomes 1-9 and Z, obtained FLpter values to generate comparative maps, and compared the BAC order to that already established in genome assemblies (chicken, rock dove) for validation of the methodology

4.2 Specific Aims

The purpose of this chapter was to produce comparative maps for 7 avian species to investigate phylogenetic relationships and lineage-specific traits arising from chromosomal rearrangements. With that in mind, the macrochromosomes of 7 different species were mapped using BACs. The specific aims of this chapter were:

- Specific aim 2a: Generate karyotypes and ideograms for 7 avian species
- Specific aim 2b: Apply a panel of 74 selected chicken BACs for the fine mapping of macrochromosomes 1-9 and Z

- Specific aim 2c: Obtain FLpter values to generate comparative maps for the macrochromosomes
- Specific aim 2d: Compare the BAC order from the 7 species to that already established in genome assemblies (rock dove) for validation of the methodology

4.3 Materials and Methods

4.3.1 Culturing Fibroblasts

DF-1 chicken embryonic fibroblasts (CEFs) derived from East Lansing line eggs (passage 0) were obtained from local suppliers (Pirbright Institute). Cell cultures were harvested at passage 3 to generate metaphase chromosomes.

Fibroblast cell lines were established by Dr. Rebecca O'Connor (University of Kent) from collagenase treatment of skin biopsies and tracheal dissections to generate metaphase chromosome suspensions for the following species: the common blackbird, Atlantic canary, Eurasian woodcock, helmeted guinea fowl, houbara bustard, mallard duck, and rock dove. The fibroblasts were cultured following protocols outlined in section 2.1. Sampling of avian tissues was reviewed and approved by the Animal Welfare and Ethics Review Board (AWERB) at the University of Kent.

4.4 Results

4.4.1 Specific aim 2a: Generate karyotypes and ideograms for 7 avian species

Conventional analysis of metaphase chromosomes from all 7 avian species revealed diploid numbers ranging between 76-96 chromosomes. Table 4.1 summarises the chromosomal findings of each species studied.

Order	Common name	Species name	2*n*
Anseriformes	Mallard duck	*Anas platyrhynchos*	80[1]
Charadriiformes	Eurasian woodcock	*Scolopax rusticola*	96[2]
Columbiformes	Rock dove	*Columba livia*	80[3]
Galliformes	Helmeted Guinea fowl	*Numida meleagris*	78[4]
Otidiformes	Houbara bustard	*Chlamydotis undulata*	76[2]
Passeriformes	Common blackbird	*Turdus merula*	80[5]
Passeriformes	Atlantic canary	*Serinus canaria*	80[6]

Table 4.1: Summary of diploid number in the 7 avian species studied. [1]) Fillon *et al.,* 2007. [2]) O'Connor *et al.,* 2018b. [3]) Damas *et al.,* 2017. [4]) Shibusawa *et al.,* 2002. [5]) Hammar, 1970. [6]) dos Santos *et al.,* 2017.

Karyotypes were completed based on existing studies (Hammar, 1970; Shibusawa *et al.,* 2002; Fillon *et al.,* 2007; dos Santos *et al.,* 2017; Damas *et al.,* 2017; O'Connor *et al.,* 2018b). The houbara bustard had conflicting karyotype data in the literature, with either a diploid number of 76 (O'Connor *et al.,* 2018b) or 78 (Mahiddine-Aoudjit, Boucekkine, and Ladjali-Mohammedi, 2019). However, karyotypes performed for this study determined a diploid number of 76. For species where no literature was present, karyotypes were completed following ISSAK classifications (Ladjali-Mohammedi *et al.,* 1999). An example of the "standard" avian karyotype (2*n*=~80) is shown in Figure 4.1A and 4.1B, representing the chicken and the helmeted guinea fowl. Figures 4.1C and 4.1D,

representing the Atlantic canary and Eurasian woodcock, demonstrate different karyotypes that vary either in diploid number or deviate from the ISSAK classification of being ordered by size (chromosome 1 and 1A being ordered before chromosome 2, the largest chromosome in the Atlantic canary).

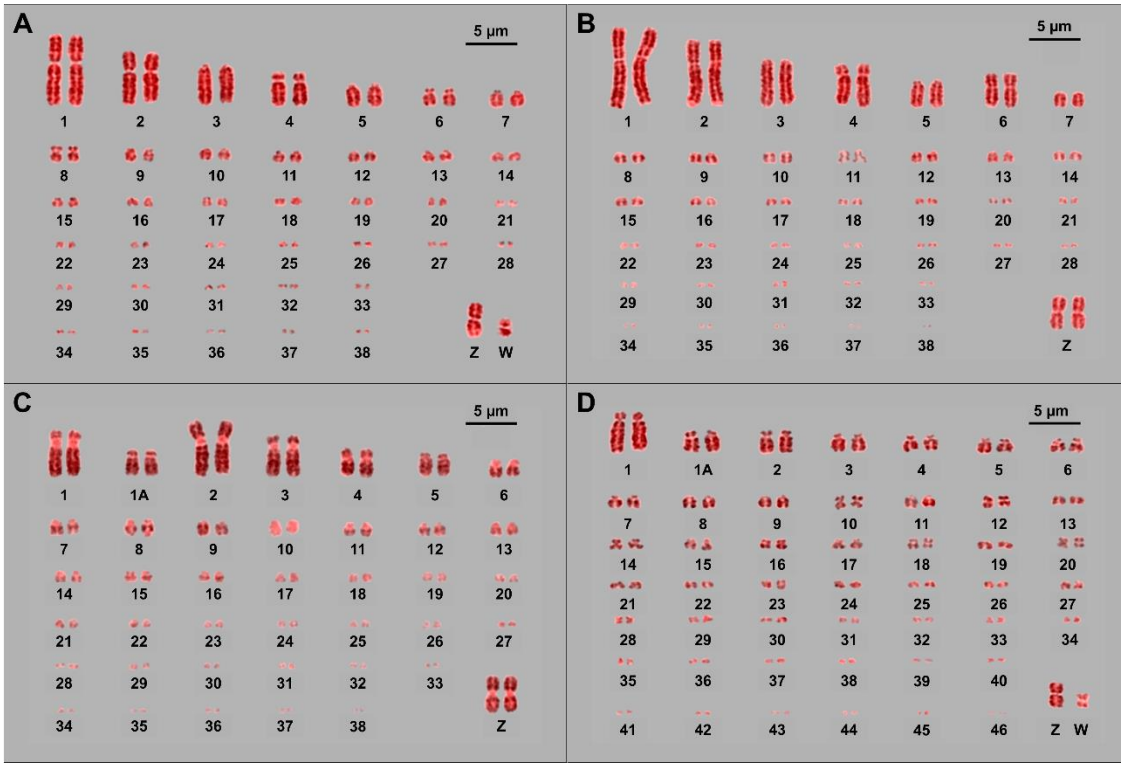


Figure 4.1: The variety of avian karyotypes observed in 4 of the 7 avian species. A) Chicken (*Gallus gallus*). B) Helmeted guinea fowl (*Numida meleagris*). C) Atlantic canary (*Serinus canaria*). D) Eurasian woodcock (*Scolopax rusticola*).

Using visual inspection and measurements of the chromosome arms, respective chromosome length, and width of bands, ideograms were generated from the karyotypes of the macrochromosomes (1-9, Z, and W). These ideograms (an example shown in Figure 4.2) allows for the inspection of differences in chromosome morphology and banding. For example, helmeted guinea fowl chromosomes in Figure 4.2B are more heavily banded than chicken chromosomes in Figure 4.2A, which may not have been apparent in the karyotype images.

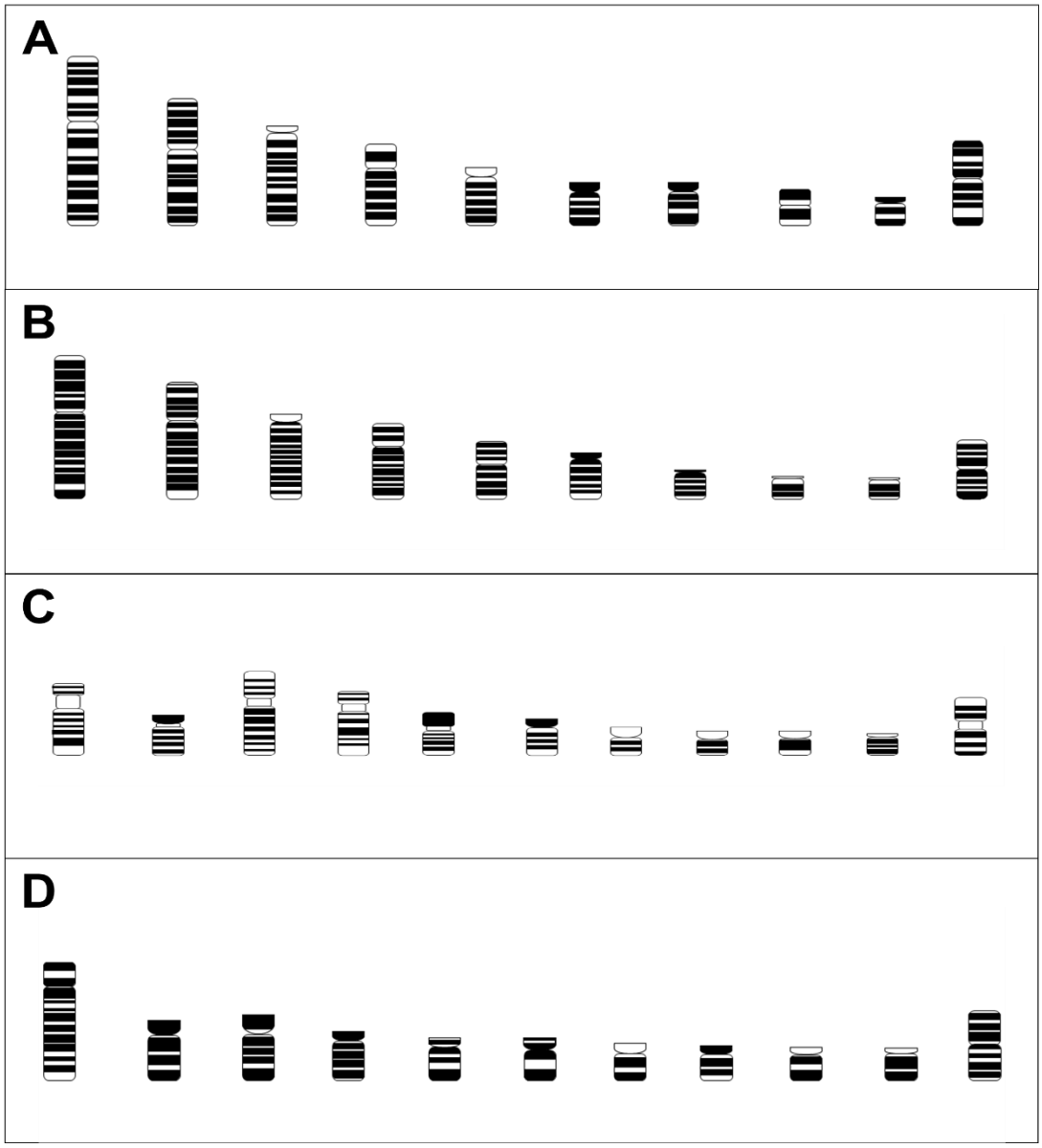


Figure 4.2: Ideograms of the macrochromosomes from 4 of the 7 avian species. A) Chicken (*Gallus gallus*). B) Helmeted guinea fowl (*Numida meleagris*). C) Atlantic canary (*Serinus canaria*). D) Eurasian woodcock (*Scolopax rusticola*).

4.4.2 Specific aim 2b: Apply a panel of 74 selected chicken BACS for the fine mapping of macrochromosomes 1-9 and Z

74 conserved BAC clones were selected based on work developed by Damas *et al.* (2017) for hybridisation to map the macrochromosomes, with the complete list of BACs and their coordinates in the chicken genome given in the appendix, Supplementary Table S4. The degree of successful hybridisations varied

between species, with an overall success rate for all 74 BACs given in Table 4.2.

Order	Common name	Success %	Divergence (mya)
Anseriformes	Mallard duck	85.14	80
Charadriiformes	Eurasian woodcock	72.97	98
Columbiformes	Rock dove	93.24	98
Galliformes	Helmeted Guinea fowl	100	47
Otidiformes	Houbara bustard	90.54	98
Passeriformes	Common blackbird	79.73	98
Passeriformes	Atlantic canary	72.97	98

Table 4.2: Percentage of successful BAC hybridisation in all 7 avian species tested.

In total, 38 of the 74 BACs successfully hybridised to metaphase chromosomes of all 7 species. A successful hybridisation was determined when there was clear signal on the chromosome (as seen in Figure 4.3), and any signal that was ambiguous (excessive non-specific hybridisation on chromosomes) were determined to be unsuccessful. The full table of BACs successfully hybridised is given in the appendix, Supplementary Table S5.

The helmeted guinea fowl, the species with the shortest evolutionary divergence of 47 million years (http://www.timetree.org; Hedges, Dudley, and Kumar, 2006) was the only bird to demonstrate successful hybridisation of all 74 BACs. The remaining species demonstrated a range of successful hybridisations despite similar divergence times. Thus, an overall correlation between evolutionary divergence and hybridisation success was not observed, as seen in Figure 4.4. To further analyse the correlation observed, a linear regression analysis revealed a correlation of 0.41 and a p-value of 0.1.

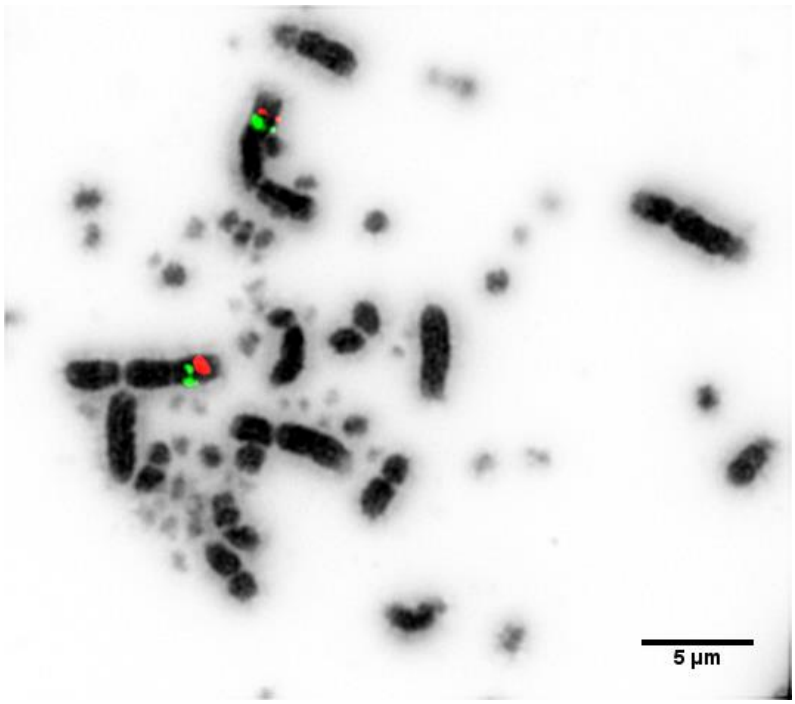


Figure 4.3: BAC clones hybridised to helmeted guinea fowl chromosome 1. The FITC (green) labelled signal represents CH261-107E2 (chicken 1 homolog), the Texas red labelled signal represents CH261-184E5 (chicken 1 homolog).

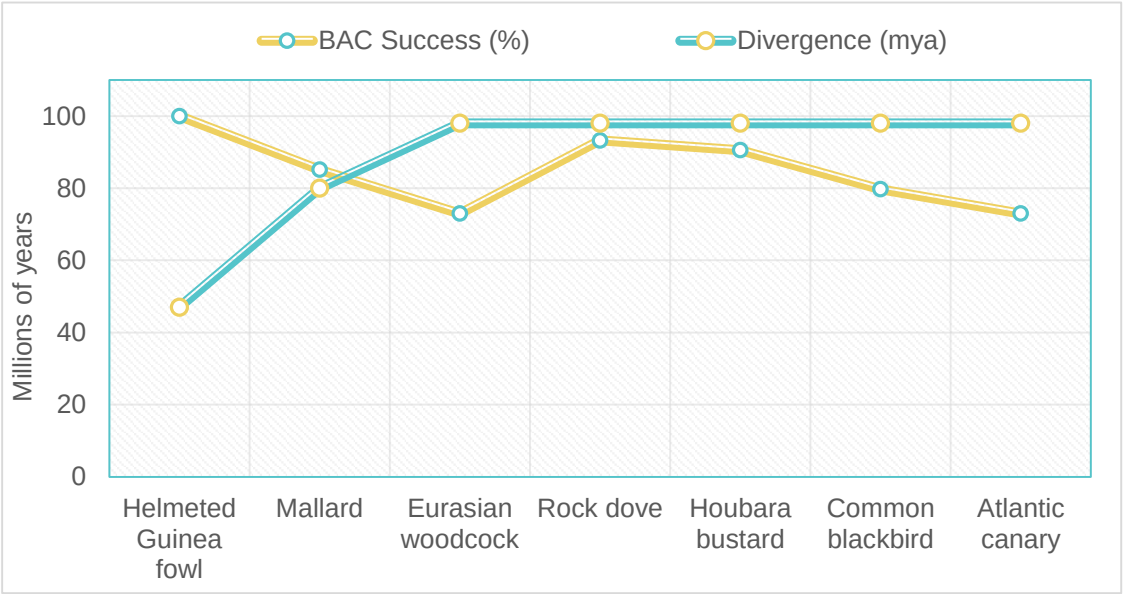


Figure 4.4: Correlation between evolutionary divergence and successful hybridisation of BACs.

4.4.3 Specific aim 2c: Obtain FLpter values to generate comparative maps for macrochromosomes 1-9 and Z

4.4.3.1 Measuring FLpter values

For BACs that were successfully hybridised, FLpter values, standard deviations, and the number of mitotic chromosomes measured were recorded. An example for chicken chromosome 1 can be seen below in Table 4.3, with the full table of results for all species given in the appendix, Supplementary Tables S6 to S13.

BAC Clone	GGA FLpter Value	S.D	n
CH261-89C18	0.0302	0.0119	65
CH261-89G23	0.1063	0.0357	39
CH261-119K2	0.1483	0.0413	36
CH261-120J2	0.1685	0.0314	37
CH261-36B5	0.3222	0.0273	40
CH261-25P18	0.3600	0.0581	30
CH261-125F1	0.4117	0.0650	38
CH261-118M1	0.5295	0.0382	38
CH261-18J16	0.5868	0.0565	32
CH261-29N14	0.6274	0.0380	35
CH261-9B17	0.7067	0.0444	38
CH261-168O17	0.7237	0.0280	31
CH261-83O13	0.7594	0.0323	38
CH261-107E2	0.8004	0.0335	29
CH261-58K12	0.8439	0.0316	38
CH261-184E5	0.8824	0.0337	35
CH261-98G4	0.9603	0.0261	40

Table 4.3: Results for chicken FLpter values, with all BACs corresponding to chicken chromosome 1. GGA = *Gallus gallus.* S.D = Standard deviation. n = number of chromosomes measured,

Relative to the chicken genome, there were 136 chromosomal rearrangements detected across all 7 species (Table 4.4). Of these, 78 were intrachromosomal and 55 were interchromosomal, and there were 3 rearrangements in the canary that were considered both inter- and intrachromosomal. Intrachromosomal

rearrangements consisted of inversions and intrachromosomal duplications. Interchromosomal rearrangements consisted of fusions, fissions, interchromosomal duplications, and translocations. The Eurasian woodcock, common blackbird, and Atlantic canary exhibited the most chromosomal rearrangements, with 30, 27, and 26 rearrangements respectively. Not only do these three species exhibit a high degree of rearrangement compared to the chicken, but each species demonstrates a particular rearrangement unique to its genome: the common blackbird demonstrated the highest number of inversions of all species tested, with a total of 20. The Eurasian woodcock displayed the highest number of interchromosomal rearrangements, with 21 in total; the majority of these interchromosomal rearrangements were fissions of the macrochromosomes, resulting in a high diploid number (2*n*=96). The Atlantic canary was the only species to exhibit duplications, both interchromosomally and intrachromosomally. Moreover, the Atlantic canary and the Eurasian woodcock are the only 2 species to demonstrate translocations of individual BACs.

The houbara bustard and the mallard duck exhibited the lowest number of chromosomal rearrangements, with 9 and 11 respectively, representing approximately 13-17% of total rearrangements. To summarise the extent of chromosomal rearrangements across all 7 species, a phylogenetic tree was produced (Figure 4.5), demonstrating the variation in chromosomal rearrangements between closely related and distantly related bird species.

Common name	Intrachromosomal Duplications	Inversions	Fusions	Fissions	Translocations	Interchromosomal Duplications
Mallard	0	8	0	3	0	0
Eurasian woodcock	0	9	0	20	1	0
Rock dove	0	16	0	3	0	0
Helmeted Guinea fowl	0	8	9	4	0	0
Houbara bustard	0	5	0	4	0	0
Common blackbird	0	20	0	7	0	0
Atlantic canary	3	14	0	5	1	3

Table 4.4: Summary of inter- and intrachromosomal rearrangements.

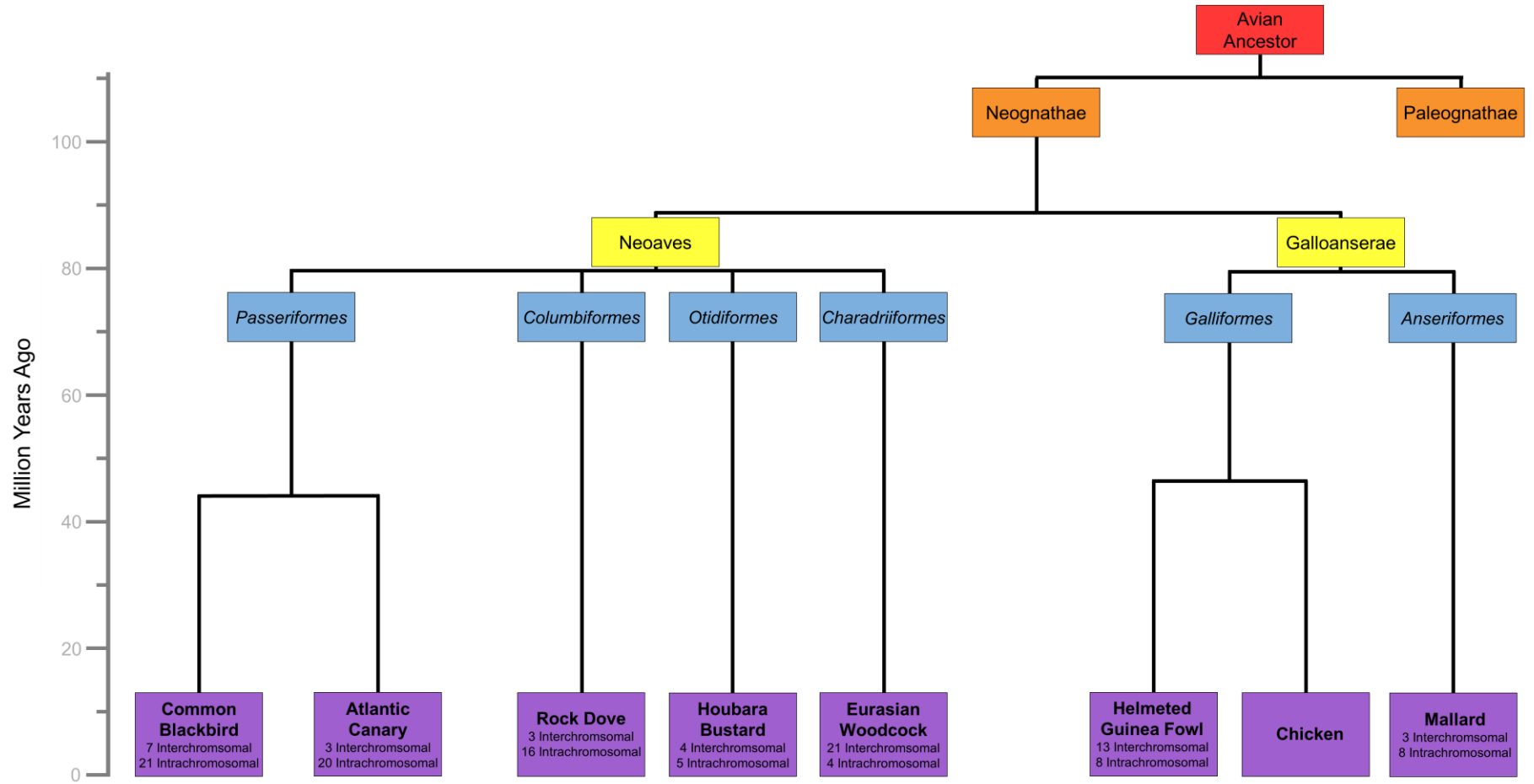


Figure 4.5: Phylogenetic tree illustrating the inter- and intrachromosomal rearrangements of the macrochromosomes for each of the 7 species tested relative to the chicken genome (modified from O'Connor *et al.*, 2018). The coloured blocks at each branch point represent different phylogenetic taxa.

4.4.3.2 Analysing FLpter values to generate comparative maps

To generate the comparative maps, the hybridisation positions of BACs were determined by measuring the FLpter values and applying these to the length of the ideogrammatic chromosomes. Figure 4.6 shows an example of comparative maps for BACs that localise to chicken chromosome 1, with all comparative maps shown in the appendix, Supplementary Figures S1 to S9.

For the BACs that showed inversions, two-tailed unpaired T-tests were conducted to calculate whether the values of the two probes were significantly different. In instances where BACs flanking the inversion had similar FLpter values, these were also included in the T-tests. An example for canary chromosome 1 can be seen below in Table 4.5, with the full table of the BACs involved in each chromosomal rearrangement and the corresponding p-values given in the appendix, Supplementary Tables S14 to S20. However, the relative order of BACs was not altered for any p-values that were not significant as the physical location of the BACs can undergo small changes that do not correlate to significant p-values, and significance can vary depending on the size of the dataset.

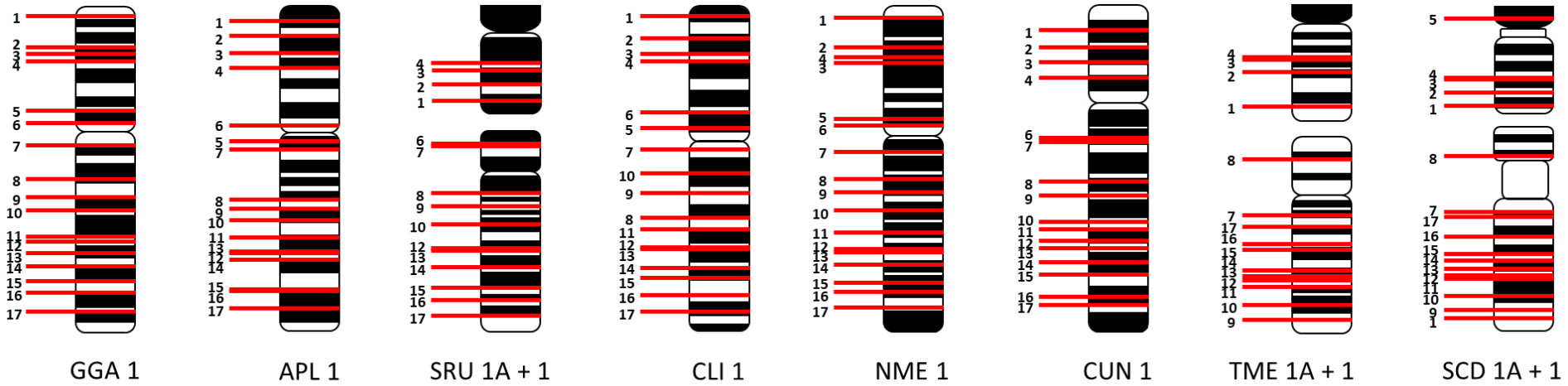


Figure 4.6: Ideograms indicating relative hybridisation positions of BACs for chicken chromosome 1, with BACs labelled 1-17 in order of position on the chicken chromosome. BAC positions are indicated for chicken (GGA) chromosome 1, mallard (APL) 1, pigeon (CLI) 1, helmeted guinea fowl (NME) 1, and houbara bustard (CUN) 1. For the common blackbird (TME), Atlantic canary (SCD), and Eurasian woodcock (SRU), BACs are indicated for chromosomes 1A (top) and 1 (bottom).

BAC Clone	FLpter Value	S.D	S.E.M	p-Value
CH261-125F1	0.4254	0.0803	0.0141	<0.0001
CH261-118M1	0.2238	0.1536	0.0215	
CH261-18J16	0.8936	0.0579	0.0100	<0.0001
CH261-29N14	0.8252	0.0626	0.0109	
CH261-29N14	0.8252	0.0626	0.0109	<0.0001
CH261-9B17	0.7362	0.0578	0.0102	
CH261-9B17	0.7362	0.0578	0.0102	0.5529
CH261-168O17	0.7303	0.0627	0.0109	
CH261-168O17	0.7303	0.0627	0.0109	0.0357
CH261-83O13	0.6943	0.0680	0.0128	
CH261-83O13	0.6943	0.0680	0.0128	0.0199
CH261-107E2	0.6544	0.0648	0.0108	
CH261-107E2	0.6544	0.0648	0.0108	0.4723
CH261-58K12	0.6422	0.0579	0.0123	
CH261-58K12	0.6422	0.0579	0.0123	<0.0001
CH261-184E5	0.5362	0.0640	0.0115	
CH261-184E5	0.5362	0.0640	0.0115	<0.0001
CH261-98G4	0.4385	0.0747	0.0181	

Table 4.5: Unpaired T-tests for the Atlantic canary, with all BACs corresponding to chicken chromosome 1. GGA = *Gallus gallus.* S.D = Standard deviation. S.E.M = Standard error of mean.

4.4.4 Specific aim 2c: Compare the BAC order from the 7 species to that already established in genome assemblies for verification of the methodology

To determine the validity of the method for accurately predicting BAC order, the rock dove was analysed in the same way presented in specific aim 2b as it possesses one of the genomes recently upgraded to a chromosome-level assembly using the combined methodology detailed in the first chapter. Consequently, not only is there sequence data for the rock dove, but there is also information on BAC positions within the genome on Evolution Highway (http://eh-demo.ncsa.uiuc.edu/birds/#/SynBlocks).

The FLpter values demonstrated that the BACs remained in the same order as compared to the genome assembly, establishing that this methodology can determine BAC order with a degree of accuracy. However, due to a limited panel of BACs applied in these experiments, some chromosomal rearrangements, and the extent of them, were undetected i.e. the full length of inversions and any rearrangements between BACs were not detected.

4.5 Discussion

4.5.1 Validity of FLpter Values

Despite comparing the BAC order, determined by FLpter values, against two genomes of known BAC order (chicken and rock dove), the validity of FLpter values can still be called into question. When measuring a chromosome, the user must determine the boundaries of the chromosome i.e. where does the p-terminus start and the q-terminus end. As the value is a fraction of the chromosome length from the p-terminus, incorrectly measuring the length of the chromosome will affect the values obtained. This problem was overcome by

applying a consistent method to measuring the chromosome, reducing the possibility of miscalculating a measurement. Moreover, the small standard deviation values obtained demonstrate the reproducibility of the FLpter values.

The validity can also be questioned with regards to correctly determining the orientation of the chromosome. With the largest of macrochromosomes, the morphology allows for the easy determination of the p and q arms. Conversely, the smaller macrochromosomes are not always readily distinguishable, and any disturbance to the morphology caused by the FISH experiments can make this difficult. To overcome these problems, any chromosomes with ambiguous orientation were not measured. If the majority of the chromosomes were ambiguous and their FLpter values omitted, additional FISH experiments were completed to obtain enough data.

Moreover, in recent genome assembly studies of the houbara bustard (co-author on paper in preparation by Poppleton *et al.*), it highlights the importance of standard deviations and the margin of error with this methodology. On chromosome 7, two BACs (CH261-112D24 and CH261-56K7) were determined to be in the same orientation as in the chicken genome. Nevertheless, a slight overlap in the standard deviation called into question the true order. These BACs were tested together and re-measured, revealing an inversion that was previously missed. It is important to note that this was the only case in which an inversion was missed within the houbara bustard genome, and this was only detected through the efforts of FISH and the RACA algorithm (Kim *et al*., 2013; Damas *et al.,* 2017; O'Connor *et al.,* 2018a).

4.5.2 Detection of Chromosomal Rearrangements

This methodology has proven to be successful in detecting several chromosomal rearrangements between the chicken genome and the 7 species tested. Where previous studies have demonstrated that microchromosomes have undergone few chromosomal rearrangements throughout evolution (O'Connor *et al.,* 2018b), the macrochromosomes exhibit both intra- and interchromosomal rearrangements, with the type of rearrangement dependent on the lineage. Nevertheless, the conservation of synteny within the macrochromosomes of many avian species is strong.

By using the chicken as a reference genome, all chromosomal rearrangements detected were those relative to the chicken. For example, chromosome 4 in the chicken is derived from the fusion of an ancestral microchromosome (Griffin *et al.,* 2007), and thus the species that do not show this fusion were described as having undergone an interchromosomal rearrangement (fission). However, the unfused chromosome 4a and 4 is a pattern seen in most avian species (Guttenbach *et al.,* 2003; Derjusheva *et al.,* 2004; Nanda *et al.,* 2011). Furthermore, a small bias is introduced regarding the degree of chromosomal rearrangements in certain species. Through the use of chicken BACs only, species further diverged (and potentially highly rearranged) from the chicken tend to have a lower success for the number of BACs hybridised (Damas *et al.,* 2017). Not only does this increase any gaps between BACs to reduce the coverage on the chromosome, but it also presents a higher percentage of chromosomal rearrangements when the BACs successfully hybridise. By using a combination of BACs derived from the chicken and other important reference

species, such as the zebra finch (*Taeniopygia guttata*), it might be possible to address this problem.

Additionally, there is a high probability that small rearrangements have been undetected due to the resolution of the BAC mapping method. An attempt was made for the mapping of BACs along the entire length of the macrochromosomes in order to provide higher coverage, but due to the selection of BACs and the paucity of BACs for the sex chromosomes, large gaps were present for some chromosomes.

4.5.3 Generating Comparative Maps for the Macrochromosomes

The availability of comparative maps provides insight into patterns of conservation and rearrangement across avian species. For some chromosomes, there were patterns of rearrangement observed between species within the same order (chromosome 1 in the common blackbird and Atlantic canary), which were absent when compared to species from other orders. Other examples can be seen in chromosome 5 for the common blackbird, Atlantic canary, and Eurasian woodcock, and chromosome 7 in the common blackbird, helmeted guinea fowl, houbara bustard, and pigeon; each of these patterns is usually in the form of an inversion of the same BACs. However, due to the varying success of BAC hybridisation, the extent in which patterns were detected in all species tested was limited. Nonetheless, systematically mapping BACs on the macrochromosomes allows us to study the genomes of multiple avian species, reducing the bias from the use of just one reference genome and one outgroup, and without the need of sequence data.

4.5.4 Implications of Chromosomal Rearrangements

Mapping BACs allows for the detection of chromosome rearrangements with the aim of identifying evolutionary breakpoints and homologous synteny blocks, both of which contribute to the evolutionary changes that result in lineage-specific traits. However, it is widely debated whether patterns of chromosome evolution are caused by fixed deleterious mutations or high mutation rates resulting in genetic drift (Burt *et al.,* 1999; Navarro and Barton, 2003; Edwards *et al.,* 2005). Nevertheless, chromosomal rearrangements have been found to play a role in speciation as a result of enhanced reproductive isolation through reduced hybrid fitness, and also due to barriers to gene flow in non-recombining regions (Noor *et al.,* 2001; Rieseberg, 2001).

Moreover, the identification of patterns between species despite divergence times of millions of years signifies an evolutionary role in promoting speciation. For example, the inversions indicate the occurrence of double stranded DNA breaks, and the recurrent use of these breakpoints are due to fragile genomic regions (Pevzner *et al.,* 2003). Larkin *et al.,* (2009) established that these evolutionary breakpoint regions have a propensity for promoting chromosomal rearrangement as they are found within gene-dense areas, in which the genes are related to lineage-specific traits (Elsik, Tellam, and Worley, 2009; Groenen *et al.,* 2012; Farré *et al.,* 2016). Furthermore, it can be assumed that the recurrent breakpoint use could generate novel combinations of genes that may help to promote adaptation. Thus, this study demonstrates a comprehensive approach to tracing evolutionary relationships of multiple distantly related bird species, providing new insight into the nature of avian genomes and genomic stability.

4.5.5 Chromosome Paints vs BAC Mapping

The generation of chromosome paints (Griffin *et al.,* 1999) was a significant breakthrough for comparative studies, allowing for the detection of large syntenic relationships between both closely and distantly related species. These chromosome paints have been tested on more than 70 different species (for example: Shetty, Griffin, and Graves, 1999; Raudsepp *et al.,* 2002; Nishida *et al.,* 2008; Nie *et al.,* 2009; Hansmaan *et al.,* 2009). However, there are many limitations with chromosome paints that restrict comparative studies: the orientation of syntenic regions cannot be established, meaning any number of inversions could be undetected. Moreover, cross-species chromosome painting can yield ambiguous results with non-specific binding, which could either be interpreted as a duplication or translocation, or if a small rearrangement is present, it could be dismissed entirely.

Some of the species studied in this chapter have had chromosome paints applied to their macrochromosomes (Guttenbach *et al.,* 2003; Derjusheva *et al.,* 2004; Shibusawa *et al.*, 2004), with the main conclusion being that there was high conservation of synteny. Whilst fissions and fusions were detected, the depth of detail provided by the paints was limited. The availability of avian genomic sequences for a well-defined library of BACs has increased the number of genetic markers, allowing for a greater detection of chromosomal rearrangements. For example, studies of the helmeted guinea fowl have shown a fusion of chromosome 6 and 7 to form chromosome 5 (when ordered by size). The BAC mapping in this chapter not only detected this fusion, but also detected whether there were any intrachromosomal rearrangements within chromosomes 6 and 7, and which orientation the chromosomes fused. Thus, the resolution of

detail provided in this study surpasses that of the chromosome painting data and provides more depth to comparative studies of avian species.

4.6 Conclusion

Using the chicken genome as a reference, the mapping of individual BAC clones to the chromosomes of multiple avian species allows for the identification of fusions, fissions, duplications, inversions, and translocations, all of which contribute to the chromosomal changes that influence speciation. Mapping of macrochromosomes 1-9 and Z revealed strong chromosome homology despite 47-98 million years of evolutionary divergence, with successful BAC hybridisation ranging from 72.97-100%. The 7 species studied exhibited chromosomal rearrangements relative to the chicken genome, with ~40% of the rearrangements being interchromosomal. This method has proven to be accurate in predicting BAC order, but is heavily dependent on the BACs to reliably work across multiple species. Nevertheless, comparative BAC mapping provides a finer resolution than chromosome paints to identify chromosomal rearrangements, and would not be possible without the selection of BACs designed by Damas *et al.* (2017) to work across multiple avian species.

5 Specific Aim 3: To investigate genome structure and conservation between avian and non-avian reptiles, comparing the chicken and two karyotypically dissimilar turtle species (yellow spotted river turtle and spiny softshell turtle) using chromosome paints and sequence conserved BACs.

5.1 Background

In order to fully elucidate the genomes and traits of vertebrates, phylogenetic studies are important in both understanding and depicting evolutionary history. The timescales for vertebrate evolution pinpoints the appearance of sauropsids approximately 310 million years ago, diverging from amphibians around 360 million years ago (Kumar and Hedges, 1998; Benton and Donoghue, 2006; http://www.timetree.org: Hedges, Dudley, and Kumar, 2006). The clade Sauropsida comprises of archosaurs (birds and crocodiles), testudines (turtles), and lepidosaurs (squamate reptiles and tuataras), and are considered to be extremely diverse, demonstrating significant morphological and physiological differences between each taxon (Pincheira-Donso *et al.,* 2013). However, the phylogenetic position of turtles has long been debated. Molecular studies using nuclear and mitochondrial DNA, coupled with phylogenomic studies, supports evidence of turtles as a sister clade to the archosaurs, forming the archelosauria (Werneberg *et al.,* 2009; Crawford *et al.,* 2012; Chiari *et al.,* 2012; Shaffer *et al.,* 2013), and rejecting the putative relationship between turtles and lepidosaurs (Lyson *et al.,* 2011).

Sauropsids typically exhibit a gross genomic structure of both macrochromosomes and microchromosomes (Takagi and Sasaki, 1974; Kuraku *et al.,* 2006; Olmo, 2008; Schield *et al.,* 2019), with crocodiles being the only

taxon lacking microchromosomes (Cohen and Gans, 1970; Kasai, O'Brien, and Fergusion-Smith, 2012; St John *et al.,* 2012). The gross genomic structure in turtle species has been shown to range from 2*n*=26 in highly rearranged genomes (Ayres *et al.*, 1969; Ventura *et al.,* 2014) to 2*n*=68 in those considered to have avian-like genomes (Barros *et al.,* 1976; Bickham, Tucker, and Legler, 1985). Overall, there are considered to be three defined groups of karyotypic structure within turtles: Species with high diploid numbers that exhibit many microchromosomes, species with low diploid numbers that exhibit very few/no microchromosomes, and species with median diploid numbers that exhibit an average number of microchromosomes compared to the first and second group. Nevertheless, the “standard” karyotype typically observed in turtles is 2*n*=52 (Bickham, Tucker, and Legler, 1985; Montiel *et al.,* 2016), with the presence or absence of microchromosomes being attributed to karyotypic diversity.

Comparative genomics provides the tools to detect the underlying rearrangements that lead to the genome evolution, and investigating these chromosomal rearrangements that occur within and between taxa allows us to infer evolutionary relationships and refine how taxa are characterised (Cardoso *et al.,* 2014). Despite the surge in sequencing and phylogenomic studies, as of July 2019, testudines have 0 complete genome assemblies, 1 chromosome-level assembly, 11 scaffold assemblies, and 1 contig assembly according to Assembly (https://www.ncbi.nlm.nih.gov/assembly). Furthermore, the lack of genomic data available limits our understanding and the studies of events that result in genomic structural diversity. Nevertheless, sequencing data reveals information that can aid our understanding of turtle genomes, such as the

conservation of gene order which has been shown to closely resemble avian genomes (Chiari *et al*., 2012; Tollis *et al.,* 2017).

With the limitations of genomic studies, cytogenetic studies have assisted in elucidating turtle genomes, with chromosome painting and BAC libraries defining gross genomic structure and synteny. Chicken chromosome paints (Griffin *et al.,* 1999) have been hybridised to numerous turtle species (for example Marshall Graves and Shetty, 2001; Kasai *et al.,* 2003; Matsuda *et al.,* 2005; Kasai, O'Brien, and Fergusion-Smith, 2012; Kasai *et al.,* 2012), demonstrating the high degree of homology between birds and turtles. Consequently, it can be assumed that the “standard” avian karyotype (the presence of macrochromosomes and microchromosomes, with a high diploid number) had a very early origin (Matsuda *et al.,* 2005). However, cross-species chromosome painting has proven to be restrictive due to the evolutionary divergence of ~250 million years between testudines and birds (O'Connor *et al.,* 2018c), and as mentioned in section 1.4.1, the resolution provided by chromosome painting is limited; chromosome paints are unable to detect intrachromosomal rearrangements, and cannot define what is encoded within these regions or why they remain conserved.

To further develop our understanding of amniote genomes, the aim of this chapter was to study the chromosomes of turtle species to trace chromosome evolution through the mapping of BACs between avian and non-avian reptiles. Using chicken chromosome paints (Griffin *et al.,* 1999) and selected BACs developed by Damas *et al.* (2017), which have been shown to effectively hybridise across long evolutionary distances, the aim of this chapter was to further genomic studies of chromosome synteny between birds and turtles.

To elucidate and fully appreciate the nature of chromosomal rearrangements between birds and turtles, two turtle species were selected with highly rearranged karyotypes. The first species is the spiny softshell turtle (*Apalone spinifera*), which has a diploid number of *2n*=66 (Badenhorst *et al.,* 2013) and has been shown to exhibit high phenotypic diversity compared to other *Apalone* species (McGaugh, Eckerman, and Janzen, 2008). The second species is the yellow spotted river turtle (*Podocnemis unifilis*), which has a low diploid number of *2n*=28 and is listed as an IUCN vulnerable specie (Tortoise and Freshwater Turtle Specialist Group, 1996).

5.2 Specific Aims

The purpose of this chapter was to investigate the relationship between avian and non-avian reptiles, using chicken chromosome paints and conserved BACs to elucidate the chromosomal rearrangements that occurred during the evolution of the archelosauria. With that in mind, the specific aims of this chapter were:

- Specific aim 3a: Apply chicken macrochromosome and microchromosome paints to detect chromosome homology between turtles and birds
- Specific aim 3b: Use sequence conserved BACs to study genome conservation between avian and non-avian reptiles

5.3 Materials and Methods

5.3.1 Chromosome Preparations

Fixed metaphase chromosomes for the yellow spotted river turtle (*Podocnemis unifilis*) and the spiny softshell turtle (*Apalone spinifera*) were kindly provided by Dr. Nicole Valenzuela (Department of Ecology, Evolution, and Organismal Biology, Iowa State University).

5.3.2 Multiprobe Devices

Microchromosome paints R1 to R9 were tested with BACs for chromosomes 10-28 to identify the chromosomal DNA present within each microchromosome pool using multiprobe devices (as seen in Figure 5.1). The paint/BAC mixture consisted of 1.5 µg chicken Hybloc (Applied Genetics Laboratories), 1.5 µL FITC labelled paints, 1.5 µL Texas Red labelled BACs, and 5.5 µL Hybridisation solution E (Cytocell Ltd). The BACs used for each multiprobe device is listed in Table 5.1.

Figure 5.1: Schematic of a multiprobe device showing the layout of microchromosome pools and BACs. Rx = Microchromosome pools labelled in FITC. Chr = Chromosome.

BAC Clone	Chicken Chromosome
CH261-71G18	10
CH261-154H1	11
CH261-4M5	12
CH261-59M8	13
CH261-69D20	14
CH261-90P23	15
CH261-97F21	16
CH261-42P16	17
CH261-72B18	18
CH261-50H12	19
CH261-10L6	20
CH261-122K8	21
CH261-18G17	22
CH261-90K11	23
CH261-65O4	24
CH261-127K7	25
CH261-170L23	26
CH261-28L10	27
CH261-72A10	28

Table 5.1: Chicken microchromosome BACs used for verification of microchromosome paints R1 to R9.

5.4 Results

Conventional analysis of metaphase chromosomes from the yellow spotted river turtle and the spiny softshell turtle revealed diploid numbers of 2*n*=28 and 2*n*=66 respectively, with no discrepancies between the diploid numbers observed in this study and those previous reportedly in the literature (Ayres *et al.*, 1969; Bickham, Tucker, and Legler, 1985; Fantin and Monjeló, 2011; Badenhorst *et al.,* 2013). The spiny softshell turtle demonstrated a karyotype typical of testudines, with the presence of macrochromosomes and microchromosomes (typically defined as small chromosomes with no distinguishable features such as shape or centromere position). In contrast, the yellow spotted river turtle did not exhibit the same karyotype, with only macrochromosomes present.

5.4.1 Specific aim 3a: Apply chicken macrochromosome and microchromosome paints to detect chromosome homology between turtles and birds

5.4.1.1 Using macrochromosome paints to detect chromosome homology

For both species, the macrochromosome assignments for both turtle species confirmed those previously reported. In the spiny softshell turtle, chromosome paints hybridised to all macrochromosomes tested (1-9). The painting data demonstrated that the chicken macrochromosomes are represented by a single-turtle counterpart (Table 5.2), with the exception of chromosome 4 which demonstrates the typical pattern of chromosome 4A and 4 as seen in most avian species.

The paint for chicken chromosome 1 hybridised to half of a macrochromosome in the spiny softshell turtle as opposed to the full length of the chromosome as detected by the other macrochromosome paints. To determine whether this was an issue with the paint hybridisation or a genuine representation of chromosome paint signal, the paint for chicken chromosome 1 was retested but yielded ambiguous results. Moreover, the paint for chicken chromosome 5 yielded ambiguous signals, in which it was difficult to determine the presence of additional signals due to the level of non-specific binding. The paint was retested but failed to clarify the ambiguous data.

Chromosome Paint	Hybridisation Location	Number of Signals (pairs)
1	Half of macrochromosome	1
2	Entirety of macro	1
3	Entirety of macro	1
4	Micro- and macrochromosome	2
5	Macrochromosome	1
6	Macrochromosome	1
7	Macrochromosome	1
8	Macrochromosome	1
9	Macrochromosome	1

Table 5.2: Chicken macrochromosome paints for chromosomes 1-9 tested on the spiny softshell turtle (*Apalone spinifera*). The location in which the paints hybridised were determined, and how many signals were present.

In the yellow spotted river turtle, only 50% of the chromosome paints hybridised to the macrochromosomes tested (Table 5.3). Unlike the spiny softshell turtle, the painting data demonstrated that the chicken macrochromosomes are represented by more than one turtle counterpart. The pattern of hybridisation remained consistent across the macrochromosomes, demonstrating blocks of synteny separated by non-syntenic regions.

Chromosome Paint	Hybridisation Location	Number of Signals (pairs)
1	2 macrochromosomes	2
2	2 macrochromosome	2
3	No signal	-
4	No signal	-
5	2 macrochromosomes	2
6	No signal	-
7	No signal	-
8	Macrochromosome	1
9	No signal	-
Z	2 macrochromosomes	2

Table 5.3: Chicken macrochromosome paints for chromosomes 1-9 tested on the yellow spotted river turtle (*Podocnemis unifilis*). The location in which the paints hybridised were determined (i.e. macrochromosome or microchromosome), and how many signals were present.

5.4.1.2 Verification of microchromosome paints

To detect chromosome homology between the chicken and both turtle species, chicken chromosome paints developed by Griffin *et al.* (1999) were used. However, in previous studies by Lithgow *et al.* (2014) the microchromosome paints R1-R9 were listed as having numerous primary (strong) and secondary (weak) signals. The nature of determining secondary signals was subjective and without further assessment, in addition to ambiguity from non-specific binding. Thus, the microchromosome paints were retested on chicken metaphase chromosomes for independent verification (an example shown in Figure 5.2). Analysis revealed that the classification of primary and secondary paints differed from that determined by Lithgow *et al.* (2014). Furthermore, no data was found in the literature to define the chromosomal DNA present (i.e. which chromosome numbers were assigned) within each microchromosome pool. Table 5.4 indicates the chromosomes which are present within each microchromosome pool, and the number of primary and/or secondary signals present.

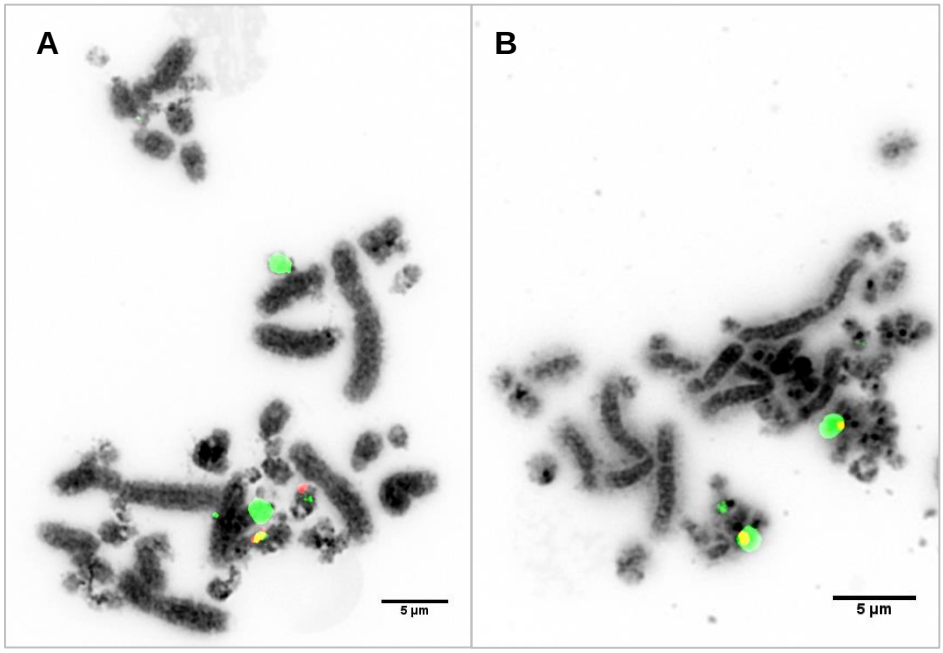


Figure 5.2: Chicken microchromosome paint R4 labelled in FITC showing the presence of primary (strong) and secondary (weak) signals. A) Microchromosome paint R4 with BAC CH261-59M8 (chicken chromosome 13) labelled in Texas Red. The BAC localises with the chromosome paint, with a primary signal on chromosome 13. B) Microchromosome paint R4 with chicken BAC CH261-69D20 (chicken chromosome 14) labelled in Texas Red. The BAC localises with the chromosome paint, with a secondary signal on chromosome 13.

Chromosome	R1	R2	R3	R4	R5	R6	R7	R8	R9
10	N	Y	N	N	N	N	N	N	N
11	Y	N	N	N	N	N	N	N	N
12	N	Y	N	N	N	N	N	N	N
13	N	N	Y	++	N	N	N	N	N
14	N	N	N	Y	N	N	N	N	N
15	N	N	N	N	Y	N	N	N	N
16	N	N	N	N	N	N	N	N	N
17	N	N	N	N	N	Y	N	N	N
18	N	N	N	N	N	Y	N	N	N
19	N	N	N	N	N	Y	N	N	N
20	N	N	N	N	Y	N	N	N	N
21	N	N	N	N	N	N	N	N	N
22	N	N	N	N	N	N	++	N	N
23	N	N	N	N	N	N	++	N	N
24	N	N	N	N	N	N	Y	N	N
25	N	N	N	N	N	N	Y	N	N
26	N	N	N	N	N	N	N	++	N
27	N	N	N	N	N	N	N	Y	N
28	N	N	N	N	N	N	N	Y	N

Table 5.4: The number of primary and secondary signals present in microchromosome paints R1 to R9. N = No signal present. Y = signal present (green). ++ = secondary signal present (yellow).

5.4.1.3 Using microchromosome paints to detect chromosome homology

In both the spiny softshell turtle and yellow spotted river turtle, ~44% of the chromosome paints hybridised to the metaphase chromosomes (Table 5.5 and 5.6). The painting data demonstrated a degree of conserved synteny between the chicken and turtle microchromosomes. However, in the spiny softshell turtle, microchromosome paint R2 demonstrated ambiguity with signals, and the presence of a secondary signal was ambiguous. Furthermore, microchromosome paint R6 exhibited 2 microchromosome pairs as opposed to 3 microchromosome pairs in the chicken. Similarly, in the yellow spotted river turtle, microchromosome paint R2 and R7 exhibited 1 and 3 microchromosome pairs as opposed to 2 and 4 microchromosome pairs in the chicken, respectively.

Chromosome Paint	Hybridisation Location	Number of Signals (pairs)	Number of Signals in Chicken (pairs)
R1	Microchromosome	1	1
R2	Microchromosome	1-2	2
R3	No signal	-	1
R4	Microchromosome	2	2
R5	No signal	-	2
R6	Microchromosome	2	3
R7	No signal	-	4
R8	No signal	-	3
R9	No signal	-	7

Table 5.5: Chicken microchromosome paints R1 to R9 tested on the spiny softshell turtle (*Apalone spinifera*). The location in which the paints hybridised were determined, and how many signals were present.

Chromosome Paint	Hybridisation Location	Number of Signals (pairs)	Number of Signals in Chicken (pairs)
R1	No signal	-	1
R2	Microchromosome	1	2
R3	No signal	-	1
R4	No signal	-	2
R5	Micro- and macrochromosome	2	2
R6	Microchromosomes	3	3
R7	Micro- and macrochromosomes	3	4
R8	No signal	-	3
R9	No signal	-	7

Table 5.6: Chicken microchromosome paints R1 to R9 tested on the yellow spotted river turtle (*Podocnemis unifilis*). The location in which the paints hybridised were determined, and how many signals were present.

5.4.2 Specific aim 3b: Use sequence conserved BACs to study genome conservation between avian and non-avian reptiles

Subtelomeric probes designed by O'Connor *et al.* (2018b) were hybridised to metaphase chromosomes for chicken homologs 1-28 and Z of the spiny softshell turtle and the yellow spotted river turtle, an example of which is shown in Figure 5.3. Probes for chicken chromosome 16 were excluded due poor coverage resulting from extreme variability from clusters of MHC genes (Miller *et al.,* 2013), in addition to the smallest microchromosomes (29-38) that as of yet do not have physically mapped sequence data assigned to them.

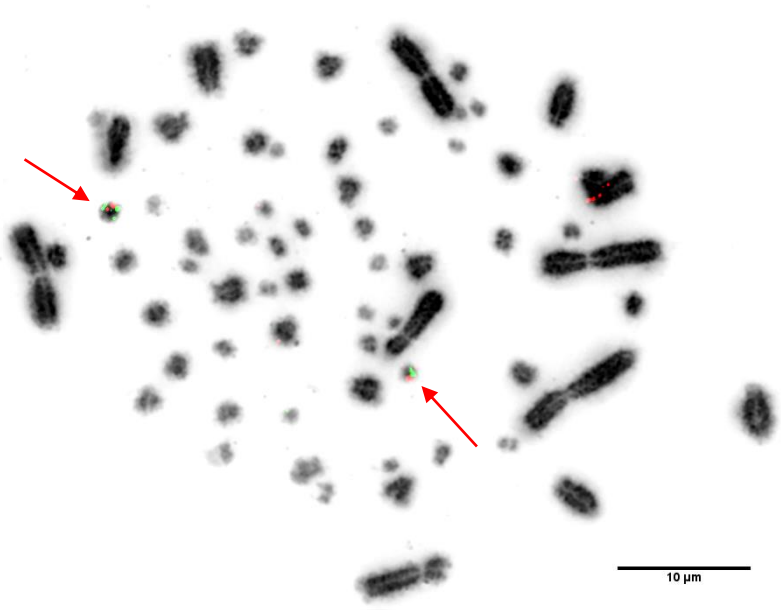


Figure 5.3: BAC clones hybridised to the spiny softshell turtle (*Apalone spinifera*). The FITC (green) labelled signal represents CH261-103F4 (chicken 24 homolog), the Texas red labelled signal represents CH261-65O4 (chicken 24 homolog).

Analysis of FISH data for the spiny softshell turtle revealed little evidence of interchromosomal rearrangements between this species and that of the chicken, with ~47% of BACs successfully hybridised to metaphase chromosomes. Moreover, of the 18 avian microchromosome orthologues (chromosomes 10-28, excluding 16), 11 orthologues appeared to remain as single microchromosomes for chromosomes 10-13, 15, 17, 19, 21, 23, 24, and 26. The yellow spotted river turtle revealed a high degree of interchromosomal rearrangements between this species and that of the chicken, however, only ~34% of BACs successfully hybridised. Of the 18 avian microchromosome orthologues, 8 orthologues were fused to macrochromosomes for chromosomes 10, 12, 14, 17, 18, 19, 21, 23, and 24 (Table 5.7). Furthermore, tandem fusions were identified between chromosomes 10, 12 and 14, 18 and 19, and 21, 23, and 24.

BAC Clone	Chicken Chromosome	PUN Chromosome
CH261-44D16	2	4
CH261-89P6	4	2
CH261-49F3	6	4
CH261-180H18	7	4
CH261-129A16	Z	5
CH261-133M4	Z	5
CH261-71G18	10	8
CH261-4M5	12	8
CH261-122H14	14	8
CH261-69D20	14	8
CH261-42P16	17	12
CH261-72B18	18	7
CH261-10F1	19	7
CH261-50H12	19	7
CH261-122K8	21	2
CH261-191G17	23	2
CH261-90K11	23	2
CH261-103F4	24	2
CH261-65O4	24	2

Table 5.7: BAC clones successfully hybridised to the yellow spotted river turtle (*Podocnemis unifilis*, PUN). Orthologues of the chicken chromosomes were determined based on karyotype and chromosome size.

The macrochromosomes yielded very few results across both species tested (Table 5.8), with only 20% of the BACs successfully hybridised in the spiny softshell turtle and 30% in the yellow spotted river turtle. The yellow spotted river turtle continued to show a high degree of interchromosomal rearrangement; BACs for chicken chromosome 2 hybridised to the second largest chromosome, and BACs for chicken chromosomes 4, 6, and 7 hybridised to the fourth largest chromosome, with 6 and 7 appearing to be fused in tandem. If excluding the macrochromosome BACs previously tested via chromosome paints, the hybridisation success of the BACs increases to ~61% in the spiny softshell turtle. However, the hybridisation success does not increase significantly for the yellow spotted river turtle, with an increase of only ~2%.

BAC Clone	Chicken Chromosome	Spiny Softshell Turtle	Yellow Spotted River Turtle
CH261-89C18	1	N	N
CH261-98G4	1	N	N
CH261-169N6	2	N	N
CH261-44D16	2	N	Y
CH261-169K18	3	Y	N
TGMCBA-295P5	3	N	N
CH261-83E1	4	Y	N
CH261-89P6	4	Y	Y
CH261-49B22	5	N	N
CH261-78F13	5	N	N
CH261-49F3	6	N	Y
TGMCBA-382J4	6	N	N
CH261-180H18	7	N	Y
CH261-56K7	7	N	N
CH261-107D8	8	N	N
TGMCBA-252A4	8	N	N
CH261-183N19	9	N	N
CH261-187M16	9	N	N
CH261-129A16	Z	N/A	Y
CH261-133M4	Z	Y	Y
CH261-115G24	10	Y	N
CH261-71G18	10	Y	Y
CH261-121N21	11	Y	N
CH261-154H1	11	N	N
CH261-4M5	12	Y	Y
CH261-60P3	12	N	N
CH261-115I12	13	N	N
TGMCBA-321B13	13	N	N

BAC Clone	Chicken Chromosome	Spiny Softshell Turtle	Yellow Spotted River Turtle
CH261-122H14	14	Y	Y
CH261-69D20	14	Y	Y
CH261-90P23	15	Y	N
TGMCBA-266G23	15	N	N
CH261-42P16	17	Y	Y
TGMCBA-375I5	17	Y	N
CH261-60N6	18	Y	N
CH261-72B18	18	Y	Y
CH261-10F1	19	Y	Y
CH261-50H12	19	Y	Y
TGMCBA-250E3	20	N	N
TGMCBA-341F20	20	N	N
CH261-122K8	21	Y	Y
CH261-83I20	21	N	N
CH261-18G17	22	N	N
CH261-40J9	22	N	N
CH261-191G17	23	Y	Y
CH261-90K11	23	Y	Y
CH261-103F4	24	Y	Y
CH261-65O4	24	Y	Y
CH261-127K7	25	N	N
CH261-59C21	25	N	N
CH261-170L23	26	Y	N
CH261-186M13	26	N	N
CH261-28L10	27	Y	N
CH261-66M16	27	Y	N
CH261-64A15	28	N	N
CH261-72A10	28	Y	N

Table 5.8: Hybridisation success of chicken and zebra finch BACs on the yellow spotted river turtle (*Podocnemis unifilis*) and spiny softshell turtle (*Apalone spinifera*). N = No signal present. Y = signal present. N/A = not tested.

5.5 Discussion

5.5.1 Genome Organisation

The two turtle species studied in this chapter demonstrate two of the three defined groups of karyotypic structure within turtles. The spiny softshell turtle exhibits a high diploid number with many microchromosomes, and the yellow spotted river turtle exhibits a low diploid number with no microchromosomes. FISH data from this chapter supports the placement of testudines as the sister clade to the archosaurs, with macrochromosomes and microchromosomes being true counterparts to those of the chicken. Furthermore, this data provides evidence that the reduction in diploid number observed in the yellow spotted river turtle is shown to be a result of genomic reorganisation, which has been influenced by multiple interchromosomal rearrangements (fusions) of avian microchromosome orthologues. This corroborates previous data which identified interstitial telomeric sequences, demonstrating the remnants of chromosomal fusions or amplification of telomere-like sequences (Noronha *et al.,* 2016).

The organisation of any genome will vary between species, and identifying chromosomal rearrangements between taxa can elucidate the lineage-specific traits that arise throughout evolution. The changes in genomic neighbourhood (i.e. changes in gene order, rearrangement of syntenic blocks) have been shown to effect the regulatory environment of genes (Ahituv *et al.,* 2005), influencing transcriptome evolution (De, Teichmann, and Babu, 2009) and thus contributing to speciation and adaption (Kirkpatrick and Barton, 2006; Hoffman and Rieseberg, 2008; Skinner and Griffin, 2012). Furthermore, genotypic traits (such as evolutionary breakpoint regions) have been shown to coevolve with changes

in genome organisation (Ruiz-Herrera *et al.,* 2005), providing that the genomic changes are capable of promoting the evolution of other traits. An example can be made of previous studies by Valenzuela and Adams (2011), demonstrating the presence of co-evolution between sex determination and diploid number in testudines. Thus, it can be inferred that the chromosomal rearrangements and differences in genomic organisation between the spiny softshell turtle and yellow spotted river turtle have contributed to the evolution of two distinct lineages.

5.5.2 Limitations of FISH

Due to the limitations of chromosome paints and the mapping of BACs, arising from sequence divergence as a result of evolutionary distance, it is possible that many chromosomal rearrangements were not detected in this study. Several of the macrochromosome and microchromosome paints failed to hybridise, restricting what can be inferred with regards to interchromosomal and intrachromosomal rearrangements. For the paints that did successfully hybridise, cryptic translocations could remain undetected due to weak chromosome paint signals. Moreover, only subtelomeric probes were used in this study, which would fail to detect any intrachromosomal event beyond the region spanned by the BAC. However, these limitations do not detract from degree of synteny observed and the overall pattern of genome organisation.

5.5.3 Conservation of microchromosomes in non-avian reptiles

The data presented in this chapter demonstrates a high degree of genome conservation between avian and non-avian reptiles, in which both macrochromosomes and microchromosomes are truly homologous to that of the chicken. The 11 orthologues which remained as microchromosomes in the spiny softshell turtle corroborates a pattern observed in numerous avian species

(Lithgow *et al.,* 2014; O'Connor *et al.,* 2018b), in which microchromosomal rearrangements are uncommon and demonstrate genome stability. However, the detection of microchromosomal rearrangements (i.e. the tandem fusion of microchromosomes in the yellow spotted river turtle), a pattern also observed in many falconiformes (Damas *et al.,* 2017; Joseph *et al.,* 2018) and the budgerigar (O'Connor *et al.,* 2018a), continues to demonstrate genomic stability as microchromosomes remain as distinct units. Furthermore, the reduced diploid number in the yellow spotted river turtle suggests that these fusions can be attributed to karyotypic evolution from the shared common ancestor.

The presence of microchromosomes has often been observed in most bird, lizard, and turtle species (reviewed by Burt, 2002); however, crocodilian species exhibit an atypical archelosaur karyotype with the noticeable absence of microchromosomes, thought to be caused by chromosomal fusion (Srikulnath, Thapana, and Muangmai, 2015). Sequencing data from the chicken genome has revealed differences in the structural composition of macrochromosomes and microchromosomes, such as the GC content, recombination rate, and gene density (International Chicken Genome Sequencing Consortium, 2004). These compositional differences have also been identified in other reptiles, such as the Japanese four-striped rat snake (Matsubara *et al.,* 2012) and the Chinese soft-shelled turtle (Kuraku *et al.,* 2006). This further supports the hypothesis that microchromosomes were fixed in the karyotype of the shared common ancestor between birds and turtles.

5.5.4 Amniote Genome Evolution Over 250 Million Years

As previously mentioned, there are numerous studies that have observed the high degree of conservation between testudine and avian species (Matsuda *et al.,* 2005; Kawai *et al.,* 2009; Uno *et al.,* 2012). Previous studies have suggested that the origin of the standard avian karyotype (2*n*=~80), in which microchromosomes are dominant, is brought about by the mechanism of fission of the macrochromosomes and microchromosomes (Takagi and Sasaki, 1974; O'Connor *et al.,* 2018c). This hypothesis is supported by previous zoo-FISH studies in which chicken chromosome paints were applied to non-avian reptiles, demonstrating the high degree of synteny (Matsuda *et al.,* 2005; Pokorná *et al.,* 2011; Pokorná *et al.,* 2012; Kasai *et al.,* 2012). Thus, it can be inferred that this karyotype of the shared common ancestor has been in place since the lineage divergence from the Lepidosauromorpha, remaining conserved for more than 250 million years (Iwabe *et al.,* 2004; Benton and Donoghue, 2006; Crawford *et al.,* 2012; O'Connor *et al.,* 2018c).

Nevertheless, each amniote taxon possesses traits unique to each lineage: avian species exhibit a large degree of phenotypic variation, increased longevity despite their high metabolic rate (Holmes and Ottinger, 2003; Munshi-South and Wilkinson, 2010), and lightweight skeletons; testudines possess a bony shell containing dermal bones not present in any vertebrate order, in which the ventral plastron and dorsal carapace as connected by lateral bridges (Gilbert *et al.,* 2001); and crocodilians exhibit an atypical karyotype which lacks microchromosomes in all 20 species. (Cohen and Gans, 1970). Moreover, the reptilian genome size has been shown to vary from an average of 1.4 Gb to 5 Gb (Figure 5.4).

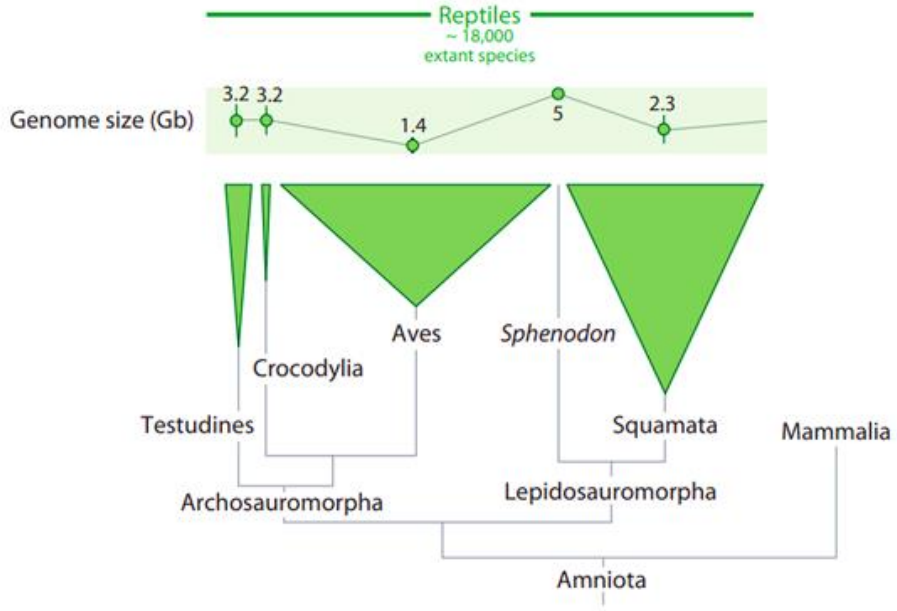


Figure 5.4: Reptilian genome size (modified from Janes *et al.,* 2010).

Since the divergence from the testudines, avian genome have remained relatively small, with a size of ~1.4 Gb (Gregory, 2002; Janes *et al.,* 2010). As with the pattern of microchromosomes, the crocodylia deviate from the archelosauromorpha, matching a larger genome size compared to that of the testudines (~3.2 Gb), in which there is a possibility that larger genomes (2.5-3 Gb) are associated with fewer chromosomes (St John *et al.,* 2012). This suggests that after the divergence of testudines over 250 million years ago, there was a reduction in genome size in avian species.

5.5.5 Genomic Stability of the Reptilian Karyotype

As aforementioned, most avian and non-avian reptiles possess a karyotype that consists of both macrochromosomes and microchromosomes. Since the divergence of testudines from the archosaurs over 250 million years ago, the standard avian karyotype commonly observed in neornithine birds has undergone a small number of interchromosomal rearrangements (O'Connor *et al.* 2018c). By supplementing bioinformatic data with that generated from this

chapter, it can be inferred that the dinosaur-theropod route that led to modern birds (Figure 5.5) remained mostly unchanged interchromosomally. Thus, the extraordinary degree of genome stability that results in the standard avian karyotype predates the emergence of early pterosaurs and dinosaurs, and possibly the evolution of flight.

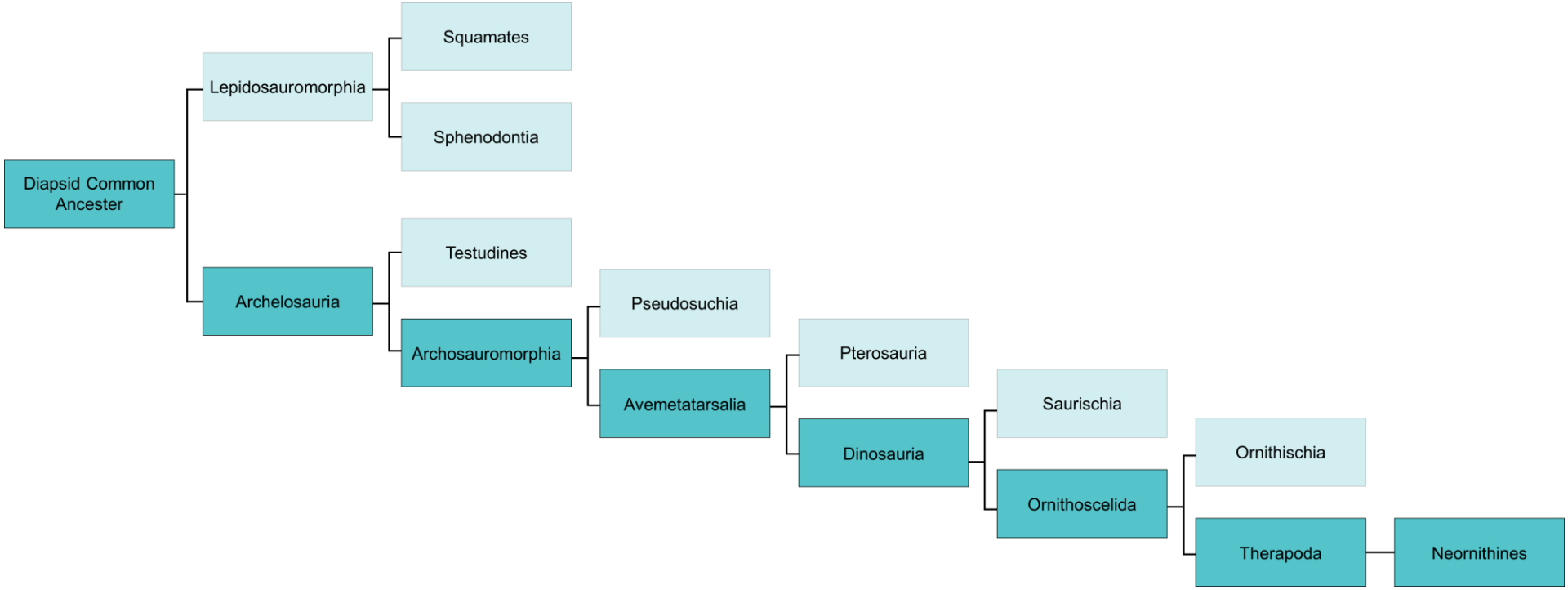


Figure 5.5: Phylogenetic tree tracing the lineage from the diapsid common ancestor, to the archosaurs, to modern birds via the dinosaur-theropod route (modified from O'Connor *et al.*, 2018c).

Considering the genomic stability of avian and non-avian reptiles and the availability of genome assemblies, it has been possible to reconstruct the karyotypes of extinct species, such as the diapsid common ancestor. Avian phenotypes considered to be ancestral correlate with the fewest number of chromosomal rearrangements, and numerous chromosomal rearrangements correlate to high degrees of phenotypic variation and adaptation (Romanov *et al.,* 2014). Thus, applying the most parsimonious sequence of events that occurred from the diapsid ancestor to extant birds via non-avian theropod dinosaur, we infer the most likely genome structure of the diapsid common ancestor. Composed of a roughly equal number of macrochromosomes and microchromosomes, the proposed karyotype of the diapsid common ancestor exhibits a diploid number of 2*n* = 36-46 (O'Connor *et al.,* 2018c). It is likely that this genome structure was a contributing factor to the survival of birds during the K-Pg mass extinction (Berv and Field, 2017), but also to the unique physiology and phenotypic diversity observed in modern birds today.

5.6 Conclusion

The spiny softshell turtle and the yellow spotted river turtle exhibit strikingly different karyotypes (i.e. a high diploid number and the presence of microchromosomes, versus a low diploid number and the absence of microchromosomes), yet there still remains an extraordinary degree of conservation between avian and non-avian reptiles. The data presented in this chapter demonstrates that the range of diploid numbers in testudine genomes (2*n*=26 to 2*n*=68) may be result of the variation in the number of microchromosomes, and more specifically as to whether they remain as individual chromosomes or undergo fusion to other chromosomes.

The standard avian karyotype appears to have been conserved for more than 275 million years since the lineage divergence from the Lepidosauromorpha. Thus, the cytogenetic analysis of testudine genomes provides an insight into amniote genome evolution and genome organisation, allowing us to facilitate evolutionary inferences across multiple taxa. Moreover, investigating the genomes of reptilian species, particularly those exhibiting typical and atypical karyotypes compared to avian species, will allow us to elucidate the origins of microchromosomes.

6 Specific Aim 4: Identify genes within the newest chicken genome assembly (*Gallus_gallus*-5.0; GCA _ 000002315.3) to generate a panel of fluorescent markers for the hitherto undiscovered microchromosomes 29 to 38.

6.1 Background

Avian karyotypes have demonstrated a remarkable degree of stability, in which most avian lineages exhibit diploid numbers of 2*n*=~80. A key feature of avian karyotypes is the presence of numerous microchromosomes, with the “standard” karyotype being composed of 10 pairs of macrochromosomes and approximately 30 pairs of microchromosomes (reviewed by Burt, 2002). Microchromosomes are GC-rich, demonstrating the highest gene densities of all avian chromosomes (Federico *et al.,* 2005), with a low density of repetitive elements (Schmid *et al.,* 2015) and approximately 50% of the total gene content (Smith *et al*., 2000; Andreozzi *et al*., 2001, ICGSC 2004; Wójcik and Smalec, 2016). The presence of microchromosomes is a trait shared amongst most avian and non-avian reptiles (Janes *et al.,* 2010), however, there are some notable deviations from this. The diploid numbers observed in avian species range from 2*n*=42 in the stone curlew (Nie *et al.,* 2009) to 2*n*=142 in the grey lourie (Christidis, 1990). Species which exhibit fewer chromosomes tend to demonstrate the fusion of microchromosomes to macrochromosomes and/or other microchromosomes, such as the *Falconiformes* (Nishida *et al.,* 2008; Damas *et al.,* 2017) and in some *Psittaciformes* (O’Connor *et al.,* 2018a). Moreover, increases in diploid number have been shown to be a result of macrochromosomal fission in recurrent breakpoint regions (Skinner and Griffin, 2012; Degrandi *et al.,* 2017).

To date, there are no sequencing or assembly methods that are capable of assembling entire avian chromosomes in a single read despite recent improvements in technology (Peona, Weissensteiner, and Suh, 2018). A plethora of projects have focussed on sequencing and assembling avian genomes, with most chromosomes having physically mapped sequence data assigned to them, yet all avian genomes are incomplete. The smallest of microchromosomes, chromosomes 29-38, do not have physically mapped sequence data as of yet. This lack of data can be attributed to the fact that the majority of avian genomes have been sequenced using short-read technologies (Kapusta and Suh, 2017), which are known to introduce bias for GC-rich sequence data (Chaisson *et al.,* 2015). Thus, the DNA composition of microchromosomes creates obstacles in genomic library preparation, and consequently the genes contained within these chromosomes remain unmapped and underrepresented in all assemblies. Hence, it has been suggested that genes thought to be absent within avian genomes remain unidentified due to the GC-richness (Hron *et al.,* 2015; Botero-Castro *et al.,* 2017; Bornelöv *et al.,* 2017) and may be present within microchromosomes.

Nevertheless, the development of long-read technologies is progressively overcoming the restrictions brought about by GC-content. The newest chicken genome assembly (*Gallus_gallus*-5.0; GCA_000002315.3) identified 240 genes previously undetected in the *Gallus_gallus*-4.0 (GCA_000002315.2) genome assembly, in addition to assigning new placements to 111 genes (Warren *et al.,* 2017). However, the lack of physical mapping data limits the degree of certainty in these predictions, with previous genome assemblies acknowledged to be incorrect in the past (Denton *et al.,* 2014; O'Connor *et al*., 2017). Moreover,

despite the reduction in missing DNA (14.1% to 2.4%), the number of scaffolds between *Gallus_gallus-5.0* and *Gallus_gallus*-4.0 increases, with 23,870 and 16,847 scaffolds respectively (https://www.ncbi.nlm.nih.gov/assembly). This increase in scaffold number is thought to be from the previously unsequenced GC-rich regions (183 Mb) that now remain partially assembled and unplaced on chromosomes (Warren *et al.,* 2017).

As mentioned in section 3.1, chromosome-level assemblies are essential for addressing biological questions, and allows for the identification of genotype to phenotype associations through the means of an established order of DNA markers. Complete chromosome-level assemblies will allow for better comparative genomic analyses, and could provide better insight into evolutionary genomics and highly rearranged karyotypes. Decoding the composition of a genome, such as the presence and/or absence of genes, gene order, repeats (and so on) allows us to understand how any living organism undergoes the processes of growth, development, and maintenance of the genome. Therefore, access to assembled and annotated genomic data allows us to investigate the coding and non-coding regions of the genome, both of which are essential for life (Li *et al.,* 2017).

Thus, the aim of this chapter was to generate fluorescent probes for microchromosomes 29-38, with the hopes of assigning sequence data to these chromosomes in addition to providing the tools to identify any interchromosomal rearrangements. Of the genes newly identified and reassigned placements, 7 genes were selected for zoo-FISH probe generation for two reasons; some genes are suspected to be located on the smallest of microchromosomes, and this study aims to determine if the new placements are indeed correct. These

genes are FUS RNA binding protein (*FUS_1*), tyrosine kinase 2 (*TYK2*), IKAROS family zinc finger 4 (*IKZF4*), Ski2-like RNA helicase (*SKIV2L*), A-kinase anchor protein 8-like (*AKAP8L*), SWI/SNF related, matrix associated, actin dependent regulator of chromatin subfamily c member 2 (*SMARCC2*), and bromodomain adjacent to zinc finger domain 2A (*BAZ2A*).

FUS_1 is a tumour suppressor gene with potent apoptotic activity for the inhibition of tumour growth in human and mouse lung cancers (Du *et al.,* 2009; Lin *et al.,* 2011). Mutations in this gene have been associated with amyotrophic lateral sclerosis type 6 (Efimova *et al.,* 2017) and Tremor, hereditary essential, 4 (Tio *et al.,* 2016). The placement of *FUS_1* is unknown in both *Gallus_gallus*-4.0 and *Gallus_gallus*-5.0 assemblies. *TYK2* encodes a member of the Janus kinase (JAK) family and is essential for cytokine signalling pathways, with mutations resulting in susceptibility to multiple infectious pathogens and hyper-immunoglobulin E syndrome (Minegishi *et al.,* 2006). The placement of *TYK2* is unknown in both *Gallus_gallus*-4.0 and *Gallus_gallus*-5.0 assemblies. *IKZF4* is a transcription factor with transcriptional repressor activity, and has been associated with immune regulation (Pan *et al.,* 2009). Mutations in this gene have been associated with polycystic ovary syndrome (Jones *et al.,* 2015) and the development of type 1 diabetes (Lempainen *et al.,* 2015). The placement of *IKZF4* was unknown in the *Gallus_gallus*-4.0 assembly, but has been assigned to chromosome 33 in the *Gallus_gallus*-5.0 assembly.

SKIV2L is a homolog of yeast *SKI2* (Lee *et al.,* 1995) and encodes a DEAD box protein, with mutations associated with age-related macular degeneration (Lu *et al.,* 2013) and trichohepatoenteric syndrome (Fabre *et al.,* 2013). The placement of *SKIV2L* was unknown in the *Gallus_gallus*-4.0 assembly, but has been

assigned to chromosome 26 in the *Gallus_gallus*-5.0 assembly. *AKAP8L* encodes anchoring proteins for the regulation of histone methylation chromatin condensation (Bieluszewska *et al.,* 2018), with mutations in this gene associated with autism (Nebel *et al.,* 2015). *AKAP8L* was previously assigned to chromosome W in the *Gallus_gallus*-4.0 assembly, but has been reassigned to chromosome 30 in the *Gallus_gallus*-5.0 assembly. *SMARCC2* encodes a protein subunit involved in chromatin remodelling complexes, which are essential for tumour suppression and normal gene expression in development, growth, and differentiation (Klochendler-Yeivin *et al.,* 2000; Weissman and Knudsen, 2009; reviewed by Ho and Crabtree, 2010). Mutations in the SWI/SNF complexes have been associated with Coffin–Siris syndrome (Tsurusaki *et al.,* 2014). The placement of *SMARCC2* was unknown in the *Gallus_gallus*-4.0 assembly, but has been assigned to chromosome 33 in the *Gallus_gallus*-5.0 assembly. *BAZ2A* encodes the large subunit of the nucleolar remodelling complex (Santoro, Li, and Grummt, 2002), with mutations in the gene associated with colorectal adenocarcinoma, prostate cancer, uterine corpus endometrioid carcinoma, and stomach adenocarcinoma (Gu *et al.,* 2015; reviewed by Chen *et al.,* 2016). The placement of *BAZ2A* was unknown in the *Gallus_gallus*-4.0 assembly, but has been assigned to chromosome 33 in the *Gallus_gallus*-5.0 assembly.

6.2 Specific Aims

The purpose of this chapter was to build upon the ongoing effort to map avian genomes through anchoring specific genes to the smallest microchromosomes, chromosomes 29-38, based upon the newest assembly of the chicken genome. With that in mind, the specific aims of this chapter were:

- Specific aim 4a: To generate fluorescent markers for the 7 selected genes in order to identify chromosomes 29-38
- Specific aim 4b: Provide information from the markers to allow for sequence data to be associated with these chromosomes

6.3 Results

6.3.1 Exploring BACs

Of the newly found genes and those with reassigned placements, genes found to be spanned by a BAC were investigated to determine whether the chromosome placements were correct (Table 6.1). These were tested with BACs from chromosomes 1-28, revealing discrepancies between the assembly data and FISH data. Some placements for the genes were found to be incorrect, for example BAC CH261-54C8, which is meant to localise to chromosome 9 according to the newest genome assembly (*Gallus_gallus*-5.0) but instead localises to chromosome 4 (Table 6.2).

BAC Clone	Gene
CH261-74E15	*CPNE5*
CH261-54C8	*IMP4*
CH261-52H12	*TCF4*
CH261-173H14	*C1QL4*

Table 6.1: BAC clones spanning genes newly found or with reassigned placements in the *Gallus_gallus*-5.0 genome assembly.

BAC Clone	*Gallus_gallus*-4.0 Location	*Gallus_gallus*-5.0 Location	Mapped Chromosome
CH261-74E15	chrUn_AADN03019059:71-504 chrUn_AADN03012208:442-504	Chromosome 26	1 and 26
CH261-54C8	chrUn_AADN03017333:351-2306	Chromosome 9	4
CH261-52H12	chrUn_AADN03013408:6304-8404	Chromosome Z	Z and 6
CH261-173H14	chrUn_random	Chromosome 33	Smaller than 28

Table 6.2: BAC clones spanning genes newly found or with reassigned placements, and the corresponding locations in the *Gallus_gallus*-4.0 and *Gallus_gallus*-5.0 genome assemblies, and the corresponding chromosome position revealed by FISH testing.

Furthermore, BACs selected to localise to unplaced scaffolds were tested to determine if they localised to microchromosomes 29 to 38. This was successful for BAC CH261-130O2, which had remained unplaced in the *Gallus_gallus*-4.0 (chrUn_AADN05000897.1) and *Gallus_gallus*-5.0 (chrUn_Scaffold5792; NT_467154) genome assemblies. FISH mapping revealed that the BAC localised to a chromosome smaller than microchromosome 28, and has now been placed on chromosome 31 in the GRCg6a chicken genome assembly. In total, 2 BACs localising to chromosomes smaller than 28 were identified and are shown in Figure 6.1.

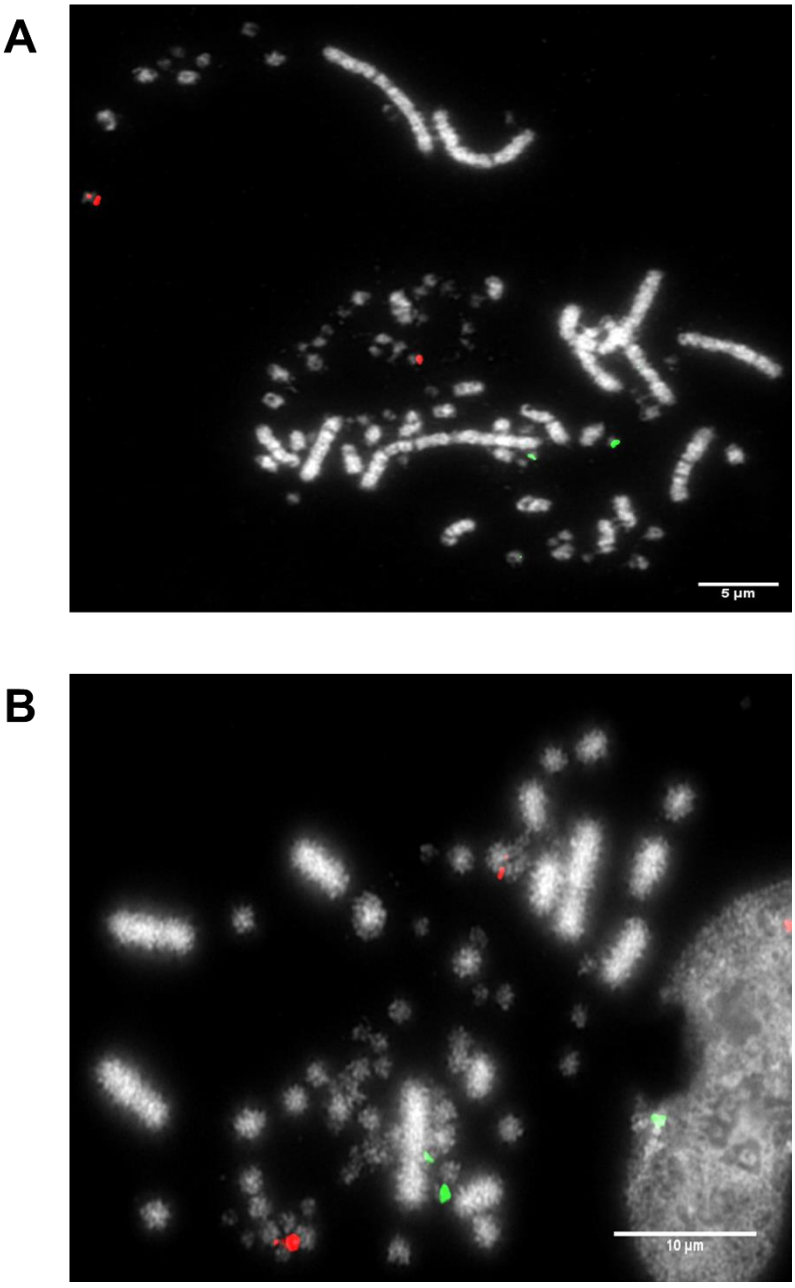


Figure 6.1: BAC clones hybridising to microchromosome 28 (Texas Red) and an unknown pair smaller than microchromosome 28 (FITC) in the chicken (*Gallus gallus*) as revealed by FISH testing. A) BAC clone CH261-101C8 for chromosome 28 labelled in Texas Red with CH261-173M14 for an unknown chromosome pair labelled in FITC. B) BAC clone CH261-101C8 for chromosome 28 labelled in Texas Red with CH261-130O2 for chromosome 31 (as defined by GRCg6a assembly) labelled in FITC.

6.3.2 PCR Optimisation

Under native conditions, most primer pairs either failed to generate PCR products or amplified non-specific DNA, an example of which can be seen in Figure 6.2. Consequently, PCR additives were used to reduce non-specific amplification. To overcome the reduction in polymerase activity, the amount of polymerase was increased for all PCR reactions with the use of additives, from 0.02 U/µL to 0.04 U/µL. Furthermore, temperature gradients were used for all PCR reactions to determine the optimal annealing temperatures both with and without additives.

In total, 31 primer pairs generated PCR products of the correct size. Betaine was found to be the most effective PCR additive, often yielding single bands of the correct weight and with an abundance of amplified DNA. Out of the 38 primer pairs tested, 19 required 1.3 M betaine, 2 required 5% formamide, and 1 required 1 x Q5 High GC-Enhancer for the significant reduction of non-specific amplification. Of the remaining primer pairs, 9 generated PCR products of the correct weight under native conditions, and 7 primer pairs failed to yield PCR products that could be purified or gel extracted (Table 6.3).

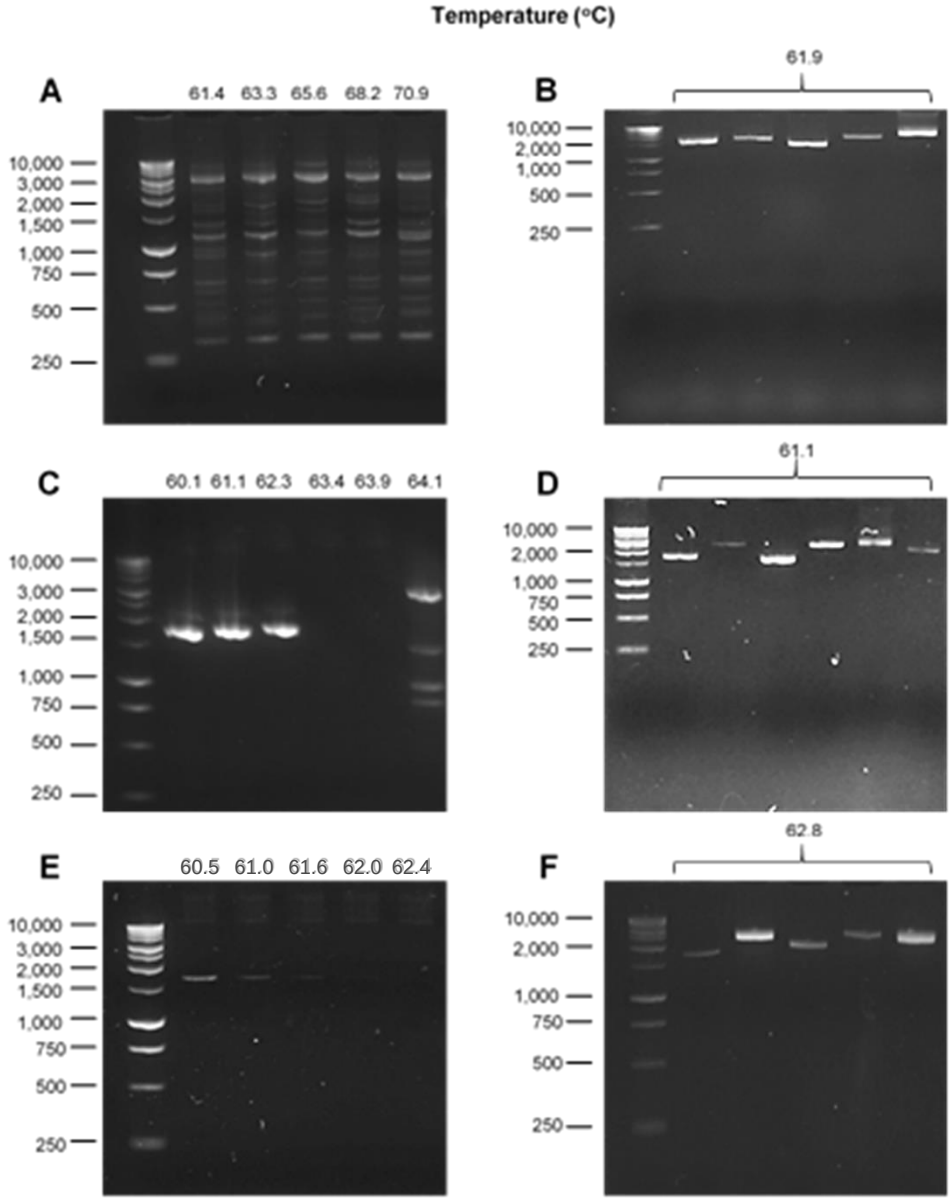


Figure 6.2: PCR amplified products for *FUS_1*, *IKZF4*, and *SKIV2L* primers, under a temperature gradient, with native and optimal conditions. A) *FUS_1* primer pair 1. B) Optimised primer pairs 1-5 for *FUS_1*. C) *IKZF4* primer pair 1. D) Optimised primer pairs 1-6 for *IKZF4*. E) *SKIV2*L primer pair 1. F) Optimised primer pairs 1-5 for *SKIV2L*. A 1 kbp ladder was used for all gels.

Gene	Primer Set	Additive	DNA Status	Concentration (ng/µL)
FUS_1	1	1.3 M Betaine	Purified - Gel extraction	44.4
FUS_1	2	1.3 M Betaine	Purified - PCR purified	18.8
FUS_1	3	1.3 M Betaine	Purified - PCR purified	28.6
FUS_1	4	1.3 M Betaine	Purified - Gel extraction	52.7
FUS_1	5	1.3 M Betaine	Purified - PCR purified	73.8
TYK2	1	-	-	-
TYK2	2	1.3 M Betaine	Purified - Gel extraction	36.9
TYK2	3	1.3 M Betaine	Purified - Gel extraction	25.1
TYK2	4	-	-	-
TYK2	5	1.3 M Betaine	Purified - Gel extraction	67.0
TYK2	6	-	-	-
IKZF4	1	Native	Purified - Gel extraction	16.0
IKZF4	2	5% Formamide	Purified - PCR purified	13.1
IKZF4	3	Native	Purified - Gel extraction	30.4
IKZF4	4	Native	Purified - Gel extraction	18.8
IKZF4	5	Native	Purified - PCR purified	23.5
IKZF4	6	5% Formamide	Purified - PCR purified	14.9
SKIV2L	1	1.3 M Betaine	Purified - PCR purified	9.8
SKIV2L	2	1.3 M Betaine	Purified - Gel extraction	27.3
SKIV2L	3	1.3 M Betaine	Purified - Gel extraction	13.3
SKIV2L	4	1.3 M Betaine	Purified - Gel extraction	10.9
SKIV2L	5	1.3 M Betaine	Purified - Gel extraction	37.3
AKAP8L	1	1.3 M Betaine	Purified - Gel extraction	22.4
AKAP8L	2	-	-	-
AKAP8L	3	-	-	-
AKAP8L	4	Native	Purified - Gel extraction	22.0
AKAP8L	5	GC Enhancer	Purified - PCR purified	15.3
AKAP8L	6	Native	Purified - Gel extraction	28.5
SMARCC2	1	1.3 M Betaine	Purified - PCR purified	15.9
SMARCC2	2	-	-	-
SMARCC2	3	Native	Purified - Gel extraction	40.5
SMARCC2	4	Native	Purified - Gel extraction	11.9
SMARCC2	5	Native	Purified - Gel extraction	25.7
BAZ2A	1	1.3 M Betaine	Purified - Gel extraction	23.3
BAZ2A	2	1.3 M Betaine	Purified - Gel extraction	75.2
BAZ2A	3	1.3 M Betaine	Purified - Gel extraction	87.6
BAZ2A	4	1.3 M Betaine	Purified - Gel extraction	82.2
BAZ2A	5	-	-	-

Table 6.3: Purified PCR products for all 7 genes tested, with the PCR additive required for each primer pair.

Combinations of PCR additives, alongside increases and decreases in annealing temperature, were tested to determine if there was any correlation between the GC content of the genes and the degree of difficulty in amplifying PCR products. There were no overall patterns observed as genes with higher GC content, such as *SMARCC2*, were able to generate PCR products under native conditions despite having the second highest GC content of all the genes tested, with an average of 65.87%.

6.3.3 Poly(A) tailing and ligation

PCR products from *FUS_1, IKZF4, and SKIV2L* were used for initial T/A cloning tests, as the three genes represented a range of GC content, from 47.85-67.70%. PCR products were poly(A) tailed prior to cloning as *KOD* Hot Start polymerase generates blunt end products, and the A-tailed products were ligated into pGem-T Easy vectors for cloning. Ligated vectors were digested with restriction enzymes (SphI and/or EcorI) to establish whether A-tailing and ligation was successful, in which gel electrophoresis revealed the presence of bands for PCR products, vector, and the ligated vector with PCR products.

6.3.4 T/A cloning

Ligated vectors were transformed into T7 Express competent *E. coli*, demonstrating successful transformations as ampicillin^{+} agar plates grew a satisfactory number of colonies and the negative controls were absent of bacterial growth. However, DNA extracted from transformed *E. coli* was screened via PCR to verify the presence of the PCR insert. Despite numerous screening attempts and transformations, an insert of the correct size was not generated for any primer pair of the 3 genes tested.

6.4 Discussion

6.4.1 Difficulties in amplifying GC-rich DNA

As previously mentioned in section 1.5.5, it has long been known that amplifying GC-rich DNA is troublesome (Nakamura *et al.,* 2011; Dabney and Meyer, 2012; Benjamini and Speed, 2012; Ross *et al.,* 2013), with many genes remaining undiscovered or underrepresented in genome assemblies. In trying to generate PCR products from 7 avian genes, several attempts were made to overcome these obstacles with limited success. In total, 31 of the 38 primer pairs successfully generated PCR products of the correct size. However, as the size of the PCR product was the only determining factor of successful amplification, it is possible that the PCR product is an incorrect gene amplicon or the sequence has been generated with polymerase errors. This is problematic as sequencing these PCR products is the easiest way to truly determine successful amplification, and only the newest technologies (PacBio in this case) were capable of sequencing these newly found genes; this is due to the efficiency in sequencing GC-rich DNA fragments as compared to other technologies, such as Illumina (Rhoads and Au, 2015). However, it is naïve to assume that the GC content is the single cause for difficulties in amplifying and/or sequencing DNA and the many incomplete genome assemblies. For example, it has been noted that the formation of stable secondary structures arising from tandem repeats containing motifs could contribute to these problems (Beauclair *et al.,* 2019).

Nevertheless, improvements in read length and resolution in repetitive and GC-rich regions continue to reduce gaps, particularly in regions that have previously been difficult to sequence. However, until sequencing and assembly methods are able to overcome these obstacles, genome assemblies will continue to be

incomplete, with 5-20% of the genome remaining unsequenced and/or unassembled (Peona, Weissensteiner, and Suh, 2018; Jain *et al.,* 2018). Thus, this introduces bias into the results of any comparative genome studies in addition to co-localisation analyses.

6.4.2 Importance of DNA markers for genome assemblies

With the increase in *de novo* genome assemblies and the generation of chromosome-level assemblies, the availability of physical mapping data is incredibly important for anchoring DNA sequences to the correct chromosomes and in the correct orientation. The abundance of BAC libraries available for avian genomes allows for the generation of physical mapping data, yet this can only be achieved for chromosomes 1 to 28, Z, and W, due to the absence of DNA markers and sequence information for chromosomes 29 to 38. Nonetheless, the newest chicken genome assembly (Warren *et al.,* 2017) managed to assign sequence data for chromosomes 31-33, although all 3 chromosomes lack physical mapping data to verify this. Furthermore, there remains 0.23 Gbp of unplaced sequence data, which remains as unplaced scaffolds within chromosomes or as unplaced scaffolds within the whole genome (Beauclair *et al.,* 2019).

The identification of 2 BACs that localise to chromosomes smaller than 28 begins the assembly of a panel of BACs for the smallest of microchromosomes, but requires further work to ensure these are suitable for comparative mapping studies and the complete assembly of all avian genomes. Nevertheless, the presence of this sequence data, placed or unplaced, provides hope of generating complete chromosome-level assemblies in which sequence data is anchored to all chromosomes. With chromosomes 29 to 38 lacking DNA

markers, all comparative genomic analyses are unable to detect the full extent of microchromosome rearrangements. In generating markers for the smallest of avian chromosomes, it could provide an in-depth perspective into patterns of chromosomal rearrangement across avian species. This would allow us to further study homologous synteny blocks and evolutionary breakpoints, both of which contribute to the evolutionary changes that result in lineage-specific traits.

6.5 Conclusion

Having generated PCR products for the development of DNA markers for chicken chromosomes 29 to 38, this chapter explores an alternative method for identifying and anchoring sequence data to chromosomes. Whilst it is uncertain if the 7 selected genes localise to the smallest of chicken microchromosomes, it nevertheless provides a means of determining whether the newly found genes and those with reassigned placements in the *Gallus_gallus*-5.0 genome assembly are indeed correct.

The improvements in sequencing technology and the refinement of algorithms, as well as a reduction in sequencing costs, have facilitated the surge in assembling *de novo* avian genomes. Moreover, these improvements yield better quality genome assemblies, with greater coverage, larger scaffolds, and less gaps. Yet, all avian genome assemblies and comparative genomic studies remain incomplete as there is a lack of sequence and physical mapping data for all avian chromosomes. In generating a panel of BACs to physically map all chicken chromosomes, this will undoubtedly increase the availability of comparable information and thus could help to reduce the gaps present in all studies.

7 General Discussion

The overall aim of this thesis was to study chromosomal rearrangements between multiple species, and how these rearrangements may have contributed to both speciation and the development of lineage-specific traits. This thesis contributes to the continuous development of avian and non-avian reptile genome assemblies, applying both cytogenetic and *in silico* tools to generate a chromosome-level assembly and physical mapping data, through which other studies may rely on.

Of the specific aims previously outlined in the introduction, this thesis was mostly successful in its pursuit of the following:

1. The application of a universal probe set to upgrade the scaffold-based budgerigar genome assembly to that of a chromosome-level, generating an assembly comparable to those generated by Sanger sequencing approaches. The combination of FISH and RACA allowed for the identification of interchromosomal rearrangements between the chicken and the budgerigar, and between the budgerigar and other *Psittaciformes*, which is indicative of lineage-specific traits.

2. The production of cytogenomic maps using a BAC-mapping approach to investigate lineage-specific traits arising from chromosomal rearrangements. In using BACs that contained evolutionary conserved regions, this approach was effective in identifying both intrachromosomal and interchromosomal rearrangements between phylogenetically distant species. The majority of species studied exhibit the standard avian karyotype, yet the presence of previously unreported rearrangements were identified, such as tandem

duplications and translocations. This highlights the importance of comprehensive mapping approaches in comparison to the commonly used chicken chromosome paints, and allows us to infer lineage-specific traits.

3. The comparison of gross genomic structure between the yellow spotted river turtle, spiny softshell turtle, and chicken highlights the degree of genome conservation between avian and non-avian reptiles. In studying these karyotypically dissimilar species, the presence of microchromosomes as whole discrete units (fused or unfused) demonstrate that the origins of microchromosomes preceded the lineage divergence of archelosaurs from the Lepidosauromorpha. Until there is an increase in the quality and number of reptile genomes sequenced and/or mapped, the chicken continues to be a reliable source for detecting synteny, which has been demonstrated through these studies (in addition to previously published data).

However, the generation of fluorescent markers for chicken microchromosomes 29 to 38 was not entirely successful, with difficulties in cloning the 7 selected chicken genes into competent *E. coli.* The method for generating these markers was PCR-based, which has previously proven to be difficult as problems arise due to high GC content and the formation of stable secondary structures (Beauclair *et al.,* 2019). Nevertheless, this approach successfully generated PCR products of the correct length and provides hope for future work.

7.1 Generating Avian Genome Assemblies

In identifying the chromosomal rearrangements that result in speciation, it is incredibly important to have high quality genome assemblies for the reference genome, an outgroup, and that of the genome being studied. To date there are

many sequenced avian genomes, of which some are considered to be that of a chromosome-level but lack any physical mapping data. This lack of data hinders the anchoring of contigs to chromosomes and the identification of mis-assembly errors, and does little to minimise the bias introduced from using reference genomes. However, the tools developed by Damas *et al.* (2017) allows for the assembly of avian genomes to that of a chromosome-level using a combination of *in silico* and molecular cytogenetic tools, helping to overcome these limitations.

In applying this methodology, the budgerigar genome was assembled to a high quality, allowing us to investigate the genome of a species exhibiting a high degree of chromosomal rearrangement. Using FISH to validate the *in silico* data, we can infer with confidence that the interchromosomal changes identified in chapter 3 are unique to the budgerigar lineage. Yet, even in generating this genome assembly and physically validating the rearrangements, the lack of chromosome-level assemblies for the majority of avian species limits our analyses of chromosomal rearrangements to determine genotype to phenotype associations and the development of lineage-specific traits.

7.2 Methods to Study Chromosome Rearrangements in Avian and Non-avian Reptiles

In the absence of sequence data, efforts to bridge the gaps in our knowledge and understanding of chromosomal rearrangements both in and between taxa continue to develop. By utilising other methods to study genome conservation, such as comparative chromosome painting and BAC mapping, we have significantly improved our knowledge on chromosome homology and karyotype evolution. A key example is discussed in chapters 5 of this thesis, whereby the

relationship of turtles to birds was largely debated (and to a degree still is). However, in using these methods, data has been generated that provides irrefutable proof of genome conservation. Furthermore, this is also demonstrated in chapter 4 of this thesis, whereby the atypical karyotype of the Eurasian woodcock (2*n*= 96) has arisen from multiple fissions of the macrochromosomes.

7.3 Tracing Chromosome Evolution

The data generated in chapter 5 of this thesis contributed to the reconstruction of the most likely ancestral karyotype of diapsids, in which the genomic structure was inferred from combining molecular cytogenetic and bioinformatic approaches. Specifically, this chapter provided FISH data (both chromosome painting and BAC mapping) to detect chromosome homology and identify rearrangements between the chicken and two turtle species. In using this methodology, the intra- and interchromosomal changes were traced from the ancestral diapsid ancestor to modern birds through the theropod dinosaur lineage.

With references to dinosaurs, a subject beloved to many science fiction fans, this research garnered attention worldwide. According to Altmetric, a website dedicated to determining the metrics of journal articles beyond that of citation-based metrics, the journal article in which this data was published is considered to be in the top 5% of all research outputs scored (as shown in Figure 7.1). Furthermore, the research was covered by 37 news outlets, with one of the most well known being the BBC (https://www.bbc.co.uk/news/science-environment-44711974, accessed on January 16th 2020).

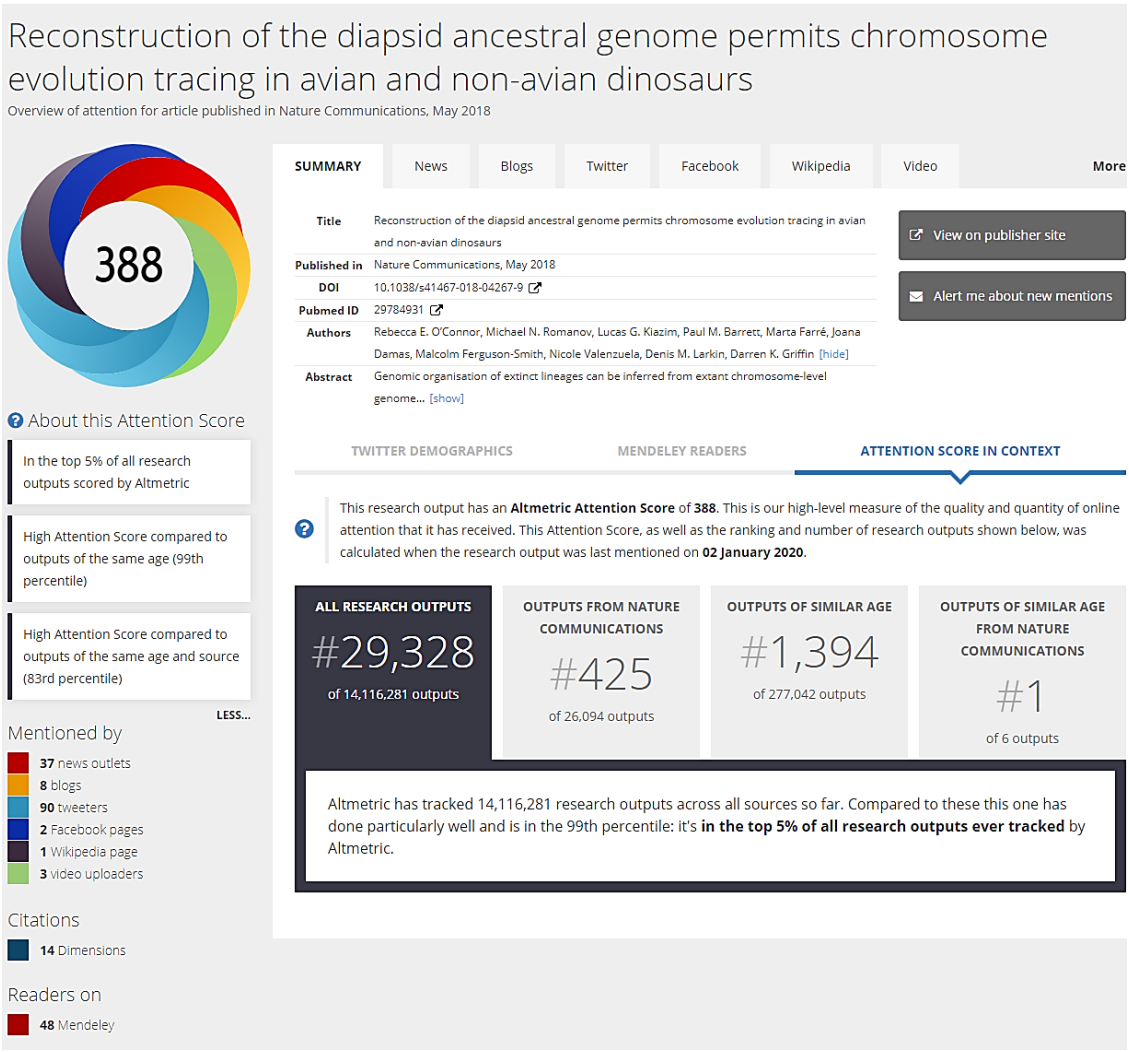


Figure 7.1: Altmetric data on the journal article "Reconstruction of the diapsid ancestral genome permits chromosome evolution tracing in avian and non-avian dinosaurs". (https://www.altmetric.com/details/42176906#score)

7.4 Generating Markers for Avian Microchromosomes

To date, the chicken genome remains the best assembled and annotated of all avian species, yet there is minimal sequence data and a complete lack of physical markers for microchromosomes 29 to 38. Whilst this lack of markers hinders our ability to validate any *in silico* data of the newest chicken genome assembly (*Gallus_gallus*-5.0; GCA _ 000002315.3), it also hinders our ability to detect the full extent of microchromosomal rearrangements between species. For example, in the assembly of the budgerigar genome, homology between the

budgerigar and chicken was only established for chromosomes 1-28 and Z (excluding chromosome 16) due to the lack/absence of sequence data, the unvalidated placement of sequence data for chicken chromosomes 30-33, and the inability to physically map these chromosomes.

Furthermore, in addition to preventing the generation of truly complete genome assemblies (in which sequence data is available for all avian chromosomes), the absence of these microchromosomal markers also impacts comparative mapping studies. Previous research focusing on microchromosome organisation by O'Connor *et al.* (2018b) studied 22 avian species from 10 orders, highlighting the extraordinary degree of conservation in the genome organisation, with exceptions in species exhibiting atypical avian karyotypes (such as the *Psittaciformes* and *Falconiformes*). However, it cannot be determined whether this pattern is truly conserved given the absence of data for microchromosomes 29 to 38, as without these physical markers there may be undetected inter- and/or intrachromosomal rearrangements. Moreover, this uncertainty may also be extended to comparative studies in non-avian reptiles, whereby avian BACs have been used to investigate genome conservation between birds and reptiles.

If it is at all possible to generate markers for chicken chromosomes 29 to 38 (given the difficulties that researchers have been facing since 2004), then all chromosome-level assemblies and comparative mapping studies would need to be revisited. These markers could allow for anchoring of unplaced sequences onto chromosomes and the anchoring of contigs assigned to microchromosomes to validate *in silico* data. Moreover, for species that exhibit atypical karyotypes or vast phenotypic diversity, the organisation of these

microchromosomes may help us more accurately elucidate which chromosomal rearrangements result in reproductive isolation and subsequent speciation, and the development of lineage-specific traits.

7.5 Future Work

Whilst this thesis has been as comprehensive as possible (bearing in mind the limitations of tools available, financial expenditure, and time constraints), there are still areas in which further research is required. Reflecting on the summation of each chapter, some key areas have been identified and are as follows:

1. The assembly of the budgerigar genome and the houbara genome have identified assembly errors due to over-aggressive scaffolding approaches, highlighting the importance of physical mapping in assembling genomes. With the availability of the avian universal probe set, this panel can be applied to detect any mis-assemblies in the numerous avian chromosome-level assemblies generated entirely using *in silico* methods. This provides the opportunity to either identify novel rearrangements or dismiss errors, refining genome assemblies whilst also furthering our understanding of avian genome evolution.

2. The comparative BAC mapping in this thesis focuses solely on the use of chicken BACs, which can be limiting in the successful hybridisation of BACs for species further diverged/rearranged. Moreover, due to the paucity of BACs on some chromosomes, rearrangements could remain undetected, and thus the use of zebra finch BACs may overcome these problems. The zebra finch and chicken lineages diverged approximately 98 million years ago, and consequently, the combination of both chicken and zebra finch BACs can be used to distinguish features that are

characteristic of all birds, and those features that differ between the Galloanserae and Neoaves.

3. Of the BACs previously developed by Damas *et al.* (2017), 73 have been shown to successfully work across multiple reptile species. However, this panel has been assembled through combining data on studies focusing on individual species (for example turtles and lizard by O'Connor *et al.* (2018c) and snakes by Singchat *et al.* (2018)). By using *in silico* methods to determine evolutionary conserved sequences, it could allow for the generation of a BAC panel to fully characterise non-avian reptile genomes.

4. The development of fluorescent markers for chicken microchromosomes 29-38 that would allow for the full characterisation and assembly of the chicken genome, which in turn would benefit comparative analyses and chromosome-level genome assemblies. To overcome the issues encountered in chapter 6, it has been speculated whether the application of pulse field gel electrophoresis could also be used to separate the smallest of avian microchromosomes. A study by Pichugin *et al.* (2001) used pulse field gel electrophoresis to estimate the size of chicken chromosomes, and by optimising this technique, it may be possible to isolate DNA from the smallest of microchromosomes.

7.6 Personal Perspectives

The last three years have been incredibly challenging yet extremely rewarding, and in writing this thesis it has highlighted the impact my PhD has had both personally and academically. Owing to my research and the efforts of everyone involved, I am a co-author on 5 published papers and 4 manuscripts in preparation, which gives me something to be endlessly proud of and something to always aspire to.

Academically, I have been able to attend national and international conferences thanks to the generosity of my supervisor and academic societies. This has provided the opportunity to collaborate with a variety of world-class research groups, disseminate my work, develop upon my interpersonal skills, and travel around Europe. Furthermore, it provided me with the confidence to train many undergraduate and postgraduate students, in addition to visiting international students and collaborators.

Personally, conducting research in a lab taught me to be resilient, motivated, unwaveringly dedicated, and thorough in everything that I do. Yet this would not have been possible without the people that I have met throughout this journey. Arguably more important than the creation of this thesis, my PhD has created friendships and memories that I will treasure forever, and I will never forget the kindness and support that everyone has shown to me.

8 References

Ahituv, N., Prabhakar, S., Poulin, F., Rubin, E.M. and Couronne, O., 2005. Mapping cis-regulatory domains in the human genome using multi-species conservation of synteny. *Human molecular genetics*, 14(20), pp.3057-3063.

Ahn, J.W., Coldwell, M., Bint, S. and Ogilvie, C.M., 2015. Array comparative genomic hybridization (array CGH) for detection of genomic copy number variants. *JoVE (Journal of Visualized Experiments)*, (96), p.e51718.

Alam, S.M.I., Sarre, S.D., Gleeson, D., Georges, A. and Ezaz, T., 2018. Did Lizards Follow Unique Pathways in Sex Chromosome Evolution? *Genes*, 9(5), p.239.

Alexander, D.J., 2000. A review of avian influenza in different bird species. *Veterinary microbiology*, 74(1-2), pp.3-13.

Alkan, C., Sajjadian, S. and Eichler, E.E., 2011. Limitations of next-generation genome sequence assembly. *Nature methods*, 8(1), p.61.

Andersson, L. and Georges, M., 2004. Domestic-animal genomics: deciphering the genetics of complex traits. *Nature Reviews Genetics*, 5(3), p.202.

Ariyadasa, R. and Stein, N., 2012. Advances in BAC-Based Physical Mapping and Map Integration Strategies in Plants. *Journal of Biomedicine and Biotechnology*, 2012, pp.1-11.

Arsham, M.S., Barch, M.J. and Lawce, H.J. eds., 2017. The AGT cytogenetics laboratory manual. *John Wiley & Sons*.

Au, K.F., Underwood, J.G., Lee, L. and Wong, W.H., 2012. Improving PacBio long read accuracy by short read alignment. *PloS one*, 7(10), p.e46679.

Ayres, M., Sampaio, M., Barros, R., Dias, L. and Cunha, O., 1969. A karyological study of turtles from the Brazilian Amazon Region. *Cytogenetic and Genome Research*, 8(6), pp.401-409.

Badenhorst, D., Stanyon, R., Engstrom, T. and Valenzuela, N., 2013. A ZZ/ZW microchromosome system in the spiny softshell turtle, Apalone spinifera, reveals an intriguing sex chromosome conservation in Trionychidae. *Chromosome Research*, 21(2), pp.137-147.

Balakrishnan, C.N., Edwards, S.V. and Clayton, D.F., 2010. The Zebra Finch genome and avian genomics in the wild. *Emu-Austral Ornithology*, 110(3), pp.233-241.

Ballif, B.C., Theisen, A., Coppinger, J., Gowans, G.C., Hersh, J.H., Madan-Khetarpal, S., Schmidt, K.R., Tervo, R., Escobar, L.F., Friedrich, C.A. and McDonald, M., 2008. Expanding the clinical phenotype of the 3q29 microdeletion syndrome and characterization of the reciprocal microduplication. *Molecular Cytogenetics*, 1(1), p.8.

Baptista, J., Prigmore, E., Gribble, S.M., Jacobs, P.A., Carter, N.P. and Crolla, J.A., 2005. Molecular cytogenetic analyses of breakpoints in apparently balanced reciprocal translocations carried by phenotypically normal individuals. *European journal of human genetics*, 13(11), p.1205.

Baptista, J., Mercer, C., Prigmore, E., Gribble, S.M., Carter, N.P., Maloney, V., Thomas, N.S., Jacobs, P.A. and Crolla, J.A., 2008. Breakpoint mapping and array CGH in translocations: comparison of a phenotypically normal and an abnormal cohort. *The American Journal of Human Genetics*, 82(4), pp.927-936.

Barros R.M., Sampaio M.M., Assis M.F., and Ayres, M., (1976) General considerations on the karyotypic evolution of Chelonia from the Amazon region of Brazil. *Cytologia* 41:559-565.

Beauclair, L., Ramé, C., Arensburger, P., Piégu, B., Guillou, F., Dupont, J. and Bigot, Y., 2019. Sequence properties of certain GC rich avian genes, their origins and absence from genome assemblies: case studies. *BMC genomics*, 20(1), pp.1-16.

Bellott, D.W., Skaletsky, H., Pyntikova, T., Mardis, E.R., Graves, T., Kremitzki, C., Brown, L.G., Rozen, S., Warren, W.C., Wilson, R.K. and Page, D.C., 2010.

Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. *Nature*, 466(7306), p.612.

Benjamini, Y. and Speed, T.P., 2012. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic acids research*, 40(10), pp.e72-e72.

Benton, M.J. and Donoghue, P.C., 2006. Paleontological evidence to date the tree of life. *Molecular biology and evolution*, 24(1), pp.26-53.

Berend, S.A., Horwitz, J., McCaskill, C. and Shaffer, L.G., 2000. Identification of uniparental disomy following prenatal detection of Robertsonian translocations and isochromosomes. *The American Journal of Human Genetics*, 66(6), pp.1787-1793.

Berggren, H., Nordahl, O., Tibblin, P., Larsson, P. and Forsman, A., 2016. Testing for local adaptation to spawning habitat in sympatric subpopulations of pike by reciprocal translocation of embryos. *PloS one*, 11(5), p.e0154488.

Berv, J.S. and Field, D.J., 2017. Genomic signature of an avian Lilliput effect across the K-Pg extinction. *Systematic biology*, 67(1), pp.1-13.

Besser, J., Carleton, H.A., Gerner-Smidt, P., Lindsey, R.L. and Trees, E., 2018. Next-generation sequencing technologies and their application to the study and control of bacterial infections. *Clinical microbiology and infection*, 24(4), pp.335-341.

Bezerra, D.M.M., de Araujo, H.F.P., Alves, Â.G.C. and Alves, R.R.N., 2013. Birds and people in semiarid northeastern Brazil: symbolic and medicinal relationships. *Journal of Ethnobiology and Ethnomedicine*, 9(1), p.3.

Bickham, J.W. and Carr, J.L., 1983. Taxonomy and phylogeny of the higher categories of cryptodiran turtles based on a cladistic analysis of chromosomal data. *Copeia*, pp.918-932.

Bickham J.W., Tucker P.K. and Legler J.M., (1985) Diploid-triploid mosaicism: An unusual phenomenon in side-necked turtles (Platemys platycephala). *Science* 227:1591-1593.

Bieluszewska, A., Weglewska, M., Bieluszewski, T., Lesniewicz, K. and Poreba, E., 2018. PKA-binding domain of AKAP 8 is essential for direct interaction with DPY 30 protein. *The FEBS journal*, 285(5), pp.947-964.

BirdLife International. 2016. Chlamydotis undulata. *The IUCN Red List of Threatened Species 2016*: e.T22728245A90341807. http://dx.doi.org/10.2305/IUCN.UK.2016- 3.RLTS.T22728245A90341807.en

Bishop, R., 2010. Applications of fluorescence in situ hybridization (FISH) in detecting genetic aberrations of medical significance. *Bioscience Horizons*, 3(1), pp.85-95.

Bornelöv, S., Seroussi, E., Yosefi, S., Pendavis, K., Burgess, S.C., Grabherr, M., Friedman-Einat, M. and Andersson, L., 2017. Correspondence on Lovell et al.: identification of chicken genes previously assumed to be evolutionarily lost. *Genome biology*, 18(1), p.112.

Botero-Castro, F., Figuet, E., Tilak, M.K., Nabholz, B. and Galtier, N., 2017. Avian genomes revisited: hidden genes uncovered and the rates versus traits paradox in birds. *Molecular biology and evolution*, 34(12), pp.3123-3131.

Boue, A. and Gallano, P., 1984. A collaborative study of the segregation of inherited chromosome structural rearrangements in 1356 prenatal diagnoses. *Prenatal diagnosis*, 4(7), pp.45-67.

Braekeleer, M.D. and Dao, T.N., 1990. Cytogenetic studies in couples experiencing repeated pregnancy losses. *Human Reproduction*, 5(5), pp.519-528.

Braña, F., Prieto, L. and González-Quirós, P., 2010. Habitat Change and Timing of Dusk Flight in the Eurasian Woodcock: A Trade-Off between Feeding and Predator Avoidance? *Annales Zoologici Fennici*, 47(3), pp.206-214.

Brancati, F., Mingarelli, R. and Dallapiccola, B., 2003. Recurrent triploidy of maternal origin. *European journal of human genetics*, 11(12), p.972.

Branco, M.R. and Pombo, A., 2006. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS biology*, 4(5), p.e138.

Brauth, S., Liang, W., Roberts, T.F., Scott, L.L. and Quinlan, E.M., 2002. Contact call-driven zenk protein induction and habituation in telencephalic auditory pathways in the budgerigar (Melopsittacus undulatus): Implications for understanding vocal learning processes. *Learning & Memory*, 9(2), pp.76-88.

Bridge, J.A., 2008. Advantages and limitations of cytogenetic, molecular cytogenetic, and molecular diagnostic testing in mesenchymal neoplasms. *Journal of Orthopaedic Science*, 13(3), p.273.

Burt, D.W., Bruley, C., Dunn, I.C., Jones, C.T., Ramage, A., Law, A.S., Morrice, D.R., Paton, I.R., Smith, J., Windsor, D. and Sazanov, A., 1999. The dynamics of chromosome evolution in birds and mammals. *Nature*, 402(6760), p.411.

Burt, D.W., 2002. Origin and evolution of avian microchromosomes. *Cytogenetic and genome research*, 96(1-4), pp.97-112.

Burt, D.W., 2005. Chicken genome: current status and future opportunities. *Genome research*, 15(12), pp.1692-1698.

Bush, G.L., Case, S.M., Wilson, A.C. and Patton, J.L., 1977. Rapid speciation and chromosomal evolution in mammals. *Proceedings of the National Academy of Sciences*, *74*(9), pp.3942-3946.

Canapa, A., Barucca, M., Biscotti, M.A., Forconi, M. and Olmo, E., 2015. Transposons, genome size, and evolutionary insights in animals. *Cytogenetic and genome research*, 147(4), pp.217-239.

Cardoso, D.C., das Graças Pompolo, S., Cristiano, M.P. and Tavares, M.G., 2014. The role of fusion in ant chromosome evolution: insights from cytogenetic

analysis using a molecular phylogenetic approach in the genus Mycetophylax. *PLoS One*, 9(1), p.e87473.

Carter, N.P., Bebb, C.E., Nordenskjo, M., Ponder, B.A. and Tunnacliffe, A., 1992. Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics*, 13(3), pp.718-725.

Casbon, J.A., Osborne, R.J., Brenner, S. and Lichtenstein, C.P., 2011. A method for counting PCR template molecules with application to next-generation sequencing. *Nucleic acids research*, 39(12), pp.e81-e81.

Chaisson, M.J., Wilson, R.K. and Eichler, E.E., 2015. Genetic variation and the de novo assembly of human genomes. *Nature Reviews Genetics*, 16(11), p.627.

Chen, M., Presting, G., Barbazuk, W.B., Goicoechea, J.L., Blackmon, B., Fang, G., Kim, H., Frisch, D., Yu, Y., Sun, S. and Higingbottom, S., 2002. An integrated physical and genetic map of the rice genome. *The Plant Cell*, 14(3), pp.537-545.

Chen, Y.C., Liu, T., Yu, C.H., Chiang, T.Y. and Hwang, C.C., 2013. Effects of GC bias in next-generation-sequencing data on de novo genome assembly. *PloS one*, 8(4), p.e62856.

Chen, L.C., Lan, H., Sun, L., Deng, Y.L., Tang, K.Y. and Wan, Q.H., 2015. Genomic organization of the crested ibis MHC provides new insight into ancestral avian MHC structure. *Scientific reports*, 5, p.7963.

Chen, J., Herlong, F.H., Stroehlein, J.R. and Mishra, L., 2016. Mutations of chromatin structure regulating genes in human malignancies. *Current Protein and Peptide Science*, 17(5), pp.411-437.

Cheng, Y. and Burt, D.W., 2018. Chicken genomics. *International Journal of Developmental Biology*, 62(1-2-3), pp.265-271.

Chi, J.X., Huang, L., Nie, W., Wang, J., Su, B. and Yang, F., 2005. Defining the orientation of the tandem fusions that occurred during the evolution of Indian muntjac chromosomes by BAC mapping. *Chromosoma*, 114(3), pp.167-172.

Chiang, T., Duncan, F.E., Schindler, K., Schultz, R.M. and Lampson, M.A., 2010. Evidence that weakened centromere cohesion is a leading cause of age-related aneuploidy in oocytes. *Current Biology*, 20(17), pp.1522-1528.

Chiatante, G., Giannuzzi, G., Calabrese, F.M., Eichler, E.E. and Ventura, M., 2017. Centromere destiny in dicentric chromosomes: new insights from the evolution of human chromosome 2 ancestral centromeric region. *Molecular biology and evolution*, 34(7), pp.1669-1681.

Chiari, Y., Cahais, V., Galtier, N. and Delsuc, F., 2012. Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). *BMC Biology*, 10(1), p.65.

Christidis, L., 1990. Animal Cytogenetics. Vol. 4, Chordata 3 B. *Aves*. The Auk, 108(3), pp.748-748.

Cohen, M.M. and Gans, C., 1970. The chromosomes of the order Crocodilia. *Cytogenetic and Genome Research*, 9(2), pp.81-105.

Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gökçümen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P. and Fitzgerald, T., 2010. Origins and functional impact of copy number variation in the human genome. *Nature*, 464(7289), p.704.

Cook Jr., E.H., Lindgren, V., Leventhal, B.L., Courchesne, R., Lincoln, A., Shulman, C., Lord, C. and Courchesne, E., 1997. Autism or atypical autism in maternally but not paternally derived proximal 15q duplication. *American journal of human genetics*, 60(4), p.928.

Crasta, K., Ganem, N.J., Dagher, R., Lantermann, A.B., Ivanova, E.V., Pan, Y., Nezi, L., Protopopov, A., Chowdhury, D. and Pellman, D., 2012. DNA breaks and chromosome pulverization from errors in mitosis. *Nature*, 482(7383), p.53.

Crawford, N.G., Faircloth, B.C., McCormack, J.E., Brumfield, R.T., Winker, K. and Glenn, T.C., 2012. More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. *Biology letters*, 8(5), pp.783-786.

Croll, D., Zala, M. and McDonald, B.A., 2013. Breakage-fusion-bridge cycles and large insertions contribute to the rapid evolution of accessory chromosomes in a fungal pathogen. *PLoS genetics*, 9(6), p.e1003567.

Dabney, J. and Meyer, M., 2012. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques*, 52(2), pp.87-94.

Damas, J., O'connor, R.E., Farré, M., Lenis, V.P.E., Martell, H.J., Mandawala, A., Fowler, K., Joseph, S., Swain, M.T., Griffin, D.K. and Larkin, D.M., 2017. Upgrading short-read animal genome assemblies to chromosome level using comparative genomics and a universal probe set. *Genome research*, 27(5), pp.875-884.

De Boer, L.E.M., 1975. Karyological heterogeneity in the Falconiformes (Aves). *Experientia*, 31(10), pp.1138-1139.

De Oliveira, E.H.C., Tagliarini, M.M., Rissino, J.D., Pieczarka, J.C., Nagamachi, C.Y., O'Brien, P.C. and Ferguson-Smith, M.A., 2010. Reciprocal chromosome painting between white hawk (Leucopternis albicollis) and chicken reveals extensive fusions and fissions during karyotype evolution of Accipitridae (Aves, Falconiformes). *Chromosome Research*, 18(3), pp.349-355.

de Villena, F.P.M. and Sapienza, C., 2001. Female meiosis drives karyotypic evolution in mammals*. Genetics*, 159(3), pp.1179-1189.

De, S., Teichmann, S.A. and Babu, M.M., 2009. The impact of genomic neighborhood on the evolution of human and chimpanzee transcriptome. *Genome research*, 19(5), pp.785-794.

Deakin, J.E. and Ezaz, T., 2019. Understanding the Evolution of Reptile Chromosomes through Applications of Combined Cytogenetics and Genomics Approaches. *Cytogenetic and genome research*, 157(1-2), pp.7-20.

DeBolt, S., 2010. Copy number variation shapes genome diversity in Arabidopsis over immediate family generational scales. *Genome biology and evolution*, 2, pp.441-453.

Deininger, P.L. and Batzer, M.A., 1999. Alu repeats and human disease. *Molecular genetics and metabolism*, *67*(3), pp.183-193.

Degrandi, T.M., del Valle Garnero, A., O'Brien, P.C., Ferguson-Smith, M.A., Kretschmer, R., de Oliveira, E.H. and Gunski, R.J., 2017. Chromosome painting in Trogon s. surrucura (Aves, Trogoniformes) reveals a karyotype derived by chromosomal fissions, fusions, and inversions. *Cytogenetic and genome research*, 151(4), pp.208-215.

Demyda-Peyrás, S., Dorado, J., Hidalgo, M. and Moreno-Millán, M., 2015. Influence of sperm fertilising concentration, sperm selection method and sperm capacitation procedure on the incidence of numerical chromosomal abnormalities in IVF early bovine embryos. *Reproduction, Fertility and Development*, 27(2), pp.351-359.

Denton, J.F., Lugo-Martinez, J., Tucker, A.E., Schrider, D.R., Warren, W.C. and Hahn, M.W., 2014. Extensive error in the number of genes inferred from draft genome assemblies*. PLoS computational biology*, 10(12), p.e1003998.

Derjusheva, S., Kurganova, A., Habermann, F. and Gaginskaya, E., 2004. High chromosome conservation detected by comparative chromosome painting in chicken, pigeon and passerine birds. *Chromosome Research,* 12(7), p.715.

Di Tomaso, M.V., Liddle, P., Lafon-Hughes, L., Reyes-Ábalos, A. and Folle, G., 2013. Chromatin damage patterns shift according to eu/heterochromatin replication. *The mechanisms of dNA replication. d S (ed.) InTech.* pp. 351-375

Dobigny, G., Britton-Davidian, J. and Robinson, T.J., 2017. Chromosomal polymorphism in mammals: an evolutionary perspective. *Biological Reviews*, 92(1), pp.1-21.

Doheny, K.F., Rasmussen, S.A., Rutberg, J., Semenza, G.L., Stamberg, J., Schwartz, M., Batista, D.A., Stetten, G. and Thomas, G.H., 1997. Segregation of a familial balanced (12; 10) insertion resulting in Dup (10)(q21. 2q22. 1) and Del (10)(q21. 2q22. 1) in first cousins. *American journal of medical genetics*, *69*(2), pp.188-193.

Dorshorst, B., Molin, A.M., Rubin, C.J., Johansson, A.M., Strömstedt, L., Pham, M.H., Chen, C.F., Hallböök, F., Ashwell, C. and Andersson, L., 2011. A complex genomic rearrangement involving the endothelin 3 locus causes dermal hyperpigmentation in the chicken. *PLoS genetics*, 7(12), p.e1002412.

dos Santos, M.D.S., Kretschmer, R., Frankl-Vilches, C., Bakker, A., Gahr, M., Ferguson-Smith, M.A. and de Oliveira, E.H., 2017. Comparative cytogenetics between two important songbird, models: The zebra finch and the canary. *PloS one*, 12(1), p.e0170997.

Du, L., Schageman, J.J., Subauste, M.C., Saber, B., Hammond, S.M., Prudkin, L., Wistuba, I.I., Ji, L., Roth, J.A., Minna, J.D. and Pertsemlidis, A., 2009. miR-93, miR-98, and miR-197 regulate expression of tumor suppressor gene FUS1. *Molecular Cancer Research*, 7(8), pp.1234-1243.

Dumas, F. and Mazzoleni, S., 2017. Neotropical primate evolution and phylogenetic reconstruction using chromosomal data. *The European Zoological Journal*, 84(1), pp.1-18.

Dumas, L., Kim, Y.H., Karimpour-Fard, A., Cox, M., Hopkins, J., Pollack, J.R. and Sikela, J.M., 2007. Gene copy number variation spanning 60 million years of human and primate evolution. *Genome research*, 17(9), pp.1266-1277.

Duriez, O., Eraud, C., Barbraud, C. and Ferrand, Y., 2005. Factors affecting population dynamics of Eurasian woodcocks wintering in France: assessing the efficiency of a hunting-free reserve. *Biological Conservation*, 122(1), pp.89-97.

Edwards, S.V., Kingan, S.B., Calkins, J.D., Balakrishnan, C.N., Jennings, W.B., Swanson, W.J. and Sorenson, M.D., 2005. Speciation in birds: genes, geography, and sexual selection. *Proceedings of the National Academy of Sciences*, 102(suppl 1), pp.6550-6557.

Efimova, A.D., Ovchinnikov, R.K., Roman, A.Y., Maltsev, A.V., Grigoriev, V.V., Kovrazhkina, E.A. and Skvortsova, V.I., 2017. the fus protein: Physiological functions and a role in amyotrophic lateral sclerosis. *Molecular Biology*, 51(3), pp.341-351.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B. and Bibillo, A., 2009. Real-time DNA sequencing from single polymerase molecules. *Science*, 323(5910), pp.133-138.

Ekblom, R., Stapley, J., Ball, A.D., Birkhead, T., Burke, T. and Slate, J., 2011. Genetic mapping of the major histocompatibility complex in the zebra finch (Taeniopygia guttata). *Immunogenetics,* 63(8), pp.523-530.

Elsik, C.G., Tellam, R.L. and Worley, K.C., 2009. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science*, 324(5926), pp.522-528.

Fabre, A., Martinez-Vinson, C., Goulet, O. and Badens, C., 2013. Syndromic diarrhea/Tricho-hepato-enteric syndrome. *Orphanet journal of rare diseases,* 8(1), p.5.

Fang, Z. and Morrell, P.L. (2016). Domestication: Polyploidy boosts domestication*. Nature Plants*, 2(8).

Fantin, C. and Monjeló, L.A.D.S., 2011. Cytogenetic studies in Podocnemis expansa and Podocnemis sextuberculata (Testudines, Podocnemididae), turtles of the Brazilian Amazon. *Caryologia*, 64(2), pp.154-157.

Farre, M., Narayan, J., Slavov, G.T., Damas, J., Auvil, L., Li, C., Jarvis, E.D., Burt, D.W., Griffin, D.K. and Larkin, D.M., 2016. Novel insights into chromosome evolution in birds, archosaurs, and reptiles. *Genome biology and evolution*, 8(8), pp.2442-2451.

Farré, M., Kim, J., Proskuryakova, A.A., Zhang, Y., Kulemzina, A.I., Li, Q., Zhou, Y., Xiong, Y., Johnson, J.L., Perelman, P.L. and Johnson, W.E., 2019. Evolution of gene regulation in ruminants differs between evolutionary breakpoint regions and homologous synteny blocks. *Genome research*, 29(4), pp.576-589.

Federico, C., Cantarella, C.D., Scavo, C., Saccone, S., Bed'Hom, B. and Bernardi, G., 2005. Avian genomes: different karyotypes but a similar distribution of the GC-richest chromosome regions at interphase. *Chromosome Research*, 13(8), pp.785-793.

Ferguson-Smith, M.A., Yang, F., Rens, W. and O'Brien, P.C.M., 2005. The impact of chromosome sorting and painting on the comparative analysis of primate genomes. *Cytogenetic and genome research*, 108(1-3), pp.112-121.

Fillon, V., Vignoles, M., Crooijmans, R.P.M.A., Groenen, M.A.M., Zoorob, R. and Vignal, A., 2007. FISH mapping of 57 BAC clones reveals strong conservation of synteny between Galliformes and Anseriformes*. Animal genetics*, 38(3), pp.303-307.

Fishman, L., Stathos, A., Beardsley, P.M., Williams, C.F. and Hill, J.P., 2013. Chromosomal rearrangements and the genetics of reproductive barriers in Mimulus (monkey flowers). *Evolution*, 67(9), pp.2547-2560.

Fitzgerald, P.H., 1975. A mechanism of X chromosome aneuploidy in lymphocytes of aging women. *Human Genetics*, *28*(2), pp.153-158.

Fong, P., 1967. Packing of the DNA molecule. *Journal of theoretical biology*, *15*(2), pp.230-235.

Frankl-Vilches, C., Kuhl, H., Werber, M., Klages, S., Kerick, M., Bakker, A., de Oliveira, E.H., Reusch, C., Capuano, F., Vowinckel, J. and Leitner, S., 2015. Using the canary genome to decipher the evolution of hormone-sensitive gene regulation in seasonal singing birds. *Genome biology*, 16(1), p.19.

Frohlich, J., Kubickova, S., Musilova, P., Cernohorska, H., Muskova, H., Vodicka, R. and Rubes, J., 2017. Karyotype relationships among selected deer species and cattle revealed by bovine FISH probes. *PloS one*, 12(11), p.e01

Fryns, J.P., Kleczkowska, A., Kubień, E. and Van den Berghe, H., 1986. Excess of mental retardation and/or congenital malformation in reciprocal translocations in man. *Human genetics*, *72*(1), pp.1-8. 87559.

Furo, I., Kretschmer, R., O'Brien, P., Pereira, J., Garnero, A., Gunski, R., Ferguson-Smith, M. and de Oliveira, E., 2018. Chromosome Painting in Neotropical Long-and Short-Tailed Parrots (Aves, Psittaciformes): Phylogeny and Proposal for a Putative Ancestral Karyotype for Tribe Arini. *Genes,* 9(10), p.491.

Gahr, M., 2000. Neural song control system of hummingbirds: comparison to swifts, vocal learning (songbirds) and nonlearning (suboscines) passerines, and vocal learning (budgerigars) and nonlearning (dove, owl, gull, quail, chicken) nonpasserines. *Journal of Comparative Neurology*, 426(2), pp.182-196.

Gall, J.G. and Pardue, M.L., 1969. Formation and detection of RNA-DNA hybrid molecules in cytological preparations*. Proceedings of the National Academy of Sciences*, 63(2), pp.378-383.

Ganapathy, G., Howard, J.T., Ward, J.M., Li, J., Li, B., Li, Y., Xiong, Y., Zhang, Y., Zhou, S., Schwartz, D.C. and Schatz, M., 2014. High-coverage sequencing and annotated assemblies of the budgerigar genome. *GigaScience,* 3(1), p.11.

Ganem, N.J., Godinho, S.A. and Pellman, D., 2009. A mechanism linking extra centrosomes to chromosomal instability. *Nature*, 460(7252), p.278.

Gazave, E., Darré, F., Morcillo-Suarez, C., Petit-Marty, N., Carreño, A., Marigorta, U.M., Ryder, O.A., Blancher, A., Rocchi, M., Bosch, E. and Baker, C., 2011. Copy number variation analysis in the great apes reveals species-specific patterns of structural variation. *Genome research*, 21(10), pp.1626-1639.

Ghurye, J. and Pop, M., 2019. Modern technologies and algorithms for scaffolding assembled genomes. *PLoS computational biology*, 15(6), p.e1006994.

Giannuzzi, G., Pazienza, M., Huddleston, J., Antonacci, F., Malig, M., Vives, L., Eichler, E.E. and Ventura, M., 2013. Hominoid fission of chromosome 14/15 and the role of segmental duplications. *Genome research*, 23(11), pp.1763-1773.

Gifalli-Iughetti, C. and Koiffmann, C.P., 2009. Synteny of human chromosomes 14 and 15 in the platyrrhines (Primates, Platyrrhini). *Genetics and molecular biology*, 32(4), pp.786-791.

Gilbert, W. and Maxam, A., 1973. The nucleotide sequence of the lac operator. *Proceedings of the National Academy of Sciences*, 70(12), pp.3581-3584.

Gilbert, S.F., Loredo, G.A., Brukman, A. and Burke, A.C., 2001. Morphogenesis of the turtle shell: the development of a novel structure in tetrapod evolution. *Evolution & development*, 3(2), pp.47-58.

Gill, F., and Donsker, D., (Eds). 2019. IOC World Bird List (v 9.2). Doi 10.14344/IOC.ML.9.2. *http://www.worldbirdnames.org/*

Giovannotti, M., Caputo, V., O'Brien, P.C.M., Lovell, F.L., Trifonov, V., Cerioni, P.N., Olmo, E., Ferguson-Smith, M.A. and Rens, W., 2009. Skinks (Reptilia: Scincidae) have highly conserved karyotypes as revealed by chromosome painting. *Cytogenetic and genome research*, 127(2-4), pp.224-231.

Gökçümen, Ö. and Lee, C., 2009. Copy number variants (CNVs) in primate species using array-based comparative genomic hybridization. *Methods*, 49(1), pp.18-25.

Goodwin, S., McPherson, J.D. and McCombie, W.R., 2016. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6), p.333.

Gordon, D., Huddleston, J., Chaisson, M.J., Hill, C.M., Kronenberg, Z.N., Munson, K.M., Malig, M., Raja, A., Fiddes, I., Hillier, L.W. and Dunn, C., 2016. Long-read sequence assembly of the gorilla genome. *Science*, *352*(6281).

Graphodatsky, A.S., Trifonov, V.A. and Stanyon, R., 2011. The genome diversity and karyotype evolution of mammals. *Molecular cytogenetics*, 4(1), p.22.

Green, E.D., 2001. Strategies for the systematic sequencing of complex genomes. *Nature Reviews Genetics*, 2(8), p.573.

Gregory, T.R. and Mable, B.K., 2005. Polyploidy in animals. *In The evolution of the genome*. Academic Press, pp. 427-517.

Gregory, T.R., 2002. A bird's-eye view of the C-value enigma: genome size, cell size, and metabolic rate in the class Aves. *Evolution*, 56(1), pp.121-130.

Gresham, D., Desai, M.M., Tucker, C.M., Jenq, H.T., Pai, D.A., Ward, A., DeSevo, C.G., Botstein, D. and Dunham, M.J., 2008. The repertoire and

dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS genetics*, 4(12), p.e1000303.

Grewal, S.I. and Moazed, D., 2003. Heterochromatin and epigenetic control of gene expression. *Science*, 301(5634), pp.798-802.

Griffin, D.K., Haberman, F., Masabanda, J., O'Brien, P., Bagga, M., Sazanov, A., Smith, J., Burt, D.W., Ferguson-Smith, M. and Wienberg, J., 1999. Micro-and macrochromosome paints generated by flow cytometry and microdissection: tools for mapping the chicken genome. *Cytogenetic and Genome Research*, 87(3-4), pp.278-281.

Griffin, D.K., Robertson, L.B.W., Tempest, H.G. and Skinner, B.M., 2007. The evolution of the avian genome as revealed by comparative molecular cytogenetics. *Cytogenetic and genome research,* 117(1), pp.64-77.

Griffin, D.K., Robertson, L.B., Tempest, H.G., Vignal, A., Fillon, V., Crooijmans, R.P., Groenen, M.A., Deryusheva, S., Gaginskaya, E., Carré, W. and Waddington, D., 2008. Whole genome comparative studies between chicken and turkey and their implications for avian genome evolution*. BMC genomics*, 9(1), p.168.

Groenen, M.A., Archibald, A.L., Uenishi, H., Tuggle, C.K., Takeuchi, Y., Rothschild, M.F., Rogel-Gaillard, C., Park, C., Milan, D., Megens, H.J. and Li, S., 2012. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*, 491(7424), p.393.

Gu, L., Frommel, S.C., Oakes, C.C., Simon, R., Grupp, K., Gerig, C.Y., Bär, D., Robinson, M.D., Baer, C., Weiss, M. and Gu, Z., 2015. BAZ2A (TIP5) is involved in epigenetic alterations in prostate cancer and its overexpression predicts disease recurrence. *Nature genetics*, 47(1), p.22.

Gurney, M.E., 1982. Behavioral correlates of sexual differentiation in the zebra finch song system. *Brain research*, 231(1), pp.153-172.

Guttenbach, M., Nanda, I., Feichtinger, W., Masabanda, J.S., Griffin, D.K. and Schmid, M., 2003. Comparative chromosome painting of chicken autosomal

paints 1–9 in nine different bird species. *Cytogenetic and Genome Research*, 103(1-2), pp.173-184.

Hall, R.N., Meers, J., Fowler, E. and Mahony, T., 2012. Back to BAC: the use of infectious clone technologies for viral mutagenesis. *Viruses*, 4(2), pp.211-235.

Hammar, B.O., 1970. The karyotypes of thirty-one birds. *Hereditas*, 65(1), pp.29-58.

Hansmann, T., Nanda, I., Volobouev, V., Yang, F., Schartl, M., Haaf, T. and Schmid, M., 2009. Cross-species chromosome painting corroborates microchromosome fusion during karyotype evolution of birds. *Cytogenetic and genome research*, 126(3), pp.281-304.

Hartmann, N. and Scherthan, H., 2004. Characterization of ancestral chromosome fusion points in the Indian muntjac deer. *Chromosoma*, 112(5), pp.213-220.

Hassold, T. and Hunt, P., 2001. To err (meiotically) is human: the genesis of human aneuploidy. *Nature Reviews Genetics*, 2(4), p.280.

Heather, J.M. and Chain, B., 2016. The sequence of sequencers: The history of sequencing DNA. *Genomics,* 107(1), pp.1-8.

Hedges, S.B., Dudley, J. and Kumar, S., 2006. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics*, 22(23), pp.2971-2972.

Hellen, E. and Kern, A., 2015. The Role of DNA Insertions in Phenotypic Differentiation between Humans and Other Primates. *Genome Biology and Evolution*, 7(4), pp.1168-1178.

Herschler, M.S. and Fechheimer, N.S., 1966. Centric fusion of chromosomes in a set of bovine triplets. Cytogenetic and Genome Research, 5(5), pp.307-312.

Hillier, L.W., Miller, W., Birney, E., Warren, W., Hardison, R.C., Ponting, C.P., Bork, P., Burt, D.W., Groenen, M.A., Delany, M.E. and Dodgson, J.B., 2004.

Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, 432(7018), pp.695-716.

Ho, L. and Crabtree, G.R., 2010. Chromatin remodelling during development. *Nature*, 463(7280), p.474-484

Hoffmann, A.A. and Rieseberg, L.H., 2008. Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation*? Annual review of ecology, evolution, and systematics*, 39, pp.21-42.

Holmes, D.J. and Ottinger, M.A., 2003. Birds as long-lived animal models for the study of aging. *Experimental gerontology*, 38(11-12), pp.1365-1375.

Hoodless, A., 1995. Studies of west palearctic birds. 195. Eurasian Woodcock Scolopax rusticola. *British Birds*, 88(12), pp.578-592.

Hron, T., Pajer, P., Pačes, J., Bartůněk, P. and Elleder, D., 2015. Hidden genes in birds. *Genome biology*, 16(1), p.164.

Hu, L., Cheng, D., Gong, F., Lu, C., Tan, Y., Luo, K., Wu, X., He, W., Xie, P., Feng, T. and Yang, K., 2016. Reciprocal translocation carrier diagnosis in preimplantation human embryos. *EBioMedicine*, 14, pp.139-147.

Huang, H. and Chen, J., 2017. Chromosome bandings. *In Cancer Cytogenetics*. Humana Press (New York, NY), pp. 59-66.

Hull, R.M., Cruz, C., Jack, C.V. and Houseley, J., 2017. Environmental change drives accelerated adaptation through stimulated copy number variation. *PLoS biology,* 15(6), p.e2001333.

Hunkapiller, T., Kaiser, R.J., Koop, B.F. and Hood, L., 1991. Large-scale and automated DNA sequence determination. *Science*, 254(5028), pp.59-67.

Iannucci, A., Svartman, M., Bellavita, M., Chelazzi, G., Stanyon, R. and Ciofi, C., 2019. Insights into Emydid Turtle Cytogenetics: The European Pond Turtle as a Model Species*. Cytogenetic and genome research*, 157(3), pp.166-171.

Ijdo, J.W., Baldini, A., Ward, D.C., Reeders, S.T. and Wells, R.A., 1991. Origin of human chromosome 2: an ancestral telomere-telomere fusion. *Proceedings of the National Academy of Sciences*, *88*(20), pp.9051-9055.

Inoue, K. and Lupski, J.R., 2002. Molecular mechanisms for genomic disorders. *Annual review of genomics and human genetics*, *3*(1), pp.199-242.

International Chicken Genome Sequencing Consortium (ICGSC), 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, 432(7018), pp.695-716.

International Human Genome Sequencing Consortium, 2001. Initial sequencing and analysis of the human genome. *Nature,* 409(6822), pp.860-921.

Itoh, Y. and Arnold, A.P., 2005. Chromosomal polymorphism and comparative painting analysis in the zebra finch. *Chromosome Research*, 13(1), pp.47-56.

Iwabe, N., Hara, Y., Kumazawa, Y., Shibamoto, K., Saito, Y., Miyata, T. and Katoh, K., 2004. Sister group relationship of turtles to the bird-crocodilian clade revealed by nuclear DNA–coded proteins. *Molecular biology and evolution*, 22(4), pp.810-813.

Jain, M., Koren, S., Miga, K.H., Quick, J., Rand, A.C., Sasani, T.A., Tyson, J.R., Beggs, A.D., Dilthey, A.T., Fiddes, I.T. and Malla, S., 2018. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature biotechnology,* 36(4), p.338.

Janes, D.E., Organ, C.L., Fujita, M.K., Shedlock, A.M. and Edwards, S.V., 2010. Genome evolution in Reptilia, the sister group of mammals*. Annual review of genomics and human genetics*, 11, pp.239-264.

Janssen, A., van der Burg, M., Szuhai, K., Kops, G.J. and Medema, R.H., 2011. Chromosome segregation errors as a cause of DNA damage and structural chromosome aberrations. *Science*, 333(6051), pp.1895-1898.

Jelesko, J.G., Carter, K., Thompson, W., Kinoshita, Y. and Gruissem, W., 2004. Meiotic recombination between paralogous RBCSB genes on sister chromatids of Arabidopsis thaliana. *Genetics*, 166(2), pp.947-957.

Ji, B., Higa, K.K., Kelsoe, J.R. and Zhou, X., 2015. Over-expression of XIST, the master gene for X chromosome inactivation, in females with major affective disorders. *EBioMedicine*, 2(8), pp.909-918.

Jiang, Y., Xie, M., Chen, W., Talbot, R., Maddox, J.F., Faraut, T., Wu, C., Muzny, D.M., Li, Y., Zhang, W. and Stanton, J.A., 2014. The sheep genome illuminates biology of the rumen and lipid metabolism. *Science*, 344(6188), pp.1168-1173.

Jin, Y., Jin, C., Salemark, L., Martins, C., Wennerberg, J. and Mertens, F., 2000. Centromere cleavage is a mechanism underlying isochromosome formation in skin and head and neck carcinomas. *Chromosoma*, 109(7), pp.476-481.

Jones, M.R., Brower, M.A., Xu, N., Cui, J., Mengesha, E., Chen, Y.D.I., Taylor, K.D., Azziz, R. and Goodarzi, M.O., 2015. Systems genetics reveals the functional context of PCOS loci and identifies genetic and molecular mechanisms of disease heterogeneity*. PLoS genetics*, 11(8), p.e1005455.

Joseph, S., O'Connor, R.E., Al Mutery, A., Watson, M., Larkin, D. and Griffin, D., 2018. Chromosome level genome assembly and comparative genomics between three falcon species reveals an unusual pattern of genome organisation. *Diversity,* 10(4), p.113.

Jourdain, E., Gunnarsson, G., Wahlgren, J., Latorre-Margalef, N., Bröjer, C., Sahlin, S., Svensson, L., Waldenström, J., Lundkvist, Å. and Olsen, B., 2010. Influenza virus in a natural host, the mallard: experimental infection data. *PloS one,* 5(1), p.e8935.

Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., Yabana, M., Harada, M., Nagayasu, E., Maruyama, H. and Kohara, Y., 2014. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome research*, 24(8), pp.1384-1395.

Kallioniemi, A., Kallioniemi, O.P., Sudar, D., Rutovitz, D., Gray, J.W., Waldman, F. and Pinkel, D., 1992. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science*, 258(5083), pp.818-821.

Kallioniemi, O.P., Kallioniemi, A., Piper, J., Isola, J., Waldman, F.M., Gray, J.W. and Pinkel, D., 1994. Optimizing comparative genomic hybridization for analysis of DNA sequence copy number changes in solid tumors. *Genes, Chromosomes and Cancer,* 10(4), pp.231-243.

Kapusta, A. and Suh, A., 2017. Evolution of bird genomes—a transposon's-eye view. *Annals of the New York Academy of Sciences*, 1389(1), pp.164-185.

Kasai, F., O'Brien, P.T.M., Martin, S. and Ferguson-Smith, M.A., 2003. Identification of the homologue of chicken Z chromosome in turtle and crocodile. *Ann Génét*, 46, p.89.

Kasai, F., O'Brien, P.C.M., Martin, S. and Ferguson-Smith, M.A., 2012. Extensive homology of chicken macrochromosomes in the karyotypes of Trachemys scripta elegans and Crocodylus niloticus revealed by chromosome painting despite long divergence times. *Cytogenetic and genome research*, 136(4), pp.303-307.

Kasai, F., O'Brien, P.C. and Ferguson-Smith, M.A., 2012. Reassessment of genome size in turtle and crocodile based on chromosome measurement by flow karyotyping: close similarity to chicken. *Biology letters*, 8(4), pp.631-635.

Kato, H. and Sandberg, A.A., 1968. Chromosome pulverization in human cells with micronuclei. *Journal of the National Cancer Institute*, *40*(1), pp.165-179.

Kato, T., Ouchi, Y., Inagaki, H., Makita, Y., Mizuno, S., Kajita, M., Ikeda, T., Takeuchi, K. and Kurahashi, H., 2017. Genomic characterization of chromosomal insertions: insights into the mechanisms underlying chromothripsis. *Cytogenetic and genome research*, 153(1), pp.1-9.

Kawagoshi, T., Uno, Y., Nishida, C. and Matsuda, Y., 2014. The Staurotypus turtles and aves share the same origin of sex chromosomes but evolved different types of heterogametic sex determination. *PLoS One*, 9(8), p.e105315.

Kawai, A., Ishijima, J., Nishida, C., Kosaka, A., Ota, H., Kohno, S.I. and Matsuda, Y., 2009. The ZW sex chromosomes of Gekko hokouensis (Gekkonidae, Squamata) represent highly conserved homology with those of avian species. *Chromosoma*, 118(1), pp.43-51.

Kehrer-Sawatzki, H. and Cooper, D.N., 2007. Structural divergence between the human and chimpanzee genomes. *Human genetics*, 120(6), pp.759-778.

Kelley, D.R. and Salzberg, S.L., 2010. Detection and correction of false segmental duplications caused by genome mis-assembly. *Genome biology*, 11(3), p.R28.

Kichigin, I.G., Giovannotti, M., Makunin, A.I., Ng, B.L., Kabilov, M.R., Tupikin, A.E., Barucchi, V.C., Splendiani, A., Ruggeri, P., Rens, W. and O'Brien, P.C., 2016. Evolutionary dynamics of Anolis sex chromosomes revealed by sequencing of flow sorting-derived microchromosome-specific DNA. *Molecular Genetics and Genomics*, 291(5), pp.1955-1966.

Kim, J., Larkin, D.M., Cai, Q., Zhang, Y., Ge, R.L., Auvil, L., Capitanu, B., Zhang, G., Lewin, H.A. and Ma, J., 2013. Reference-assisted chromosome assembly. *Proceedings of the National Academy of Sciences*, 110(5), pp.1785-1790.

Kim, S.H., Ganji, M., Kim, E., van der Torre, J., Abbondanzieri, E. and Dekker, C., 2018. DNA sequence encodes the position of DNA supercoils. *eLife*, 7, p.e36557.

Kirkpatrick, M. and Barton, N., 2006. Chromosome inversions, local adaptation and speciation. *Genetics*, 173(1), pp.419-434.

Klochendler-Yeivin, A., Fiette, L., Barra, J., Muchardt, C., Babinet, C. and Yaniv, M., 2000. The murine SNF5/INI1 chromatin remodeling factor is essential for embryonic development and tumor suppression. *EMBO reports*, 1(6), pp.500-506.

Koepfli, K.P., Paten, B. and O'Brien, S.J., 2015. The Genome 10K Project: a way forward. *Annual Review of Animal Biosciences*. Annual Reviews; 2015; Vol. 3, pp.57–111.

Kok, Y.J.D., Merkx, G.F., van der Maarel, S.M., Huber, I., Malcolm, S., Ropers, H.H. and Cremers, F.P., 1995. A duplication/paracentric inversion associated with familial X-linked deafness (DFN3) suggests the presence of a regulatory element more than 400 kb upstream of the POU3F4 gene. *Human molecular genetics*, 4(11), pp.2145-2150.

Kosztolányi, G., Méhes, K. and Hook, E.B., 1991. Inherited ring chromosomes: an analysis of published cases. *Human genetics*, 87(3), pp.320-324.

Kozarewa, I., Ning, Z., Quail, M.A., Sanders, M.J., Berriman, M. and Turner, D.J., 2009. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+ C)-biased genomes. *Nature methods*, 6(4), p.291.

Kretschmer, R., Ferguson-Smith, M.A. and De Oliveira, E.H.C., 2018. Karyotype evolution in birds: from conventional staining to chromosome painting. *Genes*, 9(4), p.181.

Kumar, S. and Hedges, S.B., 1998. A molecular timescale for vertebrate evolution. *Nature*, 392(6679), p.917.

Kuraku, S., Ishijima, J., Nishida-Umehara, C., Agata, K., Kuratani, S. and Matsuda, Y., 2006. cDNA-based gene mapping and GC 3 profiling in the soft-shelled turtle suggest a chromosomal size-dependent GC bias shared by sauropsids. *Chromosome Research*, 14(2), pp.187-202.

Ladjali-Mohammedi, K., Bitgood, J.J., Tixier-Boichard, M. and De Leon, F.P., 1999. International system for standardized avian karyotypes (ISSAK): standardized banded karyotypes of the domestic fowl (*Gallus domesticus*). *Cytogenetic and Genome Research*, 86(3-4), pp.271-276.

Lander, E.S., 2011. Initial impact of the sequencing of the human genome. *Nature,* 470(7333), p.187.

Langer, P.R., Waldrop, A.A. and Ward, D.C., 1981. Enzymatic synthesis of biotin-labeled polynucleotides: novel nucleic acid affinity probes. *Proceedings of the National Academy of Sciences*, 78(11), pp.6633-6637.

Larkin D.M., Pape, G., Donthu, R., Auvil, L., Welge, M. and Lewin, H.A., 2009. Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. *Genome Research,* 19(5), pp.770–777.

Larkin, D.M., Daetwyler, H.D., Hernandez, A.G., Wright, C.L., Hetrick, L.A., Boucek, L., Bachman, S.L., Band, M.R., Akraiko, T.V., Cohen-Zinder, M. and Thimmapuram, J., 2012. Whole-genome resequencing of two elite sires for the detection of haplotypes under selection in dairy cattle. *Proceedings of the National Academy of Sciences*, 109(20), pp.7693-7698.

Leaché, A.D., Banbury, B.L., Linkem, C.W. and de Oca, A.N., 2016. Phylogenomics of a rapid radiation: is chromosomal evolution linked to increased diversification in North American spiny lizards (Genus Sceloporus)? *BMC Evolutionary Biology*, 16(1). p.63.

Lee, S.G., Lee, I., Park, S.H., Kang, C. and Song, K., 1995. Identification and characterization of a human cDNA homologous to yeast SKI2. *Genomics*, 25(3), pp.660-666.

Lee, C., Gisselsson, D., Jin, C., Nordgren, A., Ferguson, D.O., Blennow, E., Fletcher, J.A. and Morton, C.C., 2001. Limitations of chromosome classification by multicolor karyotyping. *The American Journal of Human Genetics*, 68(4), pp.1043-1047.

Leggatt, R.A. and Iwama, G.K., 2003. Occurrence of polyploidy in the fishes. *Reviews in Fish Biology and Fisheries*, 13(3), pp.237-246.

Lempainen, J., Laine, A.P., Hammais, A., Toppari, J., Simell, O., Veijola, R., Knip, M. and Ilonen, J., 2015. Non-HLA gene effects on the disease process of type 1 diabetes: from HLA susceptibility to overt disease. *Journal of autoimmunity*, 61, pp.45-53.

Levan, A., Fredga, K. and Sandberg, A.A., 1964. Nomenclature for centromeric position on chromosomes. *Hereditas*, 52(2), pp.201-220.

Lewin, H.A., Larkin, D.M., Pontius, J. and O'Brien, S.J., 2009. Every genome sequence needs a good map. *Genome research*, 19(11), pp.1925-1928.

Li, Z., Wei, S., Li, H., Wu, K., Cai, Z., Li, D., Wei, W., Li, Q., Chen, J., Liu, H. and Zhang, L., 2017. Genome-wide genetic structure and differentially selected regions among Landrace, Erhualian, and Meishan pigs using specific-locus amplified fragment sequencing. *Scientific Reports*, 7(1), pp.1-12.

Lichtenbelt, K.D., Knoers, N.V.A.M. and Schuring-Blom, G.H., 2011. From karyotyping to array-CGH in prenatal diagnosis. *Cytogenetic and genome research*, 135(3-4), pp.241-250.

Lichter, P., Cremer, T., Borden, J., Manuelidis, L. and Ward, D.C., 1988. Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. *Human genetics*, 80(3), pp.224-234.

Lichter, P., Tang, C.J., Call, K., Hermanson, G., Evans, G.A., Housman, D. and Ward, D.C., 1990. High-resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. *Science*, 247(4938), pp.64-69.

Lieberman-Aiden, E., Van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O. and Sandstrom, R., 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950), pp.289-293.

Lin, J., Xu, K., Gitanjali, J., Roth, J.A. and Ji, L., 2011. Regulation of tumor suppressor gene FUS1 expression by the untranslated regions of mRNA in human lung cancer cells. *Biochemical and biophysical research communications*, 410(2), pp.235-241.

Lisachov, A.P., Giovannotti, M., Pereira, J.C., Andreyushkova, D.A., Romanenko, S.A., Ferguson-Smith, M.A., Borodin, P.M. and Trifonov, V.A., 2020. Chromosome painting does not support a sex chromosome turnover in

Lacerta agilis Linnaeus, 1758. *Cytogenetic and Genome Research*, *160*(3), pp.134-140.

Lithgow, P.E., O'Connor, R., Smith, D., Fonseka, G., Al Mutery, A., Rathje, C., Frodsham, R., O'Brien, P., Kasai, F., Ferguson-Smith, M.A. and Skinner, B.M., 2014. Novel tools for characterising inter and intra chromosomal rearrangements in avian microchromosomes*. Chromosome research*, 22(1), pp.85-97.

Livingstone, K. and Rieseberg, L., 2004. Chromosomal evolution and speciation: a recombination-based approach. *New Phytologist*, 161(1), pp.107-112.

Lohse, K., Clarke, M., Ritchie, M.G. and Etges, W.J., 2015. Genome-wide tests for introgression between cactophilic Drosophila implicate a role of inversions during speciation. *Evolution*, 69(5), pp.1178-1190.

Lu, F., Shi, Y., Qu, C., Zhao, P., Liu, X., Gong, B., Ma, S., Zhou, Y., Zhang, Q., Fei, P. and Xu, Y., 2013. A genetic variant in the SKIV2L gene is significantly associated with age-related macular degeneration in a Han Chinese population. *Investigative ophthalmology & visual science*, 54(4), pp.2911-2917.

Lu, S.H., Hsieh, L.L., Luo, F.C. and Weinstein, I.B., 1988. Amplification of the EGF receptor and c-myc genes in human esophageal cancers. *International journal of cancer*, *42*(4), pp.502-505.

Lubs, H.A., McKenzie, W.H., Patil, S.R. and Merrick, S., 1973. New staining methods for chromosomes. *Methods in cell biology,* Vol. 6, pp.345-380.

Lupski, J.R. and Stankiewicz, P., 2005. Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. *PLoS genetics*, 1(6), p.e49.

Lyson, T.R., Sperling, E.A., Heimberg, A.M., Gauthier, J.A., King, B.L. and Peterson, K.J., 2011. MicroRNAs support a turtle+ lizard clade. *Biology letters*, 8(1), pp.104-107.

Maas, B., Tscharntke, T., Saleh, S., Dwi Putra, D. and Clough, Y., 2015. Avian species identity drives predation success in tropical cacao agroforestry. *Journal of Applied Ecology*, 52(3), pp.735-743.

Mahiddine-Aoudjit, L., Boucekkine, O. and Ladjali-Mohammedi, K., 2019. Banding cytogenetics of the vulnerable species Houbara bustard (Otidiformes) and comparative analysis with the Domestic fowl. *Comparative cytogenetics*, 13(1), p.1.

Mak, A.C., Lai, Y.Y., Lam, E.T., Kwok, T.P., Leung, A.K., Poon, A., Mostovoy, Y., Hastie, A.R., Stedman, W., Anantharaman, T. and Andrews, W., 2016. Genome-wide structural variation detection by genome mapping on nanochannel arrays. *Genetics*, 202(1), pp.351-362.

Mandrioli, D., Belpoggi, F., Silbergeld, E.K. and Perry, M.J., 2016. Aneuploidy: a common and early evidence-based biomarker for carcinogens and reproductive toxicants. *Environmental Health*, 15(1), p.97.

Manjari, S.R., Pata, J.D. and Banavali, N.K., 2014. Cytosine Unstacking and Strand Slippage at an Insertion–Deletion Mutation Sequence in an Overhang-Containing DNA Duplex. *Biochemistry*, 53(23), pp.3807-3816.

Marques-Bonet, T., Kidd, J.M., Ventura, M., Graves, T.A., Cheng, Z., Hillier, L.W., Jiang, Z., Baker, C., Malfavon-Borja, R., Fulton, L.A. and Alkan, C., 2009. A burst of segmental duplications in the genome of the African great ape ancestor. *Nature*, 457(7231), p.877.

Marshall Graves, J.A. and Shetty, S., 2001. Sex from W to Z: evolution of vertebrate sex chromosomes and sex determining genes. *Journal of Experimental Zoology*, 290(5), pp.449-462.

Martinez, P.A., Jacobina, U.P., Fernandes, R.V., Brito, C., Penone, C., Amado, T.F., Fonseca, C.R. and Bidau, C.J., 2017. A comparative study on karyotypic diversification rate in mammals. *Heredity*, 118(4), p.366.

Masabanda, J.S., Burt, D.W., O'Brien, P.C., Vignal, A., Fillon, V., Walsh, P.S., Cox, H., Tempest, H.G., Smith, J., Habermann, F. and Schmid, M., 2004.

Molecular cytogenetic definition of the chicken genome: the first complete avian karyotype. *Genetics*, 166(3), pp.1367-1373.

Matsubara, K., Kuraku, S., Tarui, H., Nishimura, O., Nishida, C., Agata, K., Kumazawa, Y. and Matsuda, Y., 2012. Intra-genomic GC heterogeneity in sauropsids: evolutionary insights from cDNA mapping and GC 3 profiling in snake. *BMC genomics*, 13(1), p.604.

Matsuda, Y., Nishida-Umehara, C., Tarui, H., Kuroiwa, A., Yamada, K., Isobe, T., Ando, J., Fujiwara, A., Hirao, Y., Nishimura, O. and Ishijima, J., 2005. Highly conserved linkage homology between birds and turtles: bird and turtle chromosomes are precise counterparts of each other. *Chromosome Research*, 13(6), pp.601-615.

McCarroll, S.A. and Altshuler, D.M., 2007. Copy-number variation and association studies of human disease. *Nature genetics*, 39(7s), p.S37.

McGaugh, S.E., Eckerman, C.M. and Janzen, F.J., 2008. Molecular phylogeography of Apalone spinifera (Reptilia, Trionychidae). *Zoologica Scripta,* 37(3), pp.289-304.

Meltzer, P.S., Guan, X.Y., Burgess, A. and Trent, J.M., 1992. Rapid generation of region specific probes by chromosome microdissection and their application. *Nature genetics*, 1(1), p.24.

Miller, M.M., Robinson, C.M., Abernathy, J., Goto, R.M., Hamilton, M.K., Zhou, H. and Delany, M.E., 2013. Mapping genes to chicken microchromosome 16 and discovery of olfactory and scavenger receptor genes near the major histocompatibility complex. *Journal of Heredity*, 105(2), pp.203-215.

Minegishi, Y., Saito, M., Morio, T., Watanabe, K., Agematsu, K., Tsuchiya, S., Takada, H., Hara, T., Kawamura, N., Ariga, T. and Kaneko, H., 2006. Human tyrosine kinase 2 deficiency reveals its requisite roles in multiple cytokine signals involved in innate and acquired immunity. *Immunity*, 25(5), pp.745-755.

Minhas, K.M., Singh, B., Jiang, W.W., Sidransky, D. and Califano, J.A., 2003. Spindle assembly checkpoint defects and chromosomal instability in head and

neck squamous cell carcinoma. *International journal of cancer*, 107(1), pp.46-52.

Mirzaghaderi, G. and Mason, A.S., 2017. Revisiting Pivotal-Differential Genome Evolution in Wheat. *Trends in Plant Science*, 22(8), pp.674–684.

Montiel, E.E., Badenhorst, D., Lee, L.S., Literman, R., Trifonov, V. and Valenzuela, N., 2016. Cytogenetic insights into the evolution of chromosomes and sex determination reveal striking homology of turtle sex chromosomes to amphibian autosomes. *Cytogenetic and genome research*, 148(4), pp.292-304.

Moore, C.M. and Best, R.G., 2001. Chromosome Preparation and Banding. *Encyclopedia of Life Sciences*.

Morin, S.J., Eccles, J., Iturriaga, A. and Zimmerman, R.S., 2017. Translocations, inversions and other chromosome rearrangements. *Fertility and sterility*, 107(1), pp.19-26.

Moriyama, Y. and Koshiba-Takeuchi, K., 2018. Significance of whole-genome duplications on the emergence of evolutionary novelties. *Briefings in Functional Genomics*, 17(5), pp.329–338.

Morris, W.B., Stephenson, J.E., Robertson, L.B.W., Turner, K., Brown, H., Ioannou, D., Tempest, H.G., Skinner, B.M. and Griffin, D.K., 2007. Practicable approaches to facilitate rapid and accurate molecular cytogenetic mapping in birds and mammals. *Cytogenetic and genome research*, 117(1-4), pp.36-42.

Morisson, M., Denis, M., Milan, D., Klopp, C., Leroux, S., Bardes, S., Pitel, F., Vignoles, F., Gerus, M., Fillon, V. and Douaud, M., 2007. The chicken RH map: current state of progress and microchromosome mapping. *Cytogenetic and genome research*, 117(1-4), pp.14-21.

Mota-Velasco, J.C., Ferreira, I.A., Cioffi, M.B., Ocalewicz, K., Campos-Ramos, R., Shirak, A., Lee, B.Y., Martins, C. and Penman, D.J., 2010. Characterisation of the chromosome fusions in Oreochromis karongae. *Chromosome research*, 18(5), pp.575-586.

Muir, P., Li, S., Lou, S., Wang, D., Spakowicz, D.J., Salichos, L., Zhang, J., Weinstock, G.M., Isaacs, F., Rozowsky, J. and Gerstein, M., 2016. The real cost of sequencing: scaling computation to keep pace with data generation. *Genome biology*, 17(1), p.53.

Munshi-South, J. and Wilkinson, G.S., 2010. Bats and birds: exceptional longevity despite high metabolic rates. *Ageing research reviews*, 9(1), pp.12-19.

Murphy, W.J., Frönicke, L., O'Brien, S.J. and Stanyon, R., 2003. The origin of human chromosome 1 and its homologs in placental mammals. *Genome research*, 13(8), pp.1880-1888.

Murphy, W.J., Larkin, D.M., Everts-van der Wind, A., Bourque, G., Tesler, G., Auvil, L., Beever, J.E., Chowdhary, B.P., Galibert, F., Gatzke, L. and Hitte, C., 2005. Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science*, 309(5734), pp.613-617.

Nakamura, K., Oshima, T., Morimoto, T., Ikeda, S., Yoshikawa, H., Shiwa, Y., Ishikawa, S., Linak, M.C., Hirai, A., Takahashi, H. and Altaf-Ul-Amin, M., 2011. Sequence-specific error profile of Illumina sequencers. *Nucleic acids research*, 39(13), pp.e90-e90.

Nam, K. and Ellegren, H., 2012. Recombination drives vertebrate genome contraction. *PLoS genetics*, 8(5), p.e1002680.

Guttenbach, M., Nanda, I., Feichtinger, W., Masabanda, J.S., Griffin, D.K. and Schmid, M., 2003. Comparative chromosome painting of chicken autosomal paints 1–9 in nine different bird species. *Cytogenetic and Genome Research*, *103*(1-2), pp.173-184.

Nanda, I., Karl, E., Volobouev, V., Griffin, D.K., Schartl, M. and Schmid, M., 2006. Extensive gross genomic rearrangements between chicken and Old World vultures (Falconiformes: Accipitridae). *Cytogenetic and genome research*, 112(3-4), pp.286-295.

Nanda, I., Karl, E., Griffin, D.K., Schartl, M. and Schmid, M., 2007. Chromosome repatterning in three representative parrots (Psittaciformes) inferred from

comparative chromosome painting. *Cytogenetic and genome research*, 117(1-4), pp.43-53.

Nanda, I., Benisch, P., Fetting, D., Haaf, T. and Schmid, M., 2011. Synteny conservation of chicken macrochromosomes 1–10 in different avian lineages revealed by cross-species chromosome painting. *Cytogenetic and Genome Research*, 132(3), pp.165-181.

Navarro, A. and Barton, N.H., 2003. Chromosomal speciation and molecular divergence--accelerated evolution in rearranged chromosomes. *Science*, 300(5617), pp.321-324.

Nebel, R.A., Kirschen, J., Cai, J., Woo, Y.J., Cherian, K. and Abrahams, B.S., 2015. Reciprocal relationship between head size, an autism endophenotype, and gene dosage at 19p13. 12 points to AKAP8 and AKAP8L. *PLoS One*, 10(6), p.e0129270.

Neely, R.K., Deen, J. and Hofkens, J., 2011. Optical mapping of DNA: Single-molecule-based methods for mapping genomes. *Biopolymers*, 95(5), pp.298-311.

Nie, W., O'Brien, P.C., Ng, B.L., Fu, B., Volobouev, V., Carter, N.P., Ferguson-Smith, M.A. and Yang, F., 2009. Avian comparative genomics: reciprocal chromosome painting between domestic chicken (*Gallus gallus*) and the stone curlew (Burhinus oedicnemus, Charadriiformes)—An atypical species with low diploid number. *Chromosome Research*, 17(1), pp.99-113.

Nie, W., O'Brien, P.C., Fu, B., Wang, J., Su, W., He, K., Bed'Hom, B., Volobouev, V., Ferguson-Smith, M.A., Dobigny, G. and Yang, F., 2015. Multidirectional chromosome painting substantiates the occurrence of extensive genomic reshuffling within Accipitriformes. *BMC evolutionary biology*, 15(1), p.205.

Nishida, C., Ishijima, J., Kosaka, A., Tanabe, H., Habermann, F.A., Griffin, D.K. and Matsuda, Y., 2008. Characterization of chromosome structures of Falconinae (Falconidae, Falconiformes, Aves) by chromosome painting and

delineation of chromosome rearrangements during their differentiation. *Chromosome Research*, 16(1), pp.171-181.

Noor, M.A., Grams, K.L., Bertucci, L.A. and Reiland, J., 2001. Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences*, 98(21), pp.12084-12088.

Noronha, R.C.R., Barros, L.M.R., Araújo, R.E.F., Marques, D.F., Nagamachi, C.Y., Martins, C. and Pieczarka, J.C., 2016. New insights of karyoevolution in the Amazonian turtles Podocnemis expansa and Podocnemis unifilis (Testudines, Podocnemidae). *Molecular cytogenetics*, 9(1), p.73.

Nowicki, S. and Searcy, W.A., 2014. The evolution of vocal learning. *Current opinion in neurobiology*, 28, pp.48-53.

Nowicki, S., Searcy, W.A. and Peters, S., 2002. Brain development, song learning and mate choice in birds: a review and experimental test of the" nutritional stress hypothesis". *Journal of Comparative Physiology A*, 188(11-12), pp.1003-1014.

O'Connor, R.E., Fonseka, G., Frodsham, R., Archibald, A.L., Lawrie, M., Walling, G.A. and Griffin, D.K., 2017. Isolation of subtelomeric sequences of porcine chromosomes for translocation screening reveals errors in the pig genome assembly. *Animal genetics*, 48(4), pp.395-403.

O'Connor, R.E., Farré, M., Joseph, S., Damas, J., Kiazim, L.G, Jennings, R., Bennett, S., Slack, E.A., Allanson, E., Larkin, D.M. and Griffin, D.K., 2018a. Chromosome-level assembly reveals extensive rearrangement in saker falcon and budgerigar, but not ostrich, genomes. *Genome biology*, 19(1), p.171.

O'Connor, R.E., Kiazim, L.G, Skinner, B., Fonseka, G., Joseph, S., Jennings, R., Larkin, D.M. and Griffin, D.K., 2018b. Patterns of microchromosome organization remain highly conserved throughout avian evolution. *Chromosoma*, 128(1), pp.21-29.

O'Connor, R.E., Romanov, M.N., Kiazim, L.G., Barrett, P.M., Farré, M., Damas, J., Ferguson-Smith, M., Valenzuela, N., Larkin, D.M. and Griffin, D.K., 2018c.

Reconstruction of the diapsid ancestral genome permits chromosome evolution tracing in avian and non-avian dinosaurs. *Nature communications*, 9(1), p.1883.

Okada, T.A. and Comings, D.E., 1974. Mechanisms of chromosome banding. *Chromosoma*, *48*(1), pp.65-71.

Olmo, E., 2008. Trends in the evolution of reptilian chromosomes*. Integrative and Comparative Biology*, 48(4), pp.486-493.

Olson, C.R. and Mello, C.V., 2010. Significance of vitamin A to brain function, behavior and learning. *Molecular nutrition & food research*, 54(4), pp.489-495.

Ostroverkhova, N.B., Nazarenko, S.A. and Cheremnykh, A.D., 2002. Comparative genomic hybridization as a new method for detection of genomic imbalance. *Russian Journal of Genetics*, 38(2), pp.95-104.

Otto, S.P. and Whitton, J., 2000. Polyploid incidence and evolution. *Annual review of genetics*, 34(1), pp.401-437.

Pan, F., Yu, H., Dang, E.V., Barbi, J., Pan, X., Grosso, J.F., Jinasena, D., Sharma, S.M., McCadden, E.M., Getnet, D. and Drake, C.G., 2009. Eos mediates Foxp3-dependent gene silencing in CD4+ regulatory T cells. *Science,* 325(5944), pp.1142-1146.

Pan, Y., Wang, X., Liu, L., Wang, H. and Luo, M., 2016. Whole genome mapping with feature sets from high-throughput sequencing data*. PloS one*, 11(9), p.e0161583.

Parks, M.M., Lawrence, C.E. and Raphael, B.J., 2015. Detecting non-allelic homologous recombination from high-throughput sequencing data. *Genome biology*, 16(1), p.72.

Peona, V., Weissensteiner, M.H. and Suh, A., 2018. How complete are “complete” genome assemblies? - An avian perspective. *Molecular ecology resources*, 18(6), pp.1188-1195.

Pevzner, P. and Tesler, G., 2003. Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution. *Proceedings of the National Academy of Sciences*, 100(13), pp.7672-7677.

Pfau, S.J. and Amon, A., 2012. Chromosomal instability and aneuploidy in cancer: from yeast to man. *EMBO reports*, 13(6), pp.515-527.

Phippard, D., Boyd, Y., Reed, V., Fisher, G., Masson, W.K., Evans, E.P., Saunders, J.C. and Crenshaw III, E.B., 2000. The sex-linked fidget mutation abolishes Brn4/Pou3f4 gene expression in the embryonic inner ear. *Human molecular genetics*, 9(1), pp.79-85.

Pichugin, A.M., Galkina, S.A., Potekhin, A.A., Punina, E.O., Rautian, M.S. and Rodionov, A.V., 2001. Estimation of the Minimal Size of Chicken *Gallus gallus* domesticusMicrochromosomes via Pulsed-Field Electrophoresis. *Russian Journal of Genetics*, 37(5), pp.535-538.

Pincheira-Donoso, D., Bauer, A.M., Meiri, S. and Uetz, P., 2013. Global taxonomic diversity of living reptiles. *PLoS One*, 8(3), p.e59741.

Pinkel, D., Straume, T. and Gray, J.W., 1986. Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proceedings of the National Academy of Sciences*, 83(9), pp.2934-2938.

Pokorná, M., Giovannotti, M., Kratochvíl, L., Kasai, F., Trifonov, V.A., O'Brien, P.C., Caputo, V., Olmo, E., Ferguson-Smith, M.A. and Rens, W., 2011. Strong conservation of the bird Z chromosome in reptilian genomes is revealed by comparative painting despite 275 million years divergence. *Chromosoma*, 120(5), p.455.

Pokorná, M., Giovannotti, M., Kratochvíl, L., Caputo, V., Olmo, E., Ferguson-Smith, M.A. and Rens, W., 2012. Conservation of chromosomes syntenic with avian autosomes in squamate reptiles revealed by comparative chromosome painting. *Chromosoma*, 121(4), pp.409-418.

Potter, S., Bragg, J.G., Blom, M.P., Deakin, J.E., Kirkpatrick, M., Eldridge, M.D. and Moritz, C., 2017. Chromosomal speciation in the genomics era:

Disentangling phylogenetic evolution of rock-wallabies. *Frontiers in genetics*, 8, p.10.

Poultney, C.S., Goldberg, A.P., Drapeau, E., Kou, Y., Harony-Nicolas, H., Kajiwara, Y., De Rubeis, S., Durand, S., Stevens, C., Rehnström, K. and Palotie, A., 2013. Identification of small exonic CNV from whole-exome sequence data and application to autism spectrum disorder. *The American Journal of Human Genetics*, 93(4), pp.607-619.

Putnam, N.H., O'Connell, B.L., Stites, J.C., Rice, B.J., Blanchette, M., Calef, R., Troll, C.J., Fields, A., Hartley, P.D., Sugnet, C.W. and Haussler, D., 2016. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome research*, 26(3), pp.342-350.

Raccaud, M. and Suter, D.M., 2018. Transcription factor retention on mitotic chromosomes: regulatory mechanisms and impact on cell fate decisions. *FEBS letters*, 592(6), pp.878-887.

Rao, M., Morisson, M., Faraut, T., Bardes, S., Fève, K., Labarthe, E., Fillon, V., Huang, Y., Li, N. and Vignal, A., 2012. A duck RH panel and its potential for assisting NGS genome assembly. *BMC genomics*, 13(1), p.513.

Raskina, O., Barber, J.C., Nevo, E. and Belyayev, A., 2008. Repetitive DNA and chromosomal rearrangements: speciation-related events in plant genomes. *Cytogenetic and Genome Research*, 120(3-4), pp.351-357.

Raudsepp, T., Houck, M.L., O'Brien, P.C., Ferguson-Smith, M.A., Ryder, O.A. and Chowdhary, B.P., 2002. Cytogenetic analysis of California condor (Gymnogyps californianus) chromosomes: comparison with chicken (*Gallus gallus*) macrochromosomes. *Cytogenetic and genome research*, 98(1), pp.54-60.

Reid, T., 2015. Human Chromosomes. *NCI Center for Cancer Research*, Cancer Close Up 2015 collection.

Rhoads, A. and Au, K.F., 2015. PacBio sequencing and its applications. *Genomics, proteomics & bioinformatics*, 13(5), pp.278-289.

Ricklefs, R.E., 2012. Species richness and morphological diversity of passerine birds. *Proceedings of the national academy of sciences*, 109(36), pp.14482-14487.

Rieseberg, L.H., 2001. Chromosomal rearrangements and speciation. *Trends in ecology & evolution*, 16(7), pp.351-358.

Robinson, W.P., Bernasconi, F., Basaran, S., Yüksel-Apak, M., Neri, G., Serville, F., Balicek, P., Haluza, R., Farah, L.M.S., Lüleci, G. and Schinzel, A.A., 1994. A somatic origin of homologous Robertsonian translocations and isochromosomes. *American journal of human genetics*, 54(2), p.290.

Romanenko, S.A, Volobouev, V.T, Perelman, P.L, Lebedev, V.S., Serdukova, N.A, Trifonov, V.A., Biltueva, L.S, Nie, W., O'Brien, P., Bulatova, N., Ferguson-Smith, M.A, Yang, F. and Graphodatsky, A., 2007. Karyotype evolution and phylogenetic relationships of hamsters (Cricetidae, Muroidea, Rodentia) inferred from chromosomal painting and banding comparison. *Chromosome Research*, 15(3), pp.283-298.

Romanov, M.N., Farré, M., Lithgow, P.E., Fowler, K.E., Skinner, B.M., O'Connor, R., Fonseka, G., Backström, N., Matsuda, Y., Nishida, C. and Houde, P., 2014. Reconstruction of gross avian genome structure, organization and evolution suggests that the chicken lineage most closely resembles the dinosaur avian ancestor. *BMC genomics*, 15(1), p.1060.

Rosenberg, C., Knijnenburg, J., de Lourdes Chauffaille, M., Brunoni, D., Catelani, A.L., Sloos, W., Szuhai, K. and Tanke, H.J., 2005. Array CGH detection of a cryptic deletion in a complex chromosome rearrangement. *Human genetics*, 116(5), pp.390-394.

Ross, M.G., Russ, C., Costello, M., Hollinger, A., Lennon, N.J., Hegarty, R., Nusbaum, C. and Jaffe, D.B., 2013. Characterizing and measuring bias in sequence data. *Genome biology*, 14(5), p.R51.

Rubes, J., Pinton, A., Bonnet-Garnier, A., Fillon, V., Musilova, P., Michalova, K., Kubickova, S., Ducos, A. and Yerle, M., 2009. Fluorescence in situ hybridization

applied to domestic animal cytogenetics. *Cytogenetic and genome research*, 126(1-2), pp.34-48.

Rubin, C.J., Zody, M.C., Eriksson, J., Meadows, J.R., Sherwood, E., Webster, M.T., Jiang, L., Ingman, M., Sharpe, T., Ka, S. and Hallböök, F., 2010. Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature*, 464(7288), p.587.

Ruiz-Herrera, A., Garcia, F., Giulotto, E., Attolini, C., Egozcue, J., Ponsa, M. and Garcia, M., 2005. Evolutionary breakpoints are co-localized with fragile sites and intrachromosomal telomeric sequences in primates. *Cytogenetic and genome research*, 108(1-3), pp.234-247.

Ruiz-Herrera, A., Castresana, J. and Robinson, T.J., 2006. Is mammalian chromosomal evolution driven by regions of genome fragility? *Genome biology*, 7(12), p.R115.

Ruiz-Herrera, A., Farré, M. and Robinson, T.J., 2012. Molecular cytogenetic and genomic insights into chromosomal evolution. *Heredity*, 108(1), p.28.

Ryu, S.L., Murooka, Y. and Kaneko, Y., 1998. Reciprocal translocation at duplicated RPL2 loci might cause speciation of Saccharomyces bayanus and Saccharomyces cerevisiae. *Current genetics*, *33*(5), pp.345-351.

Salzberg, S.L. and Yorke, J.A., 2005. Beware of mis-assembled genomes. *Bioinformatics,* 21(24), pp.4320-4321.

Salzberg, S.L., Phillippy, A.M., Zimin, A., Puiu, D., Magoc, T., Koren, S., Treangen, T.J., Schatz, M.C., Delcher, A.L., Roberts, M. and Marçais, G., 2012. GAGE: A critical evaluation of genome assemblies and assembly algorithms. *Genome research*, 22(3), pp.557-567.

Sanger, F., Nicklen, S. and Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences*, 74(12), pp.5463-5467.

Santana, J.A., Gardner, L.I. and Neu, R.L., 1977. The X Isochromosome-X Syndrome [46, X, i (Xq)] Report of Three Cases with Review of the Phenotype. *Clinical pediatrics*, *16*(11), pp.1021-1026.

Santoro, R., Li, J. and Grummt, I., 2002. The nucleolar remodeling complex NoRC mediates heterochromatin formation and silencing of ribosomal gene transcription*. Nature genetics*, 32(3), p.393.

Sattler, M.C., Carvalho, C.R. and Clarindo, W.R., 2016. The polyploidy and its key role in plant breeding. *Planta*. 243(2), pp.281–96.

Schadt, E.E., Turner, S. and Kasarskis, A., 2010. A window into third-generation sequencing. *Human molecular genetics*, 19(R2), pp.R227-R240.

Scherthan, H., Cremer, T., Arnason, U., Weier, H.U., Lima-de-Faria, A. and Frönicke, L., 1994. Comparative chromosome painting discloses homologous segments in distantly related mammals. *Nature genetics*, 6(4), p.342.

Schield, D.R., Card, D.C., Hales, N.R., Perry, B.W., Pasquesi, G.M., Blackmon, H., Adams, R.H., Corbin, A.B., Smith, C.F., Ramesh, B. and Demuth, J.P., 2019. The origins and evolution of chromosomes, dosage compensation, and mechanisms underlying venom regulation in snakes. *Genome research*, 29(4), pp.590-601.

Schmid, M., Smith, J., Burt, D.W., Aken, B.L., Antin, P.B., Archibald, A.L., Ashwell, C., Blackshear, P.J., Boschiero, C., Brown, C.T. and Burgess, S.C., 2015. *Third report on chicken genes and chromosomes 2015. Cytogenetic and genome research*, 145(2), pp.78-179.

Schultz, R.J., 1980. The role of polyploidy in the evolution of fishes. *Polyploidy: biological relevance.* New York, NY. Plenum Publishing. pp.313–339.

Shaffer, H.B., Minx, P., Warren, D.E., Shedlock, A.M., Thomson, R.C., Valenzuela, N., Abramyan, J., Amemiya, C.T., Badenhorst, D., Biggar, K.K. and Borchert, G.M., 2013. The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome biology*, 14(3), p.R28.

Shaw, C.J. and Lupski, J.R., 2004. Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. *Human molecular genetics*, 13(suppl_1), pp.R57-R64.

Shendure, J., Balasubramanian, S., Church, G.M., Gilbert, W., Rogers, J., Schloss, J.A. and Waterston, R.H., 2017. DNA sequencing at 40: past, present and future. *Nature*, 550(7676), pp.345-353.

Shetty, S., Griffin, D.K. and Graves, J.A.M., 1999. Comparative painting reveals strong chromosome homology over 80 million years of bird evolution. *Chromosome Research*, 7(4), pp.289-295.

Shibusawa, M., Nishida-Umehara, C., Masabanda, J., Griffin, D.K., Isobe, T. and Matsuda, Y., 2002. Chromosome rearrangements between chicken and guinea fowl defined by comparative chromosome painting and FISH mapping of DNA clones. *Cytogenetic and genome research*, 98(2-3), pp.225-230.

Shibusawa, M., Nishibori, M., Nishida-Umehara, C., Tsudzuki, M., Masabanda, J., Griffin, D.K. and 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. *Cytogenetic and Genome Research*, 106(1), pp.111-119.

Singchat, W., O'Connor, R.E., Tawichasri, P., Suntronpong, A., Sillapaprayoon, S., Suntrarachun, S., Muangmai, N., Baicharoen, S., Peyachoknagul, S., Chanhome, L. and Griffin, D., 2018. Chromosome map of the Siamese cobra: did partial synteny of sex chromosomes in the amniote represent "a hypothetical ancestral super-sex chromosome" or random distribution? *BMC genomics*, 19(1), p.939.

Skinner, B.M. and Griffin, D.K., 2012. Intrachromosomal rearrangements in avian genome evolution: evidence for regions prone to breakpoints. *Heredity*, 108(1), p.37.

Skinner, B.M., Al Mutery, A., Smith, D., Völker, M., Hojjat, N., Raja, S., Trim, S., Houde, P., Boecklen, W.J. and Griffin, D.K., 2014. Global patterns of apparent

copy number variation in birds revealed by cross-species comparative genomic hybridization. *Chromosome research*, 22(1), pp.59-70.

Speicher, M.R. and Carter, N.P., 2005. The new cytogenetics: blurring the boundaries with molecular biology. *Nature Reviews Genetics*, 6(10), p.782.

Srikulnath, K., Thapana, W. and Muangmai, N., 2015. Role of chromosome changes in Crocodylus evolution and diversity. *Genomics & informatics*, 13(4), p.102.

St John, J.A., Braun, E.L., Isberg, S.R., Miles, L.G., Chong, A.Y., Gongora, J., Dalzell, P., Moran, C., Bed'Hom, B., Abzhanov, A. and Burgess, S.C., 2012. Sequencing three crocodilian genomes to illuminate the evolution of archosaurs and amniotes. *Genome biology*, 13(1), p.415.

Stankiewicz, P. and Lupski, J.R., 2002. Molecular-evolutionary mechanisms for genomic disorders. *Current opinion in genetics & development*, 12(3), pp.312-319.

Stapley, J., Birkhead, T.R., Burke, T. and Slate, J., 2008. A linkage map of the zebra finch Taeniopygia guttata provides new insights into avian genome evolution*. Genetics*, 179(1), pp.651-667.

Stiller, J. and Zhang, G., 2019. Comparative Phylogenomics, a Stepping Stone for Bird Biodiversity Studies. *Diversity*, 11(7), p.115.

Strauss, J., Barbieri, R., Yen, S. and Jaffe, R., 2009. *Yen and Jaffe's Reproductive Endocrinology*. Elsevier Health Sciences.

Striedter, G.F., 1994. The vocal control pathways in budgerigars differ from those in songbirds*. Journal of Comparative Neurology*, 343(1), pp.35-56.

Suh, A., Smeds, L. and Ellegren, H., 2018. Abundant recent activity of retrovirus-like retrotransposons within and among flycatcher species implies a rich source of structural variation in songbird genomes. *Molecular ecology*, 27(1), pp.99-111.

Sullivan, B.A. and Karpen, G.H., 2004. Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nature structural & molecular biology*, 11(11), p.1076.

Takagi, N. and Sasaki, M., 1974. A phylogenetic study of bird karyotypes. *Chromosoma*, 46(1), pp.91-120.

Theisen, A., 2008. Microarray-based comparative genomic hybridization (aCGH). *Nature Education*, 1(1), p.45.

Tio, M., Wen, R., Lim, Y.L., Wang, H., Ling, S.C., Zhao, Y. and Tan, E.K., 2016. FUS-linked essential tremor associated with motor dysfunction in Drosophila. *Human genetics*, 135(11), pp.1223-1232.

Tollis, M., DeNardo, D.F., Cornelius, J.A., Dolby, G.A., Edwards, T., Henen, B.T., Karl, A.E., Murphy, R.W. and Kusumi, K., 2017. The Agassiz's desert tortoise genome provides a resource for the conservation of a threatened species. *PloS one*, 12(5), p.e0177708.

Tortoise & Freshwater Turtle Specialist Group 1996. Podocnemis unifilis (errata version published in 2016). *The IUCN Red List of Threatened Species* 1996: e.T17825A97397562. http://dx.doi.org/10.2305/IUCN.UK.1996.RLTS.T17825A7506933.en.

Trifonov, V.A., Giovannotti, M., O'Brien, P.C., Wallduck, M., Lovell, F., Rens, W., Parise-Maltempi, P.P., Caputo, V. and Ferguson-Smith, M.A., 2011. Chromosomal evolution in Gekkonidae. I. Chromosome painting between Gekko and Hemidactylus species reveals phylogenetic relationships within the group. *Chromosome research*, 19(7), pp.843-855.

Tsipouri, V., Schueler, M.G., Hu, S., Dutra, A., Pak, E., Riethman, H. and Green, E.D., 2008. Comparative sequence analyses reveal sites of ancestral chromosomal fusions in the Indian muntjac genome. *Genome biology*, 9(10), p.R155.

Tsurusaki, Y., Okamoto, N., Ohashi, H., Mizuno, S., Matsumoto, N., Makita, Y., Fukuda, M., Isidor, B., Perrier, J., Aggarwal, S. and Dalal, A.B., 2014. Coffin–

Siris syndrome is a SWI/SNF complex disorder. *Clinical genetics*, 85(6), pp.548-554.

Tucker, J.D., 2008. Low-dose ionizing radiation and chromosome translocations: a review of the major considerations for human biological dosimetry. *Mutation Research/Reviews in Mutation Research*, 659(3), pp.211-220.

Tuke, M.A., Ruth, K.S., Wood, A.R., Beaumont, R.N., Tyrrell, J., Jones, S.E., Yaghootkar, H., Turner, C.L., Donohoe, M.E., Brooke, A.M. and Collinson, M.N., 2019. Mosaic Turner syndrome shows reduced penetrance in an adult population study. *Genetics in Medicine*, 21(4), pp.877-886.

Ullastres, A., Farré, M., Capilla, L. and Ruiz-Herrera, A., 2014. Unraveling the effect of genomic structural changes in the rhesus macaque-implications for the adaptive role of inversions. *BMC genomics*, 15(1), p.530.

Uno, Y., Nishida, C., Tarui, H., Ishishita, S., Takagi, C., Nishimura, O., Ishijima, J., Ota, H., Kosaka, A., Matsubara, K. and Murakami, Y., 2012. Inference of the protokaryotypes of amniotes and tetrapods and the evolutionary processes of microchromosomes from comparative gene mapping. *PLoS One*, 7(12), p.e53027.

Valenzuela, N. and Adams, D.C., 2011. Chromosome number and sex determination coevolve in turtles. *Evolution*, 65(6), pp.1808-1813.

Venkatesan, B.M. and Bashir, R., 2011. Nanopore sensors for nucleic acid analysis. *Nature nanotechnology*, 6(10), p.615.

Ventura, K., Moreira, C.N., Moretti, R., Yonenag-Yassuda, Y. and Rodrigues, M.T., 2014. The lowest diploid number in Testudines: Banding patterns, telomeric and 45S rDNA FISH in Peltocephalus dumerilianus, 2n= 26 and FN= 52 (Pleurodira, Podocnemididae). *Genetics and molecular biology*, 37(1), pp.61-63.

Vickrey, A.I., Bruders, R., Kronenberg, Z., Mackey, E., Bohlender, R.J., Maclary, E.T., Maynez, R., Osborne, E.J., Johnson, K.P., Huff, C.D. and Yandell, M.,

2018. Introgression of regulatory alleles and a missense coding mutation drive plumage pattern diversity in the rock pigeon. *Elife*, 7, p.e34803.

Vignal, A. and Eory, L., 2019. Avian genomics in animal breeding and the end of the model organism. *In Avian Genomics in Ecology and Evolution*. Springer, Cham, pp.21-67.

Volfovsky, N., Oleksyk, T.K., Cruz, K.C., Truelove, A.L., Stephens, R.M. and Smith, M.W., 2009. Genome and gene alterations by insertions and deletions in the evolution of human and chimpanzee chromosome 22. *BMC genomics*, 10(1), p.51.

Volik, S., Raphael, B.J., Huang, G., Stratton, M.R., Bignel, G., Murnane, J., Brebner, J.H., Bajsarowicz, K., Paris, P.L., Tao, Q. and Kowbel, D., 2006. Decoding the fine-scale structure of a breast cancer genome and transcriptome*. Genome research*, 16(3), pp.394-404.

Völker, M., Backström, N., Skinner, B.M., Langley, E.J., Bunzey, S.K., Ellegren, H. and Griffin, D.K., 2010. Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution. *Genome research*, 20(4), pp.503-511.

Wang, K., de la Torre, D., Robertson, W.E. and Chin, J.W., 2019. Programmed chromosome fission and fusion enable precise large-scale genome rearrangement and assembly. *Science*, 365(6456), pp.922-926.

Warren, W.C., Clayton, D.F., Ellegren, H., Arnold, A.P., Hillier, L.W., Künstner, A., Searle, S., White, S., Vilella, A.J., Fairley, S. and Heger, A., 2010. The genome of a songbird. *Nature*, 464(7289), p.757.

Warren, W.C., Hillier, L.W., Tomlinson, C., Minx, P., Kremitzki, M., Graves, T., Markovic, C., Bouk, N., Pruitt, K.D., Thibaud-Nissen, F. and Schneider, V., 2017. A new chicken genome assembly provides insight into avian genome structure. *G3: Genes, Genomes, Genetics*, 7(1), pp.109-117.

Waterston, R.H., Lander, E.S. and Sulston, J.E., 2002. On the sequencing of the human genome. *Proceedings of the National Academy of Sciences*, 99(6), pp.3712-3716.

Webb, D.M. and Zhang, J., 2004. FoxP2 in song-learning birds and vocal-learning mammals. *Journal of Heredity*, 96(3), pp.212-216.

Weinstock, D.M., Richardson, C.A., Elliott, B. and Jasin, M., 2006. Modeling oncogenic translocations: distinct roles for double-strand break repair pathways in translocation formation in mammalian cells. *DNA repair*, 5(9-10), pp.1065-1074.

Weirather, J.L., de Cesare, M., Wang, Y., Piazza, P., Sebastiano, V., Wang, X.J., Buck, D. and Au, K., 2017. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Research*, 6, p.100.

Weiss, M.M., Hermsen, M.A., Meijer, G.A., Van Grieken, N.C., Baak, J.P., Kuipers, E.J. and Van Diest, P.J., 1999. Comparative genomic hybridisation. *Molecular Pathology*, 52(5), p.243.

Weissensteiner, M.H. and Suh, A., 2019. Repetitive DNA: The Dark Matter of Avian Genomics. *In Avian Genomics in Ecology and Evolution*. Spring, Cham, pp. 93-150.

Weissman, B. and Knudsen, K.E., 2009. Hijacking the chromatin remodeling machinery: impact of SWI/SNF perturbations in cancer. *Cancer research*, 69(21), pp.8223-8230.

Werneburg, I. and Sánchez-Villagra, M.R., 2009. Timing of organogenesis support basal position of turtles in the amniote tree of life. *BMC Evolutionary Biology*, 9(1), p.82.

White, M.J.D., 1969. Chromosomal rearrangements and speciation in animals. *Annual review of genetics*, 3(1), pp.75-98.

Wienberg, J., Stanyon, R., Jauch, A. and Cremer, T., 1992. Homologies in human and Macasa fuscata chromosomes revealed by in situ suppression hybridization with human chromosome specific DNA libraries. *Chromosoma*, 101(5-6), pp.265-270.

Wienberg, J., Froenicke, L. and Stanyon, R., 2000. Insights into mammalian genome organization and evolution by molecular cytogenetics. *In Comparative genomics* (pp. 207-244). Springer, Boston, MA.

Wienberg, J., 2005. Fluorescence in situ hybridization to chromosomes as a tool to understand human and primate genome evolution. *Cytogenetic and genome research*, 108(1-3), pp.139-160.

Wójcik, E. and Smalec, E., 2016. Constitutive heterochromatin in chromosomes of duck hybrids and goose hybrids. *Poultry science*, 96(1), pp.18-26.

Wu, R. and Kaiser, A.D., 1968. Structure and base sequence in the cohesive ends of bacteriophage lambda DNA. *Journal of molecular biology*, 35(3), pp.523-537.

Xanthopoulou, L., Mantzouratou, A., Mania, A., Cawood, S., Doshi, A., Ranieri, D.M. and Delhanty, J.D., 2010. Male and female meiotic behaviour of an intrachromosomal insertion determined by preimplantation genetic diagnosis. *Molecular Cytogenetics*, 3(1), p.2.

Yao, H. and Schnable, P.S., 2005. Cis-effects on meiotic recombination across distinct a1-sh2 intervals in a common Zea genetic background. *Genetics,* 170(4), pp.1929-1944.

Ye, K., Wang, J., Jayasinghe, R., Lameijer, E.W., McMichael, J.F., Ning, J., McLellan, M.D., Xie, M., Cao, S., Yellapantula, V. and Huang, K.L., 2016. Systematic discovery of complex insertions and deletions in human cancers. *Nature medicine*, 22(1), p.97.

Zhan, X., Pan, S., Wang, J., Dixon, A., He, J., Muller, M.G., Ni, P., Hu, L., Liu, Y., Hou, H. and Chen, Y., 2013. Peregrine and saker falcon genome sequences

provide insights into evolution of a predatory lifestyle. *Nature genetics*, 45(5), p.563.

Zhang, G., Li, C., Li, Q., Li, B., Larkin, D.M., Lee, C., Storz, J.F., Antunes, A., Greenwold, M.J., Meredith, R.W. and Ödeen, A., 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science,* 346(6215), pp.1311-1320.

Zhang, G., Li, B., Li, C., Gilbert, M.T.P., Jarvis, E.D. and Wang, J., 2014. Comparative genomic data of the Avian Phylogenomics Project. *GigaScience,* 3(1), p.26.

9 Appendix

Supplementary Table S1: Details for all primer pairs for each of the 7 genes being tested.

Supplementary Table S2: Amplicon sequence for the primer pairs being tested.

Supplementary Table S3: The complete list of BACs and their coordinates in the budgerigar genome.

Supplementary Table S4: The complete list of BACs for the comparative mapping study and their coordinates in the chicken genome (*Gallus_gallus*-4.0).

Supplementary Table S5: The full table of BACs successfully hybridised in the common blackbird, Atlantic Canary, Eurasian woodcock, helmeted guinea fowl, houbara bustard, mallard duck, and rock dove.

Supplementary Table S6-S13: FLpter values, standard deviations, and the number of mitotic chromosomes measured for the chicken, common blackbird, Atlantic canary, Eurasian woodcock, helmeted guinea fowl, houbara bustard, mallard duck, rock dove

Supplementary Figure S1-S9: Ideograms indicating relative hybridisation positions of BACs for chicken chromosomes 2-9 and Z, with BACs labelled in order of position on the chicken chromosome. BAC positions are indicated for chicken (GGA), mallard (APL), Eurasian woodcock (SRU), pigeon (CLI), helmeted guinea fowl (NME), houbara bustard (CUN), common blackbird (TME), and Atlantic canary (SCD).

Supplementary Table S14-S20: BACs involved in each chromosomal rearrangement and the corresponding p-values for the mallard duck, Eurasian woodcock, rock dove, helmeted guinea fowl**,** houbara bustard**,** common blackbird and Atlantic canary.

Supplementary Table S1: Details for all primer pairs for each of the 7 genes being tested.

Gene Name	Primer Pair	Primer sequence	Strand	5' Position	3' Position	T_m	Max. misprime T_m	Primer dimer dG	Amplicon size	Amplicon GC
FUS_1	1	CATTGCGCCACGAGAAGTTC	+	407	426	62.94	51.94	-1.65	1670	54.31
FUS_1	1	AGCATGTCATGTCCTCTGGTT	-	2076	2056	62.86	52.8			
FUS_1	2	CATTTCTGCCTTCACACTCCA	+	4103	4123	61.92	51.35	14.17	1733	54.65
FUS_1	2	TTTGAGGTCAGAAAGTGCCAAAG	-	5835	5813	62.89	51.98			
FUS_1	3	CTTTCCACACCTCATCTTCCC	+	8186	8206	62.97	52.19	0.77	1602	53.68
FUS_1	3	TGTGAGATTAAAGCCAGGAAGG	-	9787	9766	62.63	52.15			
FUS_1	4	TGAACAACGACACCTCAGGA	+	1972	1991	62.42	52.16	3.6	2153	51.18
FUS_1	4	TTGGAGTGTGAAGGCAGAAATG	-	4124	4103	62.46	51.65			
FUS_1	5	TTTGGCACTTTCTGACCTCAAA	+	5814	5835	62.02	51.44	26.13	2391	47.85
FUS_1	5	GAAGATGAGGTGTGGAAAGAGAG	-	8204	8182	61.73	51.65			
TYK2	1	GAGATACAGAAGATGAATGGTAAGC	+	278	302	60.95	47.87	-13.25	2074	53.86
TYK2	1	TGCTGTTTGTAACCTTCCAG	-	2351	2332	60.01	49.86			
TYK2	2	GAGGGTTCTCAAGGTGGTAAA	+	4175	4195	62.6	52.12	-3.28	2313	52.83
TYK2	2	GATCCCAGCAAGTTCCGAAT	-	6487	6468	62.13	50.9			
TYK2	3	CAGATATCATGGTGGAGGAGTTT	+	8406	8428	62.16	50.89	-49.93	1915	55.82
TYK2	3	GAATGTGGTAGGGAGAATGGA	-	10320	10300	62.17	50.53			
TYK2	4	AAGGTTACAAACAGCATAGGG	+	2336	2356	61.17	49.95	-18.34	1943	55.89
TYK2	4	CATCTACGCCCTTTAAGACC	-	4278	4259	61.03	48.85			
TYK2	5	TGTTACTTGACGAGAGGAGTG	+	6315	6335	60.45	48.18	16.4	2117	59.09

Gene Name	Primer Pair	Primer sequence	Strand	5' Position	3' Position	T_m	Max. misprime T_m	Primer dimer dG	Amplicon size	Amplicon GC
TYK2	5	CACAAACTCCTCCACCATGAT	-	8431	8411	62	50.92			
TYK2	6	GAAGATCACTGTGGCCAAGC	+	8491	8510	62.66	52.58	-3.94	1851	55.48
TYK2	6	TTTAGGGTCTTCTCAGCATGG	-	10341	10321	62.08	50.15			
IKZF4	1	GTGACATCTCCGCGAACAATG	+	630739	630759	62.64	50.4	-38.78	1610	66.46
IKZF4	1	AGCAGCCCAGTTTAACCAAA	-	632348	632329	62.85	52.09			
IKZF4	2	TCAGATGGTCTCCACTGAGG	+	634482	634501	62.37	51.12	-11.03	2499	58.74
IKZF4	2	CCTTCCTCATCGCTGTACAT	-	636980	636961	61.25	50.47			
IKZF4	3	AGGCGTTCAGATCCAGATCA	+	639412	639431	62.07	48.79	-20.71	1600	63.56
IKZF4	3	CCATTCCATACTGTCTCATCC	-	641011	640991	60.18	49.7			
IKZF4	4	ACAATGCTTCGGGAAGAGAAAC	+	632103	632124	62.51	51.65	-35.39	2397	63.04
IKZF4	4	TCAGTGGAGACCATCTGATAGAC	-	634499	634477	62.65	50.57			
IKZF4	5	GATGTACAGCGATGAGGAAGG	+	636960	636980	62.22	50.76	-12.19	2482	59.47
IKZF4	5	GCAGTGCGTTTGATCTGGATC	-	639441	639421	62.66	50.92			
IKZF4	6	CTCAGAAGTTCGTGGGTAAGG	+	640748	640768	62.66	52.01	-12.2	1988	64.49
IKZF4	6	CTCAGCCAACCTTATGCTCTC	-	642735	642715	62.45	50.94			
SKIV2L	1	TTAGAGCTCTGATCCTTCTGTATCT	+	489843	489867	62.11	51.2	17.97	1642	61.81
SKIV2L	1	TTAAAGAAGGGTACAGTGAGGT	-	491484	491463	61.64	50.62			
SKIV2L	2	GACATTGGAGATGCTGTAGGG	+	493757	493777	62.81	49.88	-10.72	2440	66.23
SKIV2L	2	CTGCATGACGTAGACAGGAAAG	-	496196	496175	62.33	49.2			
SKIV2L	3	CCATTAGGGATACTGGGTTTG	+	498502	498522	62.31	51.95	-34.08	1910	67.7
SKIV2L	3	GCGGCCTTTATTGGATGTAG	-	500411	500392	62.03	50.74			

Gene Name	Primer Pair	Primer sequence	Strand	5' Position	3' Position	T_m	Max. misprime T_m	Primer dimer dG	Amplicon size	Amplicon GC
SKIV2L	4	CCTCACTGTACCCTTCTTTAAA	+	491464	491485	60.54	48.74	-16.87	2313	62.69
SKIV2L	4	CCTACAGCATCTCCAATGTCC	-	493776	493756	62.81	50.1			
SKIV2L	5	CTTTCCTGTCTACGTCATGCAG	+	496175	496196	62.33	51.22	-26.67	2349	64.15
SKIV2L	5	CCAAACCCAGTATCCCTAATG	-	498523	498503	62.31	52.2			
AKAP8L	1	CGCCTTCACTCATATCCGTAG	+	5771	5791	62.62	50.13	-18.36	1739	68.66
AKAP8L	1	CCAGAACTTTGATGATGACTGAGG	-	7509	7486	62.54	50.08			
AKAP8L	2	TGCGGCTGAACTCAATTCTTC	+	9594	9614	61.95	51.11	11.63	1870	67.49
AKAP8L	2	TGAGACTGTCTGCCTTATAGGT	-	11463	11442	62.64	51.74			
AKAP8L	3	CTCAATACGGGTCAGTGACG	+	14766	14785	62.13	50.7	-19.69	2470	66.48
AKAP8L	3	CCGGTCGATGGAGAAGTATTT	-	17235	17215	62.19	43.86			
AKAP8L	4	TTCTGGCTGAGCAGGAAATTG	+	7504	7524	62.27	50.58	-27.36	2118	61.05
AKAP8L	4	GGGCGTGGAAGAATTGAGTT	-	9621	9602	62.63	52.3			
AKAP8L	5	GAGTCCTCAGGGTCGTCAAT	+	12657	12676	63	52.16	23	2127	65.07
AKAP8L	5	TCACTGACCCGTATTGAGGT	-	14783	14764	62.97	50.57			
AKAP8L	6	GCGCACTGATCTGAAGAGAAG	+	16107	16127	61.18	48.63	11.92	1686	62.46
AKAP8L	6	ACCCACTTCTTCAGTCCTAT	-	17792	17773	60.08	49.56			
SMARCC2	1	CCAAAGGAGACCAACGCTTA	+	693672	693691	62.64	51.39	-3.74	2031	68.39
SMARCC2	1	ATCAAACTGCGGCATTTCGAG	-	695702	695682	62.62	46.4			
SMARCC2	2	TCTTCGATCTCTTTGCTCTTG	+	697821	697841	57.85	47.32	-19.36	2471	63.82
SMARCC2	2	GGTTATGAAGCGAGACAAGCA	-	700291	700271	62.03	47.69			
SMARCC2	3	GGTGTTCCTCCTTACAGCTC	+	700403	700422	62.41	52.21	-48.52	1731	64.93

Gene Name	Primer Pair	Primer sequence	Strand	5' Position	3' Position	T_m	Max. misprime T_m	Primer dimer dG	Amplicon size	Amplicon GC
SMARCC2	3	GTCAGCCAATTCGACAACGTC	-	702133	702113	62.84	47.18			
SMARCC2	4	CAGTTTGATCTCCAGTTTCTTCATC	+	695694	695718	60.71	50.04	10.28	2149	67.1
SMARCC2	4	GCAAGAGCAAAGAGATCGAAGAC	-	697842	697820	62	50.84			
SMARCC2	5	CTTGTCTCGCTTCATAACCG	+	700273	700292	60.02	46.6	-11.77	1860	65.11
SMARCC2	5	TCAGCCAATTCGACAACGTC	-	702132	702113	61.64	45.73			
BAZ2A	1	GTCCTGGTTCTGTATGTGAGG	+	1636345	1636365	62.4	51.25	8.65	1929	70.14
BAZ2A	1	ATCGCATAGAGGTCATCGGT	-	1638273	1638254	62.56	47.52			
BAZ2A	2	GAGAAGGCCAAACCCAAAGA	+	1639507	1639526	62.9	51.33	-26.75	1719	67.19
BAZ2A	2	GTCCCTCGATGATCCATTTGTT	-	1641225	1641204	62.4	47.29			
BAZ2A	3	CACCAAGTTCTTCAAGCAGATG	+	1643927	1643948	60.52	48.68	3.76	2060	70.73
BAZ2A	3	CTCACTCGCAGAAGGTCAGAT	-	1645986	1645966	62.3	50.64			
BAZ2A	4	CTCTATGCGATGGATGAGACG	+	1638263	1638283	61.46	48.37	-43.71	1750	68.46
BAZ2A	4	GACCCACCTTGGATTTCTTCT	-	1640012	1639992	62.31	51.37			
BAZ2A	5	GAACAAATGGATCATCGAGGGA	+	1641203	1641224	62.13	52.08	-11.3	1963	66.17
BAZ2A	5	CCACACAGTTCATGGTGACC	-	1643165	1643146	62.89	51.2			

Supplementary Table S2: Amplicon sequence for the primer pairs being tested.

Gene Name	Primer Pair	Amplicon Sequence
FUS_1	1	CATTGCGCCACGAGAAGTTCATATTCTCGCAGGCCCTGCCGAGGGATGGAGCCATCAGCGTGGCCTTTTTTTTGTGTTCTTTTTAGCCCGTTTTGGTCGTTTTTTGCC CCAAAAAGACACTCACGGGTTGGGGCACTTCCAGTCCCCCGCGCGCTGCTGCCCTCCGCCTCCGCTGGGGAAGCCGCCTCGGTTGCCTCCTCCGGAACCGCCGCC GCCGCCGAACCCCCCGCGTCCCATCGGGCCTGGGGGAGCAGCGGGTCGGGTTTTGGGGTGAAAAAGGCAGGTTTTGGGGCTCGCCGAAGAAGGGCTCGCGGGAT TTGGGGCTCACCTCCCCTCCCGCGGCCTCCCGCCCGCCGCCGCCGCCGCGGTTGAAGTCGGCTCGCCGGGTGGCGAAGGAAACCTTGATGGGGTTGCCGGAGAA TTCTTTGCCTAAAAAAGGGGGGTTTCAGCCCAAAATGAGTGCGAAGGGGTGGAGGGGGGGGGGCTGGAGGGGGAGGGGGGGTCGTACCGTCAAACCAATCGATGG CGGCTTTGGCGGAGGGGGGGTCGTCGAAGGACACGGTGGCTTCGCCCTTCAGCTTGCCGGTCTCGCGGTCCGTGTACAGGTTGATCATGGGCTGCCCTGTCTTCT TATTGGTCTGCGGAGAGGAACATGCTAAGGACACGCTAAGGGCACGCTAAGGGCACGGGGTGGTAAGGAAACATAGGGCGGTGCTGCCCCCTGGTGTTGGCGTCG GGGAGTTGCGAGAGCACATGGATGGCACAGTGATGCACTGCAACCACGCAACACGCTAAGGACATGCTAACACTGACACCCTAACGACGCAACATGCTAAGGACAT GCTAAGGTTGACATGTTGAGGACACGACAGGCCAAGGACACACTAACATTCATACACTAAAGACACAACATGCTAAGGACATGCTAACAGTAAGGACACAACATGCTT AGGACACGCTAACAATGAAACCCAGAGGACACGACGTGCTGAGAACACACTAACAGTGACGCCCTGAAGACTTAACATGCTAAGGACATGCTAACAGTAAGGACGCA ACATGCTTAGGACACGCTAACAATGAAACCCAGAGGACACGATGTGCTTAAGACACGCTAATGACGACACCTCAGGACACACGTTAAGTACATGCTAACAGTAAGGA CACAACATGCTTAGGACACGCTAACAATGAAACCCAGGAGACACGACGTGCTGAAAATACACTAATAGTGATGCCCTGAAGACTTAACATGCTAAGGACATGTTAGCA GTAAGGACACAACATGCTTAGGACATGCTAACAATGAAACCCAGGAGACACGACGTGCTTAAGACACGCTAACGACGACACCTCAGGACACATCATGTTAAGTACAT GCTAACAGTAAGGACACAACATGCTTAGGACACGCTAACAATGAAACCCAGAGGACACGACGTGCTGAGAACACACTAACAGCGACACCCTGAAGACTTAACATGCT AAGGACATGCTAACAGTAAGGACGCAACATGCTTAGGACACGCTAACAATGAAACCCGGAGGACACGTACTGTAGACACGTGAACAACGACACCTCAGGACACACCA CGTTAAGTACATGCTAACAGTAAGGACACAACATGCTTAGGACATGCTAACAATGAAAACCAGAGGACATGACATGCT
FUS_1	2	CATTTCTGCCTTCACACTCCAACCGTGGCGGATTTGGGGTTATTGACTGAAGTCGGTTGGGGGGGGGGGGTTATTTTTTGCCCCCCCCACCCCCAAAAATGCCACTTTT CCACCCAAATCTCCCATTGCTCACCATGGGGAGGGTGAGGGGGGGGGCAGAGCCCTTTTATTTTGCCCCTTTTCGGGCCTTTTTCCACCCAAAACCGCCGCCATCTCC CATTGCTCACCATTGGGAGGGGGGTGAGGGGGGGGCAGAGCCCTTTTATTTTGCCCCTTTTCGGGCCTTTTTCCACCCAAAACCGCCACCATCTCCCATTGCTCACTG TCAGGAGGGTGAGGGGGGGGGGGGCAGAGCCCTTTTATTCTGCCCCTTTTCAGGCCTTTTTCCACCCAAAGCCAGCACCATCTCCCATTGCTCACCATGGGGAGGGT GGGGTGGGAGGGGGGGGGCAGAGCCCTTTTATTTTGCCCCTTTTCGGGCCTTTTTCCACCCAAAGCCACCACCATCTCCCATTGCTCACCATGGGGAGGGTGGGGT GGGAGGGGGGGGCAGAGCCCTTTTATTTTGCCCCTTTTCGGGCCTTTTTCCACCCAAAACCGCCACCATCTCCCATTGCTCACTGTCAGGAGGGTGAGGGGGGGGG GCAGAGCCCTTTTATTTTGCCCCTTTTTGGGCCTTTTTCCACCCAAAACCACCACCATCTCCCATTGCTCACTGTGAGGAGGGTGAGGGGGGGCAGAGCCCTTTTATT TTGCCCCTTTTCGGGCCTTTTTCCACCCAAAACGCGGCCCCTGTTGCGGAGAGAAGTGGGGTTACAGCACACAAATACACTCTTTCCCTGCGTTGGCCGCTACGTTC CCTTTTTGGGTTCAAACGCTCAACTCTTGCCACTTACCACCAAATTTATTGAAGCCACCACGCTCACCACCGCTGAGGAGATAAATTAGAAGATATGAAGCTTTTTTTTT GGGTTTTTTTGTTTTTTTTTTGGTTTTTCGTTCATTTTATGCTTTTGAAGGGGGGGGGGGGCAGTTTTATAATCCCTGAATTGCTTTTTTTTTGTTATTATTTTTTTTGGGGG GGGTAAAACCGGCGCTTTTCCTCCATCCTGAAGCCCCTCTGTGCTTTGTTTTGGGGGGGGGGGGGGGGGGCTGTTTTGGGGTATTTTGGGGGGGGTTTTGGGGGT CTCCATCCGTCGTTGACGCTCCTCTCGCACGCAGCGCTGCCCAATGGGGGGGGTGGGGTGGGATTGGGTGGGATTTTGGGGGGGGGGGGGGGGTCAGAAAATGG CTAAAAACGCCGATTGGAGCTCTTTGCGGCCGTTCTGAAGGATCCCTGCGGGGCACCTTTGGGCCCGGCGTACAAAAAGGCTTTGGCCCCCTTTTTTTGGGGTCGTT TTCCCTGCTTTCCGCACACCCCACACGGTTGTGGGTCGTTCCCCCCCCCCCCCCCCCCCCAAAAAGGAATCCTAAAGTTGGGGGAAAAAAAAAGCAAAAAACACAAAAA TGCCCAAATCAAACGGCCGTTTGACCCTAAAAAAATGAACCCCGGGCAATGTGGCCCTTTGGGGGGCGTTTTTCTATGGGGTTTTTTCCCTGCTTTCTGCAGACCCCC CCCCTGTAGGACCGCCCCCCCCCCATCTCATTTTACTGAATCCAAAAGTTGGGAAAAACCCAAATCTGCCCTTTTTTTGGGGGGGGCCGCTCCAAAAATGGCTCTTTG GCACTTTCTGACCTCAAA

Gene Name	Primer Pair	Amplicon Sequence
FUS_1	3	CTTTCCACACCTCATCTTCCCTCCTATTTCCTCATTTTTTTATATATTTTTCTCACCTTTCCTCATTTTTTCCCACCTTTCTCCCTTCCCATTTCCTCCCACTTCCCCTTCC CAACTCCTATCCCATTCTTCTCCTTTTTTCCTCCATTCCTCCTTATTTTCCATCGTGTTTTTTTACCACCGCGGCTGTGCCACGCATCACCTTCACACCCAACGTTGGTG TCTGTCCCTAAAAACACCTTTTTCTCCCCATTTTTGGGGGGCTCTTACCACTTCGCACAGCTCCATGTGCCCCCACTGAACCTTTTTGGGTGCATTTTCGCTCCTTTTG GTCAATTCACCCATTTGCTATCGGGAGGACCCCTCCCATCAACCTCTGCCCTCAATTCTGCAGTTTCATTTTTTACCATTTTCCCCCGTTTTTTGACCACAGCCCCCAC TCCTGGGGGGTTCCCCACCACCTCTCTGTGCTTTTGGTTATTTTTGTGTACCCCTCCCCCTCTTTTTTCTTTTTCTTTTGACCGTTTTCCCCCCATTTTCAGCCACAGAA GCCCCCACCCCCAGGGGTTTCCCTCCGTCATTTTTGGGGCATTTCCCCCACCATACACCCCCATTCCCAGGGGGTTCCTCACCACCTCTCTCCCTTATATTTAACATT ATCATCATTTCTTTTTTCCACCGTTTCCCCCATTTTTGGACCATTCAAAGCACCGCTAATCGGGGGGTCTCTCCCTCATTTTTTTGGGGGGGCATTTTCCCCCATTTCTC CCCCCACTTTCCCCCTCTGGAGTCCCCCACTCCCCAGAGGGTTCCCCACCAAATCATTTTCCTTCATACCACAATTCTTTTCTTTTTTTTTTTCCCCCCATTTAACCCCA TTTCCACCTCAGGAAAGCCCCGCTATATCCAGAGCTTTCCCCCCAGTAGGGCATTATTCCCCATTTCTCCCCTCCTCACAGCACTATCAGCCCCCCACCCTCTGACATC CCCCACCACCCCTTTTCTCCCCACGCTTCCATCATCCACCACTCTCTCTTTTTTAAACCATTTCCCCCCCCAACTCCAACCACCGCTCTTGGAGAACCGGGGGTGGGGG GGGATGTCTCTCTCCCTCATTTTCGGGGCGTTCTCCCCCCATTTCACCACCTCCAGAAGCCCCCACACCCCACAGGGCTTCCCCCCCCCTCCTCCTTTTGCCCATTTTC TCCCTTTTTGGCCCCGTTTGCTGGCCCACACCTACCCCATGTTGCCTCCACGGCCTCCCCGGCTACCACGGCCACCGCGCTCGTAGCCGCCCCGGCCGTACCCTCC GCGCCCGCGGCTGCGCTCCTGCTGCCCTCCATAGCTCTGCTCCTGGTATCCTCCCGAGCTGCTCAGCGGGGCCTGCTCCTGCCCATAGCTGCTGGCTGGGGGGGG GGGAAAAAAAACCACATCACGCATCAGGACGTTCCCTTTAATGCCAGGACGTTCCCTTTAATGCATTTCCACCCCCCCGCCCCATCCACAAATGCTTCCTTTTTCAAC CCGCTCTGTTGCAAACTTCAGTTTCCAGCTCACGTTTTAAACTTAAGTCCCACCTGCCCCCTTCCTGGCTTTAATCTCACA
FUS_1	4	TGAACAACGACACCTCAGGACACACCACGTTAAGTACATGCTAACAGTAAGGACACAACATGCTTAGGACATGCTAACAATGAAAACCAGAGGACATGACATGCTGAG AACACACTAACAGTGATGCCCTGAAGACTTAACATGCTAAGGACATGCTAACAGTAAGGACGCAACATGCTTAGGACACGCTAACAGTAAGGACGCAACATGCTTAG GACACGCTAACAGTAAGGACGCAACATGCTTAGGACACGCTAACAGTGAAACCCAGAGGACATGACGTGCTGAAAACACACTAACAACAACACCTCAGGACACACCA CGTTAAGGACATGCTAACGACGATGCCTTAAGGACGCAACACGCTAAGGACACGCCAACAATGATGCCTTAAGAACACAACATACCAAGAACACGCTAACAACGACA CACTACAGCCACGAGATTCACTCCACTGTGAGCTGTGCCAATGTCTCAGCACGAGAGCAGCTCCCACACACTACAGTCACGCAACAACACGCCAAGGACACGCTAAG AACACGCCAAGGACACGCCAAGTTCACCTTAATGATGCCGATCTGCTTGAAGTAGTCGGCCACCGACTCGATGGTGACGTTCTCTCCCAGGCCTTGCACGAAGATGG TGTTGTTGTCCGAGTTGTCCTGCTCACCTGCGGGAAAGCTCGTTAGTGCTGCTCATTAGCACCGCTCATTAGCACCACTCATTAGCTCCCGTGCCCCTCCAATTAGCT CTGCCCCCCTTTGGTTCATCAGATCCCTCTGCAACAGTCCCCTCATCACCTCCCTGACAGTGCCCCCCCCCCCCCCCCCGTTCTCTCATTAGCCTCATTAAATCCCCCC TCGTTAATGGCACTTTTGGACGTTAATTAATTAACTCGCTAGCGGCCTTCAGAAATCCATTAGCTCCCCACTGACCCTCTTTAGTAATCAATTAGCCTTTCCCCTCATTA GCCCGCTCAAAATCAACCCTCCCTCTCATCAGCCCCCTCCCATCGCTCATTAACACCCTCCAGAAGTTAATTACCATCACCCACCCCCACCCCCCCCCTTGGATTAGG TTAATTAACACCCCCCCCCCGCACAAGCCTCCTCATCAATTACACCTTCACAATTAATTAACCCTGCCATTAAACCGCGCAGGAACGGTTAGCTCCCCCCCCCCCCCG TCATTAGGCTAATGAGCCGCTGACAGCCCCTAATTAGCGCCCGGCTCTAATTACCAGGGTCATGGCGCGGCCCTTGGTCCCGGGGTCCTTTCAGCACATGCAAGAA ATTAACACAGAGCGTTAATTAGCGGGGAGGGCGGGGGGGGAGGGGCGCTGCACACCCCCGCCCCGAACCCTCCAACATCACTCATAATTAATAATACTCATTAATTA ACCAAAAATAACCCCCACCCCCGCAATAATAACCTCTCACCCCCTCCCCCCCCCCAACCTGTCGCCCCACACCCCCATTTTGGGGCGAAAACCCATCGTTTTGGTGC TTTTCTTGGCGCCTCCCCCAAACGGAGGCCGTTCCGCCGTTCGCCATTTCGGGCCTTTTCAACCCAAAACCGACGCCGCCCCGGGCTTCAATTTTCCACCATTTTGG TCCAAACACACCGTTTTTTTTTGGTTTTTTTTGGTGATTTTTGTCGTGAAGGTACCAAAACCCGCCACCAACGCCCTCGGATCGGTGCAAAATCTCTCGGATTCGGTCC AAAGCCCCATTAAACACGCACTTTTTGGTAACGTTCGCCGTTCGGTCACTGCAGCGCTCACCACCACAACCACCACCCCCGTTTTTTTGGGGTGAAAAACGTCAATTT GCCCTTTTTTTTTTTTTTTTTTTTTTTTGCTTTTTGTTTTCTCCCCGTTCATAATCGCACCTCCGCTCAGCGCCACCGCGCACCCAAATTTTGCCCATTTCGGGCCTTTTTCCC CCCAAAACCGCAGCTTTCGGGCACCGTTTTTGTTTTGCTTTTAGGAACCACATGCAACCGTTTTTTTATAGTTTTTTTGTGCTCAACACCAGCTTTTTCAACCCAAAAAA AAAATACCTTTTTTTTTCTCTCTCATTTACCATTTTTTGGCTCATTTCCCCCTCCTCTCTCCCCCCCCCCCCCCCCCCACCCCCATTTCTGCCTTCACACTCCAA

Gene Name	Primer Pair	Amplicon Sequence
FUS_1	5	TTTGGCACTTTCTGACCTCAAAAATAAAAGCCCTGTGGCCCTTTTGGGGGCCTTTTTGGGGTATTTTTAGCCCTTCTGGGGTCTTTTTTGGCCATTTTGGGGCCCTTCT TGGCCTTTCTGAGGCCTTTTCTGGTCTTTCTGGGGCCTTTTCTGGCCCTTTTAGAGCCTTTTTTGGACCTTTTGGGGTCTTTGGCCCTTTTAGAGCCTTTTTTGGCCAT TTGGGGTCTTTTATGGCCTTTGTGGGACCTTTTTTGGCCTTTCTGGGGCCTTTTTTGTCCTTTCTGAGGTCTTTTCTGGCCATTTTGGGGTCCTTTTTGGCCATTTTGG GGTCCTTTTTGGCCATTTTGGGGCCTTTTCTGGCCATTTTGGGGCCCTTTTTTGGCCTCGCCGGCTGCTTGCAGCAGAGCCCCCACAGGACGCTCCCCCCTCCCTTTC TCCTCTCTCTCTTCATCGCTTTTTTTCCCCTCTTTTTCCTTATTATTATTATTTGGTCTATTTTCTCCTTTTTTTTTTTTTTTTTCCATTTGTTTTTTCCTTTTATATTTTTTTTT AATTATTTTTTCCCCCTTTCTGCCTTTTTTTTGAACCCAAAAGCTGCGGATAAAGGCGGAGCATTGCAAAGTGTCGCTGCCTTTCACACCTGCAGCACACAAAACCAGA GTTTGTTAACAGGCACAAATAAGCAACCGCCGCTCGCCCCATACTCACCTCCGCGCCCCAAAATCGGCCTTTTTCGGCCCAAAACCACACGGCAGCGACACAAACCC CCAACCTTTTGGCCATTGGTGCTGCGTGCACCGACCCACGTGCGGCTGCGGCACCGTTGTTTGCCTCTGCAGTGAGCACGCAATGCGCCGTCGGCCCCTTTTTGCC CCTCACTTTGCTCTCATTTTGCCGTCTTTCGCTGCTTTTTTGCCCCTTTTTCTTTTATTCTTTTCCGTCTTTTCCCTCTTTTTTTTTTTTTTTTCTCATCTTTATTCCAGGTTTT AATGCCATTTTTTTCTCCTCTCGATTCCATTTTCCCCTCGTTTTTTCTCACCTTGATTCCATTCTCCTCATTTTTACTCCATTTTTCCTCCATTTTTCTCTCACATTCCCTTT TTCCTCATTTTTTATTCCGTTTTTACTCCGTTTTTTCCCATCTTTATTCCGTTCTCTTTCATAATTCCGTTTTCCCCCTCGATTTCTTCTCACCTTTGTTCCATTTTGCCCCA TTTTCTGTCATTTATTCCATCTTTCCCTCATTTTCCTCAGCTTTATTCCATCCACGTTTCTGATCCCATTTCCCCCACCTTTATCCCACTGCTCCCATTCCTCTCTCCCTA TTCTTCCTTTTCTCCCCCAATTCCCCCCCATTTTCCCCTCCTTTTTCTCTCCTCTTCCCAACCTTCCCCATTTTCTCCCACCTTTCCATCCCATTTCTACCCTCCGCTCAT TTTCTCACACTTCCACTGCATTTCCTCACTCTTCCCCCTTCCTTTCCCTTTTCTCCCTCTCCGTTTCCTCTTCCCTTCCTCTCTCCCCCTCCCTTTTCTCTTTCCTTCCCC TCTCCCTTTTTCCCTTCCTCTTTCCCCTTCCTCTTTTCCTTCCTTTCTCCCTCTCCATTTCCCCTTTCCCTTCCCCTCCCTCGCCTCTTCCTTCCTCTTTCTCCCTCTCCA TTTCCTCTTTCCCCTCCTTTTCCTCTTCCTTTCTCCCCATTTCCCCTTCCCTTCCTCTCCCTATTTTCCCTCCTCCTTCCTCTTTCTCTCTCGGTTTCCTCTTCCCTTCCT CTCCCCCTCCTCCTTTTTCTTTCCTTCCTTTCCCCCTCCTTTCCCCCTCCTCTCCATTTCCCTTCCTCTTTCCCCCTCCTCTTTTCCTTCCTTTCTCCCTCTCCATTTCTT CTTTCCCCTCCTTTTCCTCTTCCTTTCTCCCCATTTCCCCTTCCCTTCCTCTCCCTATTTTCCTTCTTCCTTCCTCTTTCCCTCTCGGTTTCCTCTTTCCTCCTTCCCCTC TTCCTTCTCCATTTCCTCTTCCCTTCCTCCTCTCCTCTCCTTTTCCTCTTTCCCTTTCTCCCCCTCTATTTCCTCTTCCTTTTTCCTCTTTCTCTCTGCTTTATCTCTACTT CCTCTTTCCCCATCCTCCTTTTCCTCTTCCCCTCTCCTTATCCACTTTCTCTCCCCATTTCCTTTCCTTCCTCTCTTTTCCTCTCCAGTCTTTTCCTTTCCTTCTCCCCCAT TCCCCCCAGTTCTCATTTCCTGACCTTTCACACCCCCTCATTTCCCCCTCACATTCCCTATTTCCTCACTTCCTCCCCCGCCCCCATTCAATTTCCTGTTTCCTTTCCTT CCCCCATTTTCCTTTTCCCCCTCACCCTTCCTCATCTTCCCACTCCTCTCTTTCCACACCTCATCTTC
TYK2	1	GAGATACAGAAGATGAATGGTAAGCTGAAGGTTGCCCATGGATGCAGAATCTAGGTCACAGAGGTGCGGGACTGTTGGGTGAGAAGGTCCTGGAAGGTTGATGAGC CATTAGAGTGGGTTGAGAGGGTCCTGGAAGGTCACAGGAGGTGCGGACTGTTGGGTTGAGAGGGTCCTGGAAGGTGATGAGCCATAGAGTGGTTGAGAGGTCCTG GAAGGTCAGAGAGGTGCGGGACTGTTGGGTGAGAGGGTCCTGGAAGGTTGATGAGCCATAGAGTGGGTGAGAGGGCCTGGAAGGTCAGGAGAGGTGCGGGACTG TTGGGTTGAGAGGGTCCTGGAAGGTGTAGAGCCATCGGGGAGGGTGAGAAGGTGCTGGAAGTCAGAGAGGTGCAGGACGTTGGATTGAGAGGGTCCTGGAAGGTT GATGAGCCATAGAGTGGGTTGAGAAGGTGCTGGAAGGTCAGAGAGGTGCGGGACGGTTGGGTTGAGGGGTCCTGGAAGGTTGTAGAGCCATAGAGTGGGTTGAGA GGGTCCTGGAAGGTCAGAGAGGTGCAGGACGGTTGGGTTGAGAGGGTCCTGGAAAGGTTGTAGAGCCGTAGAGTAGGTTGAGAGAATCCTGGAAGGTCAGAGAGG TGCGGGACTGTTGGGTTGAGAAGGTCCTGGGAAGGTTGATGAGCCGATAGAGTGGTTGAGAAGGTGCTGGAAGGTCACAGAGTGCGGCGACTGTTGGTTGAGAGG GTCCTGGAAGGTTGTAGAGCCATAGAGTGGGTTGAGGAAGGTGCTGGAAGGTCACAGAACTGTGGGACTGTTGGGTTGAGAGGGTCCTGGAAGGTTGATGAGCCAT AGAGTGGGTTGAGAAGGTGCTGGAAGGTCAGAGAGGTGCGGGACGGTTGGGTTGAGAGGGTCCTGGAAGGTTGTAGAGCTATAGGTGGGGTTGAGAAGGTGCTGG AAGGTCAGAGAGGTGCGGGACTGTTGGGTGTGAGAGGGTCCTGGAAGGTTGATGAGCCATAGAGTGGGTTGAGAGGGTCCTGGAAGGTCAGAGAGGTGCGGGACT GTTGGGTTGAGAGGGTCCTGGAAGGGTTGATGAGCCCTAGAGTGGATTGAGAGGGGTCCTGGAAGGTCAGAGAGGTGCGGGACGGTTGGGTGAGAGGGTCCTGG AAGGTTGTAGAGCCGTAGAGTGATTGAGAGGGTCCTTACAGACCACAGAAGCACGGGATGGGTTGGAAAGGTCTTAACAGATCACAGAAGCGCGGGTGGGATGGAT GGAAAGGTCCTTGCAGATTAAAAAACCATCGATAAGCTGGAAGTCCTTAGGATTGCGGAACCACGCTCTCCCACCTCAGTTCCTTCCGGACTGCACGGCATGAACGA GGCGGGCCGACGTCTTCCGGACGCGCGGCACAACGAGAGCGCCGGCGGACGCGGGGGGCCGACAGAGCGTCCTTCCGTATTGTTGAGCAGGTCCTTGAGGCAGA AAGGGGGCGGTTGGTTGGAATGGCTTCAAGGTCATAGAGGGTGAAGGGTCGGGAGTCCTGGAAGGGTCGTAGAATCATGAAGCGATGCAGTGGAAGAGGCCATAA GGATGATGGAAGCACGGGATGGGTGGAGGGTCCTTGAAGGTAACAGAAGAACGGAACGATTCGGTTGAAAGGGTTCTTAAAGGTTATAGAAGCACGGAAGGGGTTA AAAAGGTCCTTAAAGGGAATGAAAACGTAGAGCGGGATGGATTGAAAGGGTGTTTAAGGATGATAAAAGCATGGAACAGCCTGGAAAGGTCTTTAATGATCACAGAG TCTGGATCTGTGGATGAGCTGGAAGGGTCCTGGAAGATAATAGAACCACAGAATGGGTTGGAAAGGTCCTGGAAGATTATAGAACGACAGAATGGGTTGGAAAGGTT CTGGGAGATAATAGAACCACAGAATGGGTTGGAAAGGTCCTGGAAGATTATAGGACCACAGAATGGGTTGGAAAGATCATGGAAAGTTATAGAACCACAGAATGGGT TGGAAAGGTCCTGGAAGAAAATAGAGCCATAGAATGGGGATGGAAAGGTCCTGGAAGGTTACAAACAGCA

Gene Name	Primer Pair	Amplicon Sequence
TYK2	2	GAGGGTTCTCAAGGTGGTAAAATGGGGTGGGGTTGGTTGGGAGGGTCCTTAAGGGGCATAGATGGGGTGGGATTGAGTTGGGAGGGTCTTAAAGGGCGTAGATGG GGTGGGTTTGGGTTGGGAGGGTCCTCAAAGGTGGTAAAATGGGGGGGGGTGAGTTGGGAGGGTCCTTGAAGGGCGTAGATGGGGTGGGTTTGGGTTGGGAGGGT CCTTGAAGGTCACAGACCCACGGGATGTGTGGGAAGGGTCCTTGAAGGTCATAGAACAACGTAACGGTTGGGTTGGAAAGGACCTCATGGCCCCCCCAACCCCAAT CCCCGCTGTTCCCAACTCTCCATTTCGCTCTGCAGAGCGTTGAGGTCATTTCCCACCGCGGATATTTCGGGCGGAGGAGCAGACGGAAGGAACCGGAACCCCCCGC ACCACCAGAGCCCCCATGGGTGCACTTCTGTGACTTCCAAGAGATCACCCACATCGTCATCGAGGAGCGCAGGGTCAGCGTCCACCGGCAGGACAACAAGTGCATG GTGAGCGCCGCGGTGGGAAGGGGGGGATTGTTCCTTATTGGGTTTTTTGGGTAGAGGTTGGATGGATTTGGGGGTGCTTGGATGCGGAAGATGGAAAAGACCGATA GGGTTCGGGTTGAGCCAAGGGTTGACCAATGGTTAACCCGGCGTTGACCAAGCATTGATCCAACACTCAACGCAGTGGTTAACTAATTAAGACCAATGGTTGACCAA AGGTTGACCCAACACTCAATGCAGTGGTTGATTTGTTAAGACCAACGATTGATGTGATTGATTAAGACCAATGGTTGAGCCAACACTCAATGCAGTGGTTGATTTGTTA AGACCAACAGTTGATGTGATTGATTAAGACCAATCGTTGACCAATGATTGACCCAACACTCAATGCAGTGGTTGATTAGTTAAGACCAACGATTGATGTGATTGATTAA GACCAATGGTTGACCCAACACTCAGTGCAGTGGTTGATTAGTTAAGACCAACGATTGATGTGATTGATTAAGACCAATGGTTGACCAATGGTTGACCCAACACTCAAT GCAGTGGTTGATTTGTTAAGACCAACGATTGATGTGATGAAGACCAATGGTTGACCAATGATTGACCCAACACTCAATGCAGTGGTTGATTTGTTAAGACCAACGATTG ATGTGATTGATTAAGACCAATGGTTGACCCAACACTCAATGCAGTGGTTGATTAGTTAAGACCAACGATTGATGTGATTAAGACCAATGGTTGACCAATGGTTGACCCA ACACTCAATGCAGTGGTTGATTTGTTAAGACCAACGATTGATGTGATTGATTAAGACCAATGGTTGACCAATGGTTGACCCAACACTCAATGCAGTGGTTGATTTGTTA AGACCAACGATTGATGTGATTGATTAAGACCAATGGTTGGCCAATGATTGCCCAGCGTTTGCCCAGCAGTTAACTCAACGGTTGACACAAAGGTTGGCCGTTGGTTGA CCCAATGGTTGACCAAAAAGCCCGCAGAGACCCAGAGGTCCCCTCTGACCCCATAACTTCTGGGAATTTGGGGCTCCGCTTCCCCACACCTCATTGCCTCCCCTCCC CGACCCCCCCCCAGGAGGTTCTCCTCCCGTCCCACGCGAGCGCCCTCTCCTTCGTCTCACTGCTGGACGGATACTTCCGACTGACGGCCGACTCCAACCACTACCTG TGCCATGATGTGGCCCCGCCGCGCCTCGTCATGAGCATCCTCAACGGCATCCACGGCCCCATGCAGTGCGTCCCCCGCCCCGCTGCGAAGCCCCCACATCCTCAG GGGGTTCCCCACAGCGTTTCCCCTTCCCCAGGGAGGAATTTGTCTTCGCCAAACTGCGTCGGGAGGAGCACGAGGAGGGGCTCTACGTCCTCCGCTGGAGCGTCC TCGACTTCGATAGGGTGATTCTGGCTGTGGCAAAGAGGGACCACCAGGAGGTACTGCAGCCGTCCCACATTGGGGGGAGTTGGGATTTGGAAATCTGCTGGTTTTG GGGTTCTGTCTGAGCTCAGGAGCTGCAGCAGAGCTCAGTGTGGGAGGTGGAAGGGAGCGGTGTTGATATGGGGCGTTCTTCTCCGCTTCCTGGGCTGTGTTGTTAA TTTGTGTTACTTGACGAGAGGAGTGGGGAAGGGGAAGGGCTCCGGGTGCCCTATTTGGGGATTGAGGTTGGGGAACCAATGGGATCCACCAGGCCTGCGAGGAAC CCCGTGGGATCCACCAGTTCCATTAGGAATACAATAGAATCTACCAGTTCCCATTCGGAACTTGCTGGGATC
TYK2	3	CAGATATCATGGTGGAGGAGTTTGTGGAGCACGGCCCTCTGGACGTCCTCCTGCGGAAGGAGAAGGGCAGAGTCCCTGTGGGGTGGAAGATCACTGTGGCCAAGC AGCTGGCCAGCGCCCTGAGCTACCTGGTGAGCGCCGGGAGGGCACGGGGGGTCCCCATGGTTGCCTCGGGGTGCAGAGAAACCCATTTTTGGGGGGTTTGACCA ATGGTTGGCCAACTGTTGACTGATGGTTGACCAGGGGTTAACTAAATGGTTGACCAGTGGTTGGCCAACTGTTGACTGATGGTTGACCAGGGGTTAACTGGTGGTTG ACCAATGGTTGGCCAACTGTTGACTGATGGTTGACCAGGGGTTAATTAATGATTGACCAACGGTTGGCCAACTGTTGGCTGATGGTTGACCAGGCGTTAACTAATGGT TGACCAGTGGTTGGCCAACTGTTGACTGATGGTTGACCAGCGGTTAACTAATGGTTGACCAATGGTTGGCCAACTGTTGACTGATGGTTGACCAGGGGTTAACTAAAT GGTTGACCAATGGTTCGCCAGCTGTTGACTGATGGGTGACCAGGGATTAACTAATGGTTGACCTATGGTTGGCCAACTGTTGACTGATGGTTGACCAGGGGTTAACT AATGGCTGACCAATGGTTGGCCAACTGTTGACTGATGGTTGACCAGGGGTTAACTAATGGTTGACCAGTGGTTGGCCAGCTGTTGACTGATGGTTGACCAGGGGTTA ACTAAATGGTTGACCAATGGTTCACCAGCTGTTGACTGATGGGTGACCAGGGGTTAACTAATGGTTGACCAATGGTTGGCCAACTGTTGACTGATGGTTGACCAGGG GTTAACTAATGGTTGACCAGTGGTTGGCCAGCTGTTGACTGATGGTTGACCAGGCGTTAACTAAATGGTTGACCAATGGTTGGCCAACTGTTGACTGATGGTTGACCA GTGGTTGGCTAACTGTTGACTAATGGTTAACCATTATTTCACCACTGGTTGACTAATGGTTTACCAATTATTGACCAACGGTTGGCCAACAGTTGCCCGTGAGCCAGCA CTGTGCCCCGGTGGCCAAAACCAACCGGGCCAGGGGGGTGGGCAGCTGACCAAACCCCACGTGGTGCCTCTCCCTGCTCCCCCAGGAGGACAAGAGCCTGGTGC ACGGCAACGTGTGTGCCAAGAACGTCCTGCTGGCGCGGACGGGGTTGTGCGACGGCACGCTGCCCTTCATCAAACTCAGTGACCCCGGGGTCAGCTTCACTGCGC TGTCCCGGGAAGGTGCGGTGTTCCTGTTGTCCTCCCTGTGTCATGGGGGTTGGGAAGGCTTTAGGGTTGGGGTTGTGGTTGTGGTTATGGTTATGGTTACGGTTAG GGTTAGGATCGCATTTGTGTTGGGGTTAGGATTCCATTTCACTTGAGGATGTGGTTGGGATTGGGTTGGGATAGGGTTAGGGTAGAGTTAGCGATGGGATCGGGTTA GGGTTGGGATTGGGTTAGGGTTGGGATCGGGTTAGGGTTGGGATCAAGTCCCATTGGAACCTTCTCCTACTGGGAGCCCCCCCTGCTACTGGGAACCCCCTCCTAC TGGGAGCCCTCACTGGGACCCCCTCATACTGGGAGCCCCTTAGCTACAGGTATGCTCTCCTGCTGGGACCCCTTCCCACCGGGACCCCCTCTCCGCACCCACCACA CTGTGGCCCCTTCCCCCCCCGCCCCCCCAGAACGCGTGGATCGGATCCCATGGATCGCCCCGGAGTGCATTCAGGACGTGGGGAATCTCCACCCGGCGGCCGACA AATGGAGCTTCGGCACCACACTGCTGGAGATCTGCTTCGACGCGACGTCCCCCTCAAGGAGCGCACCCGGCCGAGGTCCCCCATCTCCATTCTCCCTACCACATTC

Gene Name	Primer Pair	Amplicon Sequence
TYK2	4	AAGGTTACAAACAGCATAGGGGGATGGAAAGGTCCTGGAAGGTTACAGACAGGATAGTGGGTTGGAAAGGTCCTGGAAGGTTATAGGACTGTGGAATGGGTTGGAA AGGTCCTGGAAGGTCATAGGACTGTGGAATGGGGTTGGAAAGGTCCTGGAAGGTTATAGGACTGTGGAATGGGTTGAAAAGGTCCTGGAAGGTTATAGGACTGTGG AATGGGTTGGGAAGGTCCTGGAAGGTCATAGGACCAGAGGAAGGTTGGATGTGAAGCTTCCCTTCAAGCTCTCTTCCCGTCACCGTGTGGCATTGTCCCCCTTTTTG CAGGGAAAATTTGACTTCGTTAACGATGCGGCGTCGTTAACAGACTTCCGGTCGGAGCAGGAAGTGCAGCGCTTCAAAAACGAGAGCCTGGGCATGGCTGTGCTGC ACCTCTGCCACATCGCCCTCCGCAGCGGCCGCTGCCTGGAGGAGGTGGCGCGGAAACACAGGTGGGGACCCATGGGGGTGGGATGTGGGGGGGCGTGGAGGCA GATTTGGGGTGTAAAGGGGGAACGTTGGCGGTGTAGGGTGGGATTTGGGGTCGGGATTTGGGGGTTTGAGGTGGGAATGTTGAGGAGTTTGGGTGGGGATTTTGG GGTCAGGATTTGGAGATTTTGGGGTCGGGATTTGGGGATTTGGGGTTGGATTTTGGTTTGGGATTTGGAGATTTGGGGGTCAGGATTTGGGGTTGGATTTTGGTTCG GGATTTGGGATTTGGGGGTGGGATTTGGGGTTGGGGGTTGGGATTTGGGGATTTGGGGTTGGATTTTGGTTCAGGATTTGGAGATTTTGGGGTTGGGATTTGGGGA TTTGGGGTTGGATTTTGGTTCTGGATTTGGAGATTTGGGGTTGGATTTTGATTCGGTATTTGGGAATTTGGGGTCAGGATTTGGGATTTGGGGTTGGGACATTGGGGT TGGATTTTGATTCGGGATTTGGGAATTTGGGGGTTGGGATTTGGGGTTGGGATATTGGGGTTGGATTTTGGTTCGAGATTTGGGGAATTTTGGGGTCAGGATTTGTG GATTTGGGGTCAGAACATTGGGACATCCCAACGTCACTCGGGTTTTTCCCCCCCTCTTCATCCCCTCAGCTTCAAGGCGTGCATCCCGCGCTCCTTCCGGCTGGCAG ATCCAGCAGGACAGCGCCGTGACGCGGCTCCGGATGAAGAACGTCTTCCGGAAGTTCGTGCGGCGCTTCCAGCGGCACACGGTGGGCGCGGGGACGCTGACGGA GGAGGACGTCATGTTCAAATATTTGGCCACCCTGGAGCTGTTGGCGCCGCGGTTCGGCACCGAGCGCTTCGCCGCCCTCTCCTTGGACGTCTCCAATGAAGGGGAG AAGGCGCAGCCCTACATCAATGGGGGGCACGCGATGGCGGAGCACGGGGACGCGGCGGTCCCCGGGGACTGTTCGGTCAGCCATGAGGTTCTGGTCAACGGTAC CGGTGGGATCCAGTGGCGGCCCGTGCCCAGCGAGGTGAGTTTGTGGGGGGTTGGGATGGGTTTGGCTTGGGAGGGTCCTCAAAGGTGGTAAAATGAGGTGGGGT TGGGTTGGGAGGGTCATCAAAGGTGGTAAAATGGGGTGAGGTTGGGTTGGGAGGGTCCTTAAAGGGCGTAGATGGGGTGGGTTTGAGTTGGGAGGGTCCTTAAAG GTGGTAAAATGGGGTGGGTTTGGGTTGGGAGGGTCCTCAAAGGTGGTAAAATGAGTGGGCTGGGTTGGGAGGGTCTTAAGGGCGTAGATGGGGTGGGTTTGAGTT GGGAGGGTCCTTAAAGGGCATAGATGGGGTGGCTTTGAGTTGGGAGGGTTCTCAAGGTGGTAAAATGGGGTGGGGTTGGTTGGGAGGGTCCTTAAGGGGCATAGA TGGGGTGGGATTGAGTTGGGAGGGTCTTAAAGGGCGTAGATG
TYK2	5	TGTTACTTGACGAGAGGAGTGGGGAAGGGGAAGGGCTCCGGGTGCCCTATTTGGGGATTGAGGTTGGGGAACCAATGGGATCCACCAGGCCTGCGAGGAACCCCG TGGGATCCACCAGTTCCATTAGGAATACAATAGAATCTACCAGTTCCCATTCGGAACTTGCTGGGATCCACCAGGTCCTTTTCAGAACCCTGTGGGATCCACCAGTTC CATTAGGAATACAATAGAATCTACCAGTTCCCATTCAGAACTTGCTGGGATCCACCAGGTCCTTTTCGGAACCCAATGGGATCCACAAGTTCCCATTAGGAACCCAGT GGGCTCCACCAGAACCCGATGGGCTCCTCCGGGCGTTGGCTGACAGATGAGTGCCGCCATCCCAACACCACCCGTTGTCCCCCCAGGACCCCGGGGTGCCGGGA GCCCTCCAGTTCCGCCAGTTCCGCATCCAGAAGAAGGGCAGCTCCTTCGTGTTGGAGGGCTGGGAGCGCGAATTCCCCACCGTGCGGGCGCTGCTGGACGCCCTC CAGGGCTGCACGCTGCGCTCCGGCAACGACAGCTTCACGGTGAAGCGGTGCTGCCCACCCAAACCGGGAGGTACGGCCAAATTCCCCTTTCCTTCCACCTCGGAG AGGGGTGGCAGCGGTGTCGGTGCAGCTCAGCGGACGTCGGCTCCTTCTCCTGGCAGAGATCTCGGATCTGCTGATCACCACGCAGAAGGTGAAGGACAACGCGAA GCGGATCCTCAACCTGACCCAGCTCAGCTTCCACCAGATCCGCAAGGATGAGATCACTCAGGTTAGGGGGCTCAGCCCCGAATCTGCGTGGAGTGGGGGGGCGAG GAGGGAGACCCGGAGCAATCCCAGAGCCATAGGGGTCAGAGGGGACTTCTGGACATCCCCCAGTCCAACCTGCACCAAAAGTGGCCAACTGTTGGCCAACTGTTG GTCAACCGTTGGTCACCGGGACACCAGTGGCCATCTTGGCCAATGAGGGCATGCTGTCAGCTTGTGGTCAACCGTTGGCCAACCGTTGGCCACCGGGATGCCGGT GGCCATCTTGGCTAACAAGGGCACACCGTCGGCTCATGGTCAGCTGTTGCTTAGCGGTTGGTCACTGGGACGTCAGCGGCCATCATGGCCGACGTGGTCAGCTGTT GGCCAACTGTCGGCCAGCGGGATGCTAGCAGCCATCTTGGCCAATGAGAGCATGCCATTGGCTCGTAGTCAACTAAGCCACCCGGATGCCAATGGCCATCGTGGCC AATGAGGGCGCACCATCGGCTTGTGGTCAACCGTTGGCCAACCGTTGGTCACCGAGCCGCCCGTTGGCTTTCTCCTCCCACGCTCAGCGAGCCCACCTGGGCCAG GGCACCCGCACCAACATCTACGACGGCGTTCTGAACGTCCGCGGCGCCGCCGGGCCCGACGCCGACGAGGCCGAATACTTCTCCACTGAACAGAACAACAACAGC GACGGCCGCGAGATGCGCGTGGTGCTCAAAGTGTTGGACCCCACCCACCGCGACATCGCCCTGGTGAGACCCCCACCCACCACCCCACCACCTACCGGCGCCGT GCCGGGGTGGCCTCAGGTCTTTGGCACGCTCTCCCTGCAGGCGTTCTTTGAGACGGCCAGCTTGATGAGCCAAGTGTCACACGTCCACCTGGCCTTCGTGCACGGC GTCTGCGTGCGGGGATCCGAGAGTGAGTGATTCAACTCTTTTTGTCCCAAAAAGTGAACGGGGCGCCCACCCGCGGCGGTGGGGTGGGGCTGAGGGGGCTTAAGA ATATTTCTGACTGTGCTGTGTTTTGGTGGTTTTGTGGTTTTTAAGGATCTTTCCTACACAACCGATCTGTGGTTTTATGGTTCTGTGATCTTGAAGGACTTTTCCTACCC GAACCCAACTCATTTTATGACCTTGAAGGACCCTTCCAACTCACTCTGTGGTTCTGTGACTTTTAAGGACCCTTCCAACCCCAAACCAACCCATTCAATGACCTTTAAG GAATCTTCCAACTCAAACCCAACCAAATCTATCCTTCTGTGACCTTCGAGGACCTTTCCAAGCCTCCCTGTTTCCATCTGCAGATATCATGGTGGAGGAGTTTGTG

Gene Name	Primer Pair	Amplicon Sequence
TYK2	6	GAAGATCACTGTGGCCAAGCAGCTGGCCAGCGCCCTGAGCTACCTGGTGAGCGCCGGGAGGGCACGGGGGGTCCCCATGGTTGCCTCGGGGTGCAGAGAAACCC ATTTTTGGGGGGTTTGACCAATGGTTGGCCAACTGTTGACTGATGGTTGACCAGGGGTTAACTAAATGGTTGACCAGTGGTTGGCCAACTGTTGACTGATGGTTGACC AGGGGTTAACTGGTGGTTGACCAATGGTTGGCCAACTGTTGACTGATGGTTGACCAGGGGTTAATTAATGATTGACCAACGGTTGGCCAACTGTTGGCTGATGGTTG ACCAGGCGTTAACTAATGGTTGACCAGTGGTTGGCCAACTGTTGACTGATGGTTGACCAGCGGTTAACTAATGGTTGACCAATGGTTGGCCAACTGTTGACTGATGGT TGACCAGGGGTTAACTAAATGGTTGACCAATGGTTCGCCAGCTGTTGACTGATGGGTGACCAGGGATTAACTAATGGTTGACCTATGGTTGGCCAACTGTTGACTGAT GGTTGACCAGGGGTTAACTAATGGCTGACCAATGGTTGGCCAACTGTTGACTGATGGTTGACCAGGGGTTAACTAATGGTTGACCAGTGGTTGGCCAGCTGTTGACT GATGGTTGACCAGGGGTTAACTAAATGGTTGACCAATGGTTCACCAGCTGTTGACTGATGGGTGACCAGGGGTTAACTAATGGTTGACCAATGGTTGGCCAACTGTT GACTGATGGTTGACCAGGGGTTAACTAATGGTTGACCAGTGGTTGGCCAGCTGTTGACTGATGGTTGACCAGGCGTTAACTAAATGGTTGACCAATGGTTGGCCAAC TGTTGACTGATGGTTGACCAGTGGTTGGCTAACTGTTGACTAATGGTTAACCATTATTTCACCACTGGTTGACTAATGGTTTACCAATTATTGACCAACGGTTGGCCAA CAGTTGCCCGTGAGCCAGCACTGTGCCCCGGTGGCCAAAACCAACCGGGCCAGGGGGGTGGGCAGCTGACCAAACCCCACGTGGTGCCTCTCCCTGCTCCCCCA GGAGGACAAGAGCCTGGTGCACGGCAACGTGTGTGCCAAGAACGTCCTGCTGGCGCGGACGGGGTTGTGCGACGGCACGCTGCCCTTCATCAAACTCAGTGACCC CGGGGTCAGCTTCACTGCGCTGTCCCGGGAAGGTGCGGTGTTCCTGTTGTCCTCCCTGTGTCATGGGGGTTGGGAAGGCTTTAGGGTTGGGGTTGTGGTTGTGGTT ATGGTTATGGTTACGGTTAGGGTTAGGATCGCATTTGTGTTGGGGTTAGGATTCCATTTCACTTGAGGATGTGGTTGGGATTGGGTTGGGATAGGGTTAGGGTAGAG TTAGCGATGGGATCGGGTTAGGGTTGGGATTGGGTTAGGGTTGGGATCGGGTTAGGGTTGGGATCAAGTCCCATTGGAACCTTCTCCTACTGGGAGCCCCCCCTGC TACTGGGAACCCCCTCCTACTGGGAGCCCTCACTGGGACCCCCTCATACTGGGAGCCCCTTAGCTACAGGTATGCTCTCCTGCTGGGACCCCTTCCCACCGGGACC CCCTCTCCGCACCCACCACACTGTGGCCCCTTCCCCCCCCGCCCCCCCAGAACGCGTGGATCGGATCCCATGGATCGCCCCGGAGTGCATTCAGGACGTGGGGAA TCTCCACCCGGCGGCCGACAAATGGAGCTTCGGCACCACACTGCTGGAGATCTGCTTCGACGCGACGTCCCCCTCAAGGAGCGCACCCGGCCGAGGTCCCCCATC TCCATTCTCCCTACCACATTCCCATGCTGAGAAGACCCTAAA
IKZF4	1	GTGACATCTCCGCGAACAATGGCAGCGGCGGTGACAGCGGAGCTGTGCCCGCAGCCCGGGGGGGGGGCACGGGGGGCTCCGGGCTCCCGCAGCCGCTCAGCACC GGGGGGGGGGCACCACGTCCCCCCCTCCCCCACCCACCCGGTCCGGGTTTCCCGGGCTCACTCTGTCCCCACCATTGGTCCCCTCCACCCCCCACCCCCCCCCG GTGCCCCCCAGCTTCGTTTTCACTCCTCACCCCATCAGTTCTCCCTCCGCTCACAGCCGCTGTGCTGTTGGCTGGGGGGGGGGGGGTCCTGGAAAGGGAGGGGGGG GCACCGGGGGGGGACCACTGTGGGTATTGGGAGCAATGGAAGCGCCGTATGGGGTCGGATCGGTGGATCTCTGTGCGGCTGTAGGGCAGTGTGGGGTCTGGGGG GCAATGGGGCTGGTGGCCTCCATGCTGGGGGGGTGGGGGGTGGGGGTGAGCCATAGGACGGAGCTCGCAGCCCCCAGAAGATGGGATGGGGATGGGGATGGGG AGGCAGCGCTTCCCAGCAGCACAGGACGGGATGCACCACGGAGGGACTCAGCCCCAGGTCCTTGAGGATGCTTCCATCGGAACCCCGACTGCGTGGAGCCGGGA TGTGGGACCACAGCTGTCCGAAGTGCTGCTCTGCATCCTCCACCCCCCCCAGGAGCTCATCTTCCACCCCCCCCCCCCCCCCCAGGAGCGCATCATCCCGAGCAGTG CTGGAGGACCTCCCCCCGTTGGCCGCGGGGGTTATGGGGAAGCGCTGCCCTTCCATCCCACGTTGGTTGGTTGGCTTTGCTGTTGGTTGGCTTTGCTGTTGGTTGG CTTTGCTGTTGGTGACGGTGCGGCTTTTGATGCTGCACGAAGAGCCCCCAACGATGGCCTTCATTGCTTCGAGCTCCGACTTCATCCCCGGACCTCATCGGCGCTGT GCGTCCTCCATCCACGGCCCTGCGGGTGAAGTTATGGCAACGGATGCCCCAGGAGGGGGGGCTGCCCCCGTGTTGGGGCGCAGTGTGGGGTCCTGAGCTGCTGC CCACCCCCCCCTATTTGGCTTTGGGTTCGCACCCCGTGGAGTTTTCCTCCCTCCACCCCACGGAGCTCCGTCGGCCGCCCCGTTTCTGCTGCAGCGCCGGTTCAAA GTAAGGGAAGGAGAGCTGTTTGGTGATGGATAAAACACAACGGGCTGCTGCCTTTCCTCCTGGCTTTGGGGCTGCCCAGCCCCGTGGGACCGTAGAGATGGACCTC GGTGCCGTTCCAACGTCGCGGCCCTTCTTCTCTATGGCTCCGCTCGCTCCGATTCGTGGCGAAGGCGGAAAAAAAGCTCTCATGGTGGAAAACGGAGCGAATGGAA ACAATGCTTCGGGAAGAGAAACTGCAGGGATTGCGTCGTACCGCCCTCACCCCGCGTGTGAAGGCAGTGGGTCCATCCTTGGGGGGGGTCAGGGGGCTGCGAGCT GCGGGGATGGCTGCGCAGGGCCGGGATGGGGGGGCACACGCTGACCCTGCTGTCCTCAGGAGCCCCCCCATGTGTGCTCAGCACAGAGGACGTCTGGATGGCTG CGTGCTCAGCCCCGGGCTTTGGTTAAACTGGGCTGCT

Gene Name	Primer Pair	Amplicon Sequence
IKZF4	2	TCAGATGGTCTCCACTGAGGGATGTGGGACTGAGCAGTTCCAGTGTTCCCTCTCTCCCTGATGGAGCACGGAGTTGTGCAGCACTCCCCCATCAGCAGAGGAACTC CCCTGTGATTGACCCACGTGCAGTGGTGGGGACACCTCTGCTTCAGTGCTGGCGTTTCAAAGCAGCCCTTTCTGAGCCAGGGTGAACAGATGAACACCGATTTCATT CTCAGCAGGGGTAACGTCGTGAGTCACCTCCAATGGACCAGGGTGGGGGGGGGGGCATTTTTCCAGCTATTTGTTATCTATTTGCCCTCTAAAAACGCTTAGAAGCT CTGTGAAGAAATGAACCACGCTCTCCACTTCAGGCTCCTGTTGGAAGCAAAAGGAGCTGAGCACGTGGACCAGCTGAAAGGGCAGTAAGGAGCCCTTCCCCGTGAG CAGCAGGACGCCGGTGATGGCATTGCTCTGATCTGACTTGACTTCACAGCTGGTTCATCCAGCAAGCATCGTGCCCCAGAGGAGCAAGAAGATTATATAACCTGTTC TGCCAGAGGAGACGGCTCTCCAGCACGGCCGGACTCAGCGTTGGTTCTGAAAGGAGAATTACAGAACAATAGTTGGCCTTCCTCCTGCGCCCGGGAGAAAAAAACA ACTCCTCGGGGGATCTCTGATTGGGTTGGGGGGGGGGAAGGGAGAAAATCAAAGTGCAGTGAATGTCAGTGCATCCGGGCGGCTCCTGCCGCCGAGATTGGGGTG TTTTGCTTTGTGTCTGTGCTCCCATGGGGTCCCCCCCCCCGGTGCTCCCAGTGGCACTGGGCAGTGTGGAAGGAAGGTTGTGGTTCTGCCAGGAGTGCTGTGGTTGC TGAGCCCCCCCCCCCCCGGGCTGCAGCCTGCGGTGCCTTTGGCCGATGGCTCAGCAGACAGCCTGGAAGTGTCTGTGGTTGCCAAAGCTGGGCTGGATGCGGAGCT GCTGTTGCTGAGAGCCTCTCTGGGAAGGCGGTGTTGTCTGCGGGACAGGAGGGGAAGTGAGGCAGTGAGAGCTGGGATTCCTGGGCTCCTCACTCAGCTTCGTGC CCGGACCCACAGCTGAGAGAGCCCTCTGCCCCCAGCAAAGACCCAAACGAGCCCTGGGGAACACATCAGCATCGGGTCTGGAAGCTTCTGGGATTCCTGCAGGGC TGCGTTCCCATTGCTTGGATGCTGGAATCATCCCTGCAAGCAGCAGCAATCTCCTTTGTGTATGTAAAGTTCAGCCCCAACAGCTCCCCGAAGGGCTTGGAAAGATC CTCGTGCTTTAAATGGAAGAAATGGGATGTTTCAAGGGATCCCTGTGAGTGTTTGCTTTGAGGCTGTGGGCTCTGTGTGCTGCCACGTGGGTGAGCCGGGCCGGGT TTCAGAGCAGCACAGGGGAGCAGAGCAGGTGGTCCCAATGTCCACGGCACATCCTAAGGGATTCCCTCCCACTCTGCTCAGTGGGAGGATGCTTCCCAATGAGGTC AGCTGAGGACGTGCTCTGCGTGGCACTGCTCTGCCTCCAGGAGGGGTCTGGAGGTGAGGCTGAGGGAGAGAGCAGCTCTGTGGAGGTAGAGGAGGGAAGGAGGA AGGCTCCTCTCCCCATCTGTGCCAGCTCAGCCCTCGCCGCCCCATTGCCATCCTCTGGAGCCCATCTGATCTGTGGCATTGCACAGCTTGCTGGAGCAGCCAGGAA TGTAGGTCGCTTGCAGCTGCCCCGCTCCTGGTGCTGCAGGAGGAAGGAGAGATGTGTGTGGGGGGGGTCTGTGGCAGTGTTGCCCCCAACTCCATGCCCTCCAGC TCCGGCCAGCAGTGCCTGTAGCTCTGCTGGGCTCTCCTGGGCTGCCCTTTGGCTCTCCCGGGGTGGGCAGAGCCCTGGCAGGGGGGGATCTCCAGCATGTGGGA CCTTTGCCCAGCAGCCACATGGGTCGCTGTGTTGTCACCCTGGGCTCTGGTGGCTGTGTGCTCTCGTAGTGTCTGCTCCCCGCAGTGCTGTCCTGCTCGGCTTCTCT CTGTTTCCATAGTTGGATCCCAGGTTTTCCAGCAAGGGTGAGAGAACCAGCAGAGTGCTCTTCTGCTGCACGCTGGGCTTTGCTGGCCGAGGACTCACAGCGGGCT CTTCGTGCCAGGAAACGTGGAGCCTTTATTGACTCGCAGTCAACGTTCAGTGTTTTCTGCGTCCTCCCAAGTGCTTTCAGTGGCTTTTCCTTCCCCCTCTGCCTTTGG ATGAGGGGAACAAAATGCTTTACACCCACAGTCGGTCGCCAAAGCACCAAAGCGGTGGCCTTTAGGACATGCAGGCGGCCATCCAGACCATAACCCCCGTGCTCAG TGCCCTGTTCCAACATCCCCAGCAGCACAAGGAGATGCGTGCTGTCAGAGCAAAGAGCCCCATCAGTGCCTGCTCCTTTAGGAGCAAACCCTGCCTGAGCTGTGTC CTTTCTATCCCCCCCCCCTGCAGCCAGCTCCATCAAGGTGGAGATGTACAGCGATGAGGAAGG
IKZF4	3	AGGCGTTCAGATCCAGATCAAACGCACTGCGTGGTGCTGGGATAGATCAGAGGCTGTGCACCCCCACCCTGCCCCCCCCCCTCCCCCCTGGGATGGCTGTACTGAG CACTGGGGACCCCTCTACCTCACAGACCCGAGGCTGGAGGGGATTTTGGGGTGGGCTTTGCGTGCCCTATGGGTGCTGTGTGTTGCATGCCTGCACCTCGGGGTC CTGAGGAGACCCCACTGCAGCCGCTGCCGGGGAGGGGGCGGCCCCTATGCGCAGCATTTGGATGTGACAGCGAGGAGTGACAACACAGAGCTGTGTGGGGAACT CAGAGCTGAGGAAGGGATGGAGGTCAATGAGATGTGATGCGATTTGCAGCCGAGGCTTTAACAGCTAAATGGAGGGAGCGACTCCTGGTGTGCCTTCCTGTGCTGA CGCTCAGGGCCTGGCAGCACCCCCTCCTCACCCCCCACCTTTGGACCGTGACAAAGCCATGTTCAGTGCAGGAGCACCGCTGAGAGGGGAGCGGGAGGGGTGCG CTCAGCATTGGCTTCTGCTGGCTGAGCTGGAGCTCCCATAGGGAAATGCACCATGGCCCTTTTGCTGCCCTCTGGCTGGGAGCAGCGGGGAGGCAGCTTTGTGCCC ACTACCTCGGCCATGGTGATGGGGCTCCGGGGGCTTTGGGTTGGGGCTCCGGGGGCTTTGGGTTTGGGCTCTGGGGGCTTTGGGTTTGGGCTCCATCCCAGTGCT GTGAAGCATCACTTGGAGCGCACAGCAGCTTCTCAGCACTGACCCAACTACCTGTACAGAGGTGCCAGAGCAGAGCCAGCCTGGAGCTCAGTGTGTGTGGCTCGGT GGCCACCTGCACGTCCTCACGGCCTGATTTCCATCGGTGTGTTTTGCAGTCTCCTCCCCCACCGTCGGCAAGCCCTACAAGTGCAACTACTGCGGCCGCAGCTACA AGCAGCAGAGCACGCTGGAGGAGCACAAGGAGCGCTGCCACAGCTACCTGCAGAGTCTGAACACTGAGCCTCAGGCGCTGGCCGGCCAGCAAGGTGGGTGCTGG GGGTGTGCAGTGGTGGGGGGGGTCCCAGCCGGAGCCCCCCTGCTGTCCGTGCATGTCCCCACATCCATGCCCAGTGGCTATGGGGTAGAAGGTGCTTACCGGGG CTGGCACAGCGGGTGCTGCCTCACCACATCCCACTGCCTGACCTGCATCCCATTGCCTCGCCAGGTGATGAGATGCGGGACCTGGAGATCGTGCCGGACTCTCTGC TGCACCCCTCTGCCGACCGGCCGACGTTCATCGACCGCCTGGCCAACAGCCTCACCAAGAGGAAGCGTTCCACGCCTCAGAAGTTCGTGGGTAAGGGGCTGCTCC CTCCCCAGCATGGGCAGATGGGAGAGGGGGGCGGTGGGGGGTGGGGATGAGGCTCACTGATGGCTCACGCTCTGCAGTTGGGCGCTGTCGGTGCTCGCATTGCG CCTTCCACCCTGCGGGCTGGGGAGGGATGGGATAGGATGGGGTGGGATGAGATGGGATGGGGTGGCGTGGAGTAGGATGAGATGGGATGGGATGGCGTGGGGT GGTGTGGGATGAGACAGTATGGAATGG

Gene Name	Primer Pair	Amplicon Sequence
IKZF4	4	ACAATGCTTCGGGAAGAGAAACTGCAGGGATTGCGTCGTACCGCCCTCACCCCGCGTGTGAAGGCAGTGGGTCCATCCTTGGGGGGGGGTCAGGGGGCTGCGAGCT GCGGGGATGGCTGCGCAGGGCCGGGATGGGGGGGCACACGCTGACCCTGCTGTCCTCAGGAGCCCCCCCCATGTGTGCTCAGCACAGAGGACGTCTGGATGGCTG CGTGCTCAGCCCCGGGCTTTGGTTAAACTGGGCTGCTGGGGGGGGGGTTTGGTTTCCCTGAGTGCGGCCCCTGGGGGGGGCAGCCCTTCTTTTAGTGCTGAGCGCT GCTCCAGTTTCGGCGCGGCGGCGCTGAGGTCTCTGCTTGGTGGCACCGAATGCTGCGTGTGTGGAGGGGCCGGGGGTTAGAAATGAGGGGAGGGGGTCTCGGCC CCATTGCCCACAGCTCAGAGGAGCTCGGGGCTGCGTGCCCACTGCTGCCAGGCTGGGCTGTGCTCCAGAGAGGGGACTTTCCCTGCACAGCTGATGGCACTGCAT GCAGGCGATGTCTGAGTGTAATGCAGCAGAGCGAAGCAATGGGAATGGACCAATAAGAGCCACTTGGATGCAGAATGTGTGCGCTGTCCCAGCAGGGTGGGTTTCC CTCTCGCTTTGGTGTGGTGCCCGCAGACCTTTGGGCACAATGCTGGTGTTGGCATTCGCTCCTATTGCCCCACACCCAACCCTTCCCATCAGCAGAATGGCACCAGA CGTCCCCACTCTGTGCTGCCAGTGCTCCGTCTCCGTCTGTCCCTGGGGTGGGTGATGGCTCTGTGCCCACACTGTGCCCCCCCCACGGCTGAGTTTGTGTTCCCCCG GGGCTGATGGTGATGTGGGGCTGTGCTGTGGGAGCTCAGGGTGAGCAGAGCTCTGCCCCAGACGTGTTTTATTTGGGGGGGGGGGAGGGGGAGAGATGGCTGAGT ACCTCCTGAGGGCAAACTCTGACTCTGCTCTGCTCTCTCTTTCAGCAACGCACCTTAAGGATGGATATAGAAGACTGTAATGGACGGTCCTACATATCTGGTACGTCC CATCTTGATTGCCACCTGCTGTTTTGGGTGGCTGGCTTCGTAAGGTGCTCACCAGACTTCAGCATCTGGCCCTGTTCATTAAGGAGCTGGGGAGAATGAACGCTGCA GGCCTGACAAACAGCTCATTAATGAGCTGTGCCAAGCCAAGCCCTGGAGGTTTGGGCTGAAGTTTGAGTGCTCGGGGCTGTTGGATGGGGTGGAAGGGGTCAGGG CGCAGCCCAGGGCTGCTGTGCTGTTAGGGGGGGATCATTGCCTCAGGGCTGTGGGTGCTGTGCTGACCCTGCAGAGCGCTCCCTGCTGCTGGCTGTGGGGTTTGG GTGCACGGAGCTGAGCTACCCTGAAGTTGGGCTGCTGCTCTGCTTGGCTTCTCCTTCAGCCTGGAGCCTGGCTGGAGGAAAAGGCCATTGGGGCACAGCTGTGTCC CTCTCAGGGCAGCGGGGATGGGGATGCTGCCAGGCTGAGCCCATCCAGCAGCAGTGCTGGCATTGGCTGAAGCCAGGAGCTCCCCGGGGAGCCACACCGTCTGC CAGTGCCATGGGCACCTCCGTTTGTTTTTGGAGCAAAGGAAATGTGACCCAGTATTCCCTGCAGCACGCTTGGCAAGCTGTCAGCCAGGGGCACGGCGGGGAGGCT GGGGGAGGGGGGGGGGGCACAGCACTGTGGGTCCTGTCTGTCCCAAAGGGGGGCTGAGCACAAAGCTCCTGTGCTTCCCATCCCAAAGTACTTCGCCCTCCCCGC TGTGGGCTGTGGGCTGTGGGCACGGCTCTGCTGCGTGCTGCTGCCACGTCCCACCTGTGGGAATGGGCATGGGGCTGCTTCTTGGGGAGGGGGGAGGGGGACAG ATACCTGCTGGGGGTCATCCCGCTGTGTGCCCCACGCAACACTGCTGCACGCCAAGGGGTCTGCGGGGCTCAACGGCGGGGGGGGTGGGCAGCTTCAGGGCTCA CTCACATCCCTCTTCCCTCTCTGTGTTCCCAAAGGCAGCGGGGATTCCTCTCTGGAGAAGGAGTTCAGCTCGGCCATCGTGGGAGCTGTGGTGAGCACACCCAACA GCCAGCACTCCTCCCCCAGCCGTTCCCTCAGCGGTGAGTGCCCCCCAACCCCCCTTAACCCCCCCCCGAGGTCTTCATGGGTCCGCTTTTCAGGTGGCACTGAGAT CCTCCACCGGCACTGTGGGTCCGGGCTGTCCCATTGGGTGTTGTACTGCTGCCACCCCACAGTGCTGCCTGTGGGCTGAGCTCTGTGGGGCCAACGGTGGCTGAG CAGCACTCGGGGCAGCGGGTGGGTGTCCGTGCTGTGGCTGTGCTCTGCAATGGGGCCGTCTATCAGATGGTCTCCACTGA

Gene Name	Primer Pair	Amplicon Sequence
IKZF4	5	GATGTACAGCGATGAGGAAGGCAGCCGGCTGCTGGCGCAGGAGGACCGCATGCTGGAGAAGGACGACAGCGTCATCGTGGAGGATTCCCTCTCCGAGCCCCTGG GTTACTGCGATGGGCCAGGCCAGGAGCCACACTCCCCCGGTGGCATCAGGTTGCCCAATGGCAAATTGAAGTGTGACGTGTGTGGCATGGTGTGCATCGGCCCCA ACGTGCTGATGGTGCACAAGCGCAGCCACACGGGTGAGTGGGGGTCAGGGGTGGGGGGGGAGGGATGGAGCTGAGCTCTGTATTCCTGGGGGGTGGGGGGGGGGG GTGGAAGGATGGAGCTGAGCTCTGTTTTCCTGGGGGTGGGGGGGGGGGTGAAGGGATGGAGCTGAGCTCTGTTTTCCTGGGGGTGGGGGGGGGGGTGAAGGGATGG AGCTGAGCTCTGTTTTCCTGGGGGTGGGGGGGGGGTTGGAAGGATGGAGCTGAGCTCTGTTTTCCTGGGGGTGTTTCATGCTCTGCTGTGTTGTGCAGGAGAAAGGC CGTTCCACTGCAACCAATGCGGGGCTTCCTTCACCCAGAAGGGAAACCTCCTGCGTCACATCAAGCTGCACTCCGGAGAGAAGCCCTTCAAATGTCCCTTCTGCAAT TATGCCTGCCGGAGGAGGGACGCCCTCACCGGACACCTGAGGACTCACTCCGGTGGGTGCCCGGCCGGGGAGATGGGGAGGGATGCACTGAGAGGTGGGAGCT GAACCCGGCCCACCCAGCAGCACTGTACCTCGGTGGGTCAGCCTGAGCTTTCCTATGGCACGGCACCGACGTTGGGTGGGTGGTGGGTGGGGTGGTGGGGGGGT CGTTGTGTCCTGGATGTGAAGCTGGGGGTGTTGGAGGTGCTGTGCCGCGGGGTGCTGGGCTCGCTGCATCTCTTTCTCGTAAAGATAAATGTTTAATGGGTTTTATT GAGGTCCCACCCTTGGGTGTGCCCTGCCCCGATGGAAACACCGAGCACGGAGGGGTTCCGTTTGCTTTCTTCTGACTGTTGGCTTTGCTGGACGTGACGTTTGAGA GGTGGTTTTGCCTCTGGTTCTGTTGTGTCTCAGCTGGTGGGGCTGTGGGTAGCCCTGCCTGGCAGTGCTGCTGGGGCGGTTCGGGTTCAGGTGCCCAGAAGAACC CATTCCCCCTGGGAACGAGGTGAGCTTAATGGCCTCTGACACTTATGGGAAGAACAGGCAAAGAAATCCGGACCCTGGAATCTGTCCTTATCTCTGAGGATAAAAAA AAAGAAGCCAGAAGTGGGTGGCTGGAGGCAGCAGAACCCCTCCTGACATCGGCTGCGGAGAGCAAAGAGCACACGTGTGCTCCTGCTGTGCTCGGGTGTGGAGAG CAGCGCTGTCCCCATCCTGCTGGATGGCTGGGATCCCCCTTCCCGGGGTTAGATCCCCTTCCTTGGGGTCAGAACCCCAATTCCTGATATCAGATCCCCCCATCCCT GGGTTCAGTTCCCCCCATCCCTGGGGTCAGATCCCCTTTCTTGGGTCAGAACCCTAATTCCCGATGTCAGATCCCCTTTCCTAGATCAGATCCCCCCATCCCTGGGG TCAGAATCCCCCATCTCTGGCATCAGAGCCCCTCATCCCTGGGTTCAGACCCTCACTGCTGGGTTCAGTTCCCCCCATCCCAGGGGTCAGATTCCCTTTCTTGGGGT CAGAACCCCAATTCCTAATGTCAGATCCCCCCATCCCTGGGTTCAGTTCCCCCCATCCCGGGGGGTCAGATCCCCTTTCCTGGATCAGATCCCCACATCCCTGGGGTC ACATCCCCTTTCTTGGGGTCAGAACCCCAATTGTCAGATCCTGTTGTCAGATCCCCCCATCCTTGGGGTCAGATCCCCTTTCCGGGGTCAGAGCCCCTCATCCCTGG GGTTAGACCCTTCTGGGTACAGTTCCCCCCTTCCCTGGGGTCAGATCCCCTTTCCTGGGTCAGAACCCCAATTCCCAATGTCAGATCCCCCCATCCCTGGGATCAGA GCCCCTCATCCCTGAGGTCAGACCCTCACTGCTGGGTTCAGTTCCCCCCATCCCAGGGGTCAGATCCCCTTTCTTGGGGTCAGAACCCCAATTCCTGTTGTCAGAAC CCCCCATCCCTGGGACCAGAACCCAACTGGGAGCCTTGCCAGCCTCCCCCAGCTGCTGCCATATGAAGGTGTCTGTGAGCCCTCGCCGAGAGACACAGAAGTGGG ACACGTCCTCACAGCACATCAAACAGCAATCAGTCTATATGGGAGCTAATTACCGGGAAAACCTTCGTTTAATGTGGATTTCACAGGCGAGAGCCGGGCGCTTGCCC TGCCCGGCGGATGGCAGCAGCGCAGCTCTCCCCTCCACCCCTTAACAAAGCCATTTGGATCTCACACCATTCCCACGAGGAACTCAGCAGCTTTCATCTCAGATGAT GTGCTGTCCAAAAAGGAGAGGCGTTCAGATCCAGATCAAACGCACTGC
IKZF4	6	CTCAGAAGTTCGTGGGTAAGGGGCTGCTCCCTCCCCAGCATGGGCAGATGGGAGAGGGGGGCGGTGGGGGGTGGGGATGAGGCTCACTGATGGCTCACGCTCTG CAGTTGGGCGCTGTCGGTGCTCGCATTGCGCCTTCCACCCTGCGGGCTGGGGAGGGATGGGATAGGATGGGGTGGGATGAGATGGGATGGGGTGGCGTGGAGTA GGATGAGATGGGATGGGATGGCGTGGGGTGGTGTGGGATGAGACAGTATGGAATGGGATGGGGTGGAATGGGATGGGATGGGATGGAATGGGATAAGACAGGAT GGGATGGCGTGGAGTAGGATGGGATAGGGTGGGATAGGGTAGGGTGGGATGAGATAGGATGGGATGGGGTGGGATGGGATAGCAGCGGGGCTGTGCATGGGTTG GGGTGGGATGGGATGGGGTGGGATGGGGTGGAAGAGCAGCAGAGCCGTGTGTGGGGTGGGGTGAGATGGGACAGCAGTGGGGCTGTGTGTGGGATGGGGTGGG ATGGGGTGGGAAGGGATGGGGTGGGATGGGATGGGACAGCAGTGGGGTTGTGTGCAGGATAGGATGTGATGGGAAGGGATGGGATGAGTTGGGACAGTAGTGGG GCTGTGTGTGGGGCAGGATGGGAAGGGATGGGATGGGGGTGGGATGGGGATAGGGCTGTGCATGGGGATGAGATGAGATGGGATGAGATGAGACAGGATGGGGT GGGGTGGGATAGCAGTGGGGCTGTGCGTGGGGTGGGGTGGGATGGGATGGAAGAGCAGCAGGGCTGCGTGTAGGGTGGGGTGGGGTGGGATGGAATGGGATGA GATAGGATGAGATAGGATGGGAGGAGATGGGATGGGATAGCAGTGGGGCTGCGTGTGGGGTGGGATGGGACAGCAGCAGGGCCATATGCAGGATGGGGTGGGAT GGGATGGGGTGGGGTGGGATGGGATGGGACAGCAGTGGGGTTGTGTGCAGGATAAGATGTGATGGGAAGGGATGGGATGGGTTGGGACAGTAGTGGGGCTGTGT GTGGGATGGAATGGGAATGAGATGGGATGGGATTGGGATGGGATGGGGGTGGGACAGTAGTGGGGCTGTGTGTGGGATGGGATGGGAATGAGATGGGATGGGAT TGGGATGGGGGTGGGACAGTAGTGAGGCTGTGTGCAGGATGGGATGGGATGGGGTGGGGTGGGGTGGGGGTCCCGGTGCCCGCTCTGACGCTGTCTATCTCCTG CCCACAGGAGAGAAGCAGATGCGGTTCGCGCTGTCCGACCTCCCCTTCGAGGTGAACGCCGGCTTTGAGAAGGACGTGGAGATGGTGTCGGCCCACCACCCCCTG GACCCCTCCTACGGCAGCTCCCTGGCTCTGATGGGGGGCGAACACCTTCGCCCTCTCCGCCTGCCCCCCACAAACTGCATCTCAGAGGTGACCCCCGTCATCAGCT CCGTCTACACCCAACTCCAACCCCTGCCGGGCCGCCTGGAGCTGCCGGGTGGCCGCGACGTCGCCGAGGGCCACGAGGACGCGGCGGACGGGACGCAGGCGCT GTACCGGGGACGGGAGGGGGGCGGCGCGTCGCCCACCAACGGCTGCCACGACTCCACGGACACGGAGAGCATCCACGAGGAGCGGACGGCTCAGCCCCCGCCG GGAAACGGCCCCGGCAGCCGTCACAGCCCGGCGTACGCCAAAGAGGACCCCAAAGCGGCGGAGGGAGCGACGCGTTCGGCGCCGAGCGGCGGCGGCGGCGGC GGCCGGGATGCGCTGCGGGTGGTGACGGAGGACGGGGAAACGGTGCGAGCCTTTCGCTGCGAGCACTGCCGCATCCTCTTCCTGGACCACGTCATGTTCACCATC

Gene Name	Primer Pair	Amplicon Sequence
		CACATGGGCTGCCACGGCTTCAGAGACCCTTTCGAGTGCAACATCTGCGGCCATCACAGCCAGGACCGGTACGAGTTCTCCTCGCACATCGTGAGGGGAGAGCATA AGGTTGGCTGAG
SKIV2L	1	TTAGAGCTCTGATCCTTCTGTATCTTTTTGCCCACCTCCACCCCCTCAGTGCTGCCCCCCCCCGGGCCCCTGCACCTCCCCCTGGCACTGTTGGAATTGGGCTGCGCC GGGCGCTGCGAACTGCTGCCGGGTCCCGCCCCCCCCCCGCAGCACGGTGAGACTTTCCTGGACACCGCCTGACCCCTATGTAGCCTCTACCCCCCCTAACCCCCCCT CCAAGCCCTCCTCCAATGGTCCCCCCTTCTGGCGGCCCCCCAAGACCCCTCTGAGCCCCCTAAAGGACGTTGTGAGGCCCCCTCCCCAGGGACCCCCCTGACCCC CTCAGACATCCTCTGACTTCACCCAGGATACCCCCCCCATACCCCATTAACCATTCCCAGGGGGCCTTCCTGATCCCTCTTCCAGACCCCCTCTAATGTTCCCCCACCA AAGATTCCCTTTAACCCCCCCCCCCAGGGGGCTCTCAAAGGGACACCCCCGAGCCCCTCTTGCCCCTCTTGCTTGTCCCAAAAGACTCCCCCAGCCCCCTCAAATCC CCCTAAAGGACCCGCTTTCTCCCCCTTGGAATCCTCTAAGTGACCCCTCACGCCCCCCCCCCAGCTTCCCCAGGGCCTCCCGCCCTTCACTCCTCCCCTGACAGCG GAGTTGGAGCAGTGCTTCCTGGGGACCCCTGCCTGGCTGCCCCCCCACCAGCATGAGCGAGCTCAGAGGTATGGGGGGGCACTCATAATGGGAGTTGTTGGAGTG CAGTGCAGATATTTGGGGGTTCTGGACCCCTCCAGGGGAGGCTCTGAGGATGGAAGGGGAGGTGCTTATAGTGGGGGAACACTGGGGCAAAATGGGGATACTTAG GGAGCTCCAGACTCGTCAGGGGTTTCTCATCCCTCAGTGGGGCTCTGAGGGGGGAGGCAGACTGTGGGGGGCCTCTCCAACTGTTTTTGCCTCCCACCCCCACCC AGGTGCTGGACGCGGGTGCCCACCCCCCGGGCACTGTTTGTGTTGGAGCCAACCCCGGTGCACAGCAGTGTGCGGGCGCTGCGGGACCCCGGCACGGGAGCACT GCAGGGCTTTGTCGAGGTGGGGGAACAGTAATTAACTAATTGGGGGGGGTGGGGGTTGGCATGGTATAAGACCCCCCTCACACCACGCTATAACCCCAACATAGGA GCTGCATGAGGACTCAGGGCTCGCATCTTCCCCAGACGTCAGCAGCATCCCCCAATGGGAAGGTGGGCACCTCCAACCCCCCTCCCTTTTAATATCCACTGCCCCC CCCTTTTTGTGTGACCCCTAAAGCTGACCCTTACCCCCTCCTTTTTGTAACCCTCCCAGGTGGTTTGGAGGAACCTCCCCTGGAGTTGCTCCACACCCAGGGGCCAA ATGAAGAGGATTTGGACCTTGAAAATGGTACAGAGCTGTCGGGGGGCTATTTTGGGGGTGCCCCTTCTTTAATTTTATCATTTCATTGTGTGTTCCCCCCCCAACCTC AGATTTGCTGACCACCCCCCCCGGGCTGAAACGTGGTGTTGAGTTCCAGCCCAGAGGTAGGGAACAGCGAGAAGTTCCTCCTTCTCAACCCCCAAATCCGTGCCCC CCCTGCAGCACACTGAGCCCCCCCCCCCCCACCTCACTGTACCCTTCTTTAA

Gene Name	Primer Pair	Amplicon Sequence
SKIV2L	2	GACATTGGAGATGCTGTAGGGTGCTGCGAGTCTGCACCATCAGCTGCCGTAGGGAGGAACTCTGGGGGTGTGGGATGCCATGGGGGGGTCCTGGTGGGTCCTGG GGCTCTGGTGGGGGTCCCACATCCTGTCCCCCCAGCGCGGCGTGGTGTGGGAGGAGACGCTGATCCTGCTGCCTGAGCACGTGGGGCTGGTGCTGCTCAGTGCC ACCATCCCCAATGCCCTCGAGTTCGCCCAGTGGGTCGGGTAAGGGGCTGTGGGGCGGGAGGGGTGCTATGGGGCTGGGGAGGTCTGTGGGAGCAGGGACCCTAT GGAGCTGGGGCAGTCTGTGAGGGGCTCTGGGGTGGTCATGGGGCTGGGAGGGGTCTGTGGGTGATTGTGGAGGCTGGGGGGTATGCAGGGGGCTATGGAGTGG TTGTCTGTGGGTGGGTCATGGGTGGGTTCTGGGGTGGGGGGGGCTGCAGAGGGCTTGGAGGTGGTGGTGAGAGGGTCTATGGGGGGAGGGGGGCTGCAGGGTT TTGTGGGGTCGGTGTGGGATTAAGGAGATGGTCAGCTGGGGGGCTATGGGGTGTCTGTGTGGGACCTGGGGGGGCTGTGGGGTATTTCCCCCATGGTCCCCCTTC TATCCTCCCCACAGCCGCACAAAGCGGCGCTGCCTGCGGGTGCTGAGCACACGGCAGCGCCCGGTGCCCCTCGAGCACTTTCTCTACACTGGGGGCGGCGGTCC GCCCTCACCCCGTGACCTCTTCCTGCTGCTGGATGCACGGGGTGGCTTCAACACCCAGGGGTGAGCCTCCCTGTGCCCCATAGTGCCCCATGGTGCCCTACAGCCA CTCCACAGTGCCCCTACAGTCAGCTATTGTTGCCCTGTAGAGTCCAACAGCCACCCCATAGCATCTTGTGGTGCCCTACAATCATTTAATGTTTACCCACGGTGCCCT ACAACTGCCCCCATAGCGCTGCCCCCATAGCGCTGCCCCATAGCGCTGCCCATAGACACCCCATGGCTCCCACAGGTACTATGCAGCGGTGGAGGCACAGAAGCA GCGGGCGAGCAAACACACACAGAGCTTTGGGGCCAAACAGCCCCACGGGGGGGGGCAGCGGCCCCGGGCAGGTGAGGGGGGGCCCAAGTTGGAGGAGGGGGCTT CTCCTTCGGGTTTCTTTTTGGGATCCTCCCATCCCATTTTGGGCCCTCCTTTTGGGCTCTCTTGGAGCTCTTTCAGACTCACTCCACACTCCCCCCAGGACCGTGCCA TGTGGCACTCGCTGGTGGCACTGCTGCAGGCGCAGGGGCAGCTCCCGGCCGTGGCCTTCACCTTCTCCCGCGGGCGCTGCGATGCCCACGCGGCCGCCCTGGGC CGCACCGACCTCAGCTCAGCCGCCGAGAAGGGCCGCGTGCGGGGCTTCGTGCGGCGCTGCCTGGCCCGGCTCCGGGGGGGGGGACCGCCGCCTGCCCCAGGTG GGGGCTCTTGGGGGCTGTTATAGGGCTGTGGGGTGTGTGGGGAGGTACTGGGGCTTGGGGGGCTGTTAGTGGGGGGGTTGGGTGTGTTGTAGGGGATCCCTGAG AGTCTTTGAGGGCCTGGGGGGCTTTAAGGAGCCTTGGGAGCTCTTACTGGGGGACTAGGAGAGGTCTGCGGGTCTTAGGGGGGCTCTTATGGTGGTCTGTGGGTG GTTCAGGGCATCTTGGGGGGTTCTGTGTGGGGCTGGGAGGAGTCGTGGGGCGCTCTGTGGGGCGGTGTGGGTCTGCTGTGGGTCCCACGCCTCCAACCCCCTCC TTCCCCCCAGGTGCTGCAGATGTCGGAGCTCCTGGAGCGCGGCATCGGGGTGCACCACAGCGGGGTGCTGCCACTGCTCAAGGAGGTGGTGGAGATGCTCTTCAG CCAGGGGCTGGTCAAGGTATGGGGCAGCTGTGGGGCAACTGTAGGGCTGAGCCCCACACATATCCCCCGCCAACTCACCCCCCGTATCCCCCAGCTGCTCTTTGC CACTGAGACCTTCGCCATGGGGGTGAACATGCCGGCGCGCACCGTCATCTTCGACTCCATCCGCAAACACGACGGCAACAACTTCCGCGACCTGCTGCCAGGTGG GGATGTGGGGCGGGTGTGGGCAGGGCTGGGGGCTTGGGGGACGGGGGGGCAGCTTTTGGGGCTCCGGGGTGGGTTTTAGGGTCTATGGGATGGATTTGGGGGC AGGTCTGGGAGCTGTGGGATGGTTCGGGGATGTGGGGCAGGGTGAGGCTCTATGGGGCAGCCCCTCTCCCCAGGTGAGTACGTGCAGATGTCGGGGCGCGCCGG GCGCCGCGGGTTGGACCGCACCGGGACTGTCATCATCCTGTGCCGGGGGACGGTGCCCGACATGGCCGACCTGCACCGCGTCATGCTGGTGAGTGCCGCCCACA TCACCTCCTGTCACCGCTTTCCTGTCTACGTCATGCAG
SKIV2L	3	CCATTAGGGATACTGGGTTTGGGGGGGGTCTTTCCCCTGCTTGAGGCTTTTCATAGGGGGTGTCCCTGGGGGTTCTGGGGCAATGTGGCACCTTATTCCCCCTTCAT TTTCTCCTCCTCCAGGAAGGACCCCCCCAGCCCCGAGCTGCTCTCAGCCCTGCAGGAGCTGCTGCGCATGGCGGGGGGGGGCCCCGGGGGGGCTGCCCCTGCTGG ACCCAGTGGGGGCACTGCAGCTGCGGGACCCCCCTGCCGTGGAGGCGGCCGCCCGCGCCCGGAGCCTGGGGGCAGCCCTGGGGGGGTTTCGCTGTGTGCACGG CCCCCGCTTTCCTCAGCTGGTACGGGGGGGGGGGGGGTGTGGGCCACTGTGGAGCTCTTTTGTGGGATCATTTTGGTGTCGTTTGGGGATTTTGAGGGTTGTATGG GGCCATTGGGAGTCTCTTTGGGGCCATTTGGGGCTGTTTTTTGGGATCCTTCCATGTCATCTTGGGATCGTTTGGGGTCTGTTTGGGGCCATTTGGGGGCAGTTTGG GTCACTTTGGGGCCATTTTGGGGCCATTTTGGGACCACCTGGGCCATTTTGGGGCGGTTTGGGGGGATCATTTGGGGTTATTTCTGTGCAATTTGTTCCATTTTGGG GCCATCTGGGTTCCTTCGGGTTCGCTTGTCCCCGTGGTGTCCATTTTGGGGCCGTGCTCCCTTTTTTTTTGGCATTTGCTCCACCTGGGGCTGTCCCCCCCCGCAGT ACTCGCAGTTCGCTGCCCGGCGCCGCCTGCAGGCACAGGTGGAGCAGCTCCAGTACGAACTGTCGGACCGCTCCCTGCTGCTGCTGCCCGAGTACCGCCAGCGC CTCGGCGTGAGTGTCCCCAACGTCCCCCGCGTGTCCACACCGCGTCGCCCTGCTGTGTCCGTCACGCATCTCTATGGGTTGACCTGATGGCTCTTTACGCCCATAG AGTTGCATGGGTTGACCTGGTGGCTCCACGTGCTCAGCGGGGCTGACCCCCAATTTCGCCCCACAGGTGCTGCGGGCTCTGGGTTACGTGGCCGACGGGGGTGCG GTGCAGCTCCCCGGGCGTGTGGCGGCACTGCTGAGCTGCCACGAGCTGCTGCTGACGGAGCTGCTGCTGGGCAACGTGCTGAGCCCGCTGCGCCCCGAGGAGGT GGCCGCGCTGCTGTCCTGCACAGTGCACCCAGGCCGGGGGGAGCCGCCCCCCAAACTGCCTCCCAACCTGCAGCGGGTAGGGCCTCTGCGGGGCTGGGGGGCT CCGCGGGGCTGCTGGGGGCGATGGAGGCTGAGCGGGGCTGTGTGCGGCTGTGTCTGCAGGGCATGGAGCAGATCCGCGCCGTGGCGGAGCGCGTGGGGCGGC TGCAGGAGGAGTGGGGGCTGCCGCAGAGCGCCGAGGACTACGTGGGGCAGTTCGGCTTCGGGCTGGCAGAGGTCGTGTACGAGTGGGCACGCGGCATGGTGAG CACACAGCCCCCCAGTGCCCTCCAACCCCCCCTGACCCCGACAGACACCCCGGAGCCCCACACAGCCCCGCCATTCCCCTCCGGGCTCCCCGTTACCCAAAGCCC CCCCCCGCTCCCCCCATTACCCTTTAGGACCCCCCCTTCCCCACGGCCCCCGGACTGCCGCGGCCGCTCTATTCTGCTTTGGGCCCTCGTGCCCCCCCGACCCCC CGTTTCCCCCCCCGTTTCCCCGCAGCCCTTCGCCGCTCTGGCAGCCCTGGCGCCGCTGCAGGAGGGCGCGGTGGTTCGCTGCATCCAACGCCTGGAGGAGCTGTG

Gene Name	Primer Pair	Amplicon Sequence
		CCGGGAGCTGCGCCGCGCCGCGCGGATGTTGGGCGACCCGGGGCTGGCGGCCACCATGGAGGCGGCGAGCGGCTGCATCAAACGGGACATCGTCTTCGCCGCG AGTCTCTACATCCAATAAAGGCCGC
SKIV2L	4	CCTCACTGTACCCTTCTTTAAAAACAAACAGCCCCCGGAGCCCGGGGGGGGGCACCATCACTCTGTAGCGTTTTGGGGGCATTGGATTCCTTGGAGTTGGGGGGGG CAGAAGAAGAGGAGAAAAAAGAGGGGGAGGGTGCTGGGGGGGGGTACCCCCTCTCAGCCCCCCCCTGACAAAGGAGGGGCGCCTTCCCCCACACCCCCCCTGCGC CGTGCAGACAGCCTGGAGGAGCTGGTGATGGGGGTAAGTATGGGGGGAGTGGGTGGAAATAAGGGGGTCCCTGCAGCCCCCCCATGCTGAAAGCCCACCCCCCC ACACCCTCAGGAGGTGCTGGTGCCCCCCACCACAGCCACCCCCTCTGTGCCCCCCCAAAGCGAGGAGGAGGAGTGGGCGGTGGAGGAGGACTGCAGCGTCCCCG TGGAGGACTTTGAGGAGAAGATCCCAGATCCAGCATTCAAGGTGTGGGGGGGGTGCAAAATGGGGGAGGGGGGCAGTGGGTGCCCTCGGAGGAGACTGTGGGGAA ATGGGGAGCAGTTAGGGGGGCCTGGGGGGTGACTTGAGGGATGTAGAGGGTGGTTTGGGGGAGCATCTATGGATGATTTCGGGGTCCTGGGTGGGGCTCATAAGA TGGGAAGGGCGATTTGTGGGGGGTGTCCTTGGTGGGTATGGAGTTGTGGAGGGGGGGTTGGGGGTTTGGGGGTATCTGGGGGGGGTTTGAGGAACTTCTGGGGTTTGA GGGGCCATTGGGGTGCAATGTGCGATTTGGGGGGGGCAGTTTGGGGGCTCGGGGGGGGCTTGTAGGTATTGGGATCCCATGGGCACCCAGGATTGCCAGGGTGTGG GGTTCCTCTATTGGTTTGGGATGGTTTTGGGGTCCCCACATTGGGTTCCAGGGTTTTGGGTTCCCCCCCCCTTCTTGGGGCTCAGGCTCTGGGAGGGTCCTCACCCCC CCTCCCCTCCCGCAGTGGCCGTTCCGCCCCGATGCGTTCCAGCAGCGCGCGGCGCTGTGCCTGGAGCGGGGGCAGTCACTGCTGGTGGCTGCGCACACCTCTGC CGGCAAAACCGCCGTGGCTGAATACGCCATCGCGTTGGCACGGCGACACATGACACGGTGAGGGGGGCCCACAAGGGGGGGGGGGGCTGGGCAGAGGGGTGAGC CCGTACTGGGGGGCCTGAGCCGCTTCCTTGGTCGCTGAGCCCGTCCTGGGGGTGCCTCCATCTCTTCCTCCGAGCTGTGGGGGTCTCTGAACAGTTTTCAGGATCC CAAATCCTTCCTGGCGTCTCTGGGGGGTCCCCAGATCCTTTTGTGGCGTCCCTGATCCCTTCCTGGGAGTCCCCAGCCCCTTCCTGAGGTCTCTGGGGGTCCCCAA TCCCTTCCGGGAGTCCCCAGCCCCGTTCTCGGGGTCCTTATGGGGATTTCTGAACCCTTCTTGTGTTCCCCACACCCACACTGGGGTCTCCCTGTGGGGCACTGAAC CCTTCCTGTGGTCCTGGAGGGATCCTTGACCCGCTCTTTGCCGCTGACCCCCTGTGTCCCCCAGGGCCATCTACACGTCACCCATCAAAGCGCTCTCCAATCAGAAG TTCCGTGACTTCAGGGCCACCTTTGGGGACGTGGGGCTGCTGACGGGGGGATGTGCAGCTGCGCACCGACGCCTCCTGCCTCATCATGACCACCGAGATCCTGCGG TGAGCCCCAAATGGCCCCGAAAAAGTGCCCCCAACATCCCCAAAATTATCCCCAAAATGGCCTTGAAAATGGCACTGAAACTGCCCCAGAGCTGCCCTCAAACAGAC CCAAAAGTGACCCCAAGTCCCCCCAAAACCCACCTTGAACCCCTAAAAAACCCTGAACACCTGAAATCCCCACGCCCCCCCAGCTCGATGCTGTACAACGGCTCCGA GGTGCTTCGTGAGTTGGAGTGGGTCATCTTCGACGAGGTGCACTACATCAACGATGCCGAGGTGGGGGCCAGGGGGGTCCCCAAAACCAGGGGAGGTCCTTGGGG GCTGAGGAGGGTGGGGTCCCATTGTTATAGGGTAAAAGTGGAGTTTGGGGGGTGGGGTTGGGATGGTTTGGAGGGAATTGGGCATTCTCTGAGGCGAGGGGTTTG TTCATGTGGGTCTGGGGTGTCCTGATGGACCTTAGGGTGGTCTGGGGCCATGGGCTCCTGTGGGGCTGCACCATTTTCTACTGTGTGAGGGAAGGAGGGAGGGGG CTGTGCGGACCTTGGGGGGGGTTCTGGGGGTCCTGGGATGTCGTGGGGGGTCATAGGGTCATTTCAGGAGGGCGGCTGTCTGGGACATTGGAGATGCTGTAGG

Gene Name	Primer Pair	Amplicon Sequence
SKIV2L	5	CTTTCCTGTCTACGTCATGCAGCGCTGTCAGCCCAGTGTCACCCCGCGCCCCATCCCGGCGTCACCCCACCATCACTCCGCCTGGCGTCCCCTGGCTGCATCACCC CCATGTCTCCATTTCCCCCCACTCACACGTGCTGTCACCCCATGCTGCTGCTCTCACGCTGTCCCTCTGCCCGCAGGGCCGCCCGTCGGGGCTGCAGTCGCAGTTC CGCCTGACGTACGGAACCATCCTCAGCCTGCAGCGCGCGGCCGCGCTCACCGTGGAGGGGTTGATGCGCAACAGCTTCGGAGAGTTCCCCCTGCGGCGCCGCGC TGCGGTATGGGATGGGGGACGCAGGGCTGGGGGGGTGTGGCGGGGAGGGGACGGGGGGGGAGTGGGAGCAGTGGGGGCTTTGGGACGTGGGGGAGGGATGCAA TGGGGAAACAGGGAAACGTTGGGGGCATTGGGGATAAGGTGGGGACATGGGAACGTGTCGGGAGAGTTGGGGGCAGTGGGGAACATGGGGATGCTGGGGGGCG TGGGGACAATGGGGACACTCCAGGAGCAGGGAGACACTGGCGGCAATGACTGGCAGGGGGGACGCTATGGGATGTGGGACATTGGGGCAGTGTGGGGTGCGGG GCATGGGGGCAACTGGGACGTGGGGCAGTGTTGCTGCAGTGGGGCTGCGGGGCCATTCAGGTGCTGCTTTATGTCATTTTGGTGCCATTTTGTGACATCTGGGGCC GTTTCGGTGCCGTTTGATCCCATTTGGGTCCATGTGGGGCCCCCAGGCACAGCAGCGGCGCGTGGCTGAGCTGCAGCAGGAGCTGAAAGCGCTGGGAGAGCCCC CGCAGGAGGGGACCCTGGATGACCTCCCCCAGTACTACGAGGCCGTGCAGGGGCTGCTGGAGGCGCGGGCAGAGCTGCAGGTGGGGCCGGGGGATGTGGAGAC ATGGAGGGATGTGGGGGGGGTATGGGAAGGGGATGGGCTGGGGACGTTGGGGATGTTCTGATGGGGATGGAGGGACATGGGAACACTGGGGGTCATGGGGAGG GTCCTGCAAGAGAGGAAACGTCCGGAAATGGGGGGGGGGTTGGAGGCTGTGGGAGGAATGGGAGGGTCCTGGGGGTGCTGTGGGGTTCCTGATGTGGGGAGGTG GGGGGACATAATGGGGAGACCCTGGAGGAGGAGGGTACCTGGGGACGTGGCGGGGGGGTCCTGGTGCCAACCCCCATTTTTCCTCCCAGCGTCGCGTGGCGCAGT CGGTGGCAGGGCTGAAGGCGTTGGCCCCCGGGCGGGTGGTGGTGGTCTGCACCCCTCAGCACCACAATGCGCTGGGGCTCATCCTGCAGGTAATCCCCCAAAGTA ACCCCTGAACACTCAAACATAACCCTCAAACATCCCCAGGTACTTAAAATAGCCTCAGAATGACCATACACCCCCCCCCCCCAAAAAAAAACCAGAATAACCCACAGA TTTCCCCATGCTCTCTCTCCAGAATACCCCAAATGAACACACATCACCCCCAACCCCGCAACAGAACCCCACAATGCCCCCAACCCCAATGTTCCACAAAATAACCCC CCCCAAGCCCCTAAAATAACCCCCAATGCCTCCCCCCAAAAAAGCAAAATGTCCCCGATCTCTCCTGCCCTCAGTGACCCCCACATCCCTTCCCTCACCCCCTGTAC CCCCCCACAGGTGACTGCTGAAAGCGGGGGGGGGGCGCACCATCACAACAATGGTGCTGAGTGAGAAACCCACAGAGGAAGGGGGGCCGCCTCCTTCGCCCCCCC CGGACGCCCCCTACCCCGAGGACATTCTGCTGAGCCGGCTCTTCCTGCCTGAGGGTGGGTGCAGTGTGACCCCCCCCCAAAAAATGGGGTGGGAGGGGCCCAATAA GAAGGGGCTTTATGGGGGTTGGAAGCACCCAAGGAACCCCGATAGACCCCTATGGGAACCCCCCACTCATGGAGTCCATTTTAGGGTGGAGGGCACTGGAGGGCC TTTATTGGGGGGGTTATACCAAGCCCCTGGCTCCCCTTTCTGGGGGAGGGTTTGGGGTTATCTGGGGGTTATTTTGTCCCCCCCACCCCACAGGCCCCCCTGGAGC GGCGCTGGAGCAGCTGCACCCCGAAGACCTGGGGGGCATTGTGGGGCGCACACTGCGGGCAAACCCCCCCCGGCTGCTGGAGGAGCTGCGGCGGAGGCAGACG GCACGGGGAAGGTGGGGGCTTTGGGGGGAGGGGGGGGGGCTGGAGGGGTCTGATGATATTATTGGGGTCTGTGGTGGTTCCATGGGGGGGAGGGGGAATCCTTA CAGCTGTGGGGGTCCCTGGGTGGGTCCATTAGGGATACTGGGTTTGG
AKAP8L	1	CGCCTTCACTCATATCCGTAGGACGTCACTTCCGCCCGCCTTTTTCCTCACACCCGACGTCACTTCCGCCCGTGTCCTCCCCAACGCGACGCCACTTCCGGCCGCC GCACAAGATGGCGGCGCTCGCATTGCTGCGGTGTTTGGGGTCGGGCCGCGTTCCGCGGGCCTCGGGCCCGGTGCGGCCGTACCGAGCGGCCCCTCTGCGGCGC CGCGTGAGGCCTCCGCCCCTCACGGACCCGGAGGATCAACGAGCGGCGGCGAAGCGGTTCGGATGGCTCGGGGCGGCCTCGGGAGTGGCGGTGGAGCGGCTGT GGCCCAGTCCCGAGCGGCTGCGGGAGATGGAGGAGGAAGAACGGGAGTGGTGTCCGTCGCTGCGAGAGATGGAGGCGGCCCTGGAGCGGAAGGAGCAGGAAGA GCGGCGGCTGCGGGAGGAGAGGTGGGGGGGGGGGGAGGGAGCGGGGGGAGGGGAATTGGCACCGTGCTAAACGGGGGGGGTCCGCCCTGGGATGGGGGTCCC TTCCCGGGGTGGGAATGACAACAACGGGGTGGGAGGGGGAGAACCGGGAGGGGGAGGCTCCTGGGGCAAGGGGTGGCAGCAGGGCGTGAGGGTCCCCGTAGG GGCAGCACCGGAACCTTTTCCTTGGGGAGGGGAGGGGAGGGGAGGGGGGGGGGGGGCACAGCAGCGACACGAACACCTTCGGGGTGTGATTGTAGGGGGACCC GTGGGAGGATGCGGGACACGGGGGGAGAACCCTCTGTGGGGTACAGCACTGAGATGGGTGCTGGATGAGGGAACGGAGGTTCCCGGTGCCCCCATCCCCCCAGG CTCACCCCCTCCCCCTGTCCCTTAGGGAGCAGCTGGTGGCACGCAGCCTGGCAGCCATGCCTGCACGCATCGCAGCCTGGAGGCAGGAGAGGGAGCAGATGAGG GAGCGGAGCCGGCAGGACGCAGCGCGGCGGCAGAGGCTGCTGGCAGAGGCAGCTGAACGCCTGGGGGCTCCGGCCCGGCCCGGCGATCCCCGGGTTCAGGCA GTGATGCAGGACCTGGAACGGGAGCGGCGGCGCCAGGAGAAGCGGCAGCGGAGGCAGCAGCACGAGGAAGCGACCCGCAGTGCTTTGGCTGCTGCTGAGGCTG CTGCTGCATCACAGCCCCCCCGCTGTCCCCCCCAGGACCTCCGAGCCTCCCCCCGTTCCCACCCAGGACCTCTGAGCCCCCCCCCCCCCACCCCGAGTAAAGCCC CCATCCCCACAACAAAGGCGGGATATGGTTTGCTACCCTAAAACCCTGCAAAGGGACTACTGAGGCTGCGCATATTCAGCATGAACTGCCCCACCGTGACCCCTCAG CCCCTTCCCGCCCCCCCCCCCAACCCTTCCCCAGCAAACACACACTAAAAGCACCATGGGTTTAAAATTCTGCATTTAATCTCTTCCCAAGCTGGGTTTGTCCCACGG TTCTGCGCTCCCCCACAGGGAGGAACGGGGAGGAAGAAAGGGATCAGGGAACAGGGATGGGGCCACCCAGAGCACACCCGGCCTCCAAGGCAGCGGGCAGGCTC TGGGGGTGCACACCCATCACTGCCCCACGCCGTAGGGGTGAGGATGGGCCCCATTGGGCCCCACCCCACCCCGATGTCCCCTGAGGGCAGAGATGTAGGGGGGG AGGAAGGGGAGTTCTGCGTCCCCCCCACAGGGTGGGAGGAATCCACCCCGGGCCCCTCAGTCATCATCAAAGTTCTGG

Gene Name	Primer Pair	Amplicon Sequence
AKAP8L	2	TGCGGCTGAACTCAATTCTTCCACGCCCCCCACCACCAAAACCCCGTTTTCCACATCCCTTTGCCCCCTCCCCCCATCCCCAAAACCTCAAATATCCATACATATTCC CCCAACCCCCAATCTCTTCCCCCTTACCCCGTTGTAGGCTCCGTGAGGGTGAGCGGGGGGAGGTCCTGGGGGGGCTGCTCCTCACTGGGGGCAGCCGGCGGGGGG GGGACGGCGTCAGCAGCGGGGGGGGGCGCTGGGGTCGGGGCGGGCTCTGAGGTGGCAGCGGCAGAAGCCGCGGGTTCAGGAGCTGCGGTGCCCAGCGCTGCC TTGGCCTGCGGGGGAAGGGGGGGGTGAGCGGGTAAACGGGACGGAGGTGGGGGAGGGGAAGGGGGGGGGGGGGGGGGTGCGAACCTTGGTCACCATGACGACC ACGAAGTTCTTCTCGTCGATGCGGTACTCGCAGATGGGAACGTCATCGCTGAGGATCTTCCCCGCGTAGATGAGCTTCTGGCCCGCCACCGGGAACGCGTCGCTGC CCTTCTCCGCCTCGATCTTCTCCTTCAGCGCCCGCACCTGGAGGTGGGGGGGAGGCCTCGGGGTGAGCGGCATTGCGGCCCCCCCACTCACCCACCCACCCAACC TCGGGCCCGCTCCGCTCCGCTCCGCACAAACCCCCCCCCCGAGCCCCGCCCGCCCTCACCGTCTCGTGCGGCTCCATGCGGATTTTGAACGTTTGCTGCTGCAGC GTCTTCAGCGTCACGGTCACCGCCATGGCGCGGCGGCGGCGAATGCTCCCGGCAGCCCCACAACATGCAACTCACGCCGCCGCCATCTTACGCGCCCCGCTGCTT CCGGCCATGTGACCCGCGGACCGGAAGTTACGTCATACAGAGGGCGGTGCCCTCGCCCCCCCTTTGAACACGTGACGATGGACGGCTCGGACACGTGACCCTCCCC CCCCCCCCCAAAGCGACACAAAATGGCGGCGCAGAGAAAGGCGGGGTGGGGGGGATGGAAATGGCGCAGAGCGAACGCAGCGCAGCCAAATGGGGGCCGGGAC GAGCTTTAATTTAATAGAAATCGCATCGCGTCACCCGTACAGTCACGTGCGGCCGCCGGAAACGCCCCCACCCCTCCCAACAACCCCCCCGCGGCACCGGGCCCG GCCCTCAGCCCGGCCCCCTCTCCCCAACCACCCCTCTCCTCCCCCGCCGCTCCCGTCCCGGTGCCGTTCCTGAGTCCTCGGTCCCGGGGGGGGGGGCGGAGCCG GCGCGGGGGGGGGGTCAGAGGGGGGGTGGGGAGGGAGGGACCGCGGCCGGGACGCCCGTAAAAAACAAAGTCCAGGTCTCACAGAGACATTACACGGGGCCGG GGCCGGGGGGGGGCGCTTACAGCTCGTCCTTGGGGGCCCCCCCGGCCTCCTCCTCCTTCTCAGCCTCGTCCTCTTCGTCCTCTTCGTCCTCCTCGTCGTCCCCGTCG CCGTCCTCGTCCCCCTCCTCCTCTTTCCTCTGCTTGTCCTCCTCTTCCTGCTTCTGGCGCTGCTCTTCGTCCTGCTGCTCCTTCATCTTCCTCTCAGCCTCCTGCAGG GCGGGGGGGGGGGGGGTCAGCACAGCTCCCATCGCCCCCCCAGCTGCCCCTCCCTGTACCTCAATCATCCCCTCCATATCACCCCTCATATCCTGCTGGCCACCCC CCATCCTTACAGCCACCTCTGTCCCTTGCTGCCCTCCCCCCGCCCCCCCATATCGATTTGCCCCCCCTCCCCCCCCCCCATACCTTGGTGGCCCCCCAGGTCTCGTTGC CAAACTCCTCAGCAAACTTCTCATCATCGGTGATAAGGAAGTTGTCAAAGATGGTTCCAGATTTCACCTATAAGGCAGACAGTCTCA
AKAP8L	3	CTCAATACGGGTCAGTGACGTGACAGCGGTGTCGCAAAACCCGGTGATGTCATAGGGTCTGGTGACGTCACACGGAGCAGTGATGTCATATGTTGCATGCCATCCC GTGGGCTGCGTTAATGGAACCCTAAGAACTCCCTCCGTCCGTTGGGAGGGGTGAGGGGGCGTTTAACGTGACAGCCCGAAACGGGGCTGTCGTCATGACAACGTG CGGCCGTCGCGGGGCTGTCGTCAGGGCAAAAGGCCCCGGGACCCCTGGCTGCAAGGAGACGACGGGAGCCTCGGAGGGCCGCGGGCCTCGGGATAATGAGCTC CGAGATAAACGCCCAGCGCCACATGACGTCACTCCTCGGCCTCATTACGTCACAGTCCCCTCGTATCGCCCCGCGGCCCCTTTCTCCCCGCCGCACCGCGCGAAGC TCCTCCCGCCGCACTCAAGATGGCGCCGCCCTCAGCAACATGTTTTCAGGGGAGCACTTCCGGGTGAGCGCGGGAGCACTTCCGGCCGCCGCGATGTCGCCCAGC GTGGCCGAGCTGCTTCTGCAGCGCTTGGAGCGGACGCCGCCGGGCCCCGAGGGCGGCCTGTGCAGCCTGGAGGCGGCCGCCGCGCTCGGCCTCGACCACCAGA CGCTCGTGGGAGCCGTCAAGAGCCTGCAGGCGCTCGGGGAGGTGCGGGGCCCGCAGGGGAGGACGGGGAGCGCGGTCAGCGCGAGCCTCGGTGCCGGTGCCC GTGCGCGATCCTAACGCCGTGTCCCGGCGGCCGATCCCCGGCCCGGTGTCGGTGCTGGGATCCCGGTGCTACGATTGCCGTGCACGCCCCGGGTGTCTGATCGG AGCGCCGCCTCGCAGGTGATCGAGGCAGAGACGCGCGCGACCACGCGCTGGGAGCTGAGCGCCGAGGGCGAGGAAGTGCTGCGGGACGGCAGCCCCGAAGTGC GGCTCTTCCGCAGCGTCCCCAGCGAAGGGCTGCCCCAAAGCGACGCCATGGTTAGCACTGCGATGGGGGGGCGGGGGGGTGTTGGGGGGGGGGGCTTCGGGAGC ACCGCGCATCCCACACGGCCCTCAACGTTCTCCCCTCACCTCGCAGAAGCTGCCCGGTGCCCAGGTGGGCTTCAGCAAGGCCATGGCCAACAAGTGGCTGCGCCT CGACAAGGGGCACCCGGCGGCCCCCCGCATTTTCCGTGCTGTGAGTGCCCCCCCCCCTCCCCACACACACACACCCCCCCGGGCGGTGGGCTCTGACTCACCGG GGGGTGCTGCAGGTGGACGCAGTGCAGGACGTGGTGCAGAGCAGCCTGCGGCAGGTGCAGGAGGGCAATGGCAGCTCTCTGTCTGAGCGCGAGCGCACTGATCT GAAGAGAAGGAAACTGCTGCTAGAAGTGTGCGTGGGGATGGGGGCACAGCTGGATGGGGGGGGGGGGGGACGGACGATCAGGGCACCCCTTGACCTCTGCGTCC TCTTCAGCACTCTTAAGTCCTACTGGATCCGCAAGGGCAGCGCCTTCAGCACGGCTGTAGTGCGGCAGGAGACCGACCTGACACCGGAGATGATCGCCACGTGAGT GTGGGGGGACCCAGAGGCTCTCCGTGACCCCAAATCCACCCCCCCCCCCCCCCCCCGCCCCGTGACCCTCCCTGTCCCTGCAGGGGCTCCTGGAGGAAGCTCCC CTTCAAGGCCTACAACTTCTCAGCACTGGGGCTGCCGCCCACCTGCGGCCACCTGCACCCCCTGCTCAAGGTGCGCTCACAGCTCCGCCAAATCTTCCTGGAGATG GGGTGAGTGGGGGGGTGAAATGCTCAGGGAGGGGCTGGGGGGGGGAGGGGGACGCACACACAGTGGTTCCCCCTTCCCCTCAGTGACACCCCGGACCCCCAGGTT CACAGAGATGCCCACAGACAACTTTGTGGAGAGCTCCTTCTGGAACTTTGATGCCCTCTTCCAACCCCAGCAGCACCCAGCACGAGACCAGCACGACACCTTCTTCC TGCAGGGTGGGTGCAGAGGCCAGGGTGGGCACACACTCAAATCCCCACTGACCCCCCTAACATATCCATGTCCCCCCCCTCAGACCCTGCTGAAGCCCCTGAGCTA CCCGCCAACTACATGGCCAGGGTGAAGAAGGTTCACTCTCAGGGAGGCTACGGTTCACAGGGGTGAGCGCTTTGGGATGGGGTGGGGGTTTAAGGCTGGTTTGGG GTCCCCTTGGCTGATCTGAGCACCCTTCCCACCAGGTACAAGTATGAGTGGAAGGTGGAGGAAGCAAGGAAGAACCTGCTGCGCACACACACCACATCTGCCAGCG CCCGTGCTCTGTACCACCTGGCTCGCCAGGTGAAGCAGGGACAGGGGAGCATGGGGGGGGAGGGGGGGTCTGCTTGACACCTCCTCCTCCAATTTCTCACCTCAT TCACCCTCCACGTCCCCCATTCCAGGGTAAGTTCACCCCTGTGAAATACTTCTCCATCGACCGG

Gene Name	Primer Pair	Amplicon Sequence
AKAP8L	4	TTCTGGCTGAGCAGGAAATTGGCTGCCAGGTTCTCATTCTTCTCGCAGGCAAAATAGGCCTGGATGACCAAACTCTCGGGGAAGCCCAATGCCTTCAACTGTGGGCA GAGAGGAGATCAAAACCACAACGTTCTGACCAGGTACTGATCGCTGGTTGGATAACAGTTAATCCATGCTAAATGCAAGGGGAATTCCTATCGAAAAACGAGGGTGG GTTTGGGTTCCAGGGGCCCTTACAGCCCCTTCCCACTTCAGATGGTGCTGAGCTCGGGGCATCAGGGCTCCACAGCTGAACGGCTCCACGCATAACACGCATAGCT TGGAACCCTGAAGCCTGCCTCAAACTTAATGACAGCTGGAACGTAGCAATACGAATGATGACTCTTCGTGATAAAAGAGACTATTCTCATTTCTGCGTTAACTCTGCAT TCTATTGGTGACTGTAAGCCAGCAGCCTCCCCCCCATCCCCCCCATATGGGGGCTCACTTACCCTTTCTATGGCTTCTTTTTCCTGAGGCGTCACTTGGATGTAATTC ATCTGAGGGGATTCGTCCCCAATGGCCCCCATCTCCCCCTCCAGGTCCCCCAACTCCCCCAGAGGCTCATTCAGCATCTGGATGAACTGCTCCTGGTGCTGGCTGAT TTGCTGTAGGGGGCAGGCGGGGTGAGATGGGGCTGGTCCAGGGCTGGGGAAGGTCCTAGAAGGGGCGGGGGGGGTTTCCAGAAGTACCTGCAGGAGCTGTGGG TTCTCCTGGCCCAGCTGCTGGAGCAGGGCGGGCAGCAGAGCTGGGTTCTGCTGGATCACCTGTCGCATGTTCTGGAACTGGGGCTGCTCCCGTAGGAACTCCAAT GGGTTTTCCCCTGCGCACAGGAGGGGATCAGCCCAGGCCACCCCCCCTCCAACAGCACACTGCCATGCCCTGGGATGGGCAGCCAGGCCCTGCCCAAGTGGGACA CACCCAGAAAAGGGCCCCCTCCCACCCCACCTCAGCAACTGGACCCAGGTGGGGGACACTGTGCCGAGCAACATCTGGGGGGTGGGGGCAGCCTTTCACAGTGCT GAGCACCACAGAGTACCGACCCCCAAAGCGGAACAACCCCTGCACCCCCCCCCGCCCAAGGGCTGTAGCCCACCCACAGCCCCACTCACGAAGCAGATCCCACTGT AGGCCCCACCTGGCTGACCTTCAACCTGCGGCTGCTCCGGGGCTCGGCTCTCCTGCACGGGTGGGCGCTCAGCTTCAGGGCTCCCCGGGATGCCCTGCAGCAGA CAGAGAGGTGAGTGGGATCTGATTCTCCCCTCAGCCGGGTGTTATTTCCCCCCGCTACCCCCCCAACCCGGCATCCCTCTCACCGTCAGCAGGTACTCAACAGCAC GGTGGGGGTTGTTGTAGCTGGCACGCAGTGCAGCCACCACCCTCTCCCTCTCGTAGCCCATCGACATGATCTCTGTCAGCATCGTCTCGTACTCCGAGCCCGTCAC TGCCAACGGGAGGTCAACGTCAGCCTCCATGGGGGGGTGGGGGGGGGCCAGGATCCACAGCTCCCCCCTCCCAACTGCTTCATCTCATGCTCAGCCGAACCATCG GCGCTCCACCAAAGCCTGTCTGTCTGTCTGTCTGCCTCCCACCCATCCTGCACTGCAGGCTCCACTTGGCCCCCAACACCATCCCTCCCCAGACCCCCCCTCCGTGT TTCTGTCACCACCCAAATCCTCCCCCACTCCCCTCCCTCCACATTTCCCCCCCACCCCACACACATCCTCAACCTTCCCATCCTCCAACGCCCCCTTATCCCACCCCCC CCCCCCCCCCCCATTGGCCCAAACACCCACCGAGGGTGGAGGCAGCGTCGGCCGAACGCCCCGCACCACGTGGGGGGGGAACAGAACTGCGAATTCAAGAGAGAC AAATGGCAGCCGGGTCCTTTACGGGAGCCCTGGGGGTGCAGCAGCCCCCAAGCACCAACCTCCCCTTCCCCCCCCCCTACAAAGACCCCCCAGTGACACCCCTCCA TCCCTCCCTCCTCCTGCGGTCCCTTCACGTACCATTTACTGCCCCCCATTCTACGTCCCCCCTATGTTTTTCAGGTCTCTGCGGCTGAACTCAATTCTTCCACGCCC
AKAP8L	5	GAGTCCTCAGGGTCGTCAATCTTGGCACGTTCATCCCAGTCATCGGGTTTCTTGGCCTCAGGATCCTTGATCTTCTTAGGAGGCAGGAAGTCCCAATCATCCTCCAG GCTCCCCGATTCCACTTTGCTGTTATCGATCTTCACCTCATAGGTGTTGTCAGGGCGGACGATGAGGGTGTAGAGGTGGGTGAACTCATCATCCTGCAGGGATCAGT CACTCAGGGGGGGCCGGGATGCGACCCCCAAGCTGACCAGACCCCCCTGCAGCCCCCCAACCCACCTTGCAGCGGATGTCCTTGTTGATGAGCACGTTCTTCCCTT TGTAATTAAAGATAACGTGGACCTTCTTTGTGCCGGGTCCGCAGATATCAGGGCCTGTGAGGGGACACAGGGAAGTTAGAAGGGAGTGGGGGGGGGGGGGGGGGG GGGGGGGTCTGACACTGCCAGAGGGGGCCGCAGGGGGGTGAACAGGGCGCACCAAACATGATGTTGTACTCAGAGTCTCCATGCATATCCTCCTGGTTCAGGCTG GCAGGAAAAAGCTTGACGTAGCCACCGCCACAATCGATGTTCTGCTCGTGTTTGACAGTGAACTGCACCACCAACGTCTTGTCCCGGTTGCTGAAGGGCTCGAAGCG GGAGGAGAGGGCGTAGAAACGTGCATCCTGGCTGGTCTGGATCCCTGGGGGGGGGGGGGGGGCACCCAGAAACTGCCCTTGAGCCAGTATCCCTCAAAGCGAGCA CTGATGGCCCCCAGAGTCAGGGGGGGTCCTCACCTTTATCTTTCTCAGCATCGCCGTAGAACTTCCCAGCCGTGAGCACGAAGCGGCCGTAGTCTGATTTGTGCTTC GATTCCACCCACCGTTGGGTCCAAGAGTCTGCAGGAGGGAAGTGCTGAGCACTCCGTGGCCCCCCACTACTGTCCTCCCCACAGCCGTAAGGGGAAATGGTTCTGC TCCTCCCCACAGACCGGAAGTGGGGCAGCTCACGGCCACCTGCGTGGCCCCACCTGCACCTCCAACGTTAAGAGAGGAGCCCTTTGCCCCACTCCTCACGGGGAG CCCCTCCCCCCCTCACCCGGAGCCCCTCCCAAGGGCAGCACTGAGGCTGCAGGGGTGCACCTGGCTGTACAGCTGCAGGACTGGGGCTCACAGCAGAAGGCCTT CTCCGCCTCGTGGAGGAGGAATGGGGGGTAATGCTGAACCACAGTGCTTTTGCCCCCAGTGAGCAGCTCCCACCCTTTGGGGTGGGGAGGAGCGGGCACTCCTCG CAGCCCCCCCAGCCCTTTGTCTCCCGGCCCAAGCAGGCCCCGCTCCCAGGAGAGGCCTCTTCTTCCTCTTCTTTTCACACCCCCCTTAACGCCAAACCCATTCCCCC CACATCCTCCCTGTGACCCCCAACCCAAAACCACTCCTTCCCCCAGGCACCACTCTCCCCCCCCCTCCCACCCTGCAACCTACGGGTGCCTCAAGGCCCCCCCCACTT CTTATAGGCCCCTCCCCATTTCAGCATCCCCCCCCCTCCCCTTCCTCTCCCCCCCCACCTCCCTCACCTCCATCCAAGAACTCCTCCCGGAAGAACTGGGCGGGCCCG GCCGCCGTCACCGCCAGGACGGCACCGAGCAGTAGCGGGAGGCAGAGGCGGCTCATGGCGGCGGTCCCACAGCCCAGCTACGCTCACACACTCCCTCCCAGCCC GGCGGCGCCCGGCACTGCGGAGGGGGCGGCGGCAGCTCCCCCCTTTTATCCGCTCGCAAGCCTTCGTCCCGCCTTCTTCGCCACCCCATTGGCCCGACCCCGCG CTGCTCCCGCTCCCATTGGCCGCTCGCCCGCTGGCCGGCGGCGGCGGCGCCCCCGAACGTTTGCGTCCCAGATGGCCGCTTCCTATTGGCTGAACGGGGCGGCC AATCAGGAGCGACCACGCGTTCGGAGGGAGACGCCCCCCCCCCCCCAACCTCTCGCGCGCGCCCTCAGCGCGGGGCAACGGCGCGATCACGTGGGGCAGCGCGC GCGCGCGCCGGGGGGGGTGACGGGAGGGGTCGGCAGCCCATAAAGAGACCGCGGGGAGGGCCCGGGGCTGCGCGGAGCGGTCCGATCTGCAGCCCCGCGCGTG GGGAACCTCAATACGGGTCAGTGA

Gene Name	Primer Pair	Amplicon Sequence
AKAP8L	6	GCGCACTGATCTGAAGAGAAGGAAACTGCTGCTAGAAGTGTGCGTGGGGATGGGGGGCACAGCTGGATGGGGGGGGGGGGGGGACGGACGATCAGGGCACCCCTTG ACCTCTGCGTCCTCTTCAGCACTCTTAAGTCCTACTGGATCCGCAAGGGCAGCGCCTTCAGCACGGCTGTAGTGCGGCAGGAGACCGACCTGACACCGGAGATGAT CGCCACGTGAGTGTGGGGGGACCCAGAGGCTCTCCGTGACCCCAAATCCACCCCCCCCCCCCCCCCCCGCCCCGTGACCCTCCCTGTCCCTGCAGGGGCTCCTG GAGGAAGCTCCCCTTCAAGGCCTACAACTTCTCAGCACTGGGGCTGCCGCCCACCTGCGGCCACCTGCACCCCCTGCTCAAGGTGCGCTCACAGCTCCGCCAAATC TTCCTGGAGATGGGGTGAGTGGGGGGGTGAAATGCTCAGGGAGGGGCTGGGGGGGGAGGGGGACGCACACACAGTGGTTCCCCCTTCCCCTCAGTGACACCCCG GACCCCCAGGTTCACAGAGATGCCCACAGACAACTTTGTGGAGAGCTCCTTCTGGAACTTTGATGCCCTCTTCCAACCCCAGCAGCACCCAGCACGAGACCAGCAC GACACCTTCTTCCTGCAGGGTGGGTGCAGAGGCCAGGGTGGGCACACACTCAAATCCCCACTGACCCCCCTAACATATCCATGTCCCCCCCCTCAGACCCTGCTGA AGCCCCTGAGCTACCCGCCAACTACATGGCCAGGGTGAAGAAGGTTCACTCTCAGGGAGGCTACGGTTCACAGGGGTGAGCGCTTTGGGATGGGGTGGGGGTTTA AGGCTGGTTTGGGGTCCCCTTGGCTGATCTGAGCACCCTTCCCACCAGGTACAAGTATGAGTGGAAGGTGGAGGAAGCAAGGAAGAACCTGCTGCGCACACACACC ACATCTGCCAGCGCCCGTGCTCTGTACCACCTGGCTCGCCAGGTGAAGCAGGGACAGGGGAGCATGGGGGGGGGAGGGGGGGTCTGCTTGACACCTCCTCCTCCA ATTTCTCACCTCATTCACCCTCCACGTCCCCCATTCCAGGGTAAGTTCACCCCTGTGAAATACTTCTCCATCGACCGGGTCTTCCGCAATGAGAGCCTGGATGCCACG CACCTTGCTGAATTCCACCAGGTGGAGGGGGTGGTGGCCGACCGTGGGCTCACCCTGGGCCATCTGATGGGCACCCTCCAGCAGTTCTTCACCAAGCTAGGTACA GGGAGAGCACAGGGGTAGCTGGACAGGGAGCTGGAGCGCTCCCTCACTGTTATCTTCCCCTCAGGCATCAGCAAGCTGCGTTTCAAGCCAGCCTATAACCCCTACA CGGAACCCAGTATGGAGGTCTTCAGCTACCATGAAGGTGCGGGGACCCTGGGTTCTGCTTGAGCCCTGGGAGGGGTGGGGAGGGCTCCGGGGTGTGTTTTCGGG GGTGGAGGGTGCTTGGCTTTTCTCTCTCCCATGAGGCTGGGGACACTGGGACACCCCTGGGGAGCAGGGCCTGGCCTGTGACCTTCTACAAAGCACGCAGAGGTG TGGACAATAGGCAACTCTGGGGTGCCCTGCTGGTGGGGGTCCCCATCTCACCCCCTTGTGGGGTGGGGTCCCCACCCTCCTCACTCCATAGGACTGAAGAAGTGG GT
SMARCC2	1	CCAAAGGAGACCAACGCTTAAATAACATCCAGACAGTCATAAAAAAATACTGAAGCTGGGGGGGGGAAGGGGGGCCTCGAGGGGGGGGAGGGCAGGGGGGGTCC TTTGGGGGCGGGGTCCTCAGCCGCGGTGGGTGGGGGGGGGGAGAGGGGGCTCACTGTGTGGGGGGACGGGGGTGGCCGTGCTGGGGCTGGGAGCGATGGGGT CCGTCTGCAGGGGGGGGCCTTGATCTGTATGGAGAGAAAAGGAAAGAAATGGGGGGGGGTCACATCTGGGTTGAGGGGGGGGGGGGGTTGCAAGGTGTGACATACC CCTCTGCCCATGAGATGGAGCGAGCGTCCCCCCCCCATCAAACCCAATACCCCAAACCCACCATCCATCCCGATGTGCATGGGGAGGGGGGGGCCCCCCCCAAACCT CCAGCCCCCCTCCCCCTTCCCGCCCTGATATCTCACCTGCGAGCAGCCCAGGGGTGGGAAGGAGGCTCCCCTGCACGGCGGCCACGATGGCTGGGCTGTGCGCG GCTGGCGAGCCCCCCCCCCAGCAACGGCAAGGCCAGGCCGGGTGGGTGAGGAGGGGGGGGGGCAGCCCCCCGGGCGGCATGCTGCTGGCCAGGGCCCCATGCA GCGGGAGGGCGGGCGGGTTAGCGGGGAAAGTGGTGGCGATGGAGTCGGCGAGGGGGTTAGCCATGCTCAATGGGATGGCGGATGGAGGCGGAGCGAAGGGCA GCGCCGTGGGCGCATTACCGGGCACGGAGGGGTGGCCGCTGCCACCCAGGTTCCCCCCCAGGCTCTGAGACGGCGGCTGCTGCCCGGGGAACGGCACCGGGGC TGTGGGAGAGAGGAGTGGTTGGGATTGGGGGGGGGGGAAAAGGGGTTAGTATCCGAATGCTGGGTGTGCCCACACCCTGAGCCCCGCAGGTGGGGGGAGCATCA GCACTGAGCCCCATGCTGTGCCCCCCCCATAGATAGCAGCAGCAACACTGAGCCCCCTGCTGTGCTCCCCACCCCCCAGAATGAATGGGAGCATCAGCACTGAGCC CCCTGCTGTGCCCCCCCATAGGTGGCAGCAGCAACACTGAGCTCCATGCTGTGCCCCCCTCCCCATCCAAAATGGGAGCATCAGCACTGAGCTCCATGCTGTGCCC CCCCTCCATGCAATATGGATGGGAGCATCAGCACCGAGCCCCATGCCGTGCCCCCCCCCCATAGATAGCAGCAGCAACACTGAGCCCCCTGCCGTGCCCTCCCCCC CCCTAAAATGAATGGGAGCATCAGCACTGAGCCCCCTGCTGTGCCCCCCCTCCCCAAAACGGATGGCAACACCAGCACTGAGCCCCATACTGTGGCCCCCTATAGA TAGCTGACCCCCTGCTATGCCCCCCCCCATCCAAAACGGATGGCAGCACCAGCACTGAGCCCCCTGCTGTACCCCCCCCCCCAAAATGAATGGCAGCACCAGCACT GAGCCCCCTGCTGTACCCCCCCTCCCCAAAATGGATGGCAGCACCAGCACTGAGCCCCCTCCTTGCACCCACTCACCGTGCGGGATGGCCAGGCCGGGCTGCGGT GCTCCCGGTGGTGGGGGGGGCTGCGGGGGGGGGCACCCCAGGCATCAGCTGCCCCCCCCGGTTCAGCGGGGGCCACCATGGCGGGGGTGGGCGGCGCAGCCAGC GAGTGGGCGGCGGGGGCCAGGGGGGTTCCGGCGGCGGGCAGAGGTTGAGCGCCGGGGGGCAGCGGGGGGGGCTGCTGCTGCTGCTGCTGCTGCAGGTGCTGG AAGTGCTGCTGCCGGGCGCGGAGCTCAGCGTACTTCAGCTGCTCCATGTGGAAGTTCTGGCGCTCCGTCAGCAGCTGCTGCCTCTGGTATTCCAGCTGGAAAGGG GGAGGGGAAGGAGTTGGGGGGGGGGGGGGGGGGGGTCAGCGGGGTGTCACAGCGTGGGTCAGCCCCCCCCCTTATCGCCCCCAGCCCCCCCCCCCCCCTCACCG CCTCGCGCTCTCGGTCCATAATGGTTTCCAGCTCCTCGAAATGCCGCAGTTTGAT

Gene Name	Primer Pair	Amplicon Sequence
SMARCC2	2	TCTTCGATCTCTTTGCTCTTGCGGCCCGTCTTGGCGTCGCCATCGCTCTGCCGGCCCTGAGGCACAGACAAAGCTGTGCTCAGCCCCCCCCCCCCAGCACCCAGGGA AGGCGGACGGGACGCGGACGACGTTGGAACCCCCCCGGCGTGCACCCACCTGAGGCGTCTTGGGCTGCAGAGGGACCAAACCGGAGGGCGTATCGGCCAACAC GTGGAAATGGGAGGTGGGTGGGGGGCCCATGGGGGTGGGCCGGCTCTCGGCGTCCACCTGGTAGTTGATGAGGCCCCACTGCTCCAGAAAAGCATGCACTCTGC GGAGAGGGGGGTCAGAGCAGGGGGCTGGAGGGGGGGGGGGGTGGGAGGCTTTGGGGTGGGGGGGGGGGTGGGAGGTTGGGGGGTGGGCTGCACCCACCTCAT GACAGCACACACATCTCCAGTCAGGTTCCTCCTGCAGGCAGTGCTGGTCAGGTACTCCTGGGGGTTGAGGCGATAGGTGTCGATCATGAAGTTGCGGTAAGCCAGG TATCTGAGGCACAGAGGAGGCCGTCAGGGCGGCCATAGGGAGGTGAGGGGGGGGCGGAGGGAGGGGGGGGGGGTTCAGCCTCACATCTCGGGGGTCTTGGATT TGTTTTTGCCGTTGAAGAACTCGGGCAGGGCTCGGCGCTCGATGGCGTGGACGCTGCGAGGTAAAAGGTTGGGGGGGGGTCCCACACAGCTGTGATGGGAGTTGG GGGGGGGGGGTGGGAGGGATGCACGCAGCCTTAGGGGCTCCGTACCTGTTGTAGTCAAACCAGGCGGCGTAGCTGGGGATGATGATGTGGTGGGTCTGCTCCGT CACGTTGTCCTCGTGCAGGTCGGGGTTCTTGGCCTGCTCCCCTTTGCTCCCGGTGCCGTTCTCCTCCTCGTCCTGGGGGACAGTGGGGCCATGGGAGGGGAGGGG GGTAAAGGGGAGGGGGGGCTGTGAGTGCTCCTCTGCACCCTCAGGGAGCAGCGTGGAATCACAGCACAGTTTGGGGCAATAAAAGCATTCGGGATCGTCCACAGC ATGGAGATCTGCTGTGGGATTGGGGCACTTGCAGCCCCATAGAGCTGCGCTGCAGCACTGCCACAGAGAGCAAAGAGGTATTTTTCCACCCCCAGAAAGAGACAAA GAAAGGACAGATGGACGCCCCGTGGTGGTTTCTAGGATGGGAAAAGGGTCCTTGGAGGAACACAGGGACAGGGAGGATAGATGGCAGAGCAACGGTTGGCACAGC CCTGCAGCCTCACCTTGCCAGCCGTCTCCATGCTCTCATCCTCCTGTTCATCTGCAAGACAGCACAGAGGGGGGGGTGGGGGGGTGGGGGGGCAGTTTGAGCACC GCCGGCTCAGCAGCACCACGGACCCCCCGCAGCACCCAGCCTGGAGCAGCTCAGCTCACCCAGGTCCGTCATGGTGCCCCCCTTAACGGGCGCAGACTCTGAGTC CTTCTTGGTGTTGACTGTAACGGACAAAGCAGACGCTGAGCGCTGCAGTGAGGGGACGGGGTGGCCCTATGGGGGGGGGGCTGCCAGCCCTGACCCCCACCTGTT TTGGGCAGGGTGACCTCCTCCACGTTGGGGACGGGTGAGGGCTCATCCATGTCCTTGGTCAGGTCCTCCTGCTCCTCCTCGCGGTGCCCACGCTTGGATTTGTTGT ATGGGGTGGAGGGTCTGCAAAGAGCAGAGCTGCTGCTGTCAGCACCGCTGGGCTCCCACAGCCCCGGGCAGCTCCAACCCTGCGCCCCCCCCCTTTGACCCCCCCA CCCCACTCACCCCTTCTTTGCGTTCTTCTTCTTGGCTTCCGGGGTGGGTGAAGGCGAAGGCGAACGCTTTCTCTTCTTGTAGTTGCCCCCCTTCTTGTCCCTCCGATC TGAATCAGGGCTGTTCACCTATAGGGGGCACGGGGGGGGGGGGGGGGGGGTCACGGAGCACAGCTTCCAGGCTGCAACCTCCCCACCGCATCCCTGGGACAGAA ACCCCCCCCCTGGGGGAGCTGCCCCCCGACCTCATCTGTCAGAGTCTTTGCTGAGATCTTCTTCCTGCGGGCAACGGGGCTCTTCTCATCGGACACCTCATAGTCCTC CTCGTTCATCCACTCGTTGAAGGTGTCCGTGTCCAGGATCCACTTGGCGTGGACCTGCATGCAGGGGAGCAGGGAAGTCAGGGACACCAATGGTGGGCAAAGGGG AGGGGGCTGAGGGTCCTCATCACCTTCCGGGGCTTCTCCGGAGTCGGTGCATCCTCCACAGAGGCCTCGATCTCACTGGCTGGGATCCAGGTGTCGTAGCTGTAAG CACAGAGGCAGTGGGTGAGCAGATGCCCCAACTTCCAGCTCCAGCCGCATCCCTTCACCCCAAAAACAGAGCGGTGACCCCAACCCCCCCCCCCCCGGGACCCCT CACCTGTCGGGGTAGTAACCCCAATGCAGCAGCACCTGCTTGTCTCGCTTCATAACC
SMARCC2	3	GGTGTTCCTCCTTACAGCTCCGAGCACAGCAGCTCAGTCTGCACCTCCTGCATTGCACCCACCTTCCTCCAGGTTCCCAGGGACGGGGCACACAACGTGGGACGCG TTGTTTTTGTCCTCGGTCACCGTGCCCTGGGAAGGGGGGGATGGAGCACAGGGCTGAGAGCCGGGAGGATGCAGCCGATGGGACCCAACATCCCAACAGCCACAC GGACGCCCACCTGGTGCCTCTTAATGATGTCCTTCAGTTTGCTCAGCAGTTTGGGTTCGATCTCCTGGTGCAGGAAGATGTTGGGCCGGGACAGGCAGTTGTTCTGG GGAAGGAAAGGAATGCATTAAGCAGAGCAAAGAGCGAACAGAGGGCGTGGGGGAGAGGGAAGGGAAGCAGCCCCCACCTGCACCAGGGACTTCTCGATGGTCATA AACATCTCCACGTTGCGGTCCATCCGCGAGGGATTCTGGAAATCGAAACGCCGCCTTATTGGAAGAAAAGAGCGACTCTCATGAAAGCCACAGCTCCGGGACCGCC CGGCAGTCAGCCCCACTCCGCCCCTGGGGATGGGGGGACGCCAGCAGGGCTCTGTGTATCAGCTGTGCCCCACAGATGGGCGGCCAACGCAGCCCAACCCACGG CCACGGCCAGACCCGGTGTCCCATACAGCGGACAAACCTCTGCGCCACCCCCATCCTCATCCCATCCCATCCCATCCCATCCCATCCTATCCCATCCCATCCCTATC CCTATCCCTATCCCTATCACTACCCCATCCCATCCCATCCCATTCCCATTCCATCCCATCCTATCCTATCCTATCCCCATCCCATCCTATCCCCATCCCATTCCCATTCC CATTCCCATTCCCATTCCCATCCCATCCCATCCCATCCCATCCCCCCCCCACCCCCCCGGTGCCGCTCACCATCCTTGGTCGCTCTTGAACTTATAAGCAGCAG CCAGGATGTGGCACAGGGCTCCCCCCCGCCTTGAAATCCAGGAAGCATTTGATCTGCAAAGGGGAAAAAGGAACGGCTGTGGGGCTGCCCCACACCGGGACGCGTC CCTCTGCCCCCATCGCCACCCCCACTCCATCCCACGGCGCGGCCCCCCCCAAACCCCCCCCCCAGTGGGGCCGCTGCTTCCCACCCCGCACCCCCAAACCCCAGGG GGGGCTTCACGTGTCCGGGGGGGGCCGCTGTCCCCACGTCCCCCCCCCCCCCCCCATTCCAACGAGGCACTGCGGGTCCGGGGGGGCTCCAGGGGGGATCCGAGGG GGGGATCTGAGGGGGGGGTCTGAGGGGGGGGATTCGAAGGGGGGTCTGGGGGGGGTCTGAGGGGGGGGGGTTCGAAGGGGAGTCCGGGGGGGGGTCTGAGGGGG GGGGTTCGAAGGGGGGTCTGGGGGGGGTCTGAGGGGGGGGGGTTCGAAGGGGGGTCTGGGGGGGGGGTCTGAGGGGGGGGGGGCTCTTTGTCCGCCCGCACATT CACTCACCGGCAGTTTGGTGAGCGGCGCGTTGCTGACGTGCTTCCCGAACACCTCCTCCTGGAACTGCAGCAGCTGCACCACCAGACTCGACAGCGACTTATTGGT GGGGGGCTCGGCCTGGATGTACTGCGGGAGGGGGCTCAGCCGCGGGCCCGCTCCGTGCGGGGAGGGGGCCGCGAGGCCGGGCGGGGCCGGGCGTTCCGTAC CTTCTTGTAGTTCTTGCCGAGCCAAAGGCGGACGTTGTCGAATTGGCTGAC

Gene Name	Primer Pair	Amplicon Sequence
SMARCC2	4	CAGTTTGATCTCCAGTTTCTTCATCTGCGTCTCCACCAGCAGCGCCACCAAGGATTTGATCTTCCGCTCCTCCACGGCCGCCAGGTGCTGCAGGGATGGGATGCGAC GGCAGATCATGGCCTGCCCCCCCCCACCCCCCCTTCCCCAGGACCCCCCCCCAACCTCCAGCCCCTCACCTTTGCCTTCACAGCAGCGGCCGCCAGCGCAGCAGC AGCGGCAGTGGAGAGGTTCCCTTCGCCGATGTCCCGCTCCACTTTGGCTTTGCGCTCAGCTTCGGGCTCTGATGGCTCCTTCGGGGTCTCCTCGGGCGGCTCCTTC GCCTCCTTCTCCTTCTCCAGCTCCCCTGGGGTTGGGGGGCAGCGGGGCGGTCCATGGGGGGCAGCCATAGGGGCTCACCCCCCCCTCTCCAAGCCCCCATCCGG ACCCCACTCACCTGCGCTGTCGCTCTCGCCCTTCTCCACCTCCTTCTCACCCTCGGGCTCTTTGCTCTTCTCCTCCTCCTTCTTGGCCACGTCCCCCACTTTCTCCTT CACCTCCTCCTCAGCACTGCCGTCACGCTGCTCCTGGAGGGGGGGGGGTGGGTATCCAGCAGTGGTTTGTCCCCACGTAAGACCCCATCCCACCCCGCTGACCCC CCACCCCCACTACCTTCACCTCCTTCTTCTCCTCCACCGGCGCCGTCTCCGCGCGCGTCTCCTCCGCACCGCCCTCCTCTGGGACAGAAAGAAGCACGAGGTGACC CCCCCTTCCCCCCCACTGAGCTCACTGCAGACCCCCCCCCCCCTTCCCCAACACCCCCCCCTCCCGGTACCGATCCGTTCCGGCTCCTCCGAGGCGGTGCCGGCGA TGCCGCTGCTCTCCAAGCCGAACGCAGGGTCAGCTTTGCCCGTCACCTTGGCAGCCTCCTCCACCTTACGGACGTGAGCTTCCACCAACGCCGTGGGCACCTCTTC CTTCATCTTGGAGAACTCCTCTGTGTGTGGGGCAGCATTCTATCACCCACCCGGAGGGGGTGGGGGGGTGGCAAAGAGGGGCAGCCCCTCCCAGCCAGCCCCCCT GGTCCCCAACAGCCATACTGCAGCTCCCCCTCACCCCCATAGCGCTGAGCACGTGGTGCTGCCCACCCCCTTCAACACCCATTGCCCACCATGCTGTGCCCACCCA CCCCCAGCCTGCAGCCCCCCTCCGCCCCCCCAAAGCCCCTCTTACCCAGCGCCGACTTAGCAGCAGCCGAGGCGACACGGGGGTCGACGACGGAGGCCAGGAAG GCGACGGTGCTCATAACGGGGTTACCAGACTGGCTGAAGGGGATGGGCTGATAGGCCAACGGCCCCAAAGACGCCTCCGAGTCCTCCAGGTAGGGGTCCTCGATG GGCAGGCGCAGGAAATGCAGGATGCACTCATCCTGAGTGCGGCTGCCCACGTGCTCCGACACCTTGTTCCAGTCGTCCTTGTACATCTCCAGGGCCTGGGGGTGG GCAGGGGGGTGCTCAATGGGGCTGGGTGGTCGTGGTCACCCCCATCCCCCCCAGTAGCTCAGATCCTCAGCTCACCTCCAGCAGCAGCAGCGTCTCCTGCTCCGT CCATTCCCTCGTGGCGCTGGCTGCAGCTTTGCTCTGCGGTGGGGAGAGAGGGGTGAGCGGCGTGGCGTGGCATATGGGGGGGGGGGAAGAGGGGCGCGGGGGGG GCTGACCTTGGAGGGGATGTTCTTCTTGGTATACATATCGGTGCGCAGCCCGAAGTTCTGCATGTCGGCCGGCTTCTCCTTGCTCTTGTCGGGGAAGTTCAGCATCT GCTGCGAGGCCGAGCTCTGCTGCTGGAAGGAGAGCGGACGCTGGGGATGCAGCCCAAAGGGAGGGGGGGCGTTGGGCACAGCCCACGAGGACGCAGCATCCCC CCACATCCTCCACCCCCACCCCCTCCCCCCACCAATCCATCCGCTCCCACAGACGCCGTCGTGCCCCCATCTCGTCACGGTGTCCCCCCCGCGTCCCTGCTCCCCA GGGCAGGGCGGCATGCGGGGCGCGAGGGGGAGGAGGGAAGGAGCCTTCGTCGGCCAGCGGCGGGGGTCGCGCCCCACGTAGCCGCCTACCAGCTCCGGCTTC CCTTTCACCGACTCGGTGACGAGGTCTTCGATCTCTTTGCTCTTGC
SMARCC2	5	CTTGTCTCGCTTCATAACCGGCCGGACCCACTCCTCTGCAAGGGACGCAAAGAGGGGACGGTGCTGCGCCTGCCCCACTGCTGGAGCAGCCCTATGCACAGCCCT GACCCCCAACACCGCCCCACCCCGGGGTGTTCCTCCTTACAGCTCCGAGCACAGCAGCTCAGTCTGCACCTCCTGCATTGCACCCACCTTCCTCCAGGTTCCCAGG GACGGGGCACACAACGTGGGACGCGTTGTTTTTGTCCTCGGTCACCGTGCCCTGGGAAGGGGGGGGATGGAGCACAGGGCTGAGAGCCGGGAGGATGCAGCCGAT GGGACCCAACATCCCAACAGCCACACGGACGCCCACCTGGTGCCTCTTAATGATGTCCTTCAGTTTGCTCAGCAGTTTGGGTTCGATCTCCTGGTGCAGGAAGATGT TGGGCCGGGACAGGCAGTTGTTCTGGGGAAGGAAAGGAATGCATTAAGCAGAGCAAAGAGCGAACAGAGGGCGTGGGGGAGAGGGAAGGGAAGCAGCCCCCACC TGCACCAGGGACTTCTCGATGGTCATAAACATCTCCACGTTGCGGTCCATCCGCGAGGGATTCTGGAAATCGAAACGCCGCCTTATTGGAAGAAAAGAGCGACTCTC ATGAAAGCCACAGCTCCGGGACCGCCCGGCAGTCAGCCCCACTCCGCCCCTGGGGATGGGGGGACGCCAGCAGGGCTCTGTGTATCAGCTGTGCCCCACAGATG GGCGGCCAACGCAGCCCAACCCACGGCCACGGCCAGACCCGGTGTCCCATACAGCGGACAAACCTCTGCGCCACCCCCATCCTCATCCCATCCCATCCCATCCCA TCCCATCCTATCCCATCCCATCCCTATCCCTATCCCTATCCCTATCACTACCCCATCCCATCCCATCCCATTCCCATTCCATCCCATCCTATCCTATCCTATCCCCATCC CATCCTATCCCCATCCCATTCCCATTCCCATTCCCATTCCCATCCCATCCCATCCCATCCCATCCCATCCCCCCCCACCCCCCCGGTGCCGCTCACCATCCT TGGTCGCTCTTGAACTTATAAGCAGCAGCCAGGATGTGGCACAGGGCTCCCCCCGCCTTGAAATCCAGGAAGCATTTGATCTGCAAAGGGGAAAAAGGAACGGCTG TGGGGCTGCCCCACACCGGGACGCGTCCCTCTGCCCCCATCGCCACCCCCACTCCATCCCACGGCGCGGCCCCCCCAAACCCCCCCCCAGTGGGGCCGCTGCTT CCCACCCCGCACCCCCAAACCCCAGGGGGGGGCTTCACGTGTCCGGGGGGGGGCCGCTGTCCCCACGTCCCCCCCCCCCCCCCCATTCCAACGAGGCACTGCGGGTCC GGGGGGCTCCAGGGGGGATCCGAGGGGGGGATCTGAGGGGGGGGTCTGAGGGGGGGATTCGAAGGGGGGTCTGGGGGGGGTCTGAGGGGGGGGGTTCGAAG GGGAGTCCGGGGGGGGGTCTGAGGGGGGGGGGTTCGAAGGGGGGTCTGGGGGGGGGTCTGAGGGGGGGGGGTTCGAAGGGGGGTCTGGGGGGGGGTCTGAGGGGG GGGGGGCTCTTTGTCCGCCCGCACATTCACTCACCGGCAGTTTGGTGAGCGGCGCGTTGCTGACGTGCTTCCCGAACACCTCCTCCTGGAACTGCAGCAGCTGCAC CACCAGACTCGACAGCGACTTATTGGTGGGGGGCTCGGCCTGGATGTACTGCGGGAGGGGGCTCAGCCGCGGGCCCGCTCCGTGCGGGGAGGGGGCCGCGAGG CCGGGCGGGGCCGGGCGTTCCGTACCTTCTTGTAGTTCTTGCCGAGCCAAAGGCGGACGTTGTCGAATTGGCTGA

Gene Name	Primer Pair	Amplicon Sequence
BAZ2A	1	GTCCTGGTTCTGTATGTGAGGCCTTGGGCTGTCCCTATGGGGCCTGGAGGGCTGTGGGGGGCTTCCTGGGGGGGGGTCTGCTGTGTGCCCCCCAGCCTAAAATGGC CGGGTTGGGGGCTGTGCTGATCCCCCCCGGCCCTATAGCGGGGCTGCGGGTGATGGCGGTCCTCCATCCATCCCCAAACCCCCCCTTCATCCCAGCGTTGGGCCC CCCCCCCACCATTAACACCTTCGCCTCTCCCCCTCCCTCAGGTCTGAACGGGGGCATGAATGTCAACGGCTTCTCTACTGTATCTCACACCACTACTTCAGGGACCTT CACCTCCAGCACGCATTCCTCGGGGCCCCCCCCCCCACCTCCACCACCCCTACGACTGCCTCTGGGACTACACGCAGTTCCAGCCCCCCGGCGGCCTCAAGGACGG CAGCCTCGCTCAGTTCCCCCTCAACGGTGTCTCCGGGGGGTCCCGGCCCCCCTCCCCGGGGCACGGCGCGAACCTTCGGGCCGCGGGGACGGAGCTGTGGGGC AACGGCACCTCCGGCTCCATGGGGCTGAACTTCGACTCGCAGGAGCTGTACGACTCCTTCCCCGAGCAGAGCTTCGAGCTGCTGCCCAACGGCCCCACCAGCTTCT ATGCGGCCCCGCAGCCGTCCCCTCTGCTGGGCTCCGGAGAGCCGCCCTTCCCTCTGCCCGAGGAGGACTTGGGGGGCGATGCGGACGATGCCGAGGCCTCCAAG GAGCTGCAGCCCCCCAGCATGGCTGAGAACGGGGCCGGGCTGGTGGGCAGCATGGAGCTGGAGGATGCACAGCCAGGTAGGAGCGCGGCGGGGGGTGGCGGCT GCGGGGTTGGGCTGTGCTGAGTGCACCCGTAGCGGCGTCCCCATCCACCTGCAGCACTCGGGACCGCACCGCTATGGGGTTGGGACCCCTGGGTGGGAGTTGGG GGGGGGGTCTGGGCTCTGCGGGTTGGCTGATGGGCAGTGAGCGGTGGGGGGGGGGGGGATATGAGAGGGGGGGTGGAGGTGCCCGCTGCGCTCCCAGGTTTGTT CCCACTGCTTTCTGCCCCCCGCAGATCTGAAGCTCTGCGCTTACAACGGCGCGGCGCCCACCGCGGTGCCGCTGAGCCGGCAGAGCCCCGGCCTGACCCCCCCC ACGGCCGCCAGCCTGGGGTCCGCATCGCCTATGGGCGCACCGCTGGAGGACGCGCAGCTGCTGAGCGGAGACCCCCTGGAGCCCTTCGAGGCTTTGGCCAGAGG TGGGGGCTGAGGTCGTGTTGTGGGGGGGGGGGGGGCGGCCGTAGGTCGGGTGCTGCGATGGCCGCGGCGCTGCGCGGCTGTGTCGGAGGGGCGGTCCGGGAT CGCGCGGTCCCTCCTCCGCTCCGTCCTCCCTTCCCAGCGCAGTGACGAAGCTGTTCCTCAGCGCCTGAACTTCCGCGCTGCTGAACATCGCCGCCGCCGCCGCCG CTACGGCCATTTTGGGCTCCGCTTCCCAACGGAGGGGGCGGGGCCGGGGGCTGCGGGCTCCCTTAACCCTTCTCCTCCCGGTGCCCGTTTGGGTGGGGGGGGGG GGGATCGGGGGGGGGGTTCGGTCATCCCCGCTGCCCCGTGAGCAGCCATGTGGGTGTGTGTGCCCCCAATGCACTCAGAGCCGGGGGGACCCCCTGTGCACAAAC CCCCCCCCCGTGCACCGCATCCTCCCCCATGTGACCCCCTTTACACCCCCCCCCCATGCAACCCCCCTCTGTGCACACCCCCCCGTGCACCCCCTCCCATGCAACCC CCTCCGTGCACAACCCCCCCATGCAACCCCCTCTGTGCACCTCCTCTGTGCGCCCCTTGCCCCATGCAATCCCCCCCCCATGCTGCCCCACTGAGCTCTTTGCACTGT TGTGTGTGTCCCCCCCCAGACCCCGACACCGATGACCTCTATGCGAT
BAZ2A	2	GAGAAGGCCAAACCCAAAGAGCTGCCGGCCGTGAAGCGCGGCCGCGGGCGGCCCCCCAAAGTGCGGATGGTGGATCTGCTGAGCAAGACGGACGCGCGGCTGC TGAAGAGGCTGGAGGCTCAAGGTACCCCCCCCGCCCCCCCCCCCCTCAGCCCGCCCCACAGCCCGCTCGCCGTCACCAACCACCCCTTCCGTTCCCGCAGAGGTGCT CAGCGATGAGGACAAGCTGAAGATGAGCAAAATCAAGAAGAAAATGAGGCGGAAGGTAGGAGGGGTGCTTGGGGAGGGCTGGGGGCAGCCCTGGGAGCACTCCG TCCTGGTGGCTGACGTGGGGCCTTCCCCTGAGGGGGGGGGGCACAGAGCAGCCCCCTGACCTGACCGATGGGTACCCGCGTTGCTGTCAGTCCATGAGCTGCTCT TGTCGTCCCCCCCCCAGGCCAAGAACAAGCAGAAGCAGGAGGCGAAAGCTCCCAAAGCGAAGGAGGCCAAGAAGAAATCCAAGGTGGGTCCGCAGCAGCCATGGG GCAGGGAGGGAGGGAGGGGGCTGGAGGGACCACCCACCCACCCCCCCACCCCGGGGTCTGTCTCCAGGCCAAGGAGAAGAAGGGCAAAGCGGAGAAGGGCAAA GAGAAGGGGCGGCCCAAGGAGAAGAAGGGCAAAGGGCCGCGCAGGGTGGACAAAGGGCTGCTGGCCCAGCGGCGCCTGGAGGAGCGGCGACGGCAGCAGCTG ATCCTGGAGGAGATGAAGAAGCCGACGGAGGATATGTGCCTGGCAGACCACCAGGTTGGGATGGGGAGGGGGTCCGGCGGGTCTGAGCCCGCACCCCCTGCCCT GACCGCCCCGCTTCGCCCCGCAGCCGCTGCCCACCTTCTCGCGCATCCCAGGCCTGGTGCTGCCCAGCCGCGCCTTCTCCCACTGCCTGACGGTGCTGGAGTTCC TGCAGAGCTACGGCAAGGTGCTGGGATTCGACCCCAGCAGGGACGTGCCCAGCCTGAGCACGCTGCAGGAGGGGCTGCTGGGGGTGGGCGGCAGCGCCGGCGC CGTGCAGGACCTGCTGGTGCGGCTGCTGCAGGCTGCGCTCTACGACCCCGGGCTGCCCCCCTACTGCCAGGTACCGCCCCTCCTCTGTGCCTGTCCCCTGCCTTG TCCCCGTGTGCCCCCATCCCTTGCCCTGATGCCTTTGGTGCCCGTGAGGCTGCAGCTGCTGCAGGCCTGGGATGGGCGAGCTCTCCATTTCGGGACGCGCTGCCC ACGCTGCAGCCCCATTTGGGAACGCCCATCCCCAGGCTTTGGGTGCCCTCCCCATGTTGTCCCCTTCCTGCCGTGCCCCCCCCCCCAACACCTCTCTGTCCCCGCA GTCCCTGAAGATCCTGGGTGAGAAGGTGTCAGAGATCAGCCTGAACCGGGACACCGTCTCGGAGGTGCTGCGCTGCTTCCTGACGGCACACGGGGCCGAAGCTGA GCTGTGCACGGGGCTGCGCACCAAACCCTTCCAGGCGCTGCTCCCGGAGCGCAAAGCTGCCATCCTCGCCTTCCTGGTCAATGAGCTCAACAGCAGCGCCCGCAT CATCAGGTGAGCAGAGAGGGGGGAGCGCCCGCTGTGCCCCCACCCCATGGAGCTGAGCCCCCCTATCCTCCCCCCGTCGTGTTCTTGCAGTGAGATTGACAAGAC GCTGGAGAGCATGTCGGCATATAGGAAGAACAAATGGATCATCGAGGGAC

Gene Name	Primer Pair	Amplicon Sequence
BAZ2A	3	CACCAAGTTCTTCAAGCAGATGGAGCAGAAGTACCTGACGCAGCTGACGGAGCAGCCGGTTCCCCCGGGTGAGAGCACTGCCTGGGGTTTGGGGGGGGGTGATGT GAGGGGGGGTGGGCAGCTCTGCAGAACCAATATAAGCACGGTGCTCTCTTTGCACACTCCTTGCAGAGATGCGGAGCGGGTGGTGGTGGCTGCAGGACCCCGAGC AGCTGGAGGCGGTGGCACGAGCGCTGCACCCGCGTGGCATCCGGGAGAAAGCACTGCACAAACACCTGACCAAGCACAGGGAGTACCTGCGGGAGGTCTGCCTG CGTGCTGCCACAGGTACGGGCTGCGTGCGGGGTGTGGGCTGCGGGTGTGGGTCTGCGTCCCCCCTGCCGACCCCTCCGTGTGCCCCCAGACCCCATCTTCCACC CCCGCCCCGAGGCGGCCGCCGTCACCGCCGCATCTCAGGAAGCGTTGGCCCAATGGTCGGTGATGGAGAGGGCGTATGAGGCCGACCTGTCGGTGCTGCAGTGG GTGGAGGAGCTGGAGCAGCGTGTGCTGATGGCTGACCTGCAGATCCGGGTAGGTGGGGGGGTCCTCGGGTAGGTGGGGGGGTCCTCGGGGAGGTGGGGGGGTC CTCGGGGAGGTGGGGGGGTCCTCGGGGAGGTGGGGGGGTCCTCGGGGAGGTAGGGATCCTTGAGGAGGTGGGGGGATCCTCGGGGAGGTGGGGATCCTCGGG GAGGTGGGGATCCTCGGGGAGGTGGGGATCCTCGGGGAGGTGTGGGGATCCTCGGGGAGGTGTGGGGATCCTCGGGGAGGTGGGGGGATCCTCGGGGAGGTG GGGGGGTTCTCGGGTAGGTGGGGGGATCCTCGGGTAGGTGGGGGGATCCTTGAGGACGTGGGAATCCTCGGATAGGTGGGGGGTCCTCGGGGAGGTGGGGGTC CCCGTGCGCAAAGGGGGTTCTCAGCCCCACATGGGGGTCTCAGCCCTGAGCACCGCTGTCCCATAGGGCTGGACGTGCCCCAGCCCCGACTCCACGCGCACCGA CCTGCGCTACTGCGAGCACAAAGTGGAGCCTCTGGAGGACATCACGGTGCGCAGCCGGCGGGATGGGCTCCCCCCGCGCCGCGAGGACACCAACCCTCTGGACC TGGCAGTGCTGCGCCTGGCAGCGTTGGAGCAGAACGTGGAGCGGCGGTACCTGAAGGAACCGCTGTGGGCGCCGCACGAGGTGGTGCTGGAGAAGGCGGTGCT GAGCGGCCCCGAGGAGCTGAGCCCTACGGAGATGTGAGCCCCCCCCCCCCACCTTTACACCCCCGTGGGACCCCTCCCCTCCCTCGGCGCTGTGCTGAGCCGCGC CCCCCCCGCCCCCCTACAGCGCCTACGAGATCACCCCCCGCGTGCGGACGTGGCGGCAGACGTTGGAGCGGTGCCGCAGTGCAGCGCAGGTGTCGCTGTGCATC CACCAACTGGAGCGCTCCATCGCCTGGGAGAAGTCGGTCAACAAAGTGGTGAGAGCGCATGGCCGGCATGGGGGGGGGGTGGGGTTGGAGGTCCGGAGAGGGG GGGGTTGGGGACCCTCCTCCCCTCCTCCTTTCCCCTCATAGACGTGCCTGGTGTGCCGGCGTGGGGATGACGATGAGAACCTGCTGCTGTGTGATGGCTGCGACC GCGGCTGCCACCTCTACTGCCACCGCCCGCGCATGGCGGCCGTGCCCGAGGGCGACTGGTTCTGCTCCGTCTGCGTCTCACGGGTACGGCGGGGGGGGGGTCC GGGCTGTGCGGCGCCGCCGTCGGGGCCGCGTGCTGTCGGTGACCCCGTGTGTCTGTGTGTCCGTCCGTCCATCCCTCCCCCCCCCCCCCCCCTCCTCCAGGCACA GCAGTACCCGGAGCCCCTCACCCCGCGGCGGGGGAAGAAGCGGAAGCGGGGCCGCGTGCCCGGGGGGGCGCAGCGGAGGAGGACGGCAGCCCCGGCGCAGG GCGGCATCGCGGCGTCCCGACGGGGCGTCCCCGGCCAAGCGGCGCGCAGCGCCCCCGAGGGCCCACCCGGAGATCTGACCTTCTGCGAGTGAG
BAZ2A	4	CTCTATGCGATGGATGAGACGCAGCTGGTGGGCGACAAATCCCCCCTGGAGGAACCCCCCGACGACGCGCCCGGCCCCCCCCTGCCCGCCTCCAGCCCCCCCCT GCACGCCGCCAGCCCCTTCAGCCTGCTGGCAGCCCACCGCCCGGCCCCCCCGCTGCTGCCCAGCCCCGGCTCGCCGCCCCCCCTGCACGGTACGTGGAGCACCT TTATGGTGCGGGGGGGGGGGGGGCGGCGCCCTTTGGGGGTGGGGGGCTCCTATGGGAGAAGCAGGAAGGGATGGCAGTCGTGGTGACAGTGCTGCCTTTGGCC CTGCAGACAGCAGCCCGGAGCTGAGCAGCAGCCACGCAGGGCCGGCGCAGTCGGGGTCCCTGGAGCTCAGCACCGCGCTGGAGCCCGAATCCCCAAATCCTTCT GCTGAGGAAGAGGAGGAGGAAGAAGAAGAAGAAGAGGAGGAGGAGGAGGAGGTGGCTGACAGCTGCACGGAGGCTTCGGCCGCTCCGGAAGGAGAGAGCAAAG AGGAGGCTGCCCCGCTGAGCACATCGGTGGGAGGTGAGGATGGACCCCCCCCCCTCCCCCCATCACCCCCCCAGCCGGGTGCCCCCCACACCCCCCAGCTGCCC TCCCCTCTGCCCCACAGGCGACGTCCCACGCAGGCGCATCGCCACCCCCGAGGAGGTTCGCTTCCCTCTGCAGCACGGGTACGTGGCCGCCCCACATCCGCGTCC CCCGTGGCCATAAGGGACCCCACTGAGCCCCCCTCTGTGTGCAGGTGGAGGAGGGAGGTGCGCATCAAAAGGGGGAACCACCGCTGGCAGGGGGAGACGTGGTA CTATGGGCCCTGTGGGAAGAGGATGAAGCAGTTCCCGGAGGTGATCAAGGTATGGTGGGGGATGGGGGGGAGGGGGCCGGGCTCTGAGCAGCCCCCCCAGCTCC ACCAGGGTCCCTCATCCTCCTCCTTGCCCCCCCCCCCAGTACCTAAATAGGAACGTGGTGCAGGACGTCCGCAGGGAGCACTTCAGCTTCAGCCCCCGAATGCCGG TGGGGGATTTCTATGAGGAGAGGGACACACCGGAGGTGGGCACCGCGGGGTCCGGAGCTGCTCTGCGTGGGGCTGCGGGTCGTTCCCTTTGGGGCTGACGGGGT TTCCTCCCTCAGGGTCTGCAGTGGGTGAAGCTGAGCCCCGAGGAGATCCCGTCCCGCATCCAGGCCATCACCGGGAAGCGCGGCCGACCCCGCAACGCCGAGAA GGCCAAACCCAAAGAGCTGCCGGCCGTGAAGCGCGGCCGCGGGCGGCCCCCCCAAAGTGCGGATGGTGGATCTGCTGAGCAAGACGGACGCGCGGCTGCTGAAG AGGCTGGAGGCTCAAGGTACCCCCCCGCCCCCCCCCCCTCAGCCCGCCCCACAGCCCGCTCGCCGTCACCAACCACCCCTTCCGTTCCCGCAGAGGTGCTCAGCG ATGAGGACAAGCTGAAGATGAGCAAAATCAAGAAGAAAATGAGGCGGAAGGTAGGAGGGGTGCTTGGGGAGGGCTGGGGGCAGCCCTGGGAGCACTCCGTCCTG GTGGCTGACGTGGGGCCTTCCCCTGAGGGGGGGGGGCACAGAGCAGCCCCCTGACCTGACCGATGGGTACCCGCGTTGCTGTCAGTCCATGAGCTGCTCTTGTCG TCCCCCCCCAGGCCAAGAACAAGCAGAAGCAGGAGGCGAAAGCTCCCAAAGCGAAGGAGGCCAAGAAGAAATCCAAGGTGGGTC

Gene Name	Primer Pair	Amplicon Sequence
BAZ2A	5	GAACAAATGGATCATCGAGGGACGGCTGCGCCGGTGGGTGCTGCGGGGTGGGGAGGGGGAGGGGGATGTGTTGGGGGGATGTGGGTGGTGCTGTGCTGTGGGA TGGGGATAAAGCGGTGCTGTGGGATGGGGACAGGGCTGTGCCATGGGATGGGGATGGTGCTGTGCCGTGGAATGGGGACAGGGTTGGGCCATGGGGCAGTGCCG TGCTATGAGATGGGGCTGCCCCATGGGATGCAGGACAGGGCTGTGCCATGGGATGGGGACAGAGTAGTGCTGTGGGATGGGGATGGTGCTGTGCCGTGGGATGG GGATAAAGCAGTGCTGTGTGATGGGGACAGGGTTGTGCTATGGGATGGGGATGGTGCTGTGCCATGGGGCAGTGCCATGCCGTGAGATAGGGCGGTGCTGTGCCA TGGAATGGGGACAGGGTTGGGCCATGGGGCAGTGCCATGCTATGAGATGGGGCTGCCCCATGGGATGGAGGACAGGGCTGTGGCATGGGATGGGGACAGAGCAG TGCTGTGGGATGGGGATGGTGCTGTGCCGTGGGATGAGGACAGGGCTATGGGATGGGGCAGTGCTGTGGGATGGGGGACAGAGCAGTGCCATGGGATGGGGATA AAGCAGTGCTGTGTGATGAGGACGGGGCTGTGCCATGAGGCAGTGCTGTGCTGAGATGGGGCAGTGCTGTGGGATGGGGCAGTGCTGTGGGGTGAGGACAGGGC TGTGGCATGGGATGGGGACAGAGCAGTACTGTGGGATGGGGATGGTGCTGTGCCATGGGGCAGTGCCATGCCATGGGATGGGGCAGTGCCATGGGATGGGGGAC AGGGCTGTGCCATGGGGCAGTGCCGTGCTATGAGATGGGGGAGGGCTGTGTCATGGGGATGGTGCTGCGTTGTGGGATGGGGCTGTAGCATGGAGCAGTGCTAT GCCATAGGATGGTGCTGTGCTACGGAAGGGGGACAGGGTTGTGCCATGGGATGGGGATGGGGCAGTGCTCTGGGATGGGGACAGGGTTGTGCCATGGGATGGGG ATGGGGCAGTGCTGTGGGATAGGGACAAGGCTGTGCCAGGGGACGGTGCTGTGCTGTGGGACAGGGATGGGGCTGCACTCCCTACAGCCACCCCCCTGACTCGA GGGACTCTGGGCACGGAGGGAACTCCGCACGGGGTCTCCACCATCACTGTGCTGTGTCCCATCAGGTTGAAGGTGGCTCTGGCCAAGAGGACGGGCCGCCCCGA GTCGGAGCTGATGGCGTTGGAGGACGGGCGACGCCGACGCAGCTCCCGTCTGTGTGAGGAGCCCGGGCTGGAGCTGGAGGAGGAGGAGGAGGGTCGGGGCCGC CGCTCCCGCCGGGACGAGGAGGTGATGTGGGCATTGGGGGAGCGCTTTGGGATAGGATGCGGTGGGGTCTCAGGGTCCCCCCCCCGGTCCTGATGGCCTCTCGCT CCCCCCAGGCTGACACGTCCACCTCCAGCATCCCTGAGCTCGAGCGGCAGATCGAGAAGCTGGCCAAGGTGAGTGGGGGCCGCAGGTCTGCGGGTCCCCATCCC ATGGCTGTGACCCCCCCCTTCGCCCCCCCTCTCTGCAGTCCCTCACCCCGTGCCCTTCCCACAGCGGCAGATGTTCTTCCGTAAGAAGCTGCTCCACTCCTCGCAG ACGCTGCGGGCGGCCTCACTGGGCCAGGACCGGTACCGGCGGCGGTATTGGGTGCTGCCCCACCTGGGGGGGATCTTCGTGGAGGGGGCGGAAGGTGAGCGGG GTTTGGGGCTGGGGGGCGCTCTGGGGACCCCTCACGGTGCGACCCATCCTCATTCCACCCTCACAGTGCCCGAGGAGCGGCCGGAGGAGCCCCCGGAGCAGAAG GTGCTGTGCCCGGTGCCCGCGGTGAAGGAGGAGCCGCCGGAGCCCGAGGTGCCGGCGCCGGTTCCCCAACGGGTCACCATGAACTGTGTGG

Supplementary Table S3: The complete list of BACs and their coordinates in the budgerigar genome.

BAC Clone	Chicken Chromosome	Chicken Start (bp)	Chicken End (bp)	New PCF	PCF Start	PCF End	Budgerigar Chromosome	Orientation
TGMCBA-206D5	1	6,317,911	6,466,032	29b_T	5,674,707	5,822,828	6	2
CH261-89G23	1	20,319,376	20,556,326	29b_T	19,061,958	19,298,908	6	3
CH261-119K2	1	29,569,857	29,806,173	29b_T	28,312,439	28,548,755	6	4
CH261-25P18	1	71,625,759	71,868,403	29e_T	1,602,587	1,845,231	6	1
TGMCBA-146O14	1	86,207,012	86,371,517	1a_30_T	874,212	1,038,717	3	1
CH261-9B17	1	131,954,478	132,159,072	1a_30_T	39,415,315	39,619,909	3	2
CH261-168O17	1	140,200,039	140,446,061	1b_T	52,478,775	52,724,797	3	3
CH261-83O13	1	145,493,141	145,732,798	1b_T	47,192,038	47,431,695	3	4
CH261-107E2	1	155,077,290	155,306,349	1b_T	37,618,487	37,847,546	3	5
CH261-58K12	1	166,052,819	166,255,472	1b_T	26,669,364	26,872,017	3	6
CH261-98G4	1	189,398,394	189,581,262	1b_T	3,343,574	3,526,442	3	7
CH261-192C19	2	640,678	817,176	2a_T	43,272,430	43,448,928	1	8
CH261-169N6	2	5,846,212	6,041,793	2a_T	35,637,664	35,833,245	1	7
CH261-172N3	2	13,873,886	14,022,156	2a_T	27,657,301	27,805,571	1	6
CH261-177K1	2	24,103,323	24,346,901	2a_T	17,332,556	17,576,134	1	5
CH261-186J5	2	29,494,034	29,721,863	2a_T	11,957,594	12,185,423	1	4
CH261-40G6	2	38,161,264	38,349,127	2a_T	3,330,330	3,518,193	1	3
CH261-50C15	2	43,817,744	44,052,880	2b_T	2,410,017	2,645,153	1	1
CH261-123O22	2	48,112,621	48,290,667	2c_T	425,963	604,009	1	2
TGMCBA-340P4	2	66,935,741	67,090,903	2a_T	55,481,316	55,636,478	1	9
TGMCBA-78C11	2	98,646,159	98,772,400	2a_T	87,191,734	87,317,975	1	10
CH261-169E4	2	103,241,761	103,471,020	2a_T	91,787,336	92,016,595	1	11

BAC Clone	Chicken Chromosome	Chicken Start (bp)	Chicken End (bp)	New PCF	PCF Start	PCF End	Budgerigar Chromosome	Orientation
CH261-1J20	2	112,540,892	112,776,283	2a_T	101,086,467	101,321,858	1	12
CH261-44D16	2	118,894,877	119,099,952	2a_T	107,440,452	107,645,527	1	13
CH261-44H14	2	127,911,344	128,117,060	2a_T	116,456,919	116,662,635	1	14
CH261-115J5	3	62,801	245,437	3a_T	10,378,491	10,561,127	2	13
TGMCBA-295P5	3	9,939,896	10,058,922	3a_T	2,238,831	2,357,857	2	11
CH261-18I9	3	14,458,377	14,680,810	3a_T	8,006,505	8,228,938	2	12
CH261-130M12	3	20,542,045	20,707,498	3b_33a_T	55,213,319	55,378,772	2	1
CH261-160I6	3	25,180,860	25,408,942	3b_33a_T	50,511,875	50,739,957	2	2
CH261-97P20	3	40,123,190	40,359,654	3b_33a_T	35,561,163	35,797,627	2	5
CH261-17B14	3	65,251,594	65,475,612	3b_33a_T	10,445,205	10,669,223	2	6
TGMCBA-250J17	3	76,339,309	76,510,643	3c_5a_T	70,648,651	70,819,985	2	7
CH261-169K18	3	95,231,660	95,372,232	3c_5a_T	51,787,062	51,927,634	2	8
TGMCBA-64D9	3	103,642,372	103,763,457	3d_T	7,174,318	7,295,403	2	10
CH261-120H23	3	107,693,895	107,874,596	3d_T	3,777,629	3,958,330	2	9
TGMCBA-330J11	4	3,422,963	3,555,119	8b_9b_31b_T	9,012,020	9,144,176	5	6
CH261-83E1	4	4,003,197	4,175,284	8b_9b_31b_T	8,391,855	8,563,942	5	5
CH261-71L6	4	4,061,293	4,249,089	8b_9b_31b_T	8,318,050	8,505,846	5	4
TGMCBA-280M7	4	8,124,301	8,265,851	8b_9b_31b_T	4,301,288	4,442,838	5	3
CH261-111A15	4	15,329,891	15,552,198	8b_9b_31b_T	2,422,323	2,644,630	5	2
TGMCBA-200G5	4	16,276,699	16,422,023	8b_9b_31b_T	1,552,498	1,697,822	5	1
CH261-18C6	4	30,117,545	30,370,391	4g_T	1,448,979	1,701,825	7	2
CH261-93H1	4	37,167,348	37,383,279	4g_T	10,783,970	10,999,901	7	3
CH261-185L11	4	42,706,698	42,972,497	4g_T	26,593,937	26,859,736	7	4
CH261-85H10	4	64,707,690	64,953,032	4g_T	35,496,170	35,741,512	7	5

BAC Clone	Chicken Chromosome	Chicken Start (bp)	Chicken End (bp)	New PCF	PCF Start	PCF End	Budgerigar Chromosome	Orientation
CH261-89P6	4	73,638,906	73,780,597	4g_T	44,427,386	44,569,077	7	6
TGMCBA-216A16	4	84,684,219	84,838,062	4g_T	55,472,699	55,626,542	7	7
CH261-49B22	5	5,162,008	5,355,408	5b_14a_T	3,303,413	3,496,813	8	5
CH261-122F8	5	9,402,313	9,572,742	3c_5a_T	48,627,850	48,798,279	4	4
TGMCBA-145C6	5	16,566,138	16,714,631	5c_T	1,799,982	1,948,475	8	4
CH261-78F13	5	24,364,719	24,551,738	3c_5a_T	4,886,773	5,073,792	4	6
CH261-2I23	5	25,797,529	25,965,423	3c_5a_T	10,557,783	10,725,677	4	5
TGMCBA-382J4	6	2,376,226	2,513,780	6a_7a_T	7,137,994	7,275,548	4	3
CH261-49F3	6	22,289,806	22,402,526	6b_7b_T	12,499,133	12,611,853	4	1
CH261-56K7	7	3,586,775	3,748,429	7c_T	21,211	182,865	8	1
TGMCBA-224O13	7	10,237,883	10,384,434	6a_7a_T	25,193,495	25,340,046	8	3
CH261-180H18	7	19,984,877	20,218,119	6a_7a_T	15,359,810	15,593,052	8	2
CH261-186K14	7	25,589,810	25,765,994	6b_7b_T	28,184,582	28,360,766	4	2
TGMCBA-252A4	8	18,876,808	19,004,890	8b_9b_31b_T	23,028,272	23,156,354	5	10
TGMCBA-208D17	8	27,465,718	27,673,033	8b_9b_31b_T	14,360,129	14,567,444	5	11
CH261-96D24	8	28,504,295	28,697,585	8b_9b_31b_T	13,335,577	13,528,867	5	12
CH261-183N19	9	72,535	247,137	8c_9c_T	5,815,987	5,990,589	5	7
CH261-187M16	9	14,631,745	14,820,145	8a_9a_31a_T	4,268,705	4,457,105	5	8
TGMCBA-321L6	9	17,198,556	17,369,679	8a_9a_31a_T	6,835,516	7,006,639	5	9
CH261-115G24	10	11,413,550	11,625,715	10_T	11,355,471	11,567,636	9	2
CH261-71G18	10	16,137,296	16,355,683	10_T	16,079,217	16,297,604	9	1
CH261-154H1	11	16,102,045	16,317,215	11_T	3,885,689	4,100,859	7	1
CH261-60P3	12	12,073,689	12,215,683	12d_T	13,672,932	13,814,926	9	3
CH261-95H20	12	16,795,957	17,016,912	12d_T	3,229,681	3,450,636	9	4

BAC Clone	Chicken Chromosome	Chicken Start (bp)	Chicken End (bp)	New PCF	PCF Start	PCF End	Budgerigar Chromosome	Orientation
CH261-4M5	12	18,331,547	18,523,869	12d_T	4,765,271	4,957,593	9	5
TGMCBA-266O5	13	4,610,378	4,781,338	13b_T	1,261,850	1,432,810	10	1
CH261-59M8	13	13,415,153	13,615,434	13a_T	10,850,747	11,051,028	10	2
CH261-49P24	14	2,126,499	2,290,457	14c_T	4,098,394	4,262,352	8	7
CH261-122H14	14	3,691,588	3,889,704	14c_T	2,499,147	2,697,263	8	8
CH261-69D20	14	12,441,486	12,628,845	5b_14a_T	14,368,308	14,555,667	8	6
CH261-90P23	15	1,651,110	1,836,498	15_T	6,940,648	7,126,036	11	3
CH261-48M1	15	2,637,953	2,821,838	15_T	7,927,491	8,111,376	11	4
TGMCBA-266G23	15	11,679,111	11,833,972	15_T	1,757,152	1,912,013	11	1
CH261-40D6	15	12,447,806	12,640,948	15_T	2,525,847	2,718,989	11	2
TGMCBA-375I5	17	3,741,293	3,870,581	17_T	3,561,973	3,691,261	2	3
CH261-113A7	17	8,496,851	8,643,900	17_T	9,143,229	9,290,278	2	4
CH261-60N6	18	4,361,850	4,577,933	18_T	9,065,648	9,281,731	12	4
TGMCBA-263I20	18	4,517,555	4,662,859	18_T	9,221,353	9,366,657	12	5
CH261-137B21	18	6,091,090	6,242,853	18_T	3,528,409	3,680,172	12	3
CH261-118D24	18	6,845,464	7,015,909	18_T	2,755,353	2,925,798	12	2
TGMCBA-48A5	18	9,222,807	9,361,375	18_T	409,887	548,455	12	1
CH261-10F1	19	1,418,025	1,558,554	19a_T	4,575,674	4,716,203	13	4
TGMCBA-307H9	19	3,195,095	3,311,811	19a_T	2,822,417	2,939,133	13	3
CH261-50H12	19	7,445,191	7,596,603	19b_T	2,391,826	2,543,238	13	2
TGMCBA-84A3	19	7,968,445	8,093,059	19b_T	1,895,370	2,019,984	13	1
TGMCBA-250E3	20	421,457	570,315	20_T	11,456,685	11,605,543	10	3
TGMCBA-341F20	20	10,890,041	11,017,721	20_T	14,161,631	14,289,311	10	4
CH261-83I20	21	2,228,130	2,421,978	21_T	1,931,367	2,125,215	14	1

BAC Clone	Chicken Chromosome	Chicken Start (bp)	Chicken End (bp)	New PCF	PCF Start	PCF End	Budgerigar Chromosome	Orientation
CH261-122K8	21	3,753,855	3,934,968	21_T	4,117,680	4,298,793	14	2
CH261-30D24	22	13,919	194,754	22a_T	61,042	241,877	17	1
CH261-40J9	22	2,014,467	2,192,777	22b_T	2,710,236	2,888,546	17	2
CH261-18G17	22	2,432,263	2,647,085	22b_T	2,673,724	2,888,546	17	3
CH261-191G17	23	5,018	223,295	23b_T	416,089	634,366	16	1
TGMCBA-173N15	23	4,268,881	4,404,285	23c_T	409,971	545,375	16	4
CH261-49G9	23	4,628,133	4,849,754	23c_T	11,264	232,885	16	5
TGMCBA-48O8	23	5,158,595	5,284,109	23b_T	1,467,576	1,593,090	16	3
CH261-90K11	23	5,400,350	5,562,615	23b_T	1,212,562	1,374,827	16	2
TGMCBA-111K1	24	2,213,276	2,354,040	24d_T	502,608	643,372	15	1
CH261-65O4	24	3,200,064	3,351,622	24a_T	1,234,357	1,385,915	15	2
CH261-59C21	25	1,019,036	1,162,925	25_T	141,546	285,435	21	2
CH261-127K7	25	1,061,651	1,174,912	25_T	129,559	242,820	21	1
TGMCBA-332G15	26	1,263,122	1,338,812	26a_T	24,864	100,554	18	1
TGMCBA-297G21	26	1,489,981	1,598,968	26a_T	251,723	360,710	18	2
CH261-186M13	26	1,554,444	1,731,104	26a_T	271,257	447,917	18	3
CH261-170L23	26	3,688,292	3,882,865	26b_T	300,565	495,138	18	4
CH261-66M16	27	4,431,726	4,607,863	27c_T	543,477	719,614	20	3
CH261-28L10	27	4,809,590	5,025,707	27c_T	165,860	381,977	20	2
CH261-100E5	27	5,025,802	5,190,528	27c_T	1,039	165,765	20	1
CH261-72A10	28	4,134,415	4,344,571	28b_T	296,983	507,139	19	1
TGMCBA-200J22	Z	1,395,029	1,575,234	33b_T	35,473,787	35,653,992	Z	4
CH261-129A16	Z	6,524,854	6,750,728	33b_T	30,298,293	30,524,167	Z	3
CH261-133M4	Z	32,114,620	32,306,839	33b_T	4,742,182	4,934,401	Z	2

BAC Clone	Chicken Chromosome	Chicken Start (bp)	Chicken End (bp)	New PCF	PCF Start	PCF End	Budgerigar Chromosome	Orientation
TGMCBA-270I9	Z	55,607,345	55,764,093	33e_T	11,781,782	11,938,530	Z	1

Supplementary Table S4: The complete list of BACs for the comparative mapping study and their coordinates in the chicken genome (*Gallus_gallus*-4.0).

BAC Clone	Chicken Chromosome	Length (bp)	Start (bp)	End (bp)
CH261-89C18	1	171,359	875,622	1,046,980
CH261-89G23	1	236,918	20,538,145	20,775,062
CH261-119K2	1	240,022	29,623,402	29,863,423
CH261-120J2	1	232,240	34,010,417	34,242,656
CH261-36B5	1	207,564	65,968,509	66,176,072
CH261-25P18	1	237,839	71,546,463	71,784,301
CH261-125F1	1	207,294	77,066,050	77,273,343
CH261-118M1	1	229,237	98,389,770	98,619,006
CH261-18J16	1	258,356	110,513,385	110,771,740
CH261-29N14	1	196,686	120,693,003	120,889,688
CH261-9B17	1	204,575	132,642,594	132,847,168
CH261-168O17	1	261,724	140,904,296	141,166,019
CH261-83O13	1	240,258	146,261,866	146,502,123
CH261-107E2	1	228,936	155,895,248	156,124,183
CH261-58K12	1	202,909	166,741,351	166,944,259
CH261-184E5	1	233,997	172,851,270	173,085,266
CH261-98G4	1	182,677	190,251,863	190,434,539
CH261-192C19	2	177,019	644,516	821,534
CH261-169N6	2	195,579	5,892,852	6,088,430
CH261-172N3	2	148,224	13,944,048	14,092,271
CH261-177K1	2	242,791	24,003,266	24,246,056
CH261-186J5	2	227,236	29,444,922	29,672,157
CH261-40G6	2	187,835	37,719,921	37,907,755
CH261-50C15	2	235,136	43,392,223	43,627,358
CH261-123O22	2	178,051	47,697,513	47,875,563
CH261-169E4	2	229,265	103,877,781	104,107,045
CH261-1J20	2	236,575	113,204,836	113,441,410
CH261-44D16	2	205,011	119,618,244	119,823,254
CH261-44H14	2	204,003	128,738,154	128,942,156
CH261-17J16	2	180,835	143,390,457	143,571,291
CH261-115J5	3	182,630	79,367	261,996
CH261-18I9	3	222,432	15,121,477	15,343,908
CH261-130M12	3	165,410	21,229,771	21,395,180
CH261-160I6	3	230,242	25,895,426	26,125,667
CH261-97P20	3	238,286	40,923,906	41,162,191
CH261-17B14	3	223,639	65,993,327	66,216,965
CH261-169K18	3	140,443	96,076,845	96,217,287
CH261-120H23	3	180,672	108,536,641	108,717,312
CH261-183B15	4	198,902	224,440	423,341

BAC Clone	Chicken Chromosome	Length (bp)	Start (bp)	End (bp)
CH261-83E1	4	171,741	4,051,379	4,223,119
CH261-71L6	4	187,959	4,108,958	4,296,916
CH261-111A15	4	222,310	15,434,326	15,656,635
CH261-18C6	4	252,751	30,924,495	31,177,245
CH261-93H1	4	215,958	37,997,136	38,213,093
CH261-185L11	4	264,640	43,548,596	43,813,235
CH261-85H10	4	245,153	65,511,732	65,756,884
CH261-89P6	4	141,721	74,544,898	74,686,618
CH261-73F2	5	195,175	1,000,754	1,195,928
CH261-49B22	5	193,278	5,145,361	5,338,638
CH261-122F8	5	170,263	9,960,804	10,131,066
CH261-78F13	5	186,801	25,072,913	25,259,713
CH261-2I23	5	167,900	26,505,433	26,673,332
CH261-161B22	5	158,731	57,365,804	57,524,534
CH261-94G14	6	183,701	811,766	995,466
CH261-67H5	6	253,681	22,199,038	22,452,718
CH261-165L8	6	198,177	22,452,838	22,651,014
CH261-49F3	6	112,713	22,742,063	22,854,775
CH261-179F2	6	169,249	35,037,507	35,206,755
CH261-112D24	7	171,574	121,179	292,752
CH261-56K7	7	176,360	3,628,602	3,804,961
CH261-180H18	7	232,888	20,581,283	20,814,170
CH261-186K14	7	176,097	26,239,695	26,415,791
CH261-38E18	7	190,848	36,600,878	36,791,725
CH261-69H1	8	201,507	3,983,002	4,184,508
CH261-107D8	8	225,011	4,545,787	4,770,797
CH261-34H16	8	204,614	15,703,125	15,907,738
CH261-96D24	8	193,292	29,701,130	29,894,421
CH261-183N19	9	173,995	615,521	789,515
CH261-95N3	9	193,574	11,900,212	12,093,785
CH261-187M16	9	188,641	15,204,265	15,392,905
CH261-68O18	9	190,968	23,481,354	23,672,321
CH261-129A16	Z	225,875	6,665,985	6,891,859
CH261-133M4	Z	192,220	32,255,751	32,447,970
CH261-137F19	Z	208,754	81,107,116	81,315,869

Supplementary Table S5: The full table of BACs successfully hybridised in the common blackbird, Atlantic Canary, Eurasian woodcock, Helmeted guinea fowl, houbara bustard, mallard duck, and rock dove.

BAC Clone	Chicken Chromosome	Common Blackbird	Atlantic Canary	Eurasian Woodcock	Helmeted Guinea Fowl	Houbara Bustard	Mallard Duck	Rock Dove
CH261-89C18	1	Y	Y	Y	Y	Y	Y	Y
CH261-89G23	1	Y	Y	Y	Y	Y	Y	Y
CH261-119K2	1	Y	Y	Y	Y	Y	Y	Y
CH261-120J2	1	Y	Y	Y	Y	Y	Y	Y
CH261-36B5	1	N	Y	N	Y	N	Y	Y
CH261-25P18	1	N	N	Y	Y	Y	Y	Y
CH261-125F1	1	Y	Y	Y	Y	Y	Y	Y
CH261-118M1	1	Y	Y	Y	Y	Y	Y	Y
CH261-18J16	1	Y	Y	Y	Y	Y	Y	Y
CH261-29N14	1	Y	Y	Y	Y	Y	Y	Y
CH261-9B17	1	Y	Y	N	Y	Y	Y	Y
CH261-168O17	1	Y	Y	Y	Y	Y	Y	Y
CH261-83O13	1	Y	Y	Y	Y	Y	Y	Y
CH261-107E2	1	Y	Y	Y	Y	Y	Y	Y
CH261-58K12	1	Y	Y	Y	Y	Y	Y	Y
CH261-184E5	1	Y	Y	Y	Y	Y	Y	Y
CH261-98G4	1	Y	Y	Y	Y	Y	Y	Y
CH261-192C19	2	N	N	N	Y	Y	Y	N
CH261-169N6	2	Y	Y	Y	Y	Y	Y	Y
CH261-172N3	2	Y	N	N	Y	Y	Y	Y

BAC Clone	Chicken Chromosome	Common Blackbird	Atlantic Canary	Eurasian Woodcock	Helmeted Guinea Fowl	Houbara Bustard	Mallard Duck	Rock Dove
CH261-177K1	2	Y	Y	Y	Y	Y	Y	Y
CH261-186J5	2	N	Y	N	Y	Y	Y	Y
CH261-40G6	2	Y	Y	N	Y	Y	Y	Y
CH261-50C15	2	N	N	N	Y	N	N	Y
CH261-123O22	2	Y	Y	N	Y	Y	Y	Y
CH261-169E4	2	Y	Y	Y	Y	Y	Y	Y
CH261-1J20	2	Y	Y	Y	Y	Y	Y	Y
CH261-44D16	2	Y	Y	Y	Y	Y	Y	Y
CH261-44H14	2	N	N	N	Y	Y	N	Y
CH261-17J16	2	N	Y	N	Y	Y	Y	Y
CH261-115J5	3	Y	N	Y	Y	N	Y	Y
CH261-18I9	3	Y	Y	Y	Y	Y	Y	Y
CH261-130M12	3	Y	Y	Y	Y	Y	Y	Y
CH261-160I6	3	Y	Y	N	Y	Y	Y	Y
CH261-97P20	3	Y	N	Y	Y	Y	Y	Y
CH261-17B14	3	Y	Y	N	Y	Y	Y	Y
CH261-169K18	3	N	Y	Y	Y	Y	Y	Y
CH261-120H23	3	Y	Y	Y	Y	Y	Y	Y
CH261-183B15	4	N	N	N	Y	Y	N	N
CH261-83E1	4	Y	N	Y	Y	Y	Y	Y
CH261-71L6	4	Y	N	Y	Y	Y	Y	Y
CH261-111A15	4	Y	N	Y	Y	Y	Y	Y
CH261-18C6	4	Y	Y	Y	Y	Y	Y	Y
CH261-93H1	4	Y	Y	Y	Y	Y	Y	Y

BAC Clone	Chicken Chromosome	Common Blackbird	Atlantic Canary	Eurasian Woodcock	Helmeted Guinea Fowl	Houbara Bustard	Mallard Duck	Rock Dove
CH261-185L11	4	N	Y	Y	Y	Y	Y	Y
CH261-85H10	4	Y	Y	Y	Y	Y	Y	Y
CH261-89P6	4	Y	Y	Y	Y	Y	Y	Y
CH261-73F2	5	N	N	N	Y	Y	N	N
CH261-49B22	5	Y	Y	Y	Y	Y	Y	Y
CH261-122F8	5	Y	Y	Y	Y	Y	Y	Y
CH261-78F13	5	Y	Y	Y	Y	Y	N	Y
CH261-2I23	5	Y	Y	Y	Y	Y	Y	Y
CH261-161B22	5	Y	N	Y	Y	Y	Y	Y
CH261-94G14	6	N	Y	N	Y	N	N	Y
CH261-67H5	6	Y	N	N	Y	Y	Y	Y
CH261-165L8	6	N	Y	Y	Y	Y	Y	Y
CH261-49F3	6	Y	Y	Y	Y	Y	Y	Y
CH261-179F2	6	Y	N	N	Y	N	N	N
CH261-112D24	7	Y	Y	N	Y	Y	N	Y
CH261-56K7	7	Y	N	Y	Y	Y	Y	Y
CH261-180H18	7	Y	Y	Y	Y	Y	Y	Y
CH261-186K14	7	Y	Y	Y	Y	Y	Y	Y
CH261-38E18	7	A	N	Y	Y	Y	N	N
CH261-69H1	8	Y	Y	Y	Y	Y	Y	Y
CH261-107D8	8	Y	Y	Y	Y	Y	Y	Y
CH261-34H16	8	Y	Y	Y	Y	Y	Y	Y
CH261-96D24	8	Y	Y	N	Y	Y	Y	Y
CH261-183N19	9	Y	N	Y	Y	Y	Y	Y

BAC Clone	Chicken Chromosome	Common Blackbird	Atlantic Canary	Eurasian Woodcock	Helmeted Guinea Fowl	Houbara Bustard	Mallard Duck	Rock Dove
CH261-95N3	9	N	Y	Y	Y	Y	Y	Y
CH261-187M16	9	Y	Y	Y	Y	Y	Y	Y
CH261-68O18	9	N	N	N	Y	N	N	Y
CH261-129A16	Z	Y	Y	Y	Y	Y	Y	Y
CH261-133M4	Z	Y	Y	Y	Y	Y	Y	Y
CH261-137F19	Z	N	N	Y	Y	N	N	Y

Supplementary Table S6: FLpter values, standard deviations, and the number of mitotic chromosomes measured for the chicken.

BAC Clone	Chicken FLpter Value	*S.D*	n
CH261-89C18	0.0302	0.0119	65
CH261-89G23	0.1063	0.0357	39
CH261-119K2	0.1483	0.0413	36
CH261-120J2	0.1685	0.0314	37
CH261-36B5	0.3222	0.0273	40
CH261-25P18	0.3600	0.0581	30
CH261-125F1	0.4117	0.0650	38
CH261-118M1	0.5295	0.0382	38
CH261-18J16	0.5868	0.0565	32
CH261-29N14	0.6274	0.0380	35
CH261-9B17	0.7067	0.0444	38
CH261-168O17	0.7237	0.0280	31
CH261-83O13	0.7594	0.0323	38
CH261-107E2	0.8004	0.0335	29
CH261-58K12	0.8439	0.0316	38
CH261-184E5	0.8824	0.0337	35
CH261-98G4	0.9603	0.0261	40
CH261-192C19	0.0499	0.0261	35
CH261-169N6	0.0581	0.0215	31
CH261-172N3	0.1178	0.0233	63
CH261-177K1	0.1764	0.0390	35
CH261-186J5	0.1924	0.0348	35
CH261-40G6	0.2532	0.0314	38
CH261-50C15	0.2876	0.0438	31
CH261-123O22	0.3145	0.0394	39
CH261-169E4	0.7030	0.0267	36
CH261-1J20	0.7548	0.0320	40
CH261-44D16	0.7979	0.0284	36
CH261-44H14	0.8553	0.0275	35
CH261-17J16	0.9381	0.0297	37
CH261-115J5	0.0586	0.0287	34
CH261-18I9	0.1983	0.0415	36
CH261-130M12	0.2696	0.0581	34
CH261-160I6	0.2866	0.0299	37
CH261-97P20	0.4005	0.0379	32
CH261-17B14	0.6005	0.0454	42
CH261-169K18	0.8808	0.0600	27
CH261-120H23	0.9663	0.0323	41
CH261-183B15	0.0596	0.0284	34

BAC Clone	Chicken FLpter Value	*S.D*	n
CH261-83E1	0.0732	0.0257	30
CH261-71L6	0.0735	0.0325	33
CH261-111A15	0.1765	0.0592	38
CH261-18C6	0.3717	0.0405	39
CH261-93H1	0.4423	0.0453	39
CH261-185L11	0.5081	0.0605	45
CH261-85H10	0.7284	0.0494	35
CH261-89P6	0.8183	0.0514	68
CH261-73F2	0.0746	0.0242	36
CH261-49B22	0.1670	0.0579	36
CH261-122F8	0.2455	0.0666	35
CH261-78F13	0.4686	0.0488	26
CH261-2I23	0.4806	0.0478	36
CH261-161B22	0.9461	0.0452	38
CH261-94G14	0.2377	0.0449	38
CH261-67H5	0.6964	0.0767	26
CH261-165L8	0.6876	0.0629	31
CH261-49F3	0.7037	0.0726	57
CH261-179F2	0.9073	0.0638	36
CH261-112D24	0.1172	0.0366	33
CH261-56K7	0.1534	0.0485	64
CH261-180H18	0.6292	0.0679	37
CH261-186K14	0.7473	0.0649	25
CH261-38E18	0.8769	0.0647	36
CH261-69H1	0.0872	0.0360	29
CH261-107D8	0.1573	0.0640	34
CH261-34H16	0.5448	0.0744	32
CH261-96D24	0.9149	0.0644	67
CH261-183N19	0.1980	0.0559	33
CH261-95N3	0.6008	0.0774	27
CH261-187M16	0.7444	0.0764	30
CH261-68O18	0.8941	0.0653	34
CH261-129A16	0.0932	0.0444	25
CH261-133M4	0.3726	0.0474	48
CH261-137F19	0.9116	0.0480	27

Supplementary Table S7: FLpter values, standard deviations, and the number of mitotic chromosomes measured for the common blackbird.

BAC Clone	Common Blackbird FLpter Value	*S.D*	n
CH261-89C18	0.8734	0.0714	24
CH261-89G23	0.5762	0.0632	28
CH261-119K2	0.4755	0.0739	27
CH261-120J2	0.4673	0.0540	26
CH261-36B5	-	-	-
CH261-25P18	-	-	-
CH261-125F1	0.3990	0.0601	30
CH261-118M1	0.1157	0.0424	28
CH261-18J16	0.9258	0.0379	23
CH261-29N14	0.8522	0.0628	29
CH261-9B17	0.7619	0.0611	33
CH261-168O17	0.7245	0.0495	29
CH261-83O13	0.7052	0.0582	35
CH261-107E2	0.676	0.0758	19
CH261-58K12	0.5726	0.0637	34
CH261-184E5	0.5443	0.0661	39
CH261-98G4	0.457	0.0792	32
CH261-192C19	-	-	-
CH261-169N6	0.0729	0.0385	26
CH261-172N3	0.0696	0.0465	28
CH261-177K1	0.1803	0.0491	29
CH261-186J5	-	-	-
CH261-40G6	0.2109	0.0637	20
CH261-50C15	-	-	-
CH261-123O22	0.2997	0.0512	28
CH261-169E4	0.6768	0.0634	39
CH261-1J20	0.7573	0.0497	32
CH261-44D16	0.7618	0.0691	38
CH261-44H14	-	-	-
CH261-17J16	-	-	-
CH261-115J5	0.2230	0.0621	30
CH261-18I9	0.2587	0.0528	34
CH261-130M12	0.3708	0.0587	34
CH261-160I6	0.3925	0.0581	31
CH261-97P20	0.49663	0.0508	27
CH261-17B14	0.624	0.0638	36
CH261-169K18	-	-	-
CH261-120H23	0.9136	0.0300	38
CH261-183B15	-	-	-

BAC Clone	Common Blackbird FLpter Value	*S.D*	n
CH261-83E1	Micro	N/A	N/A
CH261-71L6	Micro	N/A	N/A
CH261-111A15	Micro	N/A	N/A
CH261-18C6	0.1933	0.0602	35
CH261-93H1	0.4826	0.0546	30
CH261-185L11	-	-	-
CH261-85H10	0.7526	0.0680	22
CH261-89P6	0.8394	0.0604	27
CH261-73F2	-	-	-
CH261-49B22	0.1264	0.0549	32
CH261-122F8	0.4095	0.0709	34
CH261-78F13	0.6025	0.0580	16
CH261-2I23	0.5573	0.0639	24
CH261-161B22	0.8141	0.0525	19
CH261-94G14	-	-	-
CH261-67H5	0.7334	0.0765	29
CH261-165L8	-	-	-
CH261-49F3	0.6800	0.0796	28
CH261-179F2	0.8251	0.0706	19
CH261-112D24	0.4481	0.0705	13
CH261-56K7	0.3766	0.0625	31
CH261-180H18	0.7921	0.0558	15
CH261-186K14	0.6628	0.0662	12
CH261-38E18	0.8383	0.0518	16
CH261-69H1	0.173	0.0334	26
CH261-107D8	0.3193	0.0710	27
CH261-34H16	0.6426	0.0780	16
CH261-96D24	0.8147	0.0316	19
CH261-183N19	0.2697	0.0431	17
CH261-95N3	-	-	-
CH261-187M16	0.6010	0.0785	17
CH261-68O18	-	-	-
CH261-129A16	0.8576	0.0305	20
CH261-133M4	0.5411	0.0326	20
CH261-137F19	-	-	-

Supplementary Table S8: FLpter values, standard deviations, and the number of mitotic chromosomes measured for the Atlantic canary.

BAC Clone	Atlantic Canary FLpter Value	*S.D*	n
CH261-89C18 chr1	0.9272	0.0485	53
CH261-89C18 chr1A	0.9397	0.0582	15
CH261-89G23	0.7935	0.0732	23
CH261-119K2	0.6851	0.0602	34
CH261-120J2 chr1A	0.6729	0.0605	31
CH261-120J2 chr2	0.4418	0.0553	31
CH261-36B5	0.1299	0.0540	43
CH261-25P18	-	-	-
CH261-125F1	0.4255	0.0768	32
CH261-118M1 chr1	0.1451	0.0511	39
CH261-118M1 chr2	0.4799	0.0688	17
CH261-18J16	0.8937	0.0580	33
CH261-29N14	0.8253	0.0627	23
CH261-9B17	0.7363	0.0578	32
CH261-168O17	0.7273	0.0628	33
CH261-83O13	0.6944	0.0680	28
CH261-107E2	0.6545	0.0649	36
CH261-58K12	0.6422	0.0579	22
CH261-184E5	0.5362	0.0641	31
CH261-98G4	0.4385	0.0748	17
CH261-192C19	-	-	-
CH261-169N6	0.0832	0.0405	19
CH261-172N3	-	-	-
CH261-177K1	0.1881	0.0401	19
CH261-186J5	0.2412	0.0575	20
CH261-186J5 duplication	0.4705		
CH261-40G6	0.2521	0.0599	15
CH261-40G6 duplication	0.4802	0.0424	25
CH261-50C15	-	-	-
CH261-123O22	0.3011	0.0677	19
CH261-123O23 duplication	0.5011	0.0601	17
CH261-169E4	0.7566	0.0438	34
CH261-1J20	0.7976	0.0454	23
CH261-44D16	0.8153	0.0516	31
CH261-44H14	-	-	-
CH261-17J16	0.4601	0.0516	23
CH261-115J5	-	-	-
CH261-18I9	0.1704	0.0451	19

BAC Clone	Atlantic Canary FLpter Value	*S.D*	n
CH261-130M12	0.4756	0.0557	40
CH261-160I6	0.5000	0.0526	25
CH261-97P20	-	-	-
CH261-17B14	0.6992	0.0474	28
CH261-169K18	0.8465	0.0413	29
CH261-120H23	0.9151	0.0406	30
CH261-183B15	-	-	-
CH261-83E1	-	-	-
CH261-71L6	-	-	-
CH261-111A15	-	-	-
CH261-18C6	0.4042	0.0562	33
CH261-93H1	0.5175	0.0698	30
CH261-185L11	0.5821	0.0596	26
CH261-85H10	0.6570	0.0632	33
CH261-89P6	0.7727	0.0664	23
CH261-73F2	-	-	-
CH261-49B22	0.1727	0.0376	22
CH261-122F8	0.3859	0.0489	25
CH261-78F13	0.5655	0.0653	20
CH261-2I23	0.5282	0.0308	21
CH261-161B22	-	-	-
CH261-94G14	0.6072	0.0524	22
CH261-67H5	-	-	-
CH261-165L8	0.6930	0.0592	18
CH261-49F3	0.4292	0.0634	19
CH261-179F2	-	-	-
CH261-112D24	0.373	0.0539	16
CH261-56K7	-	-	-
CH261-180H18	0.5784	0.0564	20
CH261-186K14	0.7565	0.0640	18
CH261-38E18	-	-	-
CH261-69H1	0.2751	0.0629	18
CH261-107D8	0.2982	0.0601	23
CH261-34H16	0.6922	0.0611	24
CH261-96D24	0.777	0.0505	17
CH261-183N19	-	-	-
CH261-95N3	0.2729	0.0655	28
CH261-187M16	0.7137	0.0583	17
CH261-68O18	-	-	-
CH261-129A16	0.1407	0.0500	27
CH261-133M4	0.3701	0.0431	20
CH261-137F19	-	-	-

Supplementary Table S9: FLpter values, standard deviations, and the number of mitotic chromosomes measured for the Eurasian woodcock.

BAC Clone	Eurasian Woodcock FLpter Value	*S.D*	n
CH261-89C18	0.8767	0.0357	26
CH261-89G23	0.7234		
CH261-119K2	0.5986	0.0596	22
CH261-120J2	0.5499	0.0596	18
CH261-36B5	-	-	-
CH261-25P18	0.0669	0.0310	16
CH261-125F1	0.0760	0.0302	13
CH261-118M1	0.3115	0.0451	17
CH261-18J16	0.3709	0.0487	14
CH261-29N14	0.4644	0.0480	11
CH261-9B17	-	-	-
CH261-168O17	0.5820	0.0576	14
CH261-83O13	0.5955	0.0573	26
CH261-107E2	0.6775	0.0666	30
CH261-58K12	0.7741	0.0668	7
CH261-184E5	0.8337	0.0529	20
CH261-98G4	0.9156	0.0300	31
CH261-192C19	-	-	-
CH261-169N6	Micro	N/A	N/A
CH261-172N3	-	-	-
CH261-177K1	Micro	N/A	N/A
CH261-186J5	-	-	-
CH261-40G6	-	-	-
CH261-50C15	-	-	-
CH261-123O22	-	-	-
CH261-169E4	0.7556	0.0585	27
CH261-1J20	0.8576	0.0366	22
CH261-44D16	0.8769	0.0283	20
CH261-44H14	-	-	-
CH261-17J16	-	-	-
CH261-115J5	Micro	N/A	N/A
CH261-18I9	Micro	N/A	N/A
CH261-130M12	0.388	0.0573	21
CH261-160I6	-	-	-
CH261-97P20	0.7225	0.0557	14
CH261-17B14	-	-	-
CH261-169K18	Micro	N/A	N/A
CH261-120H23	Micro	N/A	N/A
CH261-183B15	-	-	-

BAC Clone	Eurasian Woodcock FLpter Value	*S.D*	n
CH261-83E1	Micro	N/A	N/A
CH261-71L6	Micro	N/A	N/A
CH261-111A15	Micro	N/A	N/A
CH261-18C6	0.4772		
CH261-93H1	0.2012	0.0468	10
CH261-185L11	0.7448	0.0597	14
CH261-85H10	0.8303	0.0552	10
CH261-89P6	0.8656	0.0403	37
CH261-73F2	-	-	-
CH261-49B22	0.4574	0.0528	11
CH261-122F8	Micro	N/A	N/A
CH261-78F13	0.6074	0.0449	19
CH261-2I23	0.4710	0.0664	12
CH261-161B22	0.8362	0.0517	15
CH261-94G14	-	-	-
CH261-67H5	-	-	-
CH261-165L8	0.7282	0.0549	14
CH261-49F3	Micro	N/A	N/A
CH261-179F2	-	-	-
CH261-112D24	-	-	-
CH261-56K7	0.123	0.0531	14
CH261-180H18	0.6282	0.0544	13
CH261-186K14	0.7416	0.0587	15
CH261-38E18	0.8004	0.0619	19
CH261-69H1	0.399	0.0394	17
CH261-107D8	0.3638	0.0583	25
CH261-34H16	0.6279	0.0545	14
CH261-96D24	-	-	-
CH261-183N19	0.2244	0.0566	24
CH261-95N3	0.5520	0.0425	16
CH261-187M16	0.5868	0.0611	26
CH261-68O18	-	-	-
CH261-129A16	0.1148	0.0220	17
CH261-133M4	0.3912	0.0472	17
CH261-137F19	On Chr W	-	-

Supplementary Table S10: FLpter values, standard deviations, and the number of mitotic chromosomes measured for the helmeted guinea fowl.

BAC Clone	Helmeted Guinea Fowl FLpter Value	*S.D*	n
CH261-89C18	0.0407	0.0202	99
CH261-89G23	0.1220	0.0342	40
CH261-119K2	0.1758	0.0285	33
CH261-120J2	0.1598	0.0324	36
CH261-36B5	0.3481	0.0267	38
CH261-25P18	0.3678	0.0346	40
CH261-125F1	0.4476	0.0393	34
CH261-118M1	0.5318	0.0391	35
CH261-18J16	0.5715	0.0557	36
CH261-29N14	0.6281	0.0335	39
CH261-9B17	0.6946	0.0382	39
CH261-168O17	0.7444	0.0224	34
CH261-83O13	0.7505	0.0226	39
CH261-107E2	0.7917	0.0251	39
CH261-58K12	0.8478	0.0294	36
CH261-184E5	0.8755	0.0212	38
CH261-98G4	0.9237	0.0232	39
CH261-192C19	0.0415	0.0171	31
CH261-169N6	0.0576	0.0225	28
CH261-172N3	0.1112	0.0335	72
CH261-177K1	0.1748	0.0306	37
CH261-186J5	0.2001	0.0283	36
CH261-40G6	0.2341	0.0410	35
CH261-50C15	0.2885	0.0329	27
CH261-123O22	0.3104	0.0392	36
CH261-169E4	0.6765	0.0348	37
CH261-1J20	0.7314	0.0322	39
CH261-44D16	0.7701	0.0368	37
CH261-44H14	0.8282	0.0301	35
CH261-17J16	0.9126	0.0271	36
CH261-115J5	0.0687	0.0288	29
CH261-18I9	0.1612	0.0461	33
CH261-130M12	0.2417	0.0420	35
CH261-160I6	0.2813	0.0486	30
CH261-97P20	0.3928	0.0378	34
CH261-17B14	0.5938	0.0477	35
CH261-169K18	0.8556	0.0439	25
CH261-120H23	0.9394	0.0326	36
CH261-183B15	Micro	N/A	N/A

BAC Clone	Helmeted Guinea Fowl FLpter Value	*S.D*	n
CH261-83E1	Micro	N/A	N/A
CH261-71L6	Micro	N/A	N/A
CH261-111A15	Micro	N/A	N/A
CH261-18C6	0.4117	0.0391	37
CH261-93H1	0.4807	0.0452	33
CH261-185L11	0.5491	0.0542	22
CH261-85H10	0.7549	0.0395	36
CH261-89P6	0.8331	0.0419	35
CH261-73F2	0.1210	0.0410	32
CH261-49B22	0.1995	0.0516	34
CH261-122F8	0.2910	0.0561	31
CH261-78F13	0.4248	0.0532	31
CH261-2I23	0.4885	0.0659	33
CH261-161B22	0.8955	0.0380	34
CH261-94G14	0.3892	0.0390	37
CH261-67H5	0.2147	0.0452	29
CH261-165L8	0.1969	0.0396	23
CH261-49F3	0.1875	0.0407	51
CH261-179F2	0.0914	0.0314	30
CH261-112D24	0.6041	0.0420	29
CH261-56K7	0.5625	0.0457	66
CH261-180H18	0.7847	0.0463	20
CH261-186K14	0.8268	0.0465	29
CH261-38E18	0.9157	0.0341	35
CH261-69H1	0.3487	0.0489	32
CH261-107D8	0.2971	0.0417	24
CH261-34H16	0.3539	0.0630	30
CH261-96D24	0.8397	0.0692	34
CH261-183N19	0.2375	0.0546	37
CH261-95N3	0.1783	0.0376	25
CH261-187M16	0.1069	0.0359	31
CH261-68O18	0.0680	0.0239	33
CH261-129A16	0.8873	0.0405	34
CH261-133M4	0.6121	0.0387	69
CH261-137F19	0.4627	0.0471	31

Supplementary Table S11: FLpter values, standard deviations, and the number of mitotic chromosomes measured for the houbara bustard.

BAC Clone	Houbara Bustard FLpter Value	*S.D*	n
CH261-89C18	0.0784	0.0328	54
CH261-89G23	0.1311	0.0485	29
CH261-119K2	0.175	0.0451	27
CH261-120J2	0.2238	0.0611	30
CH261-36B5	-	-	-
CH261-25P18	0.4063	0.0573	28
CH261-125F1	0.4176	0.0497	29
CH261-118M1	0.538	0.0624	26
CH261-18J16	0.5831	0.0522	30
CH261-29N14	0.662	0.0473	49
CH261-9B17	0.6828	0.0646	29
CH261-168O17	0.7185	0.0653	20
CH261-83O13	0.7415	0.0563	25
CH261-107E2	0.784	0.0545	34
CH261-58K12	0.8191	0.0470	26
CH261-184E5	0.8875	0.0442	45
CH261-98G4	0.9094	0.0446	32
CH261-192C19	0.0573	0.0237	21
CH261-169N6	0.0722	0.0345	28
CH261-172N3	0.1397	0.0514	57
CH261-177K1	0.2022	0.0408	35
CH261-186J5	0.2084	0.0378	28
CH261-40G6	0.2608	0.0569	30
CH261-50C15	-	-	-
CH261-123O22	0.3450	0.0460	28
CH261-169E4	0.7070	0.0572	31
CH261-1J20	0.7565	0.0502	37
CH261-44D16	0.7899	0.0443	29
CH261-44H14	0.8338	0.0502	24
CH261-17J16	0.8985	0.0486	27
CH261-115J5	-	-	-
CH261-18I9	0.1208	0.0429	25
CH261-130M12	0.4273	0.0504	27
CH261-160I6	0.4044	0.0589	24
CH261-97P20	0.277	0.0503	30
CH261-17B14	0.6181	0.0649	25
CH261-169K18	0.8293	0.0515	33
CH261-120H23	0.9229	0.0447	28
CH261-183B15	Micro	-	-

BAC Clone	Houbara Bustard FLpter Value	*S.D*	n
CH261-83E1	Micro	-	-
CH261-71L6	Micro	-	-
CH261-111A15	Micro	-	-
CH261-18C6	0.1994	0.0654	33
CH261-93H1	0.3787	0.0625	28
CH261-185L11	0.5653	0.0628	30
CH261-85H10	0.6314	0.0592	26
CH261-89P6	0.7756	0.0529	26
CH261-73F2	0.1509	0.0698	20
CH261-49B22	0.2503	0.0601	22
CH261-122F8	0.3576	0.0730	26
CH261-78F13	0.4421	0.0787	24
CH261-2I23	0.4911	0.0808	32
CH261-161B22	0.8527	0.0546	28
CH261-94G14	-	-	-
CH261-67H5	0.6627	0.0673	27
CH261-165L8	0.6647	0.0681	24
CH261-49F3	0.7277	0.0722	21
CH261-179F2	-	-	-
CH261-112D24	0.2505	0.0591	21
CH261-56K7	0.3037	0.0413	19
CH261-180H18	0.5902	0.0684	14
CH261-186K14	0.6292	0.0630	27
CH261-38E18	0.8485	0.0500	26
CH261-69H1	0.2339	0.0557	23
CH261-107D8	0.4254	0.0774	44
CH261-34H16	0.6178	0.0815	19
CH261-96D24	0.7612	0.0434	32
CH261-183N19	0.305	0.0604	22
CH261-95N3	0.2111	0.0500	27
CH261-187M16	0.6588	0.0713	31
CH261-68O18	-	-	-
CH261-129A16	0.8509	0.0437	25
CH261-133M4	0.591	0.0512	16
CH261-137F19	-	-	-

Supplementary Table S12: FLpter values, standard deviations, and the number of mitotic chromosomes measured for the mallard duck

BAC Clone	Mallard Duck FLpter Value	*S.D*	n
CH261-89C18	0.047	0.0225	21
CH261-89G23	0.0916	0.0412	27
CH261-119K2	0.1469	0.0407	34
CH261-120J2	0.1905	0.0500	25
CH261-36B5	0.4207	0.0388	32
CH261-25P18	0.3632	0.0462	33
CH261-125F1	0.4468	0.0511	21
CH261-118M1	0.5689	0.0535	27
CH261-18J16	0.6206	0.0459	30
CH261-29N14	0.656	0.0361	32
CH261-9B17	0.7082	0.0476	41
CH261-168O17	0.7575	0.0372	34
CH261-83O13	0.7506	0.0444	38
CH261-107E2	0.7777	0.0475	26
CH261-58K12	0.8676	0.0488	28
CH261-184E5	0.872	0.0452	23
CH261-98G4	0.9267	0.0468	29
CH261-192C19	0.0464	0.0211	14
CH261-169N6	0.0727	0.0349	24
CH261-172N3	0.1153	0.0207	46
CH261-177K1	0.1878	0.0407	29
CH261-186J5	0.1918	0.0480	21
CH261-40G6	0.2593	0.0435	17
CH261-50C15	-	-	-
CH261-123O22	0.3972	0.0371	27
CH261-169E4	0.7512	0.0573	33
CH261-1J20	0.7656	0.0479	35
CH261-44D16	0.7865	0.0424	34
CH261-44H14	-	-	-
CH261-17J16	0.9274	0.0439	28
CH261-115J5	0.104	0.0350	21
CH261-18I9	0.1239	0.0404	26
CH261-130M12	0.2491	0.0315	27
CH261-160I6	0.4134	0.0397	28
CH261-97P20	0.3908	0.0352	35
CH261-17B14	0.6065	0.0683	30
CH261-169K18	0.8421	0.0449	21
CH261-120H23	0.9114	0.0456	27
CH261-183B15	-	-	-

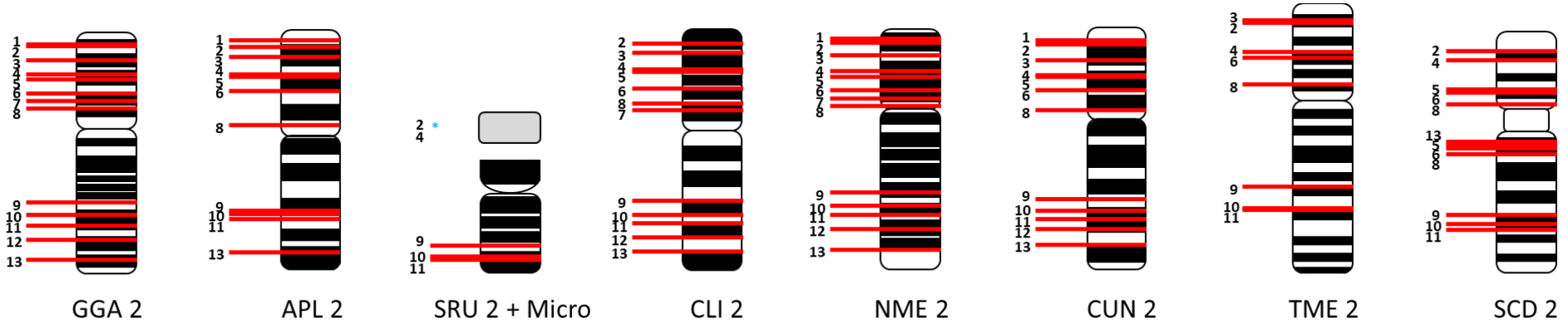
BAC Clone	Mallard Duck FLpter Value	*S.D*	n
CH261-83E1	Micro	N/A	N/A
CH261-71L6	Micro	N/A	N/A
CH261-111A15	Micro	N/A	N/A
CH261-18C6	0.5235	0.0574	28
CH261-93H1	0.1765	0.0476	30
CH261-185L11	0.3778	0.0637	24
CH261-85H10	0.6530	0.0583	29
CH261-89P6	0.7875	0.0564	29
CH261-73F2	-	-	-
CH261-49B22	0.1587	0.0526	25
CH261-122F8	0.2125	0.0546	27
CH261-78F13	-	-	-
CH261-2I23	0.4829	0.0510	25
CH261-161B22	0.8693	0.0572	26
CH261-94G14	-	-	-
CH261-67H5	0.6162	0.0541	17
CH261-165L8	0.6978	0.0346	14
CH261-49F3	0.7167	0.0377	22
CH261-179F2	-	-	-
CH261-112D24	-	-	-
CH261-56K7	0.2003	0.0579	22
CH261-180H18	0.6058	0.0548	22
CH261-186K14	0.6217	0.0478	18
CH261-38E18	-	-	-
CH261-69H1	0.2915	0.0424	19
CH261-107D8	0.3033	0.0535	27
CH261-34H16	0.6137	0.0573	22
CH261-96D24	0.7894	0.0509	33
CH261-183N19	0.2256	0.0501	16
CH261-95N3	0.5946	0.0496	20
CH261-187M16	0.619	0.0565	22
CH261-68O18	-	-	-
CH261-129A16	0.8669	0.0448	20
CH261-133M4	0.5716	0.0560	21
CH261-137F19	-	-	-

Supplementary Table S13: FLpter values, standard deviations, and the number of mitotic chromosomes measured for the rock dove.

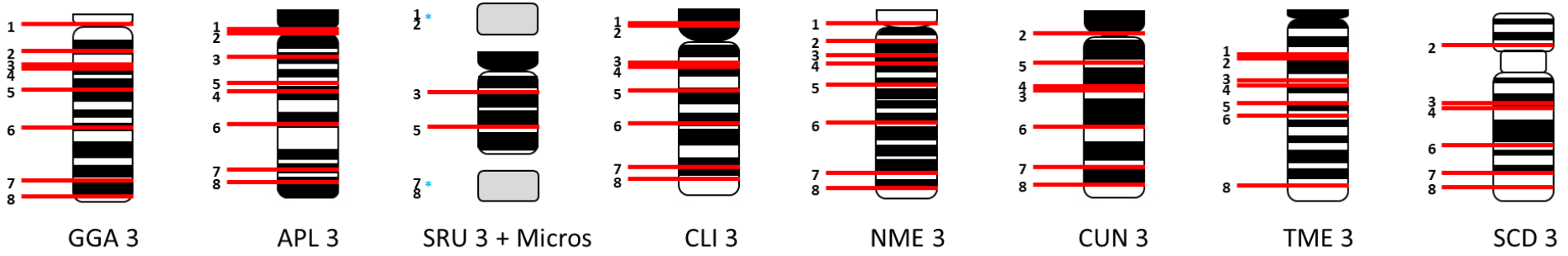
BAC Clone	Rock Dove FLpter Value	*S.D*	n
CH261-89C18	0.0295	0.0172	29
CH261-89G23	0.1025	0.0245	38
CH261-119K2	0.1498	0.0293	28
CH261-120J2	0.1718	0.0327	30
CH261-36B5	0.3750	0.0423	23
CH261-25P18	0.3257	0.0481	35
CH261-125F1	0.4409	0.0475	30
CH261-118M1	0.4409	0.0417	36
CH261-18J16	0.4409	0.0425	18
CH261-29N14	0.4409	0.0378	30
CH261-9B17	0.4409	0.0611	34
CH261-168O17	0.4409	0.0432	31
CH261-83O13	0.4409	0.0378	37
CH261-107E2	0.4409	0.0373	35
CH261-58K12	0.4409	0.0397	31
CH261-184E5	0.4409	0.0434	35
CH261-98G4	0.4409	0.0348	35
CH261-192C19	0.4409	-	-
CH261-169N6	0.4409	0.0227	26
CH261-172N3	0.4409	0.0367	46
CH261-177K1	0.4409	0.0386	36
CH261-186J5	0.4409	0.0375	24
CH261-40G6	0.4409	0.0416	26
CH261-50C15	0.4409	0.0382	28
CH261-123O22	0.4409	0.0324	30
CH261-169E4	0.4409	0.0378	33
CH261-1J20	0.4409	0.0370	32
CH261-44D16	0.4409	0.0436	37
CH261-44H14	0.4409	0.0537	17
CH261-17J16	0.4409	0.0446	31
CH261-115J5	0.4409	0.0318	23
CH261-18I9	0.4409	0.0417	26
CH261-130M12	0.4409	0.0641	37
CH261-160I6	0.4409	0.0588	29
CH261-97P20	0.4409	0.0531	26
CH261-17B14	0.4409	0.0497	33
CH261-169K18	0.4409	0.0417	31
CH261-120H23	0.4409	0.0420	32
CH261-183B15	0.4409	-	-

BAC Clone	Rock Dove FLpter Value	*S.D*	n
CH261-83E1	0.4409	N/A	N/A
CH261-71L6	0.4409	N/A	N/A
CH261-111A15	0.4409	N/A	N/A
CH261-18C6	0.4409	0.0509	34
CH261-93H1	0.4409	0.0510	37
CH261-185L11	0.4409	0.0478	20
CH261-85H10	0.4409	0.0493	32
CH261-89P6	0.4409	0.0609	32
CH261-73F2	0.4409	-	-
CH261-49B22	0.4409	0.0545	26
CH261-122F8	0.4409	0.0488	29
CH261-78F13	0.4409	0.0418	19
CH261-2I23	0.4409	0.0686	30
CH261-161B22	0.4409	0.0397	28
CH261-94G14	0.4409	0.0427	14
CH261-67H5	0.4409	0.0389	28
CH261-165L8	0.4409	0.0464	29
CH261-49F3	0.4409	0.0488	20
CH261-179F2	0.4409	-	-
CH261-112D24	0.4409	0.0351	26
CH261-56K7	0.4409	0.0273	21
CH261-180H18	0.4409	0.0498	33
CH261-186K14	0.4409	0.0496	20
CH261-38E18	0.4409	-	-
CH261-69H1	0.4409	0.0574	22
CH261-107D8	0.4409	0.0475	26
CH261-34H16	0.4409	0.0585	28
CH261-96D24	0.4409	0.0441	13
CH261-183N19	0.4409	0.0529	20
CH261-95N3	0.4409	0.0447	17
CH261-187M16	0.4409	0.0661	21
CH261-68O18	0.4409	0.0539	22
CH261-129A16	0.4409	0.0483	23
CH261-133M4	0.4409	0.0480	32
CH261-137F19	0.4409	0.0512	17

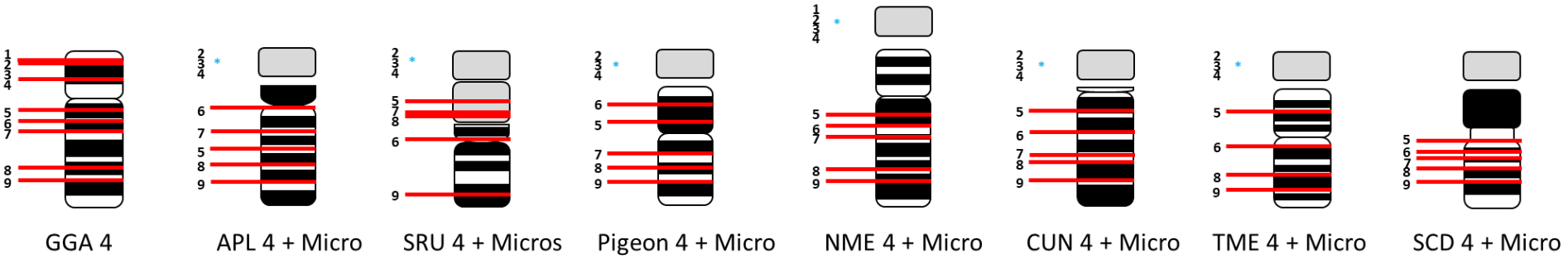
Supplementary Figure S1: Ideograms indicating relative hybridisation positions of BACs for chicken chromosome 2, with BACs labelled in order of position on the chicken chromosome. BAC positions are indicated for chicken (GGA), mallard (APL), Eurasian woodcock (SRU), pigeon (CLI), helmeted guinea fowl (NME), houbara bustard (CUN), common blackbird (TME), and Atlantic canary (SCD).



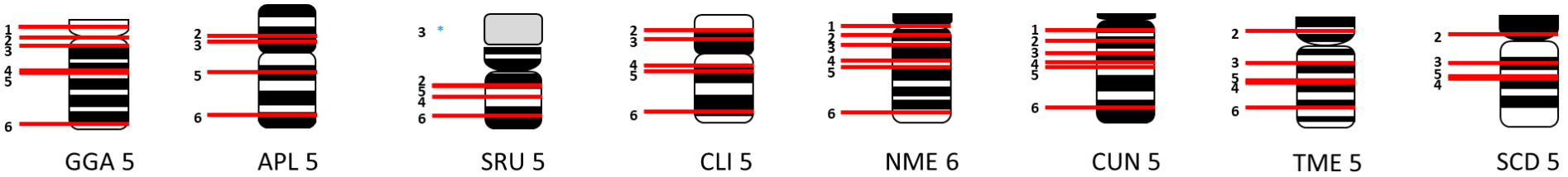
Supplementary Figure S2: Ideograms indicating relative hybridisation positions of BACs for chicken chromosome 3, with BACs labelled in order of position on the chicken chromosome. BAC positions are indicated for chicken (GGA), mallard (APL), Eurasian woodcock (SRU), pigeon (CLI), helmeted guinea fowl (NME), houbara bustard (CUN), common blackbird (TME), and Atlantic canary (SCD).



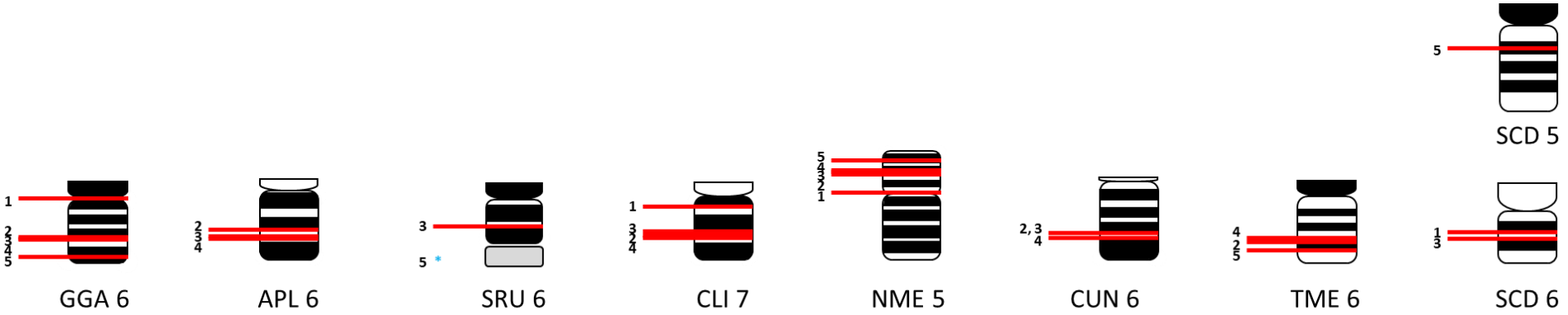
Supplementary Figure S3: Ideograms indicating relative hybridisation positions of BACs for chicken chromosome 4, with BACs labelled in order of position on the chicken chromosome. BAC positions are indicated for chicken (GGA), mallard (APL), Eurasian woodcock (SRU), pigeon (CLI), helmeted guinea fowl (NME), houbara bustard (CUN), common blackbird (TME), and Atlantic canary (SCD).



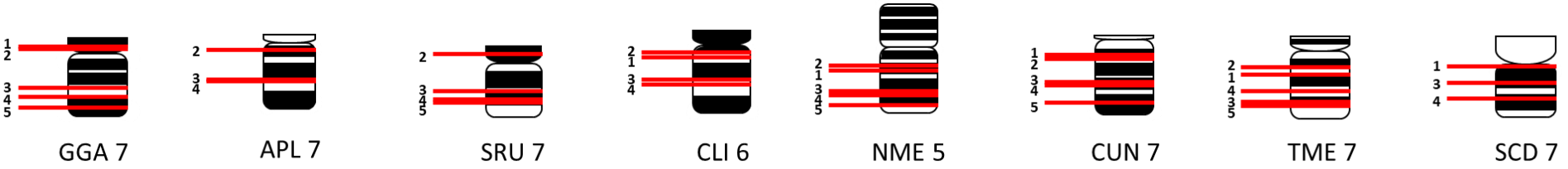
Supplementary Figure S4: Ideograms indicating relative hybridisation positions of BACs for chicken chromosome 5, with BACs labelled in order of position on the chicken chromosome. BAC positions are indicated for chicken (GGA), mallard (APL), Eurasian woodcock (SRU), pigeon (CLI), helmeted guinea fowl (NME), houbara bustard (CUN), common blackbird (TME), and Atlantic canary (SCD).



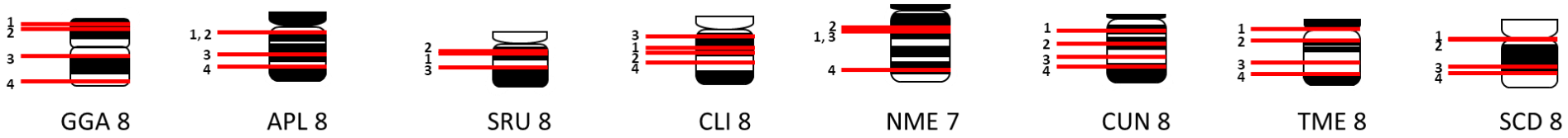
Supplementary Figure S5: Ideograms indicating relative hybridisation positions of BACs for chicken chromosome 6, with BACs labelled in order of position on the chicken chromosome. BAC positions are indicated for chicken (GGA), mallard (APL), Eurasian woodcock (SRU), pigeon (CLI), helmeted guinea fowl (NME), houbara bustard (CUN), common blackbird (TME), and Atlantic canary (SCD).



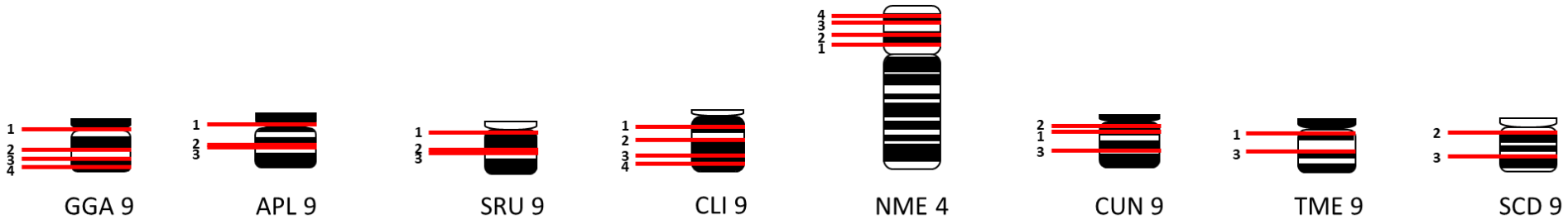
Supplementary Figure S6: Ideograms indicating relative hybridisation positions of BACs for chicken chromosome 7, with BACs labelled in order of position on the chicken chromosome. BAC positions are indicated for chicken (GGA), mallard (APL), Eurasian woodcock (SRU), pigeon (CLI), helmeted guinea fowl (NME), houbara bustard (CUN), common blackbird (TME), and Atlantic canary (SCD).



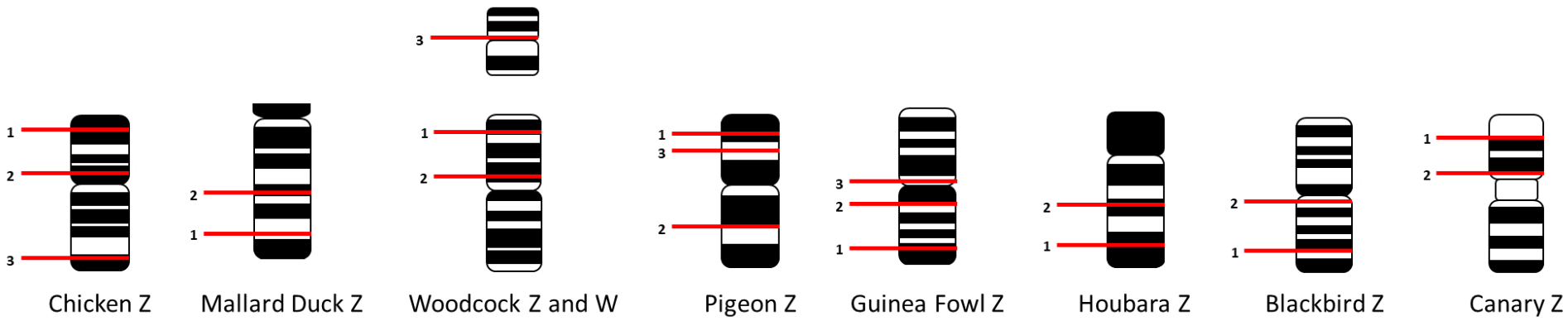
Supplementary Figure S7: Ideograms indicating relative hybridisation positions of BACs for chicken chromosome 8, with BACs labelled in order of position on the chicken chromosome. BAC positions are indicated for chicken (GGA), mallard (APL), Eurasian woodcock (SRU), pigeon (CLI), helmeted guinea fowl (NME), houbara bustard (CUN), common blackbird (TME), and Atlantic canary (SCD).



Supplementary Figure S8: Ideograms indicating relative hybridisation positions of BACs for chicken chromosome 9, with BACs labelled in order of position on the chicken chromosome. BAC positions are indicated for chicken (GGA), mallard (APL), Eurasian woodcock (SRU), pigeon (CLI), helmeted guinea fowl (NME), houbara bustard (CUN), common blackbird (TME), and Atlantic canary (SCD).



Supplementary Figure S9: Ideograms indicating relative hybridisation positions of BACs for chicken chromosome Z, with BACs labelled in order of position on the chicken chromosome. BAC positions are indicated for chicken (GGA), mallard (APL), Eurasian woodcock (SRU), pigeon (CLI), helmeted guinea fowl (NME), houbara bustard (CUN), common blackbird (TME), and Atlantic canary (SCD).



Supplementary Table S14: BACs involved in each chromosomal rearrangement and the corresponding p-values for the mallard duck.

BAC Clone	Chicken Chromosome	FLpter Value for the Mallard Duck	Standard Deviation	Standard Error of Mean	p Value
CH261-36B5	1	0.4084	0.0388	0.0079	0.0003
CH261-25P18	1	0.3632	0.0462	0.0080	
CH261-25P18	1	0.3632	0.0462	0.0080	<0.0001
CH261-125F1	1	0.4467	0.0510	0.0111	
CH261-9B17	1	0.6984	0.0499	0.0070	<0.0001
CH261-168O17	1	0.7516	0.0372	0.0063	
CH261-168O17	1	0.7516	0.0372	0.0063	0.8706
CH261-83O13	1	0.7532	0.0443	0.0072	
CH261-160I6	3	0.4134	0.0397	0.0075	0.0197
CH261-97P20	3	0.3908	0.0352	0.0060	
CH261-18C6	4	0.5235	0.0574	0.0108	<0.0001
CH261-93H1	4	0.1765	0.0476	0.0087	
CH261-93H1	4	0.1765	0.0476	0.0087	<0.0001
CH261-185L11	4	0.3778	0.0637	0.0130	

Supplementary Table S15: BACs involved in each chromosomal rearrangement and the corresponding p-values for the Eurasian woodcock.

BAC Clone	Chicken Chromosome	FLpter Value for the Eurasian Woodcock	Standard Deviation	Standard Error of Mean	p Value
CH261-78F13	5	0.5947	0.0548	0.0125	<0.0001
CH261-2I23	5	0.4710	0.0663	0.0191	
CH261-69H1	8	0.3989	0.0393	0.0095	0.0360
CH261-107D8	8	0.3637	0.0583	0.0116	

Supplementary Table S16: BACs involved in each chromosomal rearrangement and the corresponding p-values for the rock dove.

BAC Clone	Chicken Chromosome	FLpter Value for the Rock Dove	Standard Deviation	Standard Error of Mean	p Value
CH261-36B5	1	0.3750	0.0423	0.0088	0.0002
CH261-25P18	1	0.3257	0.0480	0.0081	
CH261-118M1	1	0.6497	0.0417	0.0069	<0.0001
CH261-18J16	1	0.5778	0.0424	0.0100	
CH261-18J16	1	0.5778	0.0424	0.0100	<0.0001
CH261-29N14	1	0.5135	0.0378	0.0069	
CH261-50C15	2	0.3377	0.0382	0.0072	0.0059
CH261-123O22	2	0.3112	0.0324	0.0059	
CH261-18C6	4	0.2981	0.0509	0.0087	<0.0001
CH261-93H1	4	0.1532	0.0510	0.0084	
CH261-67H5	6	0.6742	0.0389	0.0073	0.0002
CH261-165L8	6	0.6290	0.0464	0.0086	
CH261-112D24	7	0.3339	0.0351	0.0070	0.0001
CH261-56K7	7	0.2721	0.0624	0.0136	
CH261-69H1	8	0.4589	0.0573	0.0122	<0.0001
CH261-107D8	8	0.5386	0.0475	0.0093	

Supplementary Table S17: BACs involved in each chromosomal rearrangement and the corresponding p-values for the helmeted guinea fowl

BAC Clone	Chicken Chromosome	FLpter Value for the Helmeted Guinea Fowl	Standard Deviation	Standard Error of Mean	p Value
CH261-119K2	1	0.1757	0.0284	0.0049	0.0337
CH261-120J2	1	0.1597	0.0323	0.0053	
CH261-112D24	7	0.6040	0.0420	0.0078	0.0004
CH261-56K7	7	0.5630	0.0534	0.0065	
CH261-69H1	8	0.3486	0.0488	0.0086	0.0001
CH261-107D8	8	0.2971	0.0416	0.0085	

Supplementary Table S18: BACs involved in each chromosomal rearrangement and the corresponding p-values for the houbara bustard

BAC Clone	Chicken Chromosome	FLpter Value for the Houbara Bustard	Standard Deviation	Standard Error of Mean	p Value
CH261-130M12	3	0.4273	0.0503	0.0096	0.1404
CH261-160I6	3	0.4043	0.0589	0.0120	
CH261-160I6	3	0.4043	0.0589	0.0120	<0.0001
CH261-97P20	3	0.2770	0.0502	0.0091	
CH261-183N19	9	0.3050	0.0604	0.0129	<0.0001
CH261-95N3	9	0.2111	0.0500	0.0096	

Supplementary Table S19: BACs involved in each chromosomal rearrangement and the corresponding p-values for the common blackbird.

BAC Clone	Chicken Chromosome	FLpter Value for the Common Blackbird	Standard Deviation	Standard Error of Mean	p Value
CH261-125F1	1	0.3990	0.0601	0.0109	<0.0001
CH261-118M1	1	0.1157	0.0424	0.0080	
CH261-18J16	1	0.9258	0.0379	0.0070	<0.0001
CH261-29N14	1	0.8522	0.0627	0.0116	
CH261-29N14	1	0.8522	0.0627	0.0116	<0.0001
CH261-9B17	1	0.7618	0.0610	0.0106	
CH261-9B17	1	0.7618	0.0610	0.0106	0.8002
CH261-168O17	1	0.7582	0.0495	0.0090	
CH261-168O17	1	0.7582	0.0495	0.0090	0.0003
CH261-83O13	1	0.7052	0.0581	0.0098	
CH261-83O13	1	0.7052	0.0581	0.0098	0.1203
CH261-107E2	1	0.6760	0.0757	0.0173	
CH261-107E2	1	0.6760	0.0757	0.0173	<0.0001
CH261-58K12	1	0.5725	0.0637	0.0109	
CH261-58K12	1	0.5725	0.0637	0.0109	0.0682
CH261-184E5	1	0.5443	0.0660	0.0106	
CH261-184E5	1	0.5443	0.0660	0.0106	<0.0001
CH261-98G4	1	0.457	0.0791	0.0139	
CH261-169N6	2	0.0729	0.0384	0.0075	0.7767
CH261-172N3	2	0.0695	0.0464	0.0087	
CH261-78F13	5	0.5814	0.0525	0.0131	0.2172
CH261-2I23	5	0.5572	0.0639	0.0130	
CH261-67H5	6	0.7334	0.0764	0.0142	0.0124

BAC Clone	Chicken Chromosome	FLpter Value for the Common Blackbird	Standard Deviation	Standard Error of Mean	p Value
CH261-49F3	6	0.6799	0.0795	0.0150	
CH261-112D24	7	0.4481	0.0705	0.0195	0.0051
CH261-56K7	7	0.3766	0.0743	0.0133	
CH261-180H18	7	0.7921	0.0558	0.0149	<0.0001
CH261-186K14	7	0.6627	0.0662	0.0191	

Supplementary Table S20: BACs involved in each chromosomal rearrangement and the corresponding p-values for the Atlantic canary.

BAC Clone	Chicken Chromosome	FLpter Value for the Atlantic Canary	Standard Deviation	Standard Error of Mean	p Value
CH261-125F1	1	0.4254	0.0803	0.0141	<0.0001
CH261-118M1	1	0.2238	0.1536	0.0215	
CH261-18J16	1	0.8936	0.0579	0.0100	<0.0001
CH261-29N14	1	0.8252	0.0626	0.0109	
CH261-29N14	1	0.8252	0.0626	0.0109	<0.0001
CH261-9B17	1	0.7362	0.0578	0.0102	
CH261-9B17	1	0.7362	0.0578	0.0102	0.5529
CH261-168O17	1	0.7303	0.0627	0.0109	
CH261-168O17	1	0.7303	0.0627	0.0109	0.0357
CH261-83O13	1	0.6943	0.0680	0.0128	
CH261-83O13	1	0.6943	0.0680	0.0128	0.0199
CH261-107E2	1	0.6544	0.0648	0.0108	
CH261-107E2	1	0.6544	0.0648	0.0108	0.4723
CH261-58K12	1	0.6422	0.0579	0.0123	
CH261-58K12	1	0.6422	0.0579	0.0123	<0.0001
CH261-184E5	1	0.5362	0.0640	0.0115	
CH261-184E5	1	0.5362	0.0640	0.0115	<0.0001
CH261-98G4	1	0.4385	0.0747	0.0181	
CH261-78F13	5	0.5654	0.0653	0.0146	0.0289
CH261-2I23	5	0.4978	0.1187	0.0247	

10 Publications Arising from This Work

Original Research Manuscripts:

O'Connor, R.E., Farré, M., Joseph, S., Damas, J., **Kiazim, L.,** *et al.* (2018). Chromosome-level assembly reveals extensive rearrangement in saker falcon and budgerigar, but not ostrich, genomes. *Genome Biology*

O'Connor R.E., Romanov M.N, **Kiazim, L.,** *et al.* (2018). Reconstruction of the diapsid ancestral genome permits chromosome evolution tracing in avian and non-avian dinosaurs. *Nature Communications*

nature COMMUNICATIONS

ARTICLE

DOI: 10.1038/s41467-018-04267-9 OPEN

Reconstruction of the diapsid ancestral genome permits chromosome evolution tracing in avian and non-avian dinosaurs

Rebecca E. O'Connor [1], Michael N. Romanov [1], Lucas G. Kiazim [1], Paul M. Barrett [2], Marta Farré [3], Joana Damas [3], Malcolm Ferguson-Smith[4], Nicole Valenzuela [5], Denis M. Larkin [3] & Darren K. Griffin [1]

Genomic organisation of extinct lineages can be inferred from extant chromosome-level genome assemblies. Here, we apply bioinformatic and molecular cytogenetic approaches to determine the genomic structure of the diapsid common ancestor. We then infer the events that likely occurred along this lineage from theropod dinosaurs through to modern birds. Our results suggest that most elements of a typical 'avian-like' karyotype (40 chromosome pairs, including 30 microchromosomes) were in place before the divergence of turtles from birds ~255 mya. This genome organisation therefore predates the emergence of early dinosaurs and pterosaurs and the evolution of flight. Remaining largely unchanged interchromosomally through the dinosaur–theropod route that led to modern birds, intrachromosomal changes nonetheless reveal evolutionary breakpoint regions enriched for genes with ontology terms related to chromatin organisation and transcription. This genomic structure therefore appears highly stable yet contributes to a large degree of phenotypic diversity, as well as underpinning adaptive responses to major environmental disruptions via intrachromosomal repatterning.

[1] School of Biosciences, University of Kent, Canterbury, Kent CT2 7NJ, UK. [2] Department of Earth Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK. [3] Department of Comparative Biomedical Sciences, Royal Veterinary College, University of London, London NW1 0TU, UK. [4] Department of Veterinary Medicine, Cambridge University, Cambridge CB3 0ES, UK. [5] Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Iowa, IA 50011, USA. These authors contributed equally: Denis M. Larkin, Darren K. Griffin. These authors jointly supervised this work: Denis M. Larkin, Darren K. Griffin. Correspondence and requests for materials should be addressed to D.K.G. (email: d.k.griffin@kent.ac.uk)

In the absence of cellular material and DNA from biological samples of long-extinct, early diverging lineages, data from genome sequence assemblies of extant species can nonetheless facilitate the reconstruction of gross genome structures (karyotypes). This can be achieved provided those assemblies are at, or close to, chromosome level, i.e. one scaffold per chromosome[1]. In a previous study, we analysed (close to) chromosome-level assemblies from six extant birds (and a lizard outgroup) to determine the most likely karyotype of the neornithine ancestor for the macrochromosomes and the neognathe ancestor for the microchromosomes[2]. Recreating the most parsimonious sequence of events that might have led to contemporary genome structures (karyotypes), we determined that chicken (*Gallus gallus*) was the closest karyotypically to the reconstructed ancestral pattern, with zebra finch (*Taeniopygia guttata)* and budgerigar (*Melopsittacus undulatus)* undergoing the most intra- and interchromosomal rearrangements, respectively. In the current study, to reconstruct the most likely karyotype of the diapsid common ancestor (DCA >255 mya), we applied similar approaches, i.e. the multiple-genome rearrangement and analysis (MGRA2) tool. We focussed on the best-quality chromosome-level assemblies of avian and reptilian genomes and a mammalian outgroup. Supplementing bioinformatic data with novel molecular cytogenetic approaches on turtle metaphases, we tested the hypothesis that the typical karyotype seen in neornithine birds underwent few interchromosomal rearrangements since the divergence of turtles from archosaurs (birds and crocodilians) <255 mya. Combining both sets of data, we thence inferred the most parsimonious sequence of events that occurred from the diapsid ancestor, to the archelosaur ancestor[3], and thence via non-avian theropod dinosaurs to extant birds (see Supplementary Note 1 for divergence times).

Studies of the best-assembled genomes (including chromosome level) also indicate that evolutionary breakpoint regions (EBRs) lie in gene-dense loci, enriched with genes related to lineage-specific biology, transposable elements and other repetitive sequences[4–7], while sequences that stay together during evolution (homologous synteny blocks, HSBs) are enriched in developmental genes and regulatory elements[7]. While a contribution of random breakage in chromosome evolution[9] cannot be excluded for all (especially smaller) rearrangements that might have neutral effects on phenotypes, multiple evidence has accumulated to suggest that at least the largest HSBs and some EBRs in animal genomes are maintained non-randomly[4,5,10,11]. Differences in the composition of their DNA features suggest that, although chromosome aberrations in germ cells may indeed occur in regions more prone to breakage (e.g. recombination hotspots or open chromatin areas), those breaks not disturbing essential genes or providing a selective advantage will more likely be fixed in populations, becoming EBRs[4]. In the current study, we investigated gene content of those EBRs and HSBs identified as being involved in the karyotypic changes from the diapsid ancestor to modern birds and identified genes that may indicate adaptive (EBRs) or conserved (HSBs) phenotypic features or those likely to be involved in gross genomic rearrangement.

Here, we analysed data from the genome assemblies of Carolina anole lizard (*Anolis carolinensis)*[6], chicken (*G. gallus)*[7,12], mallard (*Anas platyrhynchos*)[8], zebra finch (*T. guttata*)[13], and grey short-tailed opossum (*Monodelphis domestica*)[14]. All of these species have robust chromosome-level assembled genomes. Moreover, *M. domestica* has a karyotypic structure thought to resemble the mammalian ancestor most closely[14]. Among other genome assemblies that might have proved useful in our analyses, those generated from alligators and turtles[15,16] were discounted as they are too fragmented, i.e. not close to chromosome level. Also, near-chromosome-level assemblies for turkey, budgerigar and ostrich were ultimately excluded because our cytogenetic studies in these species (not shown) revealed that the level of fragmentation or misassembles in these genomes had the potential to introduce false breakpoint regions (Supplementary Note 2). Finally, we discounted crocodilians from our molecular analysis, partly because of a relative lack of fluoresence in situ hybridisation (FISH) success of multiple attempts, and partly because all crocodilian species studied have an atypical archelosaur karyotype with no microchromosomes, mostly brought about by fusion[17].

Our results suggest that most features of a typical 'avian-like' karyotype were in place before the divergence of birds and turtles, that the predominant mechanism of change thereafter was intrachromosomal rearrangement and that EBRs were enriched for GO terms associated with chromatin modification and chromosome organisation.

Results

Summary of results. Pairwise sequence alignments permitted visualisation of 397 multispecies HSBs, their orientation in each genome and EBRs between them (listed in Supplementary Data 2). Using MGRA2, we generated 19 contiguous ancestral regions (CARs), roughly correlating to chromosomes, in the most likely ancestral karyotype of the DCA. CAR sizes are given in Supplementary Data 1 and the analysis pipeline is described in Methods. Our analysis of chromosomes from the turtle genome with one of the largest chelonian diploid numbers ($2n$) (spiny soft-shelled turtle (*Apalone spinifera)* ($2n = 66$)) revealed little or no evidence of interchromosomal differences between this species and chicken. That is, chicken-derived fluorescent probes highly selected to hybridise across large evolutionary distances[5] plus chromosome painting data provided evidence that most chicken chromosomes 1–28 + Z are each represented by a single-turtle counterpart (see Methods and Supplementary Data 3). Successful hybridisation to the chromosomes of red-eared slider (*Trachemys scripta*) ($2n = 50$) revealed a karyotype with microchromosomal homeologues either having fused to macrochromosomes or, more likely, having retained the ancestral state of the DCA. Indeed, one of the main technical advances made in this study was the isolation of a probe set that would hybridise directly across species that diverged ~255 mya. More limited success in hybridising to *A. carolinensis* metaphases ($2n = 36$) (Fig. 1) revealed some broad similarities to the DCA established by the bioinformatic approach (see Supplementary Data 3).

FISH analysis suggests avian and turtle chromosomes are precise counterparts. Cross-species hybridisation (zoo-FISH) of chicken bacterial artificial chromosome (BAC) probes previously designed to work in multiple avian species and located subtelomerically[5], as well as chicken chromosome paints were successfully hybridised to the chromosomes of *A. carolinensis* (Carolina anole lizard), as well as two turtles *T. scripta* and *A. spinifera*. Results, for the most part, provide evidence that, for avian chromosomes 1–28 + Z (with rare exceptions, e.g. 16 for which we could not generate data), each chicken chromosome is syntenic to the turtle with the largest diploid number, *A. spinifera* ($2n = 66$). That is, we found little or no evidence of interchromosomal rearrangement, with the exception of rare events, e.g. fusions of chromosome 4 in chicken and chromosome 22 in turtles. For the macrochromosomes and some pools of microchromosomes, chromosome paints[18] produced signals cross-species and, for the microchromosomes, selected BAC probes[5] provided strong BAC signals (note, macrochromosomal BACs,

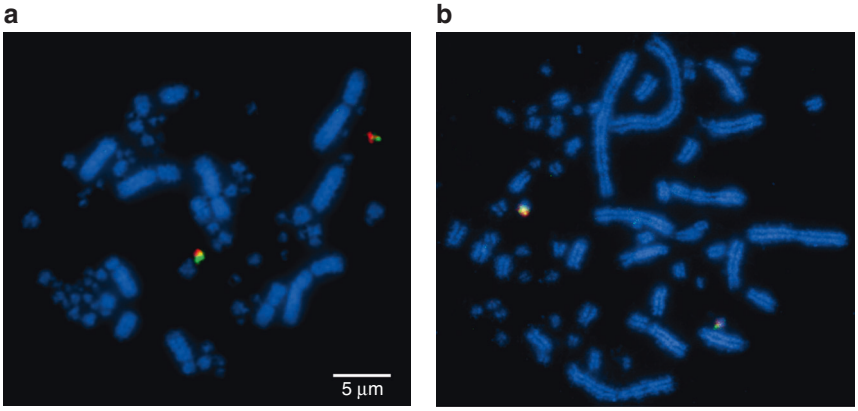


Fig. 1 Cross-species hybridisation (zoo-FISH) results. Hybridisation of chicken chromosome 23 sub-telomeric BACs to **a** chicken ($2n = 78$) and **b** turtle (*Apalone spinifera* $2n = 66$) metaphases. This is an example of how most chromosomes studied in birds and turtles (species examined with the highest diploid number) are precise counterparts of one another. All syntenic chromosomes were of similar sizes and morphologies. All chromosomes are labelled in blue (DAPI) with BAC probes labelled in red (Texas Red) and green (FITC), respectively (where signals overlap a little, a yellow/orange colour is seen). Successful strong hybridisation across large evolutionary distances was one of the technical advances of this project. Scale bar applies to both images

despite working well on other avian species, did not produce many successful hybridisations on non-avian reptiles). Specifically, of the microchromosomal BACs, 29 of the original 36 (81%) worked successfully (chromosomes 10–15 and 17–28; 16 not included due to lack of sequence coverage and chromosomes smaller than 28 do not have sequences assigned to them) in both turtle species (all but chromosome 20 had at least one BAC signal) and 17 of 36 (47%) worked in Carolina anole lizard (chromosomes 14, 18, 20, 25, 27 and 28 were not represented). Of those that worked on all species, results revealed that the orthologues of chicken chromosomes 12 and 13 were fused and chromosome 26 attached to chromosome 4 in the red-eared slider and Carolina anole but represented as separate microchromosomes in the spiny soft-shelled turtle (*A. spinifera*). The chromosome 22 orthologue appeared as a separate chromosome in the Carolina anole, but as fused to the centre of a macrochromosome in the two turtle species. Eight avian microchromosome orthologues for chromosomes 10, 11, 15, 17, 19, 21, 23 and 24, appeared to be conserved as single microchromosomes in all three reptiles studied, and in red-eared slider (*T. scripta*), chromosomes 14, 18, 25 and 28 were also represented as single microchromosomes, with chromosomes 12, 13 and 26 also as single microchromosomes for spiny soft-shelled turtle. All macrochromosomal assignments confirmed those previously reported for the lizard and for the two turtles. In order to establish that there were no rearrangements between microchromosomes of birds and reptiles, working BACs as close to the telomere as possible were used, and all microchromosomal sizes were measured in comparison to known macrochromosomal size by ImageJ. We cannot preclude the possibility that some interchromosomal events were not detected by our approach; for instance, cryptic translocations could possibly give chromosome paint signals too weak to detect microscopically or be subtelomeric to the BAC used. However the fact that no additional rearrangements other than those already known were detected between *A. spinifera* ($2n = 66$) and avian ($2n = 80$) karyotypes supports our central assertion of identity by descent, in place ~255 mya, with ~7 fissions (presumably at least some, but possibly all) occurring before the emergence of the dinosaurs.

Sequence alignments, multispecies HSBs and EBRs. Pairwise bioinformatic alignments of the chicken, mallard, zebra finch, Carolina anole and grey short-tailed opossum genomes using the Evolution Highway chromosome browser allowed for the visualisation of multispecies HSBs (msHSBs) and their orientation in each genome as well as the identification of EBRs between these msHSBs. These alignments were screened for blocks shared in the five genomes compared, and a total of 397 msHSBs were found. These were distributed across 19 of the 28 sequenced chicken chromosomes available on Evolution Highway: i.e. chromosomes 1–9, 11–13, 15, 18, 20, 24, 26, 27 and Z. The 397 msHSBs were also dispersed on 19 duck chromosomes, 21 zebra finch chromosomes, 10 anole chromosomes, and 8 opossum chromosomes. If we compare the total size of 397 msHSBs relative to the chicken 28 chromosomes available on Evolution Highway, they represented 49% of the total genome length; if we compare these msHSBs to the size of the above 19 chicken chromosomes, they represented about 53% of their combined length.

Using the MGRA2 algorithm, we produced a series of CARs representing the most likely ancestral karyotype of the diapsid common ancestor (DCA). While we cannot be entirely sure that each CAR represents a whole chromosome as MGRA2 will inherently 'break' the chromosome if it cannot find synteny, a total of 19 diapsid ancestral CARs were found, probably representing fewer chromosomes. The number of msHSBs per CAR varied between 2 and 59. A total of 17 chicken chromosomes were aligned to these CARs (chromosomes 1–8, 11–13, 15, 18, 24, 26, 27 and Z) meaning that some microchromosomes were not represented.

Chromosome inversions from DCA to chicken. Reconstructed CARs derived from MGRA2 were subsequently mapped to the extant genomes. The rearrangements between the DCA and chicken were then modelled using maximum parsimony. A total of 49 inversions were identified between the DCA and the chicken genome along with 10 interchromosomal changes (see Supplementary Note 3). Of the interchromosomal rearrangements found, a translocation was identified between orthologues of chicken chromosomes 5 and 20, consistent with the FISH results (Fig. 2, Supplementary Fig. 1). Comparison of the data generated for the DCA and a putative archelosaur common ancestor revealed that the majority of interchromosomal rearrangements occurred to form this basic structure with most intrachromosomal rearrangements (inversions) after (for simplicity, on Fig. 2, all intrachromosomal changes are shown after formation of the basic (archosauromorph common ancestor) pattern) although we cannot rule out the

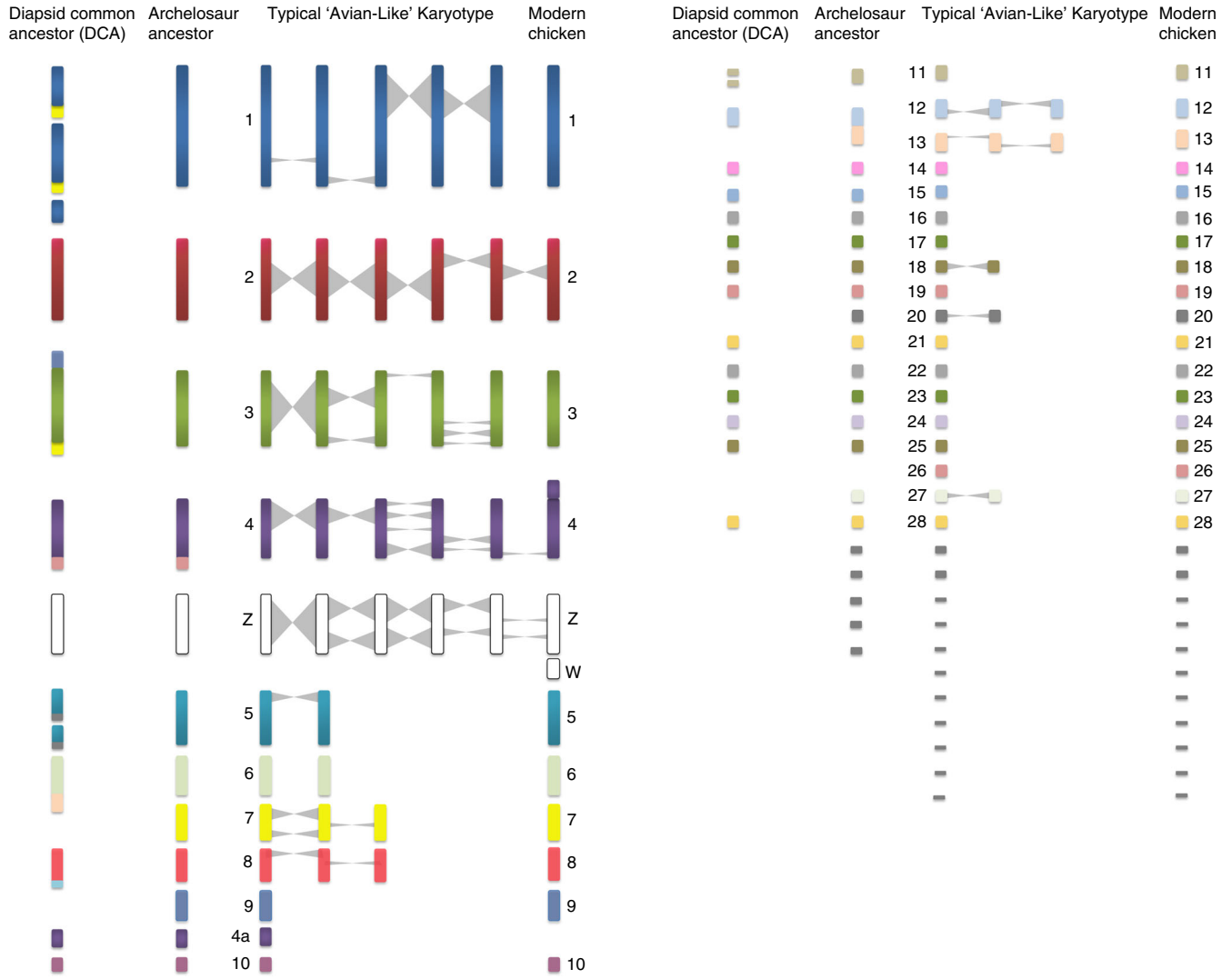


Fig. 2 Overall inferred karyotypic changes from the diapsid common ancestor (DCA), through the archosauromorph ancestor, modern chicken ($2n = 78$) arising out of theropod dinosaurs, collating all available lines of evidence, bioinformatic and molecular cytogenetic. The left side shows the 19 CARs including the ancestral microchromosomes, which gives an impression of a higher diploid number greater than 46 (total chromosome number predicted) as we cannot necessarily determine the nature of all the fusions. For simplicity, the intrachromosomal rearrangements (inversions) are all depicted after the archelosaur ancestor, however, a small proportion may have occurred before. The colour scheme is randomly assigned with each chromosome for which we have data given a different colour. Chromosomes for which we do not have FISH data are depicted in grey (e.g. chromosomes 16, 20, 22 and the very smallest microchromosomes). Diapsid common ancestor (DCA): 255 mya; likely $2n = 36$–46; 19 CARs identified (some CARs likely fused as single chromosomes, hence this diagram appears as apparently more chromosomes; some chromosomes not covered by sequence assembly; Chromosome 7 orthologue (yellow) attached to three chromosomes, chromosome 26 (dark pink) to the orthologue of chromosome 4). Archelosaur ancestor: <255 mya; likely ~$2n = 66$; most chromosomes syntenic to modern birds; chromosomes without direct evidence (including chromosome 16, 20, 22) depicted in grey. Typical 'Avian-Like' karyotype: likely ~$2n = 80$ (numbering according to chicken genome); ~7 fissions; some inversions may have occurred before the pattern was established interchromosomally, but intrachromosomal changes shown separately for clarity. Modern chicken: $2n = 78$; known fusion of chromosome 4 shown (not present in most birds); sex chromosome evolution post Palaeognathae–Neognathae divergence

possibility that some occurred before. Between the DCA and the archosauromorph common ancestor, a fusion most likely occurred to form chromosome 1 and translocations/fissions occurred between avian ancestral CARs that became chromosome 7). Most rearrangements between these two ancestors were nonetheless intrachromosomal with a total of 49 inversions that appear to have occurred between the DCA and the extant chicken genome.

Enrichment of gene ontology terms in msHSBs. Within the 397 multispecies HSBs (msHSBs), significant enrichments were observed for gene ontology (GO) terms relevant to amino acid transmembrane transport (symport; Supplementary Data 4) and signalling (group enrichment scores, 2.44–2.50 as produced using DAVID Functional Annotation Clustering tool; single GO terms as produced using DAVID Functional Annotation Chart, $p < 0.05$, false-discovery rate (FDR) < 5%). Other msHSB-specific GO term enrichments were related to synapse/neurotransmitter transport, nucleoside metabolism and use, cell morphogenesis and cytoskeleton and sensory organ development (group enrichment scores, 1.72–3.02––Supplementary Data 4). For the functional analysis of EBRs, we produced a set of 234 EBRs that were intrinsic to the archosaur common ancestor. Within these, we identified significant enrichments in genes and single GO terms relevant to

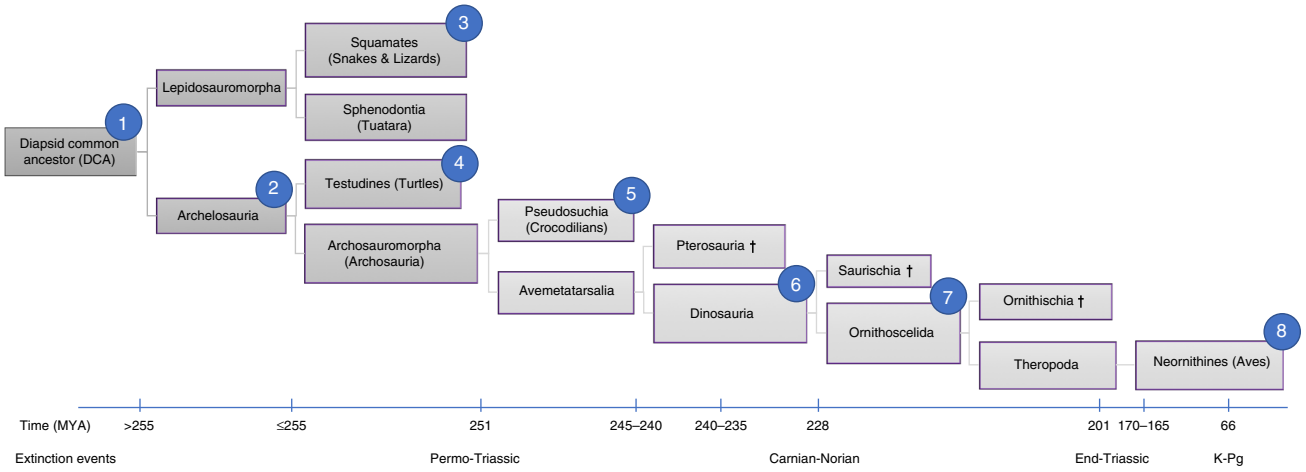


Fig. 3 Representative phylogenetic tree illustrating the lineage investigated in this analysis. For ease of reading, the timelines are not to scale. The major extinction events and the principal findings of this study are highlighted in the context of this phylogeny as follows. (1) DCA karyotype reconstructed by MGRA2 analysis. (2) Cytogenetic analysis revealed the basic 'avian' pattern $2n = 66$ mostly in place before archelosauria divergence. (3) One lizard (*A. caroliensis*) studied genomically and cytogenetically. (4) Two turtles (*T. scripta* and *A. spinifera*) studied cytogenetically. (5) Crocodilian genomes not considered suitable for analysis because of fragmented genome assemblies and fused chromosomes. (6) Early emerging dinosaurs and pterosaurs probably had at least $2n = 66$ and up to $2n = 80$ with typical avian pattern. (7) Theropod dinosaurs and possibly other groups probably had close to $2n = 80$ with typical avian pattern. (8) Three avian species $2n = 78$–80 studied cytogentically and using chromosomal-level genome assemblies. † extinct lineage

chromatin modification and chromosome organisation, as well as proteasome/signalosome structure ($p < 0.01$, FDR < 5%, Supplementary Data 5). In particular, the first annotation cluster (enrichment score, 4.30) consisted of six genes related to proteasome/signalosome structure within EBRs located on six chicken chromosomes. The second annotation cluster (enrichment score, 1.64) included 15 genes relevant to chromatin modification within 12 EBRs on seven chicken chromosomes, these genes and their functions are listed in Supplementary Data 5.

Discussion

Combining all lines of available evidence (bioinformatic and molecular cytogenetic from this study along with previous findings), a picture emerges of an inferred DCA karyotype with a chromosome number of $2n = 36$–46, consistent with that previously proposed[19] and with most other non-avian, non-testudinate reptiles. Roughly half would have been macro- and half microchromosomes[6,19]. Alföldi et al.[6] found direct synteny at the sequence level between the microchromosomes of *A. carolinensis* and *G. gallus*, with all but one lizard microchromosome corresponding to a single chicken microchromosome. Given that the *A. carolinensis* genome has 12 microchromosomal pairs (that are mostly syntenic to chicken chromosomes[6]) compared to the 28–30 seen in most birds, the most likely explanation is that these were present in the DCA, the remainder evolving thereafter by fission[6] to at least $2n = 66$. Our current data suggest that interchromosomal rearrangement largely ceased thereafter, with the exception of ~7 fissions that explain the difference between the pattern in *A. spinifera* ($2n = 66$) and most modern birds ($2n\sim = 80$). Indeed, even the karyotype of *T. scripta* ($2n = 50$) has broad similarities to the avian pattern, but with more microchromosomal homeologues attached to larger chromosomes. This would either indicate that *T. scripta* has a karyotype that represents an earlier stage of differentiation to the "bird-like" turtle pattern, or that it subsequently underwent a series of fusions (such as in the crocodilians); the former being the most likely since it requires fewer events. The considerable range of diploid numbers in turtles ($2n = 26$ to $2n = 68$), with most being more 'avian-like' than their other reptilian counterparts, suggests that further study of this group will provide greater insight into the sequence of events that led to the establishment of a highly successful mode of genome organisation. We cannot preclude the possibility that some interchromosomal rearrangements were beyond the sensitivity of our detection, e.g. the weak chromosome painting signals may not have detected subtle changes, or cryptic chromosome translocations may have occurred that were telomeric to our fluorescent probes. This does not, however, detract from our assertions that a broad overall pattern of genome organisation was in place at least 255 mya in the archelosaur ancestor and changed little in the majority of living species. Moreover, our findings are consistent with previous studies using chicken macrochromosome paints on Chinese soft-shelled turtle (*Pelodiscus sinensis*) chromosomes ($2n = 66$)[20], *T. scripta*[21] and painted turtle (*Chrysemys picta*) chromosomes (both $2n = 50$)[22]. Thus, similar studies on more turtles may reveal species with greater diploid numbers and a pattern resembling that of birds even more (Figs. 2 and 3).

Determining the precise timing of events that led from the archelosaur ancestor ($2n > 66$) to modern birds ($2n\sim = 80$) is beyond the resolution of this study (Fig. 3). A similar rate of fission continuing beyond 255 mya, would have established a near-complete neornithine-like karyotype by ~240 mya, roughly coinciding with the emergence of the earliest dinosaurs and pterosaurs[23]. Equally, a dramatic slowdown or halt in fission events at ~255 mya would suggest that the forebears of the earliest dinosaurs and pterosaurs had a pattern more similar to that of *A. spinifera*. In both scenarios (or an intermediate), however, a pattern very similar to that of most birds would have been present. Burt[24] proposed that most avian microchromosomes were present in the avian common ancestor >80 mya[25], arguing that it probably had the small genome size characteristic of birds and a karyotype of around $2n = 60$. Our results, however, suggest a much earlier emergence of the typical avian pattern with genome size reduction occurring later in

archosaur or theropod evolution[24]. Indeed Uno et al.[26] proposed that the archosauromorph ancestor may have had microchromosomes similar to the turtle but did not specify their nature.

Our results suggest that, aside from ~7 fissions, the primary mechanism for chromosomal rearrangement in the avian stem lineage after 255 mya was intrachromosomal (e.g. inversions). Reconstructed CARs, when compared to extant genomes, resulted in the identification of rearrangements between the DCA and chicken (*G. gallus)* genomes, modelling 49 inversion events. This, however, is almost certainly an underestimate due to the paucity of sequence coverage in some areas, including the smallest avian microchromosomes. Rates of change are not easy to determine, but there is some evidence of intrachromosomal change speeding up in modern birds, even in the chicken, which is thought to be very similar chromosomally to the avian common ancestor[27]. Increased intrachromosomal change has been reported in specific groups, with several studies suggesting that the greatest rates would be found in passerines (compared to other birds)[1,4,28], the group containing most avian species. It is perhaps reasonable to speculate therefore that bursts of speciation may have also been accompanied by increased rates of intrachromosomal rearrangement in other non-avian dinosaur lineages.

Within the multispecies HSBs, we identified significant enrichments for GO terms relevant to transmembrane transport (symport) and signalling, synapse/neurotransmitter transport, nucleoside metabolism and use, cell morphogenesis and cytoskeleton and sensory organ development (Supplementary Data 4). HSBs are often enriched for GO terms related to phenotypic features that remain constant[29] and the results presented here are consistent with this hypothesis. Sankoff[30] however stated that EBRs are where the 'action' in genome evolution lies and, previously, we found that GO terms in avian EBRs associated with specific adaptive features, e.g. enrichment for forebrain development in the budgerigar EBRs (consistent with vocal-learning)[4]. In the current study, we identified significant enrichments in genes and single GO terms relevant to chromatin modification, chromosome organisation and proteasome/signalosome structure in EBRs (Supplementary Data 5). This illustrates some parallels to recent findings in rodents where chromosomal changes were associated with open chromatin[27]. Most of the 15 genes found in this GO term were related to control of gene expression. Transcription factors modify chromatin by making it accessible during transcription and this is noteworthy because EBRs rearrange transcription factor genes. This might affect expression of other genes of the same pathway. Interestingly, two of these genes showed a different expression pattern between birds and mammals, HDAC8 involved in early embryo development[31] and PRMT8 expressed in brain[32]. These results suggest a correlative link between chromosomal and morphological changes among species, mediated by rearranging genes controlling the expression of developmental pathways.

Estimates of non-avian dinosaur genome size, based on osteocyte sizes inferred from bone histology, identified a distinction between small, characteristically avian, genomes in theropods and sauropodomorphs vs. much larger ornithischian genomes[33]. These results were interpreted as supporting the hypothesis that small genome size and low repeat content was a genomic exaptation that preceded and facilitated the endothermic metabolic demands of birds, e.g. for flight[34]. It was further hypothesised that the avian karyotype evolved in response to a reduction in genome size in birds[34]. This theory was subsequently challenged by a study that suggested that a decline in overall genome size occurred in non-volant dinosaurs[33]. Here, we propose further that an avian-like karyotype not only predated the origin of flight but evolved well before, and independent of any, purported genome size reduction. Nonetheless, we note that there may be an association between genomes with fewer chromosomes (and no microchromosomes) and larger genome sizes around 2.5–3 Gb, as in mammals[35] and crocodilians[15]. More repetitive elements could provide substrates for interchromosomal rearrangement, which is commonplace in mammals but rare in birds, and it has been suggested that an avian karyotype provides fewer opportunities for interchromosomal rearrangement due to the existence of fewer recombination hotspots (despite an overall higher recombination rate)[36,37], repeat structures[12,38,39] and endogenous retroviruses[2,4,40]. Therefore, although flight evolution might be correlated with smaller genome size (consider pterosaurs vs. other avemetatarsalians[41]; bats vs. other mammals[34] and strong vs. weak flying/ flightless birds[42]), other mechanisms are clearly involved. Specifically, the formation of an avian-like karyotype long before the evolution of flight suggests neither a causative, nor a correlative link between the two.

Stasis of this karyotypic structure for >255 million years nonetheless suggests a highly successful mode of genome organisation that might have provided a blueprint for evolutionary success. The reasons for its persistence are speculative but might be due to its ability, facilitated by many chromosomes, including microchromosomes with high recombination rates, to generate variation, the substrate of natural selection. A larger number of small chromosomes inherently generate variation through increased genetic recombination and increased random chromosome segregation. Burt[24] suggested that a higher recombination rate has also contributed to the unique genomic features seen in microchromosomes such as high GC-content, low repeat content and high gene-density, which subsequently led to its maintenance. Variation, in turn, facilitates rapid adaptation and may therefore have contributed to wide phenotypic variation, in extant animals represented by >10,000 species of birds ($2n\sim=80$), >300 species of turtles ($2n<68$) and, quite possibly, a large number of non-avian dinosaurs also. Of course, a karyotype with many, tiny chromosomes is not the only means by which variation can be generated (genic, epigenetic and interchromosomal variation all are mechanisms reported in other groups): indeed, amphibians display enormous phenotypic variation but possess relatively few chromosomes. Nonetheless, the above may explain the apparent paradox of a group with very little interchromosomal change, but incredible phenotypic diversity.

In conclusion, any 'dinosaur genomics' effort of this type is limited to reconstructing common ancestors (e.g. of birds and crocodilians), along with other nodes that have extant descendants (e.g. Archelosauria, Diapsida etc.) and inferring the most parsimonious set of events that led to extant animals. With this in mind, few studies have attempted to infer the nature of the gross structural genomic changes that occurred from the DCA, to the archelosaur ancestor, to birds. Given our data, it is perhaps not an unreasonable speculation that, if we had the opportunity to make metaphase chromosomes from tissue of non-avian theropods, both karyotypic and molecular cytogenetic analysis (genome size aside) would reveal little difference from a modern chicken, duck or ostrich (or at least a spiny soft-shelled turtle), i.e. $2n = 66–80$ in the majority of species. Of course, we cannot preclude the possibility that certain groups of non-avian dinosaurs underwent significant interchromosomal change, as these are known to occur among extant avian dinosaurs (kingfishers[43] (fissions), parrots[44] and falcons[5] (fusions) are modern examples). Rather than being simply interesting descriptions of inferred karyotypes, therefore, we propose that the overall genome organisation and evolution of dinosaur chromosomes (inclusive of the avian radiation) might have been a major contributing factor to their morphological

disparity, physiology, high rates of morphological change[45] and ultimate survival. In other words, we have an apparent paradox of a highly stable karyotype that is rarely changing, nonetheless contributing to great morphological diversity. We already believed this to be the case for the great phenotypic diversity we see in birds; the current results however suggest that the karyotype may have contributed to species diversity in non-avian dinosaurs also. Moreover, the evidence that the karyotype had deeper origins than previously appreciated is congruent with other recent discoveries about dinosaur morphology, demonstrating that features previously thought to be characteristic of crown-group birds only (e.g. feathers and pneumatised skeletons) arose first among more ancient dinosaur or archosaurian ancestors[23,46]. Dinosaurs have pervaded popular culture and the creative arts, perpetuated, in part, through film, television, press and literature. Their dominance for many millions of years, their radiations following two mass extinction events and, despite being almost wiped out by a third (the K–Pg meteor impact), their persistence as a highly diverse and speciose clade (extant birds)[47] has fascinated scientists since the very earliest discoveries. Of course, many of the evolutionary changes were in response to a rapidly changing environment. Whether a disproportionate advantage was offered by having an 'avian-like' karyotype will be the subject of future studies and speculation.

Methods

Cell culture and chromosome preparation. Chromosome preparations were established from fibroblast cell lines of the Caroline anole (*A. carolinensis*) ($2n = 36$), red-eared slider (*T. scripta*) ($2n = 50$) and spiny soft-shelled turtle (*A. spinifera*) ($2n = 66$). Cells were cultured at 30 °C and 5% CO_2 in Alpha MEM (Fisher), supplemented with 10% fetal bovine serum (Gibco) and 1% Pen-Strep-L-gutamine (Sigma). Chromosome suspension preparation followed standard protocols, briefly, mitostatic treatment with colcemid at a final concentration of 5.0 µg/ml for 1 h at 40 °C was followed by hypotonic treatment with 75 mM KCl for 15 min at 37 °C and fixation with 3:1 methanol:acetic acid.

Selection of BACs. Chicken and zebra finch BACs were chosen for interspecies FISH experiments according to a range of criteria, including the proportion of conserved elements shared across multiple avian species. Due to the high degree of apparent genome conservation observed between avian and reptilian species, this set of BACs was applied to chromosome suspensions of the birds in this study and from *A. carolinensis*, *T. scripta* and *A. spinifera*.

Preparation of BAC clones for FISH. BAC clone DNA was isolated using the Qiagen Miniprep Kit (Qiagen) prior to amplification and direct labelling by nick translation. Probes were labelled with Texas Red-12-dUTP (Invitrogen) and FITC-Fluorescein-12-UTP (Roche) prior to purification using the Qiagen Nucleotide Removal Kit (Qiagen).

Fluorescence in situ hybridisation. Metaphase preparations were fixed to slides and dehydrated through an ethanol series (2 min each in 2×SSC, 70%, 85% and 100% ethanol at room temperature). Probes were diluted in a formamide buffer (Cytocell) with Chicken Hybloc (Insight Biotech) and applied to the metaphase preparations on a 37 °C hotplate before sealing with rubber cement. Probe and target DNA were simultaneously denatured on a 75 °C hotplate prior to hybridisation in a humidified chamber at 37 °C for 72 h. Slides were washed post hybridisation for 30 s in 2 × SSC/0.05% Tween 20 at room temperature, and then counterstained using VECTASHIELD anti-fade medium with DAPI (Vector Labs). Images were captured using an Olympus BX61 epifluorescence microscope with cooled CCD camera and SmartCapture (Digital Scientific UK) system. Fissions and/or translocations were detected if cross-species signals appeared on two different chromosomes on the species of interest, and fusions were identified where the signals appeared on a noticeable larger chromosome than the (usually chicken) chromosome from which it was derived. This was achieved by visual inspection aided by ImageJ analysis for the smaller chromosome.

Reconstruction of the DCA karyotype. In order to reconstruct the hypothetical DCA, we selected the following sequenced extant amniote genomes: four diapsids (three birds, one lizard) and one basal mammalian representative as an outgroup (grey short-tailed opossum, *M. domestica*; assembly MonDom5). The three avian genomes, chicken (*G. gallus*; assembly galGal4), mallard (*A. platyrhynchos*; assembly BGI_duck_1.0; chromosome-level assembly, Faraut et al., personal communication) and zebra finch (*T. guttata*; assembly taeGut1), were also used to reconstruct the avian ancestor, in this case using the Carolina anole genome (*A. carolinensis*; assembly anoCar2) as the outgroup. In order to compare reptilian and avian genomes cytogenetically, we selected chicken BAC probes designed to work in FISH experiments on all avian and reptilian chromosomes[5]. By combining novel bioinformatic and FISH data produced for the current study with that of previous studies, we established, by inference, the most parsimonious explanation of the available data regarding the nature of the dinosaurian genome. Initial experiments designed to establish whether the sequenced crocodile genomes[48] were suitable outgroup species met with only limited success, given the fragmented nature of these assemblies. This, and the fact that the microchromosomes are fused in extant crocodilians ($2n = 30$) meant that this genome assembly was not a suitable reference for our work; there was also no sufficiently well-assembled chromosome-level turtle genome available for analysis at the time of writing.

Alignment of multiple genomes and identification of HSBs and EBRs. Results from this study were generated from the alignment of the three best avian genomes assembled at a chromosomal level (chicken, mallard and zebra finch) along with the best-assembled reptile genome available assembled to a partial chromosomal level (Carolina anole) and one mammalian outgroup, grey short-tailed opossum (all genomes were aligned against chicken). The whole-genome sequences of the species of interest were aligned using LastZ and visualised using the interactive genome browser Evolution Highway[4, 49]. Pairwise blocks of synteny were identified relative to chromosomes of the chicken, which served as a reference genome (galGal4). Genome alignments for the five species as inferred from sequence orthology maps were mapped against chicken chromosomes. The start and end coordinates of the contiguously aligned orthologous regions observed in all the species compared were used to define msHSBs at the 300-kb resolution. These msHSBs were assigned to and subsequently sorted in individual chromosomes in each species according to their location, orientation and sequential order.

Arrangement of ancestral diapsid and avian karyotypes. To reconstruct a putative ancestral DCA karyotype, the Multiple Genomes Rearrangements and Ancestors tool version 2 (MGRA2[50]; http://mgra.bioinf.spbau.ru/), was used as follows: based on pairwise alignments for mallard, zebra finch, Carolina anole, and grey short-tailed opossum visualised relative to the chicken, a set of respective msHSBs was generated as referred to above. In this case, the orthology map of the opossum was used as an input for the MGRA programme and included in the analysis as an outgroup. The five species-specific msHSB sets served as MGRA2 inputs for individual genomes which then produced a series of CARs representing the most likely ancestral configuration for the species identified in both hypothetical diapsid and avian ancestors.

Genome rearrangement analysis. To reconstruct the chromosomal changes that occurred between the groups, we used the MGR and Genome Rearrangements In Man and Mouse (GRIMM tools[51]; http://grimm.ucsd.edu/). MGRA2 outputs served as MGR/GRIMM inputs to trace the most parsimonious scenarios for evolutionary changes in two scenarios: first, the intrachromosomal and interchromosomal rearrangements that might have occurred from the hypothetical diapsid ancestor to the avian one and second, those rearrangements that may have occurred between the avian ancestor and the extant species.

Identification of gene ontology enrichment terms in HSBs and EBRs. Gene lists for msHSBs and archosaur EBRs were extracted from the Ensembl BioMart data system using chicken as the reference. With the chromosome-level assemblies available on Evolution Highway, definitions of EBRs are only possible for the archosaur ancestor. Since human genes are best annotated, the gene lists derived from chicken were matched to orthologous human genes and filtered for homology type and orthology confidence, leaving only those genes that were one-to-one orthologues. Background gene lists were also generated using all chicken–human orthologues with the maximum orthology confidence. The first two background gene lists tested covered all assembled chicken chromosomes and the second two lists only included results for 19 of the chicken chromosomes where the msHSBs and EBRs were found. In addition, in order to test whether genes with low gene identity matches affected the GO analysis, thresholds of 70, 60 and 50% homology at nucleotide level for the orthologue gene lists were set and the resulting GO outputs were compared. Based on these tests, the 70% gene identity threshold was selected for generating the msHSB/EBR gene lists, and the 19-chromosome list with all orthologous genes was used for the background GO analysis list. Gene lists were used as inputs for the web-based functional annotation tool DAVID[52] using human Ensembl Gene ID as the list identifier and subsequently analysed using the Functional Annotation Clustering tool. Cluster data from each gene list output were downloaded into Microsoft Excel and filtered using an enrichment score of ≥ 1.3 and a p-value < 0.05 to edit the list for clusters considered to be significant. In addition, Functional Annotation Chart reports containing single GO terms and their associated genes were generated using the same gene lists. The latter information was also taken into account to identify significant GO terms for the tested gene lists, especially in situations when the Functional Annotation Clustering

analysis did not result in any significant gene–GO term enrichment groups. In order to correct for multiple sampling error, a FDR threshold of 5% was used. Finally, individual lists of genes that fit the GO criteria were manually curated and their function established, initially from Ensembl and thereafter from the original publications that described their isolation and analysis.

Data availability. The authors declare that the data supporting the findings of this study are available within the paper (and its supplementary information files).

Received: 4 December 2017 Accepted: 12 April 2018
Published online: 21 May 2018

References

1. Zhang, G. et al. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**, 1311–1320 (2014).
2. Romanov, M. N. et al. Reconstruction of gross avian genome structure, organization and evolution suggests that the chicken lineage most closely resembles the dinosaur avian ancestor. *BMC Genomics* **15**, 1060 (2014).
3. Benton, M. J. et al. Constraints on the timescale of animal evolutionary history. *Palaeontol. Electron.* **18**, 1–106 (2015).
4. Farré, M. et al. Novel insights into chromosome evolution in birds, archosaurs, and reptiles. *Genome Biol. Evol.* **8**, 2442–2451 (2016).
5. Damas, J. et al. Upgrading short-read animal genome assemblies to chromosome level using comparative genomics and a universal probe set. *Genome Res.* **27**, 875–884 (2017).
6. Alföldi, J. et al. The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* **477**, 587–591 (2011).
7. Hillier, L. W. et al. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716 (2004).
8. Rao, M. et al. A duck RH panel and its potential for assisting NGS genome assembly. *BMC Genomics* **13**, 513 (2012).
9. Nadeau, J. H. & Taylor, B. A. Lengths of chromosomal segments conserved since divergence of man and mouse. *Proc. Natl Acad. Sci. USA* **81**, 814–818 (1984).
10. Pevzner, P. & Tesler, G. Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution. *Proc. Natl Acad. Sci. USA* **100**, 7672–7677 (2003).
11. Larkin, D. M. et al. A cattle–human comparative map built with cattle BAC-ends and human genome sequence. *Genome Res.* **13**, 1966–1972 (2003).
12. Warren, W. C. et al. A new chicken genome assembly provides insight into avian genome structure. *G3* **7**, 109–117 (2017).
13. Warren, W. C. et al. The genome of a songbird. *Nature* **464**, 757–762 (2010).
14. Mikkelsen, T. S. et al. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* **447**, 167–177 (2007).
15. St John, J. A. et al. Sequencing three crocodilian genomes to illuminate the evolution of archosaurs and amniotes. *Genome Biol.* **13**, 415 (2012).
16. Shaffer, H. B. et al. The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biol.* **14**, R28 (2013).
17. Srikulnath, K., Thapana, W. & Muangmai, N. Role of chromosome changes in *Crocodylus* evolution and diversity. *Genomics Inform.* **13**, 102 (2015).
18. Griffin, D. K. et al. Micro- and macrochromosome paints generated by flow cytometry and microdissection: tools for mapping the chicken genome. *Cytogenet. Cell Genet.* **87**, 278–281 (1999).
19. Beçak, W., Beçak, M. L., Nazareth, H. R. S. & Ohno, S. Close karyological kinship between the reptilian suborder Serpentes and the class Aves. *Chromosoma* **15**, 606–617 (1964).
20. Matsuda, Y. et al. Highly conserved linkage homology between birds and turtles: bird and turtle chromosomes are precise counterparts of each other. *Chromosom. Res.* **13**, 601–615 (2005).
21. Kasai, F., O'Brien, P. C. M., Martin, S. & Ferguson-Smith, M. A. 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 (2012).
22. Badenhorst, D. et al. Physical mapping and refinement of the painted turtle genome (*Chrysemys picta*) inform amniote genome evolution and challenge turtle-bird chromosomal conservation. *Genome Biol. Evol.* **7**, 2038–2050 (2015).
23. Baron, M. G., Norman, D. B. & Barrett, P. M. A new hypothesis of dinosaur relationships and early dinosaur evolution. *Nature* **543**, 501–506 (2017).
24. Burt, D. W. Origin and evolution of avian microchromosomes. *Cytogenet. Genome Res.* **96**, 97–112 (2002).
25. Cracraft, J. et al. Response to Comment on "Whole-genome analyses resolve early branches in the tree of life of modern birds". *Science* **349**, 1460 (2015).
26. Uno, Y. et al. Inference of the protokaryotypes of amniotes and tetrapods and the evolutionary processes of microchromosomes from comparative gene mapping. *PLoS ONE* **7**, e53027 (2012).
27. Capilla, L. et al. Mammalian comparative genomics reveals genetic and epigenetic features associated with genome reshuffling in Rodentia. *Genome Biol. Evol.* **8**, 3703-3717 (2016).
28. Skinner, B. M. & Griffin, D. K. Intrachromosomal rearrangements in avian genome evolution: evidence for regions prone to breakpoints. *Heredity* **108**, 37–41 (2011).
29. Larkin, D. M. et al. Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. *Genome Res.* **19**, 770–777 (2009).
30. Sankoff, D. The where and wherefore of evolutionary breakpoints. *J. Biol.* **8**, 66 (2009).
31. Murko, C. et al. Expression of class I histone deacetylases during chick and mouse development. *Int. J. Dev. Biol.* **54**, 1527–1537 (2010).
32. Wang, Y.-C. et al. Identification, chromosomal arrangements and expression analyses of the evolutionarily conserved *prmt1* gene in chicken in comparison with its vertebrate paralogue *prmt8*. *PLoS ONE* **12**, e0185042 (2017).
33. Organ, C. L., Shedlock, A. M., Meade, A., Pagel, M. & Edwards, S. V. Origin of avian genome size and structure in non-avian dinosaurs. *Nature* **446**, 180–184 (2007).
34. Hughes, A. L. & Hughes, M. K. Small genomes for better flyers. *Nature* **377**, 391–391 (1995).
35. Kapusta, A., Suh, A. & Feschotte, C. Dynamics of genome size evolution in birds and mammals. *Proc. Natl Acad. Sci. USA* **114**, E1460–E1469 (2017).
36. Kawakami, T. et al. A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. *Mol. Ecol.* **23**, 4035–4058 (2014).
37. Smeds, L., Mugal, C. F., Qvarnström, A. & Ellegren, H. High-resolution mapping of crossover and non-crossover recombination events by whole-genome re-sequencing of an avian pedigree. *PLoS Genet.* **12**, e1006044 (2016).
38. Gao, B. et al. Low diversity, activity, and density of transposable elements in five avian genomes. *Funct. Integr. Genom.* **17**, 427–439 (2017).
39. Mason, A. S., Fulton, J. E., Hocking, P. M. & Burt, D. W. A new look at the LTR retrotransposon content of the chicken genome. *BMC Genomics* **17**, 688 (2016).
40. Cui, J. et al. Low frequency of paleoviral infiltration across the avian phylogeny. *Genome Biol.* **15**, 539 (2014).
41. Organ, C. L. & Shedlock, A. M. Palaeogenomics of pterosaurs and the evolution of small genome size in flying vertebrates. *Biol. Lett.* **5**, 47–50 (2009).
42. Gregory, T. R. in *The Evolution of the Genome* (ed. Gregory, T. R.) 3–87 (Elsevier, New York City, NY, 2005).
43. Christidis, L. Animal Cytogenetics 4: Chordata 3 B: Aves. (Gebrüder Borntraeger, Berlin, 1990).
44. Nanda, I., Karl, E., Griffin, D. K., Schartl, M. & Schmid, M. Chromosome repatterning in three representative parrots (Psittaciformes) inferred from comparative chromosome painting. *Cytogenet. Genome Res.* **117**, 43–53 (2007).
45. Berv, J. S. & Field, D. J. Genomic signature of an avian Lilliput Effect across the K-Pg extinction. *Syst. Biol.* **67**, 1–13 (2018).
46. Zhou, Z. The origin and early evolution of birds: discoveries, disputes, and perspectives from fossil evidence. *Naturwissenschaften* **91**, 455–471 (2004).
47. Barrowclough, G. F., Cracraft, J., Klicka, J. & Zink, R. M. How many kinds of birds are there and why does it matter? *PLoS ONE* **11**, e0166307 (2016).
48. Green, R. E. et al. Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. *Science* **346**, 1254449 (2014).
49. Murphy, W. J. et al. Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science* **309**, 613–617 (2005).
50. Avdeyev, P., Jiang, S., Aganezov, S., Hu, F. & Alekseyev, M. A. Reconstruction of ancestral genomes in presence of gene gain and loss. *J. Comput. Biol.* **23**, 150–164 (2016).
51. Bourque, G. & Pevzner, P. A. Genome-scale evolution: reconstructing gene orders in the ancestral species. *Genome Res.* **12**, 26–36 (2002).
52. Dennis, G. Jr et al. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* **4**, P3 (2003).

Acknowledgements

This work was supported in part by the Biotechnology and Biological Sciences Research Council (BB/K008161/1 to D.K.G. and BB/K008226/1 and BB/J010170/1 to D.M.L.).

Author contributions

R.E.O'.C. was the post-doc lead on the project, providing the first draft of the manuscript and performing the bulk of the experimental work. L.G.K. assisted with the cytogenetic work, while M.N.R., J.D. and M.F. contributed to the bioinformatic analysis. P.M.B., M. F.-S. and N.V. provided much needed insight into the nature of chromosome changes, essential material (M.F.-S. and N.V.) and insight into diapsid taxonomy and biology (P. M.B.). D.M.L. and D.K.G. were the academic leads on the manuscript, D.M.L. with an

emphasis on bioinformatics and D.K.G. on molecular cytogenetics. D.K.G. was the custodian of the manuscript and first point of contact dealing with reviewer comments. All authors contributed to writing the manuscript and contributing on drafts.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-04267-9.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

O'Connor et al. Genome Biology (2018) 19:171
https://doi.org/10.1186/s13059-018-1550-x

Genome Biology

RESEARCH Open Access

Chromosome-level assembly reveals extensive rearrangement in saker falcon and budgerigar, but not ostrich, genomes

Rebecca E O'Connor[1†], Marta Farré[2†], Sunitha Joseph[1], Joana Damas[2], Lucas Kiazim[1], Rebecca Jennings[1], Sophie Bennett[1], Eden A Slack[2], Emily Allanson[2], Denis M Larkin[2†] and Darren K Griffin[1*†]

Abstract

Background: The number of de novo genome sequence assemblies is increasing exponentially; however, relatively few contain one scaffold/contig per chromosome. Such assemblies are essential for studies of genotype-to-phenotype association, gross genomic evolution, and speciation. Inter-species differences can arise from chromosomal changes fixed during evolution, and we previously hypothesized that a higher fraction of elements under negative selection contributed to avian-specific phenotypes and avian genome organization stability. The objective of this study is to generate chromosome-level assemblies of three avian species (saker falcon, budgerigar, and ostrich) previously reported as karyotypically rearranged compared to most birds. We also test the hypothesis that the density of conserved non-coding elements is associated with the positions of evolutionary breakpoint regions.

Results: We used reference-assisted chromosome assembly, PCR, and lab-based molecular approaches, to generate chromosome-level assemblies of the three species. We mapped inter- and intrachromosomal changes from the avian ancestor, finding no interchromosomal rearrangements in the ostrich genome, despite it being previously described as chromosomally rearranged. We found that the average density of conserved non-coding elements in evolutionary breakpoint regions is significantly reduced. Fission evolutionary breakpoint regions have the lowest conserved non-coding element density, and intrachromomosomal evolutionary breakpoint regions have the highest.

Conclusions: The tools used here can generate inexpensive, efficient chromosome-level assemblies, with > 80% assigned to chromosomes, which is comparable to genomes assembled using high-density physical or genetic mapping. Moreover, conserved non-coding elements are important factors in defining where rearrangements, especially interchromosomal, are fixed during evolution without deleterious effects.

Keywords: Chromosome-level genome assembly, Genome evolution, CNE, EBR

Background

The number of de novo (new species) genome sequence assemblies is increasing exponentially (e.g., [1, 2]). Improved technologies are generating longer reads, greater read depths, and ultimately assemblies with fewer, longer contigs per genome [3, 4]; however, the ability to assemble a genome with the same number of scaffolds or contigs as chromosomes ("chromosome-level" assembly) remains the ultimate aim of a de novo sequencing effort. This is for several reasons, among them the requirement for an established order of DNA markers as a pre-requisite for revealing genotype-to-phenotype associations for marker-assisted selection and breeding, e.g., in species regularly bred for food production, companionship, or conservation purposes [5].

Chromosome-level assemblies were rapidly established for agricultural animals (chicken, pig, cattle, sheep) [6–9] in part because they were assembled as maps prior to (e.g., Sanger) sequencing. Species used for food consumption in developing countries (e.g., goat, camel, yak, buffalo,

* Correspondence: d.k.griffin@kent.ac.uk
†Rebecca E O'Connor, Marta Farré, Denis M Larkin and Darren K Griffin contributed equally to this work.
[1]School of Biosciences, University of Kent, Canterbury, UK
Full list of author information is available at the end of the article

ostrich, quail); animals bred for conservation (e.g., falcons and parrots), and companion animals (e.g., pet birds) are still however poorly represented, in part because they were initially assembled using NGS data alone. New techniques, e.g., optical mapping [10], BioNano [11], Dovetail [12], and PacBio long-read sequencing [13], make significant steps towards this. Recent progress on the goat genome for instance resulted in a chromosome-level assembly using PacBio long-read sequencing [2]; others however encounter technical issues: BioNano contigs fail to map across multiple DNA nick site regions, centromeres, or large heterochromatin blocks, and PacBio requires starting material of hundreds of micrograms of high molecular weight DNA, thereby limiting its usage. To achieve a chromosome-level assembly therefore often requires a combination of technologies to integrate the sequence data, e.g., Hi-C [14], linkage mapping, pre-existing chromosome-level reference assemblies, and/or molecular cytogenetics [15, 16]. To this end, we made use of bioinformatic approaches, e.g., the Reference-Assisted Chromosome Assembly (RACA) algorithm [17]. RACA however is limited in needing a closely related reference species for comparison [17] and further mapping of superscaffolds physically to chromosomes. We therefore recently developed an approach where RACA produces sub-chromosome-sized predicted chromosome fragments (PCFs) which are subsequently verified and mapped to chromosomes using molecular methods [15]. In so doing, we previously established a novel, integrated approach that allows de novo assembled genomes to be mapped directly onto the chromosomes of interest and displayed the information in an interactive browser (Evolution Highway) to allow direct, chromosome-level comparison. To date however, only two genomes—the pigeon (*Columba livia*) and the Peregrine falcon (*Falco peregrinus*)—have been assembled in this way [15].

In the current study, we focused on generating chromosome-level assemblies for three further avian genomes. These are the following: The saker falcon (*Falco cherrug*—FCH), classified as endangered [18], is phenotypically remarkable for its visual acuity [19] and acceleration speeds [20]. It has an atypical avian genomic structure ($2n = 52$) with fused microchromosomes [21]. Secondly, we selected the common budgerigar (*Melopsittacus undulatus*—MUN) which also has a highly rearranged karyotype with multiple fusions ($2n = 62$). As a member of the order *Psittaciformes* (parrots), the budgerigar is one of the world's most popular companion animals as well as a highly valued model for studies into vocal learning [22]. Finally, we selected the ostrich (*Struthio camelus*—SCA), the largest extant bipedal land animal [23]. The ostrich is able to travel long distances with a remarkable degree of metabolic economy [24]. Apparently possessing a typical avian karyotype ($2n = 80$), with a large degree of homology with the chicken (like other ratite birds) revealed by cross species chromosome painting [25–27], it however purportedly has 26 previously undetected interchromosomal rearrangements when compared to the ancestral avian karyotype as revealed by sequence assembly analysis of optical mapping data [28]. For these three species, we used our previously described approach combining computational algorithms for ordering scaffolds into predicted chromosome fragments (PCFs) which we then physically mapped directly to the chromosomes of interest using a set of avian universal bacterial artificial chromosome (BAC) probes [15].

Chromosome-level assemblies also inform studies of evolution and speciation given that inter-species differences arise from chromosomal changes fixed during evolution [29–35]. In recent studies, we have used (near) chromosome-level assemblies to reconstruct ancestral karyotypes and trace inter- and intrachromosomal changes that have occurred to generate the karyotypes of extant species [28, 36]. Theories explaining the mechanisms of chromosomal change in vertebrates include a role for repetitive sequences used for non-allelic homologous recombination (NAHR) in evolutionary breakpoint regions (EBRs) [37] and the proximity of DNA regions in chromatin [38]. During gross genome (karyotype) evolution, unstable EBRs delineate stable homologous synteny blocks (HSBs) and we have established that the largest HSBs are maintained non-randomly and highly enriched for conserved non-coding elements (CNEs) [9–11, 15, 39]. We recently proposed the hypothesis that a higher fraction of elements under negative selection involved in gene regulation and chromosome structure in avian genomes (~ 7%) [40] compared to mammals (~ 4%) [41] could contribute to some avian-specific phenotypes, as well as the evolutionary stability of the overall organization of most avian genomes [39]. We further studied the fate of CNEs in the EBRs flanking interchromosomal rearrangements of a highly rearranged avian genome, finding that, in the peregrine falcon, interchromosomal EBRs contain 12 times fewer CNEs than intrachromosomal ones [15].

In order to investigate the role of CNEs in chromosome rearrangements further, we therefore concentrated on species that had previously been reported as highly chromosomally rearranged. Studying these highly rearranged genomes at this resolution provided insight into the mechanisms of chromosomal rearrangement.

Results

Predicted chromosome fragments for three new species

Predicted chromosome fragments were generated for fragmented saker falcon, budgerigar, and ostrich whole-genome sequences using RACA [17]. The zebra finch and the chicken chromosome assemblies were used as reference and outgroup respectively for all

reconstructions, except for ostrich. For saker falcon, we generated 95 PCFs representing 97.26% of the original genome, while for ostrich and budgerigar, 100 and 84 PCFs were produced (Table 1). These initial PCF sets contained ~10% putatively chimeric scaffolds for both ostrich and saker falcon, while for budgerigar, ~31% of the scaffolds were split by RACA due to insufficient read and/or comparative evidence to support their structures.

We then tested the split scaffold regions by PCR to assess their existence in the target genome. Only the split regions defined to < 6 kbp in the target genomes were tested, representing 36%, 71%, and 28% of all split scaffolds in the saker falcon, ostrich, and budgerigar assemblies, respectively (Table 1). Of these, 11, 20, and 32 resulted in amplicons of expected length in saker falcon, budgerigar, and ostrich genomic DNA, respectively. For the split regions with negative PCR results, we tested an alternative (RACA-suggested) order of the flanking syntenic fragments (SFs). Out of these, amplicons were obtained for 5/11 in saker falcon, 11/23 in budgerigar, and 7/17 in ostrich, confirming the chimeric nature of the original scaffolds properly detected in these cases. As in our previous publication [15], to estimate which of the remaining split regions (> 6 kb; 39 in falcon, 111 in budgerigar, and 20 in ostrich PCFs) were likely to be chimeric, we empirically identified the genome-wide minimum physical coverage [42] levels for each species in the SF joining regions for which the PCR results were most consistent with original scaffold structures. A physical coverage of 379×, 216×, and 239× were estimated for saker falcon, budgerigar, and ostrich to produce the highest agreement between scaffolds and PCR results. Finally, we used the adjusted physical coverage thresholds to reconstruct a new set of PCFs for all three species (Table 1). To do so, we re-ran RACA by updating the MIN_INTRACOV_PERC parameter with the new physical coverage thresholds (Table 1) and including scaffolds with the structures confirmed by PCR as additional inputs. This resulted in an increased number of PCFs, a reduction of the N50, and a lower fraction of chimeric scaffolds for all species.

Chromosome-level assemblies for three new species

We successfully generated chromosome-level assemblies for the three avian species of interest, with coverage similar to Sanger sequencing assembled genomes. Our method involves (a) construction of PCFs for fragmented assemblies based on the comparative and sequence read data implemented in the RACA algorithm, (b) PCR and computational verification of a limited number of scaffolds that are essential for revealing species-specific chromosome structures, (c) creation of a refined set of PCFs using the verified scaffolds and adjusted adjacency thresholds in RACA, and (d) the use of a panel of "universal" BAC clones to anchor PCFs to chromosomes in a high-throughput manner (see Fig. 1 for representative image) and is reported in detail elsewhere [15]. Using this approach, for the ostrich ($2n = 80$), the N50 of the original NGS genome was improved approximately eightfold, with over 79% of the genome placed onto chromosomes with 71.26% of the original assembly fully oriented (see Table 2). Chromosome-level assembly was accomplished for all GGA (chicken) homologs with the exception of chromosome GGA16 for which BAC clones were not available. PCFs were generated ranging in size from 350 kb to 82 Mb; the second largest of which (80.5 Mb) represented the entire p-arm of chromosome 1. For the budgerigar ($2n = 62$), FISH mapping (e.g., Fig. 1) resulted in 21 pairs of budgerigar autosomes and the Z chromosome being assembled with a fourfold improvement on the scaffold N50 from 11 to 38 Mb. 93.56% of the original assembly was placed onto chromosomes, and 77.93% was fully oriented. For the Saker falcon ($2n = 52$), in total, 19 autosomes and the Z chromosome were assembled to chromosome level, with a fivefold N50 improvement, resulting in 90.12% of the original assembly assigned to chromosomes and 67.52% of the assembly fully oriented. Assembly statistics for all three genomes

Table 1 Statistics for the scaffold split regions tested by PCR

Statistics	Saker falcon	Ostrich	Budgerigar
Pair-end read physical coverage within tested scaffolds	135–524	2–604	0–631
No. split SF adjacencies by RACA (default param.)	61	69	154
No. tested scaffold split regions	22 (100%)	49 (100%)	43 (100%)
No. amplified split regions (confirmed SF joints)	11 (50%)	32 (65%)	20 (46%)
No. non-amplified split regions	11 (50%)	17 (35%)	23 (54%)
No. tested RACA-suggested adjacencies	11	8	18
No. amplified adjacencies (chimeric SF joints)	5	7	11
Final no. ambiguous SF joints from tested split regions	6	10	12
Selected pair-end read spanning threshold	379	239	216

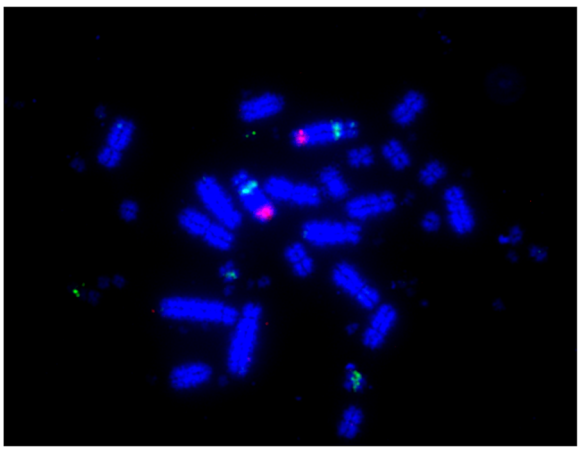


Fig. 1 BAC clones hybridized to budgerigar chromosome two (MUN2). The green (FITC labeled) signal represents TGMCBA-375I5 (GGA17 homolog) and maps to PCF 17, and the Texas red labeled signal represents CH261-169K18 (GGA3 homolog) and maps to PCF 3c_5a

are listed in Table 2. In the course of the FISH experiments performed, we did not detect any BAC spanning breakpoints. A representative screenshot (Fig. 2) of chromosomes homologous to ancestral chromosome 3 is given (BACs, scaffolds, and PCFs shown), and the whole dataset is freely available on http://eh-demo.ncsa.uiuc.edu/birds/.

Comparative genomics with chicken

All three species were aligned against the chicken (*Gallus gallus*—GGA) genome assembly. Chicken is the most characterized avian genome at sequence depth and chromosome level [6], and the species considered to be most similar chromosomally to the avian ancestor [28].

Homology between the ostrich and the chicken (as illustrated in Fig. 3) was confirmed interchromosomally between all chromosomes tested, with the exception of GGA4 which is homologous to ostrich chromosome 4 *plus* one microchromosome (a fusion thought to have occurred in the chicken lineage [43]). Contrary to our previous study [28], we found no further evidence of interchromosomal rearrangement compared to the chicken. A total of 14 intrachromosomal differences were identified in the ostrich when compared to the chicken listed in Additional file 1: Table S1.

Homologies between the budgerigar and the chicken were identified for all mapped chicken chromosomes (GGA1-28, excluding 16, plus Z). Fusions of ten homologs were identified with three budgerigar chromosomes (MUN4, 5, and 8), exhibiting the fusion of three chicken homologs each (Fig. 4). The fusion of two chicken homologs was demonstrated in three budgerigar chromosomes (MUN2, 9, and 10). Three fissions were evident where the GGA1 homolog split to form MUN3 and 6 with no evidence of further fusion; GGA5 and GGA7 homologs split and fused as separate chromosomes (MUN4 and 8). The GGA4 homolog exhibited the pattern seen in most other birds where the p-arm of GGA4 is in fact a fused ancestral microchromosome. Where previously assigned, the budgerigar chromosomes were numbered according to Nanda et al. [44]. Where no previous assignment had been given, the chromosomes were numbered according to decreasing PCF size. A representative ideogram illustrating the gross genomic structure and the chicken homologies is shown in Fig. 4. In total, of the 18 mapped chicken microchromosome homologs, 7 were fused to other chromosomes, while 11 remained intact as microchromosomes. Given the deviation from the typical avian pattern, these interchromosomal changes are thought to be unique to the budgerigar lineage. A total of 16 intrachromosomal rearrangements were identified between budgerigar and chicken, none of which were seen in the ostrich-chicken comparison, nor in the 14 chicken-specific intrachromosomal changes reported by Farre et al. [39], suggesting that these arose after the Galloanserae-Neoaves divergence (illustrated in Additional file 1: Table S2).

Extensive interchromosomal genome rearrangement was evident in the saker falcon where, in total, 12 fusions and 5 fissions were detected when compared to the chicken genome. Each of the largest chicken macrochromosome homologs (GGA1 to GGA5) were represented by two saker falcon chromosomes indicating fission in the falcon lineage for chromosomes 1, 2, 3, and 5 but the commonly reported chicken lineage fusion for GGA4. Both the GGA6 and GGA7 homologs were found as single blocks fused with other chicken homologs while GGA8, GGA9, and GGAZ were represented as individual chromosomes. Of the 17 mapped chicken microchromosomes, regions homologous to GGA microchromosomes 10, 12, 13, 14, 15, 17, 18, 19, 20, 21, 23, and 28 were fused to GGA macrochromosome homologs, leaving GGA 11, 22, 24, 26, and 27 conserved as intact microchromosomes. The overall genomic structure is illustrated in Fig. 5, with saker falcon chromosomes numbered according to size. A total of 36 intrachromosomal differences were identified when compared to the chicken, none of which were evident in the ostrich-chicken comparison, nor in the 14 chicken-specific intrachromosomal changes reported by Farre et al. [39], suggesting that these are probably unique to the falcon lineage, arising after the Galloanserae-Neoaves divergence. These are illustrated in Additional file 1: Table S3.

Rearrangements from the avian ancestor

The overall pattern of chromosomal rearrangement evident in the three species is illustrated in Table 3 and

Table 2 Assembly statistics from original NGS genome to RACA assembly and combined RACA and FISH assembly

Original assembly			
Stats	Budgie	Ostrich	Saker falcon
No. scaffolds longer 10 kbp	1138	1179	731
Total length (Gbp)	1.08	1.22	1.17
N50 (Mbp)	11.41	3.64	4.16
Default RACA assembly			
Stats	Budgerigar PCFs	Ostrich PCFs	Saker falcon PCFs
No. PCFs	84	100	95
Total length (Gbp)	1.04	1.17	1.14
N50 (Mbp)	46.54	37.95	39.38
No. chimeric scaffolds	80 (31%)	58 (10%)	50 (10%)
No. used scaffolds	254	588	458
% original assembly	96.29	95.90	97.26
RACA + PCR assembly			
Stats	Budgerigar PCFs	Ostrich PCFs	Saker falcon PCFs
No. PCFs	95	136	103
Total length (Gbp)	1.04	1.17	1.14
N50 (Mbp)	37.96	28.09	22.28
No. chimeric scaffolds	55 (21%)	31 (5%)	25 (5%)
No. used scaffolds	254	588	458
% original assembly	96.29	96.02	97.26
RACA + FISH assembly			
Stats	Budgerigar chromosomes	Ostrich chromosomes	Saker falcon chromosomes
No. PCFs placed	46	53	64
No. PCFs oriented	28	37	37
Disagreements RACA-FISH	4	0	0
Length placed (bp)	1,013,720,408	969,537,146	1,055,312,481
Length oriented (bp)	844,433,024	869,521,333	790,725,803
% original assembly placed	93.56	79.45	90.12
% original assembly oriented	77.93	71.26	67.52

Fig. 6 by divergence from the inferred avian ancestor. Given the similarity interchromosomally of chicken and ostrich, and the prior knowledge that GGA4 arose from the fusion of two ancestral chromosomes, the single interchromosomal difference (GGA4 fusion) is easily derived. For the intrachromosomal changes, using ostrich as an outgroup infers the changes since the divergence of the Neognathe ancestor (see above). In the absence of a chromosomally assembled outgroup genome for all birds in this study, it is not easy to determine whether the intrachromosomal differences are ancestral or derived in chicken and ostrich respectively. For this reason, in the far-right hand column of Table 3, the differences between chicken and ostrich are noted but without any conclusions as to which is the ancestor.

There were two fissions common to both the budgerigar and the saker falcon. The first of these involved the chicken chromosome 1 homolog (FCH3 and 5; MUN3 and 6) where the fission point (between GGA ~ 72 and ~ 86 Mb) corresponds to the breakpoint seen in the chromosomally assembled zebra finch genome (between GGA ~ 74 and ~ 75 Mb), and probably in all Passerines according to zoo-FISH studies [45]. The second was a fission that occurred in the homolog of chicken chromosome 5, the derivative products of which went on to form budgerigar chromosomes 4 and 8 and saker chromosomes 7 and 10. Finally, a fission present in falcon but not in budgerigar (chromosome 2 centric) is also observed in turkey, but is probably an example of homoplasy given that centromeres are prone to fission.

In the budgerigar genome, 13 chicken homologs showed no evidence of fission or fusion and in the saker, 8 homologs showed no evidence of interchromosomal rearrangement. The Z chromosome was the only

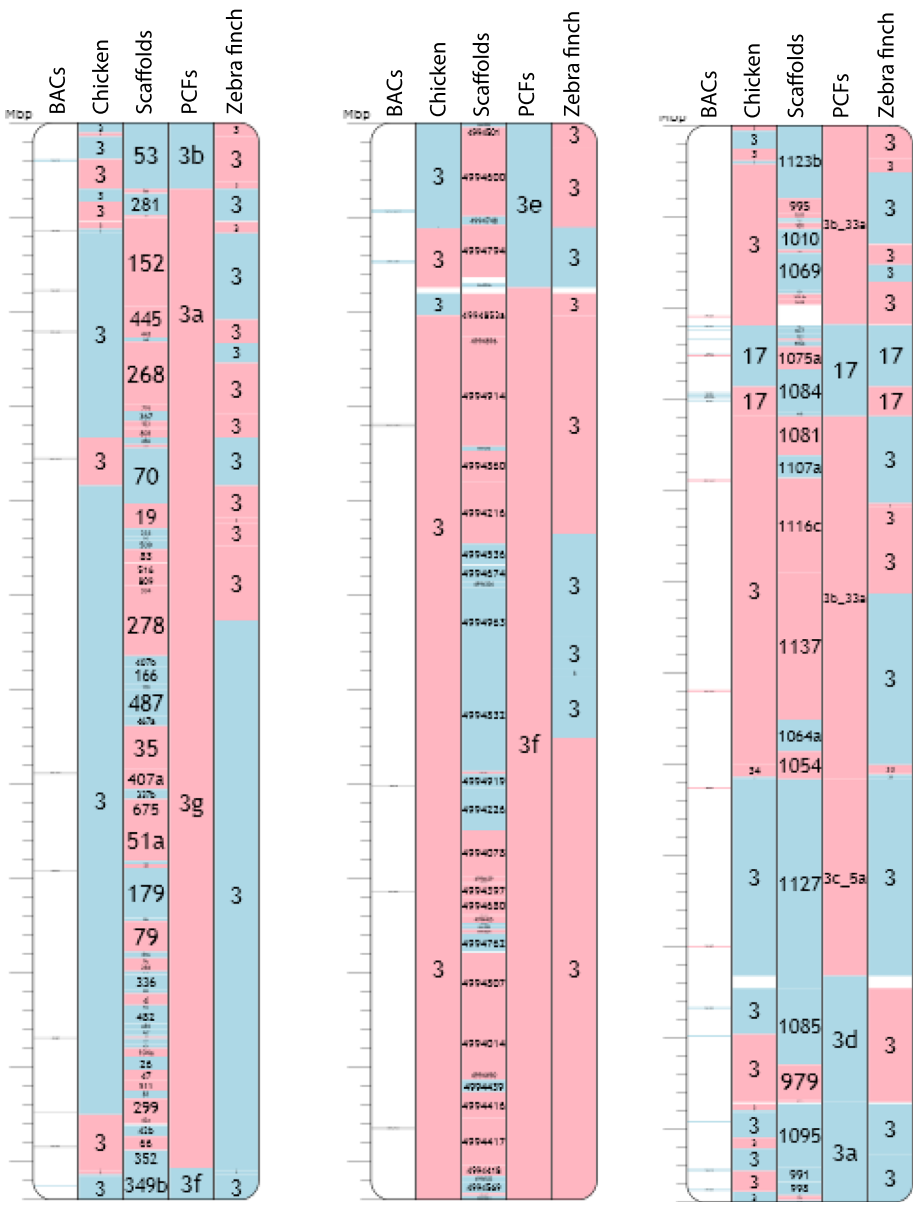


Fig. 2 Chromosomes homologous to chicken (ancestral) chromosome 3 with mapped BACs, scaffolds, PCFs, and zebra finch homologies shown. SCA3 = ostrich chromosome 3, FCH6 = saker falcon chromosome 6, MUN6 = budgerigar chromosome 2. The full dataset can be found on the interactive browser Evolution Highway at the following link: http://eh-demo.ncsa.uiuc.edu/birds/

macrochromosome that did not rearrange interchromosomally in all species tested.

The ostrich was revealed to have the lowest number of intrachromosomal differences relative to chicken, with a total of 14 identified—three of which were on chromosome 3. The budgerigar, although highly rearranged interchromosomally, appeared to have a similar number of intrachromosomal rearrangements, with evidence of 16 inversions, 3 of which were on the homolog of GGA3 (albeit different from those seen in the ostrich). The saker falcon, however, while also highly rearranged at an interchromosomal level, also exhibited a very large number of intrachromosomal changes relative to the other species with 36 inversions. No intrachromosomal rearrangement was evident in the homologs of chromosomes 19, 21, and 25 (Additional file 2).

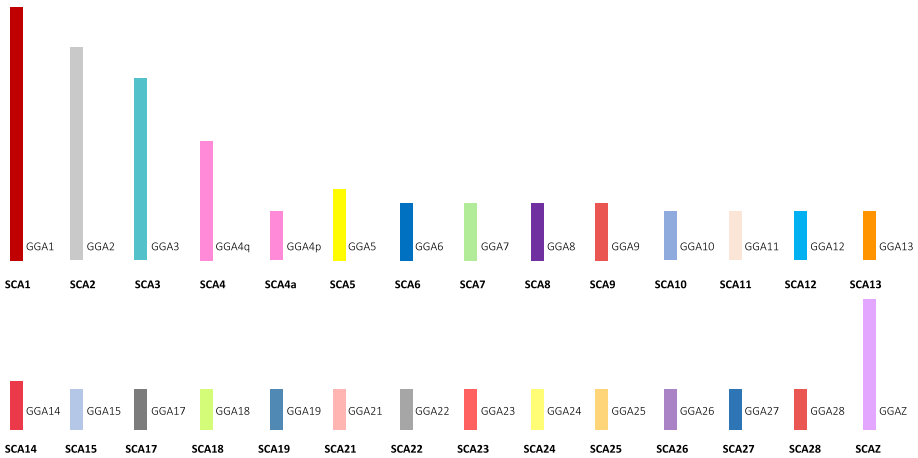


Fig. 3 Ideogram representation of the gross genomic structure of the ostrich (*Struthio camelus*—SCA) with chicken homologies per chromosome. Each GGA (chicken) homolog is represented as a different color—randomly assigned. Intrachromosomal differences are not shown here but listed in Additional file 1: Table S1

CNEs in avian inter- and intrachromosomal EBRs

Analysis of the three new avian genomes, previously thought to have undergone significant interchromosomal rearrangement compared to most avian genomes, allowed us to investigate the role of conserved non-coding elements (CNEs) in inter- vs intrachromosomal rearrangement. Our results determined that only two genomes were in fact highly rearranged interchromosomally. A total of 27 inter- and 146 intrachromosomal EBRs were identified in the three genomes (listed in Additional file 3: Tables S7–S9). We calculated densities of CNEs [39] in both types of EBRs using chicken genome as a reference. Intra- and interchromosomal EBRs were defined to ≤ 100 kb in the chicken genome. Avian EBRs had a significantly lower fraction of CNEs than their two adjacent chromosome intervals of the same size each (up- and downstream; $P = 0.01$). Moreover, the interchromosomal EBRs (fusions and fissions) had, on average, approximately

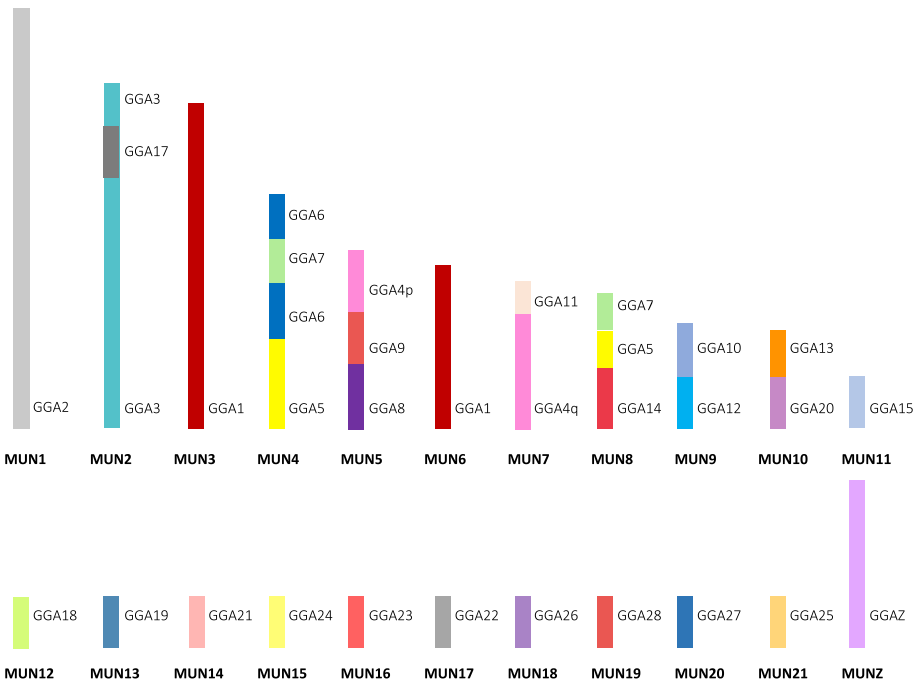


Fig. 4 Ideogram representation of the gross genomic structure of the budgerigar (*Melopsittacus undulatus*—MUN) with chicken homologies per chromosome. Each GGA (chicken) homolog is represented as a different color as assigned in Fig. 3. Intrachromosomal differences are not shown here but listed Additional file 1: Table S2

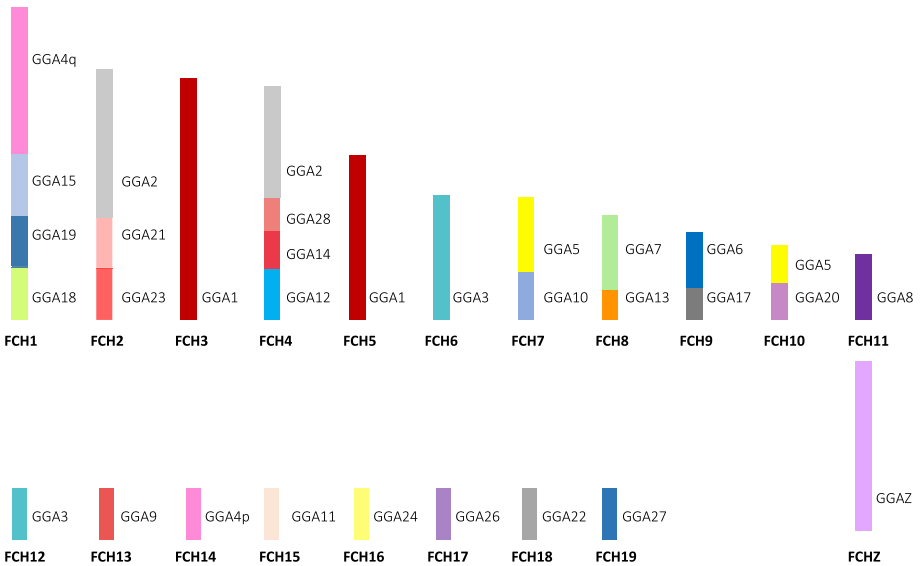


Fig. 5 Ideogram representation of the gross genomic structure of the saker falcon (*Falco cherrug*—FCH) with chicken homologies per chromosome. Each GGA (chicken) homolog is represented as a different color as assigned in Fig. 3. Intrachromosomal differences are not shown here but listed in Additional file 1: Table S3

2.2 times lower density of CNEs than the intrachromosomal EBRs ($P = 2.40 \times 10^{-5}$). The lowest density of CNEs was observed in the fission breakpoints ($P = 0.04$). In order to identify the CNE densities and the distribution associated with avian EBRs at the genome-wide level, we further counted CNE bases in 1-kb windows overlapping EBRs and avian multi-species HSBs (msHSBs) > 1.5 Mb [39]. The genome-wide CNE density was 0.087, close to the density observed in msHSBs. The average density of CNEs in the EBR windows was significantly lower (0.022) than that in the msHSBs (0.107, $P < 0.01$). Fission EBRs had the lowest density of CNEs observed, approximately zero CNE bases, while in the intrachromomosomal, EBRs had the highest among the EBR regions (0.026, $P = 0.035$, Table 4).

Discussion

Increasing numbers of newly sequenced genomes require tools that facilitate inexpensive, efficient chromosome-level assembly for the reasons described above. The tools used here and developed in our previous study [15] have generated chromosome-level assemblies for previously published but highly fragmented sequenced genomes. The assemblies generated using this approach now have > 80% of their genomes placed on chromosomes, making them highly comparable to genomes assembled using Sanger sequencing techniques and high-density physical or genetic mapping [29]. The method used here is less expensive and requires fewer resources than pre-existing approaches, in part thanks to the ability to generate predicted chromosome fragments of a sub-chromosomal size using comparative genome and not end existing read pair information only. The subsequent use of BAC probes designed to work equally well on a large number of highly diverged avian species creates a resource for physical mapping that is transferrable potentially to all avian species.

The ostrich genome

Avian interchromosomal rearrangements are rare, except in cases (e.g., *Psittaciformes* and *Falconiformes*) where it is evident that karyotypes are highly rearranged [15, 45, 46]. In the case of the ostrich and other ratites (emu and rhea), avian-typical patterns have been illustrated using comparative chromosome painting [25–27]. However, results presented in our previous study [28], based on NGS assemblies enhanced with newer third-generation technologies, suggested that the ostrich is in fact the exception to this pattern. Our older data that included optical map-enhanced NGS assembly [47] indicated the presence of 26 ostrich interchromosomal rearrangements compared to the avian ancestor. The data presented in the current study however contradicts these findings and confirms the original chromosome painting data that the ostrich genome is in fact a "typical" avian genome in terms of overall karyotypic structure. The most likely explanation for these erroneously called interchromosomal rearrangements is errors in either the optical map data itself or the original Illumina scaffolds that were enhanced by the map, again, highlighting the importance of anchoring genome sequences to the chromosomes directly, rather than relying purely on a sequence-based

Table 3 Patterns of fusion and fission revealed in the budgerigar, saker falcon, and the ostrich using the chicken genome as a reference

Ancestral chromosome (numbered according to chicken)	Budgerigar		Saker falcon		Ostrich	Chicken	Chicken-ostrich differences
	Inter-	Intra-	Inter-	Intra-	Inter-	Inter-	Intra-
1	Fission	1	Fission	4	0	0	0
2	–	2	Fission and fusion to GGA21 and 28	0	0	0	0
3	Fusion to GGA17	3	Fission	4	0	0	3
4a	–	0	–	2	0	Fusion	1
4b	Fusion to GGA9	0	Fusion to GGA15	4	0		2
5	Fission and fusion to GGA6	1	Fission and fusion to GGA10 and 20	2	0	0	1
6	Fusion to GGA5	0	Fusion to GGA17	3	0	0	0
7	Fission and fusion to GGA6 and 5	1	Fusion to GGA13	2	0	0	3
8	Fusion to GGA9	0	–	0	0	0	1
9	Fusion to GGA8	1	–	0	0	0	0
10	Fusion to GGA12	0	Fusion to GGA5	0	0	0	0
11	Fusion to GGA4q	0	–	0	0	0	0
12	Fusion to GGA10	1	Fusion to GGA14	0	0	0	0
13	Fusion to GGA20	0	Fusion to GGA7	0	0	0	0
14	Fusion to GGA5	1	Fusion to GGA12 and 28	1	0	0	1
15	–	2	Fusion to GGA4q and 19	2	0	0	0
16	No data	No data	No data	No data	No data	No data	No data
17	Fusion to GGA3	0	Fusion to GGA6	1	0	0	0
18	–	1	Fusion to GGA19	2	0	0	1
19	–	0	Fusion to GGA15 and 18	0	0	0	0
20	Fusion to GGA13	0	Fusion to GGA5	0	0	0	0
21	–	0	Fusion to GGA2 and 23	0	0	0	0
22	–	0	–	2	0	0	1
23	–	2	Fusion to GGA21	2	0	0	0
24	–	0	–	1	0	0	0
25	No data	No data	No data	No data	No data	No data	No data
26	–	0	–	2	0	0	0
27	–	0	–	1	0	0	0
28	–	0	Fusion to GGA2 and 14	1	0	0	0
Z	–	0	–	0	0	0	0

The left-hand column represents the ancestral avian chromosome, with the subsequent columns indicating the number of inter- and intrachromosomal changes detected that have led to each extant species. For the intrachromosomal differences between ostrich and chicken, in the absence of an outgroup, the direction of change cannot be determined and thus only differences between the two species is noted

and single mapping approach. In this regard, therefore, our previous results generated somewhat of a paradox in that ostrich molecular branch lengths appeared short but the ostrich "genome rearrangement branch length" appeared relatively long. The results presented here however resolve this paradox by providing a new assembly in which there are fewer rearrangements in the ostrich genome.

The saker falcon and the budgerigar genomes

Among the *Psittaciformes* and *Falconiformes*, few studies of karyotype structure have been performed. Only one zoo-FISH study for each order [21, 44] has attempted to characterize the overall genome structure, finding the limited success common to most zoo-FISH studies. The chromosome painting study on the falcons revealed similar patterns of rearrangement between the peregrine

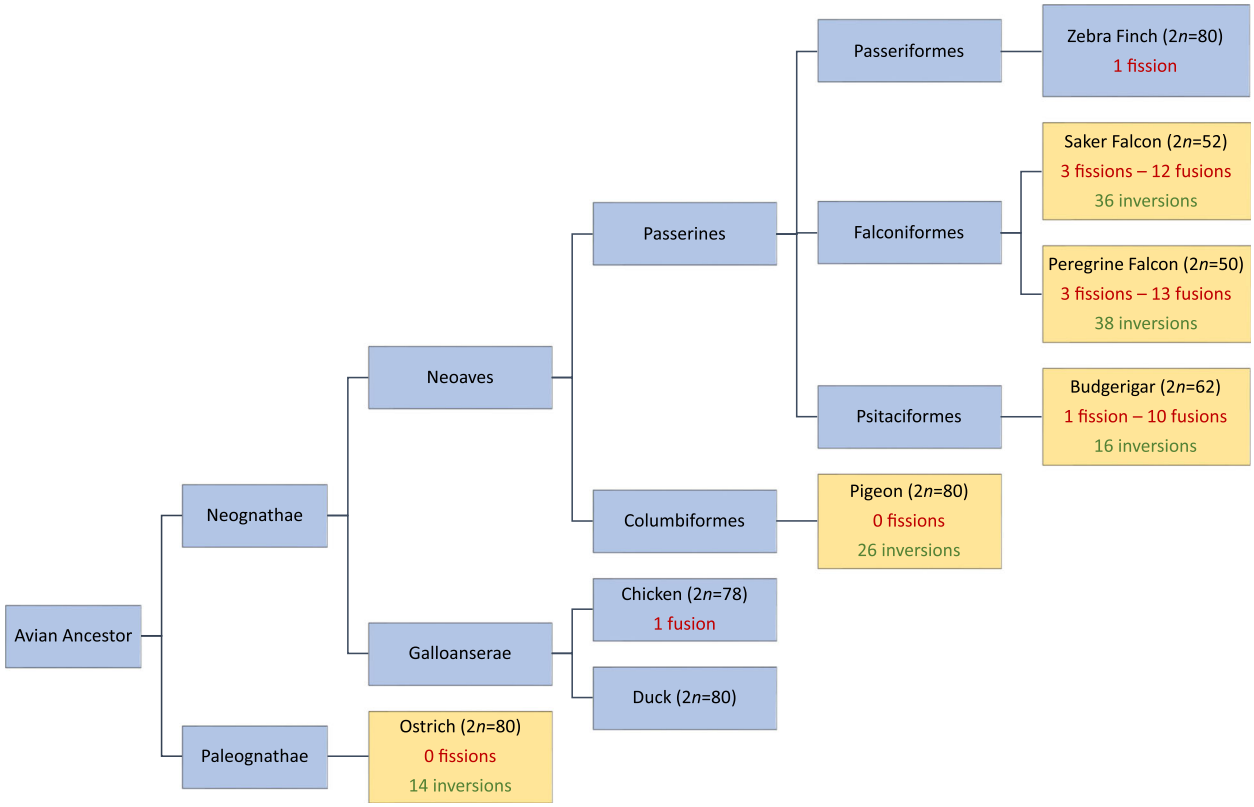


Fig. 6 Phylogenetic tree highlighting the relationship of species analyzed here and in our previous study [15] demonstrating the number of inter- and intrachromosomal rearrangements that have occurred relative to the avian ancestor (interchromosomal) and the chicken (intrachromosomal). Species investigated using our approach are highlighted in yellow with other species (chicken, duck, and zebra finch) represented for context. Phylogeny is based on Burleigh et al. [70]

falcon and the common kestrel (*Falco tinnunculus*) (2*n* = 52), but less similarity in the merlin (*Falco columbarius*) (2*n* = 40). The study focusing on *Psittaciformes* revealed patterns of similarity between the budgerigar, the cockatiel, and the peach-faced lovebird [44]. Common to both of these studies was a pattern of rearrangement that was similar among closely related species within the same order; however, when comparing the orders against each other, there were few parallels between them. In both studies, a lack of available tools capable of detecting the microchromosomes in the genome reorganization meant that results were limited to patterns involving the macrochromosomes only. Conversely, results presented here reveal previously undetected rearrangements involving microchromosomes, demonstrating that fusion is the most common mechanism of interchromosomal rearrangement, i.e., there was no evidence of reciprocal translocation. In some examples, particularly in the falcon genomes, multiple microchromosomes have fused together, but have still remained intact as discrete regions of conserved synteny, albeit fused to larger chromosomes. Also revealed through the chromosomal assembly of these genomes is a common breakpoint in the homolog of GGA1. Occurring in the same genomic region in both the budgerigar and the saker falcon, this breakpoint also occurs in the same region of the closely related zebra finch genome suggesting that this occurred in the Australavian ancestor of all three birds (zebra finch, falcon, and budgerigar) and was therefore already fixed in these three descendant lineages.

Table 4 Statistics for CNE density in 1-kb windows for avian EBRs, msHSBs, and genomewide

	Average no. CNE bases	Average density of CNE bases
Genome	86.85	0.087
msHSB	106.81	0.107
Intra	26.14	0.026
Fusion	13.76	0.013
Fission	5.48	0.005
EBR*	23.76	0.024

*Fission, fusion, and intrachromosomal EBRs combined

Intrachromosomal rearrangements

A comparison of the number of intrachromosomal rearrangements between the species tested here and those

assembled in our previous study [15] revealed that the fewest changes (when compared to the chicken genome as the reference) occurred in the ostrich, with evidence of only 14 inversions across the karyotype. Sharing a common ancestor over 100 mya [48], both the ostrich and the chicken are considered to be the most ancestral extant representatives of modern birds. These results suggest that their genomes also exhibit this ancestral pattern with little change between the two species. At the other extreme, the saker falcon examined here and peregrine falcon described previously [15] both exhibit a remarkably large number of changes (with an average of 37 inversions) consistent with the highly rearranged nature of the falcon genome. Surprisingly however, the budgerigar exhibits only 16 intrachromosomal rearrangements (similar to ostrich), suggesting that (unlike the falcons) the chromosomal rearrangement is limited to a pattern of overall interchromosomal change that once fixed, changed relatively little intrachromosomally. The difference between the number of inversions seen in the falcons and the budgerigar is surprising given that they have both been subject to so much chromosomal change. It may be that there is some biological advantage to this gross genomic structure in falcons that does not offer a selective advantage to the parrots, perhaps due to the high metabolic demands required by birds of prey.

Most studies into EBRs and HSBs have focused on mammals, many of which illustrate that EBRs tend to appear in gene-dense regions [49]. These "EBR genes" appear to be related to biological features specific to individual lineages [7, 8, 49]. A pattern of EBR reuse is also evident with some regions of the genome being particularly prone to chromosomal breakage [50, 51]. In fact, among birds (chicken, turkey, and zebra finch), it appears that breakpoint reuse occurs more often than is seen in mammals [43, 52], with previous data produced (comparing chicken and zebra finch) suggesting a key role for recombination-based mechanisms in the generation of chromosome rearrangements [53]. Larkin and colleagues argue that the presence of HSBs across multiple species is the result of a selective advantage to keeping particular combinations of genes together [49], with evidence of gene ontology enrichment for terms related to organismal development and the central nervous system, although some authors refute the notion that these proximity patterns occur or that there is any adaptive significance when they do (e.g., [54, 55]).

Here, however, we focus our studies on CNE distribution, indicating that CNEs are more depleted in EBRs generally but particularly in interchromosomal rearrangements—especially fission. Compared to our previous study [15] based on one interchromosomally rearranged genome (peregrine falcon), in this study, we used two additional genomes including one of which is phylogenetically distant from the peregrine falcon—the budgerigar. Our findings are, however, in line with what we found previously, demonstrating that in avian genomes the CNEs are important factors defining where rearrangements (especially the interchromosomal ones) are able to be fixed in evolution without leading to deleterious effects. This is further reinforced by the fact that chromosomal fissions in both studies are associated with genome intervals having no CNEs at all.

Species that exhibit a high degree of interchromosomal rearrangement (mammals, non-avian reptiles, and amphibians) all tend to have large, repeat-rich genomes that appear to correlate with a higher rate of rearrangement. The results presented here suggest that some avian lineages (such as the falcons and the parrots) also undergo a similar degree of chromosomal change but without the correspondingly large, repeat-rich genome. Instead, comparisons of the zebra finch and the budgerigar suggest that the high chromosomal mutation rates seen in both lineages may in fact be changes that have occurred in response to the exploitation of evolutionary niches, which ultimately end in fixed interchromosomal rearrangements. In the majority of other bird species however, it appears that such fixation is prevented, resulting in maintenance of an overall stable avian karyotype. A large number of CNEs in avian chromosomes (about twice as high as in the mammalian genomes) could form regulatory networks that cannot be altered, contributing to stability of chromosomes.

Why some rearrangements become fixed, and others do not, is a relatively understudied field, although clues may lie in the study of gene ontology terms present in EBRs. Farré and colleagues found a correlation between EBRs and specific avian adaptive features in individual species, including forebrain development in the budgerigar (one of the species investigated here), consistent with this species being not only a vocal learner but having distinctive neuronal connections compared to other vocal learners [39]. As more genomes become available with better assemblies, these analyses may well point to adaptive phenotypic features of individual orders and families.

Conclusions

By combining comparative sequence analysis, targeted PCR, and high-throughput molecular cytogenetics, the results presented here provide further evidence for an approach that is theoretically applicable to any animal genome as a cost-effective means of transforming fragmented scaffold-level assemblies to chromosomal level. The N50 of each genome was improved significantly, and a series of intra- and interchromosomal rearrangements that were previously undetectable were identified. Most bird genomes remain remarkably conserved in terms of their chromosome number (in 60–70% of

species $2n = \sim 80$) [43, 45, 46, 53], and interchromosomal changes are relatively rare, but when they do occur, they tend to be lineage specific, e.g., in *Psittaciformes* (parrots), *Falconiformes* (falcons), and *Sphenisciformes* (penguins) [15, 45, 56]. Fusion is the most common mechanism of change, there is no evidence yet of reciprocal translocation, and all microchromosomes remain "intact," even when fused to larger chromosomes. Why some groups exhibit a high degree of interchromosomal rearrangement remains unclear; some (e.g., kingfishers) have an unusually high ($2n = 130+$) number and both higher and lower deviations from the typical ($2n = \sim 80$) organization can occur in the same group. For instance, the Adélie penguin ($2n = 96$) and the emperor penguin ($2n = 72$) suggest that similar mechanisms can cause both a rapid reduction and a rapid increase in chromosome number. The short time period over which these changes occur in the penguins and the rearranged karyotypes of the *Falconiformes* and the *Psittaciformes* (but not the sister group, the *Passeriformes*) suggest that these changes can happen quickly. Vertebrates with large, repeat-rich genomes (such as mammals and amphibians) frequently demonstrate rapid intra- and interchromosomal rearrangements [31]. The results presented here suggest that birds too can undergo similar changes in certain groups although there is little evidence that these highly rearranged avian genomes are particularly large or more repeat rich than other avian genomes.

Methods

Avian genome assemblies, repeat masking, and gene annotations

The chicken (Gallus_gallus 4.0 [6]) and zebra finch (WUGSC 3.2.4 [57]) chromosome assemblies were downloaded from the UCSC Genome Browser [58]. The assemblies of saker falcon ostrich and budgerigar were provided by the Avian Phylogenomics Consortium [59]. All sequences were repeat-masked using Window Masker [60] with *-sdust* option and Tandem Repeats Finder [61]. Chicken gene (version of 27/04/2014) and repetitive sequence (version of 11/06/2012) annotations were downloaded from the UCSC Genome Browser [62]. Chicken genes with a single ortholog in the human genome were extracted from Ensembl Biomart (v.74 [63]).

Pairwise and multiple genome alignments, nucleotide evolutionary conservation scores, and conserved elements

Pairwise alignments using chicken and zebra finch chromosome assemblies as references and other assemblies as targets were generated with *LastZ* (v.1.02.00; [64]) and converted into the UCSC "chains" and "nets" alignment formats with the Kent-library tools ([58]). Conserved non-coding elements obtained from the alignments of 48 avian genomes were used [39].

Reference-assisted chromosome assembly of avian genomes

Saker falcon, budgerigar, and ostrich PCFs were generated using the Reference-Assisted Chromosome Assembly (RACA [17]) tool. We chose the zebra finch genome as reference and chicken as outgroup for the saker falcon and the budgerigar based on the phylogenetic distances between the species [65]. For the ostrich, we used chicken as the reference and zebra finch as outgroup and vice versa experiments were performed as the ostrich is phylogenetically equally distant from chicken and zebra finch. Two rounds of RACA were done for both species. The initial run was performed using the following parameters: *WINDOWSIZE=10 RESOLUTION=150000 MIN_INTRA-COV_PERC=5*. Prior to the second run of RACA, we tested the scaffold split during the initial RACA run using PCR amplification across the split intervals (see below) and adjusted the parameters accordingly as previously described [15].

PCR testing of adjacent SFs

Primers flanking split SF joints within scaffolds or RACA-predicted adjacencies were designed using Primer3 software (v.2.3.6 [66]). To avoid misidentification of EBRs or chimeric joints, we selected primers only within the sequences that had high-quality alignments between the target and reference genomes and found in adjacent SFs. Due to alignment and SF detection settings, some of the intervals between adjacent SFs could be > 6 kb and primers could not be chosen for a reliable PCR amplification. Whole blood was collected aseptically from adult saker falcon, ostrich, and budgerigar. DNA was isolated using DNeasy Blood and Tissue Kit (Qiagen) following standard protocols. PCR amplification was performed according to the protocol described in [15]. Briefly, PCR amplification was performed in a total volume of 10 μL as follows: 5 μL of DreamTaq Master Mix (Fermentas), 1 μL of each primer at 2 μM, and ≈ 30 ng DNA. PCR amplification was carried out in a T100 Thermal Cycler (BioRad) using the following profile: initial denaturation at 95 °C for 3 min, 32 cycles for 30 s at 95 °C, 1 min at 59 °C, and 1 min/kb at 72 °C. PCR products were stained with SYBR Safe (Invitrogen), separated in a 1.5% agarose gel, and visualized in a ChemiDOC MP system (Biorad).

Preparation of BAC clones for fluorescence in situ hybridization (FISH)

The full set of BAC clones reported in Damas et al. [15] as suitable for inter-species hybridization in birds were used for hybridization with saker falcon, budgerigar, and

ostrich metaphase chromosomes. All experiments were dual color. BAC clone DNA was isolated using the Qiagen Miniprep Kit (Qiagen) prior to amplification and direct labelling by nick translation. Probes were labeled with Texas Red-12-dUTP (Invitrogen) and FITC-Fluorescein-12-UTP (Roche) prior to purification using the Qiagen Nucleotide Removal Kit (Qiagen).

Cell culture and chromosome preparation

Chromosome preparations were established from fibroblast cell lines generated from collagenase treatment of 5- to 7-day-old embryos or from skin biopsies. Cells were cultured at 40 °C, and 5% CO_2 in Alpha MEM (Fisher), supplemented with 10% fetal bovine serum (Gibco), 1% Pen Strep/L-glutamine (Sigma). Chromosome suspension preparation followed the standard protocols, and brief mitostatic treatment with colcemid at a final concentration of 5.0 μg/ml for 1 h at 40 °C was followed by hypotonic treatment with 75 mM KCl for 15 min at 37 °C and fixation with 3:1 methanol:acetic acid.

Fluorescence in situ hybridization (FISH)

Metaphase preparations were fixed to slides and dehydrated through an ethanol series (2 min each in 2 × SSC, 70%, 85%, and 100% ethanol at room temperature). Probes were diluted in a formamide buffer (Cytocell) with Chicken Hybloc (Insight Biotech) and applied to the metaphase preparations on a 37 °C hotplate before sealing with rubber cement. Probe and target DNA were simultaneously denatured on a 75 °C hotplate prior to hybridization in a humidified chamber at 37 °C for 72 h. Slides were washed post-hybridization for 30 s in 2 × SSC w/ 0.05% Tween 20 at room temperature, then counterstained using VECTASHIELD anti-fade medium with DAPI (Vector Labs). Images were captured using an Olympus BX61 epifluorescence microscope with a cooled CCD camera and SmartCapture 3 (Digital Scientific UK) system.

EBR detection and CNE density analysis

Pairwise synteny blocks were defined using the maf2synteny tool [67] at 100, 300, and 500 kb resolution using the pairwise alignments obtained by lastZ. Using chicken as the reference genome, EBRs were detected and classified using the ad hoc statistical approach described previously [39]. All well-defined (or flanking oriented PCFs) fusion and fission points were identified from pairwise alignments with the chicken genome. Only the EBRs ≤ 100 kb were used for the CNE analysis. EBRs smaller than 1 kb were extended ± 1 kb. For each EBR, we defined two windows upstream (+ 1 and + 2) and two downstream (− 1 and − 2) of the same size as the EBR. We calculated the fraction of bases within CNEs in each EBR site, upstream and downstream windows. Differences in CNE densities were tested for significance using the Kruskall-Wallis test followed by Mann-Whitney *U* test. The CNEs analyzed were identical to those reported in Damas et al. [15].

Comparing CNE densities in EBRs and msHSBs

Chicken chromosomes (excluding GGA16, W and Z) were divided into 1-kb non-overlapping intervals. Only windows with > 50% of their bases with chicken sequence data available were used in this analysis. All intervals were assigned either to msHSBs > 1.5 Mb [39], avian EBR flanking: fusions, fissions, intrachromosomal EBR, and the intervals found in the rest of the chicken genome. We estimated the average CNE density for each window type and the distance, in number of 1-kb windows, between each window with the lowest CNE density (0 bp) and the nearest window with the average msHSB CNE density or higher. CNE densities were obtained using bedtools (v.2.20-1 [68]). Differences in distances between the two window types in msHSBs and EBRs were tested for significance using the Kruskall-Wallis test followed by Mann-Whitney *U* test. Thus, although the CNEs were the same as in Damas et al. [15], they were analyzed in the context of the new EBRs and mHSBs reported in this study.

Additional files

Additional file 1: Intrachromosomal rearrangements: BAC IDs and chromosomal orientation of clones (with start and stop coordinates from the chicken genome). The order of clones from the top to the bottom represents the order in which that appears on the chromosomes of the species of interest. Text in red indicates the p- (short) arm of the chromosome (where it is discernable). Data is listed in supplementary tables as follows: **Table S1.** Ostrich genome; **Table S2.** Budgerigar genome; **Table S3.** Saker falcon genome. (ZIP 58 kb)

Additional file 2: Chromosomal coordinates and orientation of mapped scaffolds and PCFs are listed by chromosome for each species. Data is listed in supplementary tables as follows: **Table S4.** Ostrich genome; **Table S5.** Budgerigar genome; **Table S6.** Saker falcon genome. (ZIP 87 kb)

Additional file 3: EBRs detected and genome position in relation to the chicken genome. Data is listed in supplementary tables as follows: **Table S7.** Ostrich genome; **Table S8.** Budgerigar genome; **Table S9.** Saker falcon genome. (ZIP 61 kb)

Funding

This work was supported in part by the Biotechnology and Biological Sciences Research Council [BB/K008226/1 and BB/J010170/1 to D.M.L, and BB/K008161/1 to D.K.G].

Availability of data and materials

O'Connor R.E*., Farré M*., Joseph S., Damas J., Kiazim L.G., Jennings R., Bennett S., Slack E.A., Allanson E., Larkin D.M., and Griffin, D.K. "Chromosome-level assembly reveals extensive rearrangement in saker falcon, budgerigar but not ostrich genomes." Data sets: doi:https://doi.org/10.5061/dryad.q70q40m [69].

Authors' contributions

REO'C was the post-doc lead on the project, providing the first draft of the manuscript and designing and performing the experimental cytogenetic analysis while SJ, LGK, RJ, and SB assisted with the cytogenetic work. MF was

the post-doc lead on the bioinformatic work while JD, EAS, and EA assisted with the bioinformatic analysis. DML and DKG conceived and designed the study, DML with an emphasis on bioinformatics, and DKG on molecular cytogenetics. DKG was the custodian of the manuscript and the first point of contact dealing with reviewer comments. All authors approved the final manuscript.

Ethics approval and consent to participate

No ethical approvals were required during the course of this project.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Author details

[1]School of Biosciences, University of Kent, Canterbury, UK. [2]Department of Comparative Biomedical Sciences, Royal Veterinary College, University of London, London, UK.

Received: 12 January 2018 Accepted: 24 September 2018
Published online: 24 October 2018

References

1. Foote AD, Liu Y, Thomas GWC, Vinař T, Alföldi J, Deng J, et al. Convergent evolution of the genomes of marine mammals. Nat Genet. 2015;47:272–5.
2. Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, et al. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. Nat Genet. 2017; 49:643–50.
3. Gordon D, Huddleston J, Chaisson MJP, Hill CM, Kronenberg ZN, Munson KM, et al. Long-read sequence assembly of the gorilla genome. Science. 2016;352:aae0344.
4. Koepfli K-P, Paten B, O'Brien SJ. The Genome 10K Project: a way forward. Annu Rev Anim Biosci Annu Rev. 2015;3:57–111.
5. Andersson L, Georges M. Domestic-animal genomics: deciphering the genetics of complex traits. Nat Rev Genet. 2004;5:202–12.
6. Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, et al. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature. 2004;432:695–716.
7. Groenen MAM, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, et al. Analyses of pig genomes provide insight into porcine demography and evolution. Nature. 2012;491:393–8.
8. Elsik CG, Tellam RL, Worley KC, Gibbs RA, Muzny DM, Weinstock GM, et al. The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science. 2009;324:522–8.
9. Jiang Y, Xie M, Chen W, Talbot R, Maddox JF, Faraut T, et al. The sheep genome illuminates biology of the rumen and lipid metabolism. Science. 2014;344:1168–73.
10. Teague B, Waterman MS, Goldstein S, Potamousis K, Zhou S, Reslewic S, et al. High-resolution human genome structure by single-molecule analysis. Proc Natl Acad Sci U S A. 2010;107:10848–53.
11. Mak ACY, Lai YYY, Lam ET, Kwok T-P, Leung AKY, Poon A, et al. Genome-wide structural variation detection by genome mapping on nanochannel arrays. Genetics. 2016;202:351–62.
12. Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, et al. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res. 2016;26(3):342–50.
13. Rhoads A, Au KF. PacBio sequencing and its applications. Genomics Proteomics Bioinformatics. 2015;13:278–89.
14. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009;326:289–93.
15. Damas J, O'Connor R, Farré M, Lenis VPE, Martell HJ, Mandawala A, et al. Upgrading short-read animal genome assemblies to chromosome level using comparative genomics and a universal probe set. Genome Res. 2017; 27:875–84.
16. Larkin DM, Daetwyler HD, Hernandez AG, Wright CL, Hetrick LA, Boucek L, et al. Whole-genome resequencing of two elite sires for the detection of haplotypes under selection in dairy cattle. Proc Natl Acad Sci U S A. 2012; 109:7693–8.
17. Kim J, Larkin DM, Cai Q, Asan ZY, Ge R-L, et al. Reference-assisted chromosome assembly. Proc Natl Acad Sci U S A. 2013;110:1785–90.
18. BirdLife International. Falco cherrug (amended version of 2016 assessment). The IUCN Red List of Threatened Species 2017: e.T22696495A110525916. http://dx.doi.org/10.2305/IUCN.UK.2017-1.RLTS.T22696495A110525916.en. 2017. Downloaded on 05 October 2018.
19. Wu Y, Hadly EA, Teng W, Hao Y, Liang W, Liu Y, et al. Retinal transcriptome sequencing sheds light on the adaptation to nocturnal and diurnal lifestyles in raptors. Sci Rep. 2016;6:33578.
20. Ferguson-Lees J, Christie DA. Raptors of the world: Houghton Mifflin Harcourt; 2001.
21. Nishida C, Ishijima J, Kosaka A, Tanabe H, Habermann FA, Griffin DK, et al. Characterization of chromosome structures of Falconinae (Falconidae, Falconiformes, Aves) by chromosome painting and delineation of chromosome rearrangements during their differentiation. Chromosom Res. 2008;16:171–81.
22. Bradbury JW, Balsby TJS. The functions of vocal learning in parrots. Behav Ecol Sociobiol. 2016;70:293–312.
23. Schaller NU, D'Août K, Villa R, Herkner B, Aerts P. Toe function and dynamic pressure distribution in ostrich locomotion. J Exp Biol. 2011; 214(Pt 7):1123–30.
24. Watson RR, Rubenson J, Coder L, Hoyt DF, Propert MWG, Marsh RL. Gait-specific energetics contributes to economical walking and running in emus and ostriches. Proc R Soc. 2011;278:2040–6.
25. Shetty S, Griffin DK, Graves JAM. Comparative painting reveals strong chromosome homology over 80 million years of bird evolution. Chromosom Res. 1999;7:289–95.
26. Guttenbach M, Nanda I, Feichtinger W, Masabanda JS, Griffin DK, Schmid M. Comparative chromosome painting of chicken autosomal paints 1-9 in nine different bird species. Cytogenet Genome Res. 2003;103:173–84.
27. Nishida-Umehara C, Tsuda Y, Ishijima J, Ando J, Fujiwara A, Matsuda Y, et al. The molecular basis of chromosome orthologies and sex chromosomal differentiation in palaeognathous birds. Chromosom Res. 2007;15:721–34.
28. Romanov MN, Farré M, Lithgow PE, Fowler KE, Skinner BM, O'Connor R, et al. Reconstruction of gross avian genome structure, organization and evolution suggests that the chicken lineage most closely resembles the dinosaur avian ancestor. BMC Genomics. 2014;15:1060.
29. Lewin HA, Larkin DM, Pontius J, O'Brien SJ. Every genome sequence needs a good map. Genome Res. 2009;19:1925–8.
30. Dobigny G, Britton-Davidian J, Robinson TJ. Chromosomal polymorphism in mammals: an evolutionary perspective. Biol Rev. 2017;92:1–21.
31. Eichler EE, Sankoff D. Structural dynamics of eukaryotic chromosome evolution. Science. 2003;301:793–7.
32. Molina WF, Martinez PA, Bertollo LAC, Bidau CJ. Evidence for meiotic drive as an explanation for karyotype changes in fishes. Mar Genomics. 2014;15:29–34.
33. Pardo-Manuel de Villena F, Sapienza C. Female meiosis drives karyotypic evolution in mammals. Genetics. 2001;159:1179–89.
34. Rieseberg LH. Chromosomal rearrangements and speciation. Trends Ecol Evol. 2001;16:351–8.
35. White MJD. Chromosomal rearrangements and speciation in animals. Annu Rev Genet Annual Rev. 1969;3:75–98.
36. O'Connor RE, Romanov MN, Kiazim LG, Barrett PM, Farré M, Damas J, et al. Reconstruction of the diapsid ancestral genome permits chromosome evolution tracing in avian and non-avian dinosaurs. Nat Commun. 2018;9:1883.
37. Murphy WJ, Larkin DM, Everts-van der Wind A, Bourque G, Tesler G, Auvil L, et al. Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. Science. 2005;309:613–7.
38. Branco MR, Pombo A. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. PLoS Biol. 2006;4:e138.
39. Farré M, Narayan J, Slavov GT, Damas J, Auvil L, Li C, et al. Novel insights into chromosome evolution in birds, archosaurs, and reptiles. Genome biol Evolution. 2016;8:2442–51.
40. Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, et al. Comparative genomics reveals insight into avian genome evolution and adaptation. Science. 2014;346:1311–21.
41. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature. 2005;438:803–19.

42. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. Nat Rev Genet. 2010;11:685–96.
43. Skinner BM, Griffin DK. Intrachromosomal rearrangements in avian genome evolution: evidence for regions prone to breakpoints. Heredity (Edinb). 2011; 108:37–41.
44. Nanda I, Karl E, Griffin DK, Schartl M, Schmid M. Chromosome repatterning in three representative parrots (Psittaciformes) inferred from comparative chromosome painting. Cytogenet Genome Res. 2007;117:43–53.
45. Griffin DK, Robertson LBW, Tempest HG, Skinner BM. The evolution of the avian genome as revealed by comparative molecular cytogenetics. Cytogenet Genome Res. 2007;117:64–77.
46. Christidis L. Animal cytogenetics, Vol. 4: Chordata 3. B. Aves. Berlin, Stuttgart: Gebrüder Borntraeger; 1990.
47. Zhang J, Li C, Zhou Q, Zhang G. Improving the ostrich genome assembly using optical mapping data. Gigascience. 2015;4:24.
48. Zheng Y, Wiens JJ. Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. Mol Phylogenet Evol. 2016;94:537–47.
49. Larkin DM, Pape G, Donthu R, Auvil L, Welge M, Lewin HA. Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. Genome Res. 2009;19:770–7.
50. Sankoff D. Genome rearrangement with gene families. Bioinformatics. 1999; 15:909–17.
51. Stankiewicz P, Lupski JR. Molecular-evolutionary mechanisms for genomic disorders. Curr Opin Genet Dev. 2002;12:312–9.
52. Lithgow PE, O'Connor R, Smith D, Fonseka G, Al Mutery A, Rathje C, et al. Novel tools for characterising inter and intra chromosomal rearrangements in avian microchromosomes. Chromosom Res. 2014;22(1):85–97.
53. Völker M, Backström N, Skinner BM, Langley EJ, Bunzey SK, Ellegren H, et al. Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution. Genome Res. 2010;20:503–11.
54. Singer GAC, Lloyd AT, Huminiecki LB, Wolfe KH. Clusters of co-expressed genes in mammalian genomes are conserved by natural selection. Mol Biol Evol. 2005;22:767–75.
55. Sémon M, Duret L. Evolutionary origin and maintenance of coexpressed gene clusters in mammals. Mol Biol Evol. 2006;23:1715–23.
56. Schmid M, Smith J, Burt DWDW, Aken BLBL, Antin PBPB, Archibald ALAL, et al. Third report on chicken genes and chromosomes 2015. Cytogenet Genome res. 2015;145:78–179.
57. Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Künstner A, et al. The genome of a songbird. Nature. 2010;464:757–62.
58. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res. 2002;12:996–1006.
59. Zhang G, Li B, Li C, Gilbert MTP, Jarvis ED, Wang J. Comparative genomic data of the Avian Phylogenomics Project. Gigascience. 2014;3:26.
60. Morgulis A, Gertz EM, Schäffer AA, Agarwala R. A fast and symmetric DUST implementation to mask low-complexity DNA sequences. J Comput Biol. 2006;13:1028–40.
61. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27:573–80.
62. Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, et al. The UCSC Genome Browser database: 2015 update. Nucleic Acids Res. 2015;43:D670–81.
63. Kinsella RJ, Kähäri A, Haider S, Zamora J, Proctor G, Spudich G, et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database (Oxford). 2011;2011:bar030.
64. Harris RS. Improved pairwise alignment of genomic dna: Pennsylvania State University; USA, 2007.
65. Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science. 2014;346:1320–31.
66. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3--new capabilities and interfaces. Nucleic Acids Res. 2012;40:e115.
67. Kolmogorov M, Raney B, Paten B, Pham S. Ragout--a reference-assisted assembly tool for bacterial genomes. Bioinformatics. 2014;30:i302–9.
68. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.
69. O'Connor R.E*., Farré M*., Joseph S., Damas J., Kiazim L.G., Jennings R., Bennett S., Slack E.A., Allanson E., Larkin D.M. and Griffin, D.K. 'Chromosome-level assembly reveals extensive rearrangement in saker falcon, budgerigar but not ostrich genomes'. Data sets: https://doi.org/10.5061/dryad.q70q40m.
70. Burleigh JG, Kimball RT, Braun EL. Building the avian tree of life using a large-scale, sparse supermatrix. Mol Phylogenet Evol. 2015;84:53–63.