



The
University
Of
Sheffield.

Evolutionary Genomics of the Zebra Finch *Z* Chromosome Inversion Polymorphism

Jake Andrew Pepper

A thesis submitted in partial fulfilment of the requirement for the degree of Doctor of
Philosophy

The University of Sheffield
Faculty of Science
School of Biosciences

September 2022

Acknowledgements

I would first like to thank the University of Sheffield School of Biosciences & the Adapting to the Challenges of a Changing Environment doctoral training partnership (ACCE DTP) funded by the Natural Environment Research Council (NERC) for their financial support and opportunities over the last four years.

Secondly, I would like to thank my PhD supervisors Professor Jon Slate and Dr Alison Wright for their support and advice during this project. Particularly, Jon for all his time and effort spent helping me produce this thesis.

Also thank you to my lab group, I've enjoyed learning about your research for the past four years. Special thanks to my future co-author Peter Price, whose mutual interest in this project has improved my own work.

Thank you to Dr Kai Zeng for his expert advice in interpreting the results of my analyses in Chapter 3.

Fran, thank you for helping me complete this project with my sanity intact. Thank you for listening to me talk for hours about my work and pushing me to write when I really didn't want to. This thesis wouldn't have gotten finished without you.

To my parents Hayley and Richard, and my brother Sam, thank you for always supporting me throughout my academic journey.

And finally, thank you to my pet finches (Eddie, Eleanor, Stephen, John, Betty, Canute, Aethelred, Emma, Sweyn, Henry, Harold, Geoffrey & Matilda) for reminding me that zebra finches are more than just their genomes.

Table of Contents

Abstract.....	8
Declaration.....	9
1. General Introduction.....	11
1.1. Inversion polymorphisms.....	11
1.2. Origins of supergenes.....	12
1.3. Selection pressures and evolutionary implications of supergenes	13
1.4. Examples of supergenes	15
1.5. Inversion polymorphisms in birds.....	17
1.6. A supergene in the zebra finch.....	18
1.7. The zebra finch; a model organism.....	20
1.8. Availability of zebra finch genome sequence	20
1.9. Thesis structure	21
2. Defining the Z chromosome karyotype of the zebra finch reference genome	24
2.1. Abstract	24
2.2. Introduction	25
2.3. Methods and Materials	29
2.3.1. Reference Genomes.....	29
2.3.2. Diagnostic SNPs.....	31
2.3.3. Identification of zebra finch reference genome inversion haplotypes using diagnostic SNPs and Principal component analysis.....	34
2.3.4. Using diagnostic SNPs to determine inversion karyotypes of publicly available sequence.....	35
2.4. Results	37
2.4.1. Reference genome inversion haplotypes	37
2.4.2. Inversion typing of assembly scaffolds	40
2.4.3. Identification of inversion haplotypes for other sequenced zebra finches	44
2.5. Discussion	48
2.6. Supplementary Material	51
3. Relaxation of purifying selection within the zebra finch Z chromosome inversion polymorphism.....	52
3.1. Abstract	52
3.2. Introduction	53

3.3. Methods and Materials	58
3.3.1. Sequence data	58
3.3.2. Reference genome	59
3.3.3. Variant calling	59
3.3.4. Calculation of summary statistics.....	59
3.3.5. Identifying synonymous and non-synonymous SNPs	60
3.4. Results	62
3.4.1. SNP density	62
3.4.2. Pi Diversity Statistic	64
3.4.3. Fixation Index (F_{ST}).....	67
3.4.4. Neutrality test statistics.....	69
3.4.5. Synonymous and non-synonymous SNPs	72
3.4.6. Fixed differences in coding regions between inversion haplotypes	74
3.5. Discussion	77
3.6. Supplementary Material	83
4. Estimating the age of the zebra finch Z chromosome inversion polymorphism	84
4.1. Abstract	84
4.2. Introduction	85
4.3. Methods and Materials	87
4.3.1. Sequence data	87
4.3.2. Identifying regions within the inversion polymorphism	88
4.3.3. Regions identified in previous studies.....	88
4.3.4. Regions identified using F_{ST} and using distribution of fixed differences between A and B haplotypes	92
4.3.5. Phylogenetic inference	93
4.3.6. PSMC analysis.....	94
4.4. Results	96
4.4.1. Regions identified using F_{ST}	96
4.4.2. Regions identified using distribution of fixed differences between A and B haplotypes.....	98
4.4.3. Phylogenetic inference	99
4.4.4. PSMC analysis.....	107
4.5. Discussion	109
4.6. Supplementary Material	113
5. General discussion	114

5.1. Key Findings	114
5.2. Divergence between supergene haplotypes.....	115
5.3. Limitations and future directions	116
5.4. Concluding Remarks	119
6. References.....	120
7. Supporting Material	129
Table S2.1:.....	129
Table S2.2:.....	155
Table S2.3:.....	169
Table S2.4:.....	173
Figure S2.1:	179
Figure S2.2:	180
Table S3.1:.....	181
Table S3.2:.....	186
Table S3.3:.....	191
Figure S3.1:	194
Figure S3.2:	195
Figure S3.3:	196
Figure S4.1:	197
Figure S4.2:	198
Figure S4.3:	199
Figure S4.4:	200
Figure S4.5:	201

List of Tables

Table 2.1: Summary of zebra finch reference genomes.	30
Table 2.2: A summary of inversion karyotype diagnostic SNPs.	32
Table 2.3: Counts of inversion karyotype diagnostic SNPs.	33
Table 2.4: Percentage supports for inversion haplotypes in each reference genome.	37
Table 2.5: Support for haplotypes with varying numbers of diagnostic SNPs.	41
Table 2.6: Inversion haplotypes of bTG1.4_alt scaffolds.	42
Table 2.7: Inversion haplotypes of birds from Singhal et al., 2015.	47
Table 3.1: SNP density on the Z chromosome.	64
Table 3.2: Nucleotide diversity (π) on the Z chromosome.	65
Table 3.3: Synonymous and nonsynonymous polymorphism and fixed differences across the Z chromosome.	72
Table 3.4: Genes containing >3 nonsynonymous fixed differences.	76
Table 4.1: Regions of the Z chromosome inferred by Kim et al., 2017.	90
Table 4.2: Corresponding bTG1.4 sequence for regions identified by Kim et al.,	92
Table 4.3: Sub-regions of the Z chromosome inversion.	97
Table 4.4: Summary of phylogenetic inferences.	105

List of Figures

Figure 1.1: A domesticated male zebra finch, <i>Taeniopygia guttata</i>.	19
Figure 2.1: Principal component analysis (PCA) of Z chromosome SNP genotypes.	34
Figure 2.2: PCA including reference assemblies.	39
Figure 2.3: Positions of diagnostic SNPs (A vs B) on Z chromosome.	43
Figure 2.4: PCA including 24 zebra finches from Singhal et al., 2015.	46
Figure 3.1: SNP density on the Z chromosome.	63
Figure 3.2: Nucleotide diversity (π) on the Z chromosome and across the genome.	66
Figure 3.3: Fixation index (F_{ST}) on the Z chromosome and across the genome.	68
Figure 3.4: Neutrality statistics on the Z chromosome.	70
Figure 3.5: Neutrality statistics across the genome.	71
Figure 4.1: Corresponding bTG1.4 sequence for regions identified by Kim et al.,	91
Figure 4.2: F_{ST} and inferred sub-regions of the inversion.	97
Figure 4.3: A vs B nonsynonymous fixed differences across the Z chromosome.	98
Figure 4.4: Phylogenetic inference using all 24 zebra finches and all SNPs within the inversion region (6.5 – 71.5Mb).	100
Figure 4.5: Phylogenetic inference excluding heterozygous birds.	101
Figure 4.6: Phylogenetic inference using SNPs within sub-regions of the inversion region.	105
Figure 4.7: N_e of inversion haplotypes inferred by PSMC.	108

Abstract

A large (~60-70Mbp) inversion polymorphism found on the Z chromosome of the model organism *Taeniopygia guttata* (the zebra finch), has previously been shown to have profound effects on sperm traits. There are three main haplotypes for this polymorphism (A, B & C). However, relatively little is known about how gene sequence and selection pressures vary between these inversion haplotypes or the phylogenetic relationship between them. In this thesis I explore the structure and molecular evolution of this inversion polymorphism. I produced a list of diagnostic single nucleotide polymorphisms (SNPs) which were used to determine the inversion karyotype of zebra finch Z chromosome sequence. This list was used to show that the male bird used to construct the zebra finch reference genome is heterozygous for the inversion polymorphism. However, I demonstrated that the two inversion haplotypes on the reference genome have not been successfully assembled or accurately phased. I used publicly available zebra finch sequence data to show evidence for a relaxation of purifying selection within the inversion region, particularly on the B haplotype, which has allowed sequence divergence between the haplotypes to occur. Finally, I showed, using phylogenetic inferences, that the inversion polymorphism emerged less than 1 million years ago, and the ancestral form is likely to be Haplotype A, with the other two haplotypes derived from separate inversion mutations. Multiple sub-regions of the inversion region were identified, which may indicate the locations of each of the inversions that established the derived haplotypes. The work presented in this thesis adds to the existing literature on supergenes, which are of increasing interest due to their dramatic impact on phenotypic variation and adaptive evolution in a wide variety of species.

Declaration

The following people were involved in this research project:

Jake Pepper, Jon Slate, Alison Wright

Chapter 2: Defining the Z chromosome karyotype of the zebra finch reference genome

J.P & J.S conceived the idea for the study;

J.P performed the analyses and wrote the manuscript;

J.S commented on drafts of the manuscript.

Chapter 3: Relaxation of purifying selection within the zebra finch Z chromosome inversion polymorphism

J.P & J.S conceived the idea for the study (with input from A.W);

J.P performed the analysis and wrote the manuscript;

A.W provided guidance on methodology;

J.S commented on drafts of the manuscript.

Chapter 4: Estimating the age of the zebra finch Z chromosome inversion polymorphism

J.P & J.S conceived the idea for the study;

J.P performed the analysis and wrote the manuscript;

J.S commented on drafts of the manuscript.

I, the author, confirm that this thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not been previously presented for an award at this, or any other, university.

A finch asks his mother...
“Mum, why does my beak look
different to yours?”
she replies,
“Well son, I hate to break it
to you, but you’re adapted”

1. General Introduction

1.1. Inversion polymorphisms

Chromosome inversions are a class of genetic mutation defined as the rearrangement of DNA sequence produced by excision of a DNA fragment between two breakpoints within the same chromosome, and the subsequent reinsertion of this fragment in reverse orientation. In most cases, the rearrangement of a chromosome region in this way will not affect the function of any genes within that region. Therefore, inversions do not usually have any phenotypic consequence unless a critical gene is disrupted, either through directly disrupting its sequence, or by altering its position relative to either a regulatory sequence or an imprinting centre (Anton *et al.*, 2005).

Even if no critical genes are affected, inversions may still be deleterious because of reduced fitness of heterozygotes due to the impact inversions have on gametogenesis. During meiosis, the method of sequence pairing between inversion segments in heterozygotes depends on the length of the inversion. Very short inversions are not able to align their homologous regions and often remain as asynaptic “balloons” (Winsor *et al.*, 1978). Larger inversions will achieve homologous pairing by twisting and folding the inverted segment into a structure known as an inversion loop (Anton *et al.*, 2005). In the case of the largest of inversions (those involving most of the length of the chromosome) synapsis will either occur across the length of the inversion region, leaving terminal noninverted regions unpaired, or each terminal region of the chromosome will pair, leaving the inversion region unpaired with no inversion loop formed (Anton *et al.*, 2005). When homologous sequence does pair, recombination events may occur; and if chiasmata form within an inverted region, large deletions and duplications of sequence can occur. In the case of paracentric inversions (where the centromere is located outside the inverted region), a chiasmata within the inverted segment would result in a dicentric and an acentric fragment, and consequently unbalanced gametes (Anton *et al.*, 2005; Miller, 2020) without a full haploid set of genes. See Figure 1 of Yapan *et al.*, (2013) for a diagram of an inversion loop in a paracentric inversion. In summary, inversions may be highly deleterious due to their tendency to produce unbalanced gametes, large duplications, deletions, and their potential to cause meiotic failure. Most inversion mutations are highly selected against and therefore they will not persist within a population. The larger the inversion, the more likely it is to be harmful.

Despite their potential deleterious effects, inversion mutations may still segregate within a population alongside a non-inverted, ancestral sequence, resulting in an inversion polymorphism. These inversion polymorphisms are found in many different species, with over 1000 predicted inversions in the human genome (Martínez-Fundichely *et al.*, 2014; Puig *et al.*, 2015), and many experimentally proven inversion polymorphisms in *Drosophila* and other dipterans (Hoffmann, Sgrò and Weeks, 2004). In addition, inversion polymorphisms are prevalent in many plant species (Huang and Rieseberg, 2020), and are also found in many species of birds (Lamichhaney *et al.*, 2015; Zinzow-Kramer *et al.*, 2015; Knief *et al.*, 2016; Kim *et al.*, 2017).

1.2. Origins of supergenes

One of the most prominent features of inversion polymorphisms is their ability to suppress recombination within the inverted region. This occurs because synapsis is inhibited or because of the creation of unbalanced gametes in inversion heterozygotes (Anton *et al.*, 2005). This reduced recombination leads to high linkage disequilibrium across the region, which preserves alternate combinations of alleles in each inversion haplotype (Hoffmann and Rieseberg, 2008). The resulting tightly linked loci can then be inherited together as if they were a single Mendelian locus, a concept known as a supergene (Thompson and Jiggins, 2014; Charlesworth, 2016). Supergenes have been studied for almost a hundred years, with early observations noting that genes responsible for colour-pattern morphs in certain species of grouse locusts formed tightly linked groups, with complete suppression of crossing over among the genes (Nabours, 1929; Nabours, Larson and Hartwig, 1933; Darlington and Mather, 1950).

Of course, early studies of supergenes took place before the structure of DNA was even known. More recently, advances in genomic sequencing and bioinformatics have facilitated research into the structure of supergenes (Gutiérrez-Valencia *et al.*, 2021). Characterising this structure has helped us to better comprehend the underlying genetic mechanisms of recombination suppression between supergene haplotypes (Berdan, Flatt, *et al.*, 2022). Whilst inversion polymorphisms can reduce recombination and produce a supergene, there are other genetic mechanisms such as chromosomal translocations which can also reduce effective recombination across a region and produce a supergene (Morel *et al.*, 2004; Talukdar, 2010). A large region of suppressed recombination can encompass many whole genes, causing elevated linkage disequilibrium between loci separated by large physical distances. Potentially this can facilitate co-adaptation between these loci, because recombination is un-likely to move

alleles onto different haplotypes to those on which they arose (Hoffmann and Rieseberg, 2008). In this way supergenes are able to control complex phenotypes, whilst maintaining the inheritance pattern of a single Mendelian locus (Thompson and Jiggins, 2014).

1.3. Selection pressures and evolutionary implications of supergenes

The influence of different types of selection on a supergene polymorphism determines its phenotypic impact and evolutionary trajectory. One of the defining features of a supergene is the presence of multiple alternate complex phenotypes within a single population (Thompson and Jiggins, 2014). These alternate phenotypes are often caused by positive selection acting on each supergene haplotype (combinations of linked alleles), which drives the evolution of separate adaptations within each haplotype. These alternative adaptive alleles can accumulate and spread within a haplotype, but the lack of recombination prevents them from spreading beyond the haplotype they arose in, therefore driving sequence divergence. In this way, supergenes act as a mechanism to allow multiple advantageous trait combinations to persist separately within one population, which would normally be impossible due to the effects of recombination.

Suppressed recombination also allows increased co-adaptation of genes within an inversion. Since recombination events between co-adapted alleles will split up advantageous haplotypes, suppression of recombination means alleles are less likely to be removed from the haplotypes in which they arose and therefore favourable combinations of alleles can be maintained (Hoffmann and Rieseberg, 2008). It has been theorised that these alternative adaptations caused by supergenes could facilitate speciation, as the large adaptive differences between haplotypes could lead to impaired fitness of heterozygotes. Consequently, homozygous carriers of the inverted haplotype can become an emerging species (Rieseberg, 2001).

Whilst the reduced rate of recombination between supergene haplotypes can facilitate the acquisition of beneficial alleles, it may also speed up the accumulation of deleterious mutations. Normally, deleterious mutations would be removed from populations through the action of purifying selection. However, suppressed recombination reduces the efficacy of purifying selection, meaning linked deleterious alleles are able to persist within supergenes (Gutiérrez-Valencia *et al.*, 2021; Berdan, Flatt, *et al.*, 2022). Purifying selection's efficacy is

diminished as the reduction in recombination within a supergene region creates a localised (in a genomic context) form of population substructure. As opposite supergene haplotypes cannot recombine, the effective population size for each arrangement of the supergene is reduced, thereby reducing efficacy of purifying selection, especially in the rarer haplotype. (Berdan, Blanckaert, *et al.*, 2022). As deleterious mutations accumulate, they can disrupt genes and other important genetic sequence, resulting in degeneration of one of the haplotypes. It is theorised that this degeneration could destabilise polymorphic supergene systems, and result in loss of the polymorphism entirely, as one haplotype is selected against and removed (Berdan, Blanckaert, *et al.*, 2022; Berdan, Flatt, *et al.*, 2022). It has also been suggested that reduced recombination due to an inversion polymorphism could be the driving force in the evolution of sex chromosomes from autosomes, and degeneration of one sex chromosome could be the result of relaxed purifying selection allowing accumulation of deleterious alleles (Bourke and Mank, 2013; Wright *et al.*, 2016).

Sequence divergence between haplotypes, driven either by positive selection or by relaxed purifying selection will maintain polymorphism in the overall population. Under positive selection, each haplotype may acquire, through mutation, different alleles which confer increased fitness and are driven to fixation. Under relaxed purifying selection, new deleterious alleles may not be purged, and instead will persist, potentially even reaching fixation, within the haplotype they arose in. Intuitively, one would expect haplotypes carrying a large number of deleterious alleles to be purged from the population. However, many known supergenes are over one million years old (Gutiérrez-Valencia *et al.*, 2021). For supergene polymorphisms to be maintained for this long, the polymorphism needs to be protected by balancing selection; this can occur in multiple forms such as overdominance (heterozygote advantage), antagonistic selection and negative frequency-dependent selection (Berdan, Flatt, *et al.*, 2022).

In summary, the reduced recombination inherent in supergenes not only allows the formation and maintenance of multi-locus complexes, but also puts them at risk of degeneration. In addition, reduced recombination has an impact on the multiple forms of selection pressures supergenes are subjected to, including positive, balancing, and purifying selection. Understanding which of these selection pressures is acting on a given supergene polymorphism is crucial to understanding its evolutionary trajectory and its potential impact on the future of natural populations.

1.4. Examples of supergenes

There are many prominent examples of supergenes in nature, responsible for variation in a variety of complex phenotypes which would usually be controlled by many loci. Supergenes control significant aspects of morphology and behaviour and are found in a wide variety of taxa.

An example of a complex behavioural phenotype controlled by a supergene can be found in fire ants *Solenopsis invicta*. A supergene-containing chromosome is responsible for two alternate forms of social organisation. The first is a colony in which there is only a single queen ant, while the second can be comprised of dozens of reproductive queens (Wang *et al.*, 2013; Purcell *et al.*, 2014). The social structure of a colony is determined by a large (~13 Mb) chromosome region, characterised by a pair of ‘social chromosomes’, SB and Sb (Wang *et al.*, 2013). Recombination is suppressed between the two variants of this region. Depending on the environmental conditions either of these two forms of social structure can be favoured by selection. In saturated and harsher environments multiple-queen colonies (where all queens are SB/Sb) are favoured to single queen colonies (where queens are SB/SB). However, single queen colonies are able to undergo long-distance dispersal to colonise unsaturated and resource-rich environments (Ross and Keller, 1995; Pracana *et al.*, 2017). As a result of this, positive selection acts on both supergene haplotypes in an environment-specific context, which maintains the polymorphism. Multiple-queen colonies will kill the queen in single-queen colonies if the social forms come into contact. However, the Sb/Sb karyotype is always lethal meaning the SB variant would be maintained even if multiple-queen colonies were always favoured.

Given that Sb/Sb is a lethal karyotype, Sb is comparable to Y or W chromosomes in diploid species, as it is non-recombining over much of its length. Only regions outside the supergene are able to recombine with comparable sequence on the SB chromosome, similar to a pseudo-autosomal region on a sex chromosome (Wang *et al.*, 2013; Pracana *et al.*, 2017). The supergene seen in fire ants is therefore a good example of how reduced recombination caused by inversion polymorphisms could result in the formation of sex chromosomes (Wright *et al.*, 2016).

Another example of a supergene controlling behaviour is found in Atlantic cod, where two different ecotypes coexist. These ecotypes correspond to haplotypes of an inversion on Chromosome 1, where fish possessing one haplotype are ‘stationery’ and fish with the opposite

haplotype or which are heterozygous, exhibit migratory behaviour (Berg *et al.*, 2016, 2017; Matschiner *et al.*, 2022). As recombination is suppressed within this supergene region, mildly deleterious mutations are more likely to persist within the inverted region. This accumulation of mutational load would lead to reduced fitness of individuals which are homozygous for either inversion haplotype. However, this is mitigated by gene conversion and double crossover, which allow some exchange of alleles between the two inversion haplotypes during meiosis (Matschiner *et al.*, 2022). The supergene polymorphism is therefore maintained. Even with the rare genetic exchange, the sequence has diverged within the region sufficiently for mating between ecotypes to now only occur at a low frequency (Bradbury *et al.*, 2014; Berg *et al.*, 2017; Matschiner *et al.*, 2022). This example illustrates how a supergene polymorphism could facilitate speciation.

Supergenes are also found in plants, for example in the yellow monkeyflower, *Mimulus guttatus*, in which an inversion polymorphism controls the life-history of an individual plant (Lowry and Willis, 2010). Depending on which arrangement of the inversion it possesses, the plant can either be an annual ecotype which completes its life cycle within one growing season, or a perennial ecotype which lives for multiple years (Lowry and Willis, 2010). The annual ecotype is commonly found in locations with reduced water availability in the summer (such as the Mediterranean) whilst the perennial ecotype is found in regions with high soil moisture year-round (Lowry and Willis, 2010). This suggests that each ecotype has higher fitness depending on environmental conditions, which maintains the supergene polymorphism, and suggests a link between supergenes and evolutionary adaptation (Lowry and Willis, 2010).

A particularly striking example of behavioural and morphological polymorphism due to a supergene exists in the ruffs, *Philomachus pugnax*, a lek-breeding wading bird. A large supergene region on chromosome 11 determines three alternative male mating strategies known as ‘independent’, ‘satellite’ and ‘faeder’ (Küpper *et al.*, 2015; Lamichhaney *et al.*, 2015). The independent strategy corresponds with the ancestral inversion haplotype, and involves males with black feather ornamentation aggressively defending territories and females, whilst the derived inversion haplotype corresponds with ‘faeder’ males, which mimic females in order to copulate with females in the courts of independent males (Küpper *et al.*, 2015; Lamichhaney *et al.*, 2015). A third haplotype, ‘satellite’ is believed to have originated through recombination or gene conversion between the inverted faeder and non-inverted independent alleles (Küpper *et al.*, 2015). This haplotype results in semi-cooperative males with white ornamental plumage. All three male mating morphs coexist as a balanced

polymorphism and are a good example of the high complexity of phenotypes which can be determined by a supergene.

A similar example is found in another bird species, white-throated sparrows *Zonotrichia albicollis*, where a supergene is responsible for an alternative pigmentation pattern and components of social behaviour, resulting in two distinct morphs coexisting in the sparrow population (Huynh, Maney and Thomas, 2011). This supergene is found on chromosome 2, and the chromosome exists in two different arrangements ZAL2 and ZAL2^m, which determine variation in plumage, social behaviour and mate choice (Huynh, Maney and Thomas, 2011; Zinzow-Kramer *et al.*, 2015). Sparrows which are homozygous for the ZAL2 arrangement have tan-coloured head plumage and engage in more parental care, whilst sparrows which are heterozygous for the supergene have white-coloured head plumage, are more aggressive and engage in less parental care (Zinzow-Kramer *et al.*, 2015). Birds mate disassortatively with respect to morph. Individuals which are homozygous for the ZAL2^m arrangement are extremely rare, because mate pairings between white-coloured birds (which are necessary to produce ZAL2^m/ZAL2^m offspring) are extremely rare. This disassortative mating means the ZAL2^m arrangement is almost always found in heterozygous birds and therefore recombination within the supergene region is extremely suppressed (Huynh, Maney and Thomas, 2011), leading to genetic differentiation and variation in gene expression within the region (Zinzow-Kramer *et al.*, 2015). There is also evidence to show that ZAL2^m is degrading due to the extreme lack of recombination; this process is reminiscent of the theorised origin of Y and W sex chromosomes (Tuttle *et al.*, 2016).

These examples illustrate the wide variety of taxa which possess supergene polymorphisms, as well as the complexity of phenotypes which can be determined by them. The variety of evolutionary consequences which can result from a supergene polymorphism is also apparent, with adaptation to local environments as seen in the yellow monkeyflower, the degradation of a chromosome as seen in white-throated sparrows, the formation of pseudo-sex chromosomes as seen in fire ants, the potential facilitation of speciation as seen in Atlantic cod and the balancing of extreme phenotypic variation within a single species as seen in ruffs.

1.5. Inversion polymorphisms in birds

Inversions are generally deleterious due to their structural under-dominance (i.e. lower fitness of heterozygotes), and so new inversions should be selected against and therefore be rare, especially large-scale ones. However, in some taxa, inversion rearrangements evolve more

frequently, for example in Estrildid finches (Hooper and Price, 2015). In general, large-scale genomic mutations persist infrequently in avian species. Inter-chromosomal rearrangements such as fusions, fissions and translocations all occur relatively rarely (Ellegren, 2013; Zhang *et al.*, 2014) meaning the diploid chromosome number for birds only varies between $2n = 76$ and 80. Synteny (Shetty, Griffin and Graves, 1999; Ellegren, 2010), and to a lesser extent gene order (Stapley *et al.*, 2008), is highly conserved. In contrast to the generally conserved chromosome number and the rarity of inter-chromosomal rearrangements, inversions occur much more frequently (Hooper and Price, 2015). The average rate of pericentric inversion fixation across species of Estrildid finch is once every 2.26 million years (Hooper and Price, 2017). This high rate of inversions relative to inter-chromosomal rearrangements is a pervasive feature of all passerine birds, an avian order which contains ~60% of all bird species, and so it has been suggested that inversions are common across all avian taxa (Hooper and Price, 2017).

The relatively high prevalence of inversions occurring during the evolution of bird lineages would suggest that supergenes in extant species should also be common. As expected, there are multiple prominent examples of supergenes in birds, including in ruffs (Lamichhaney *et al.*, 2015), in white-throated sparrows (Zinzow-Kramer *et al.*, 2015) and in the zebra finch *Taeniopygia guttata* (Kim *et al.*, 2017; Knief *et al.*, 2017). The zebra finch (Figure 1.1) is an Australian songbird (order Passeriformes, suborder oscine) of the *Estrildidae* family, and part of a sub-clade of species which has been suggested as a sub-family named *Poephilinae* (Olsson and Alström, 2020). This sub-clade was found to have the highest rate of inversion fixation across all estrildid finches (Hooper and Price, 2017), and so it is perhaps not surprising that a member of this clade possesses segregating supergenes.

1.6. A supergene in the zebra finch

The zebra finch genome is known to contain a number of large inversion polymorphisms (Itoh *et al.*, 2011; Knief *et al.*, 2016). One of these inversions encompasses the majority of the Z chromosome (approximately 60Mbp), and was discovered using cytological studies (Itoh *et al.*, 2011). It appears to be segregating at high frequencies in all wild and captive zebra finch populations that have been studied to date (Knief *et al.*, 2016, 2017; Kim *et al.*, 2017). Since this supergene is located on a sex chromosome, recessive alleles will always be expressed in females (ZW), but less often in males (ZZ). This could mean that selection against deleterious recessive alleles affecting females will be stronger than against deleterious recessive alleles affecting males. Therefore, segregating variation might be more likely to

affect traits expressed in males. In fact, previous work on the zebra finch supergene has identified three distinct inversion haplotypes on the Z chromosome (A, B & C), which are responsible for 67–90% of the additive genetic variance in sperm morphology within male finches. AA males are characterised by sperm with long overall length, AB and AC males have intermediate length sperm, and BB, BC and CC males have shorter sperm (Kim *et al.*, 2017; Knief *et al.*, 2017).

The midpiece to tail size ratio is an important factor in explaining variation in sperm motility (Bennison *et al.*, 2016), with an intermediate ratio producing sperm with the fastest swimming speed or motility, (in this case males with haplotype AB and AC). Heterozygote advantage due to this increased sperm motility may result in balancing selection, maintaining the supergene polymorphism within the zebra finch population (Kim *et al.*, 2017; Knief *et al.*, 2017). Other selection pressures may also be acting on the zebra finch inversion polymorphism, and the sequence variation underlying sperm variation, and potentially other phenotypic variation is yet to be discovered.

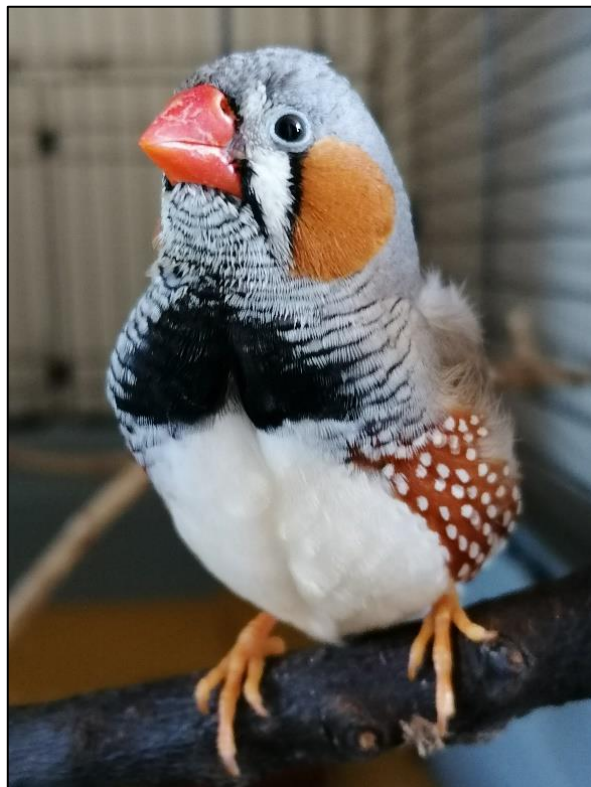


Figure 1.1: A domesticated male zebra finch, *Taeniopygia guttata*.
Photo credit: J. Pepper

1.7. The zebra finch; a model organism

Biological research involving zebra finches began around 70 years ago (Morris, 2008), and since then the species has become an important model organism. For example, due to their ability to learn vocalisations, they have been used to understand the neurobiological basis of vocal learning in birds (Mello, 2014). These learned vocalisations show many parallels to humans, and so zebra finches became an important model for human neurobiology (Doupe and Kuhl, 1999). In addition, zebra finches live in large flocks, are highly social, and have a high complexity in their social interactions, meaning they are widely studied in behavioural ecology (Galoch and Bischof, 2007; Guillette and Healy, 2014). They are also highly sexually dimorphic, varying in size, weight, feather and beak colouration, and behaviour, meaning studies of sexual selection in the species are common (Birkhead, Fletcher and Pellatt, 1998; Simons and Verhulst, 2011).

The status of the zebra finch as a model organism explains why there are efforts to understand as much as possible of the species genetics. The supergene found on the zebra finch Z chromosome could have an impact on phenotypic variation within the species, and subsequently impact any number of studies which utilise the zebra finch. For this reason, it is relevant to understand the origin and implications of the supergene.

1.8. Availability of zebra finch genome sequence

The volume and wide variety of research involving zebra finches resulted in the species being chosen as only the second avian species and first passerine to have its genome sequenced (Warren *et al.*, 2010). This reference genome was produced using the Sanger method to produce sequence reads, but in the following decade, the introduction of next generation sequencing (NGS) technologies has caused the field of genomics to grow exponentially (Goodwin, McPherson and McCombie, 2016). The source of DNA for the first zebra finch reference genome (TaeGut3.2.4) was a single domesticated male zebra finch, “Black17” (Warren *et al.*, 2010). An updated version of the zebra finch reference genome (bTG1.4), also using DNA from “Black17”, was produced in 2021 by the Vertebrate Genomes Project (VGP) (Rhie *et al.*, 2021). This updated zebra finch assembly used NGS technologies and modern assembly techniques, rather than the Sanger sequencing technology used to produce the original reference. The VGP also produced a second reference genome (bTG2) using DNA obtained from a female zebra finch, “Blue55”. Both the updated male reference genome (bTG1.4) and the female reference (bTG2) were produced using the VGP’s assembly pipeline,

which is intended to generate high-quality, complete reference genomes for all of the roughly 70,000 extant vertebrate species. The pipeline utilises PacBio long-read sequence data and complements it with Linked Reads, Hi-C and long-range optical mapping (Rhie *et al.*, 2021). This updated methodology resulted in dramatic improvements to the old reference, including placing 68.5Mb of previously unplaced sequence into chromosomes, and assembling seven new chromosomes (Chr 30-36). Like most passerines, the diploid number of chromosomes in the zebra finch genome is $2n = 80$ (Santos *et al.*, 2017), and this updated reference genome has assemblies of all of these chromosomes (39 autosomes and two sex chromosomes, Z and W). The assembly also corrects many mis-joins, false gene duplications and other substantial errors from the previous iteration. This newly updated set of reference genomes could be an invaluable tool in investigating the zebra finch Z chromosome supergene.

In addition to the reference genome, a previous study into recombination hotspots in the zebra finch and the long-tailed finch has made sequence reads for multiple individuals publicly available (Singhal *et al.*, 2015). These zebra finch sequence reads could be repurposed for studies investigating the Z chromosome.

1.9. Thesis structure

In this thesis I will determine the inversion karyotypes of the zebra finch reference genomes and a number of individuals whose genome sequences are publicly available. I will use this publicly available sequence to determine the approximate boundaries of the Z chromosome supergene and investigate the types of selection that have acted on the inversion polymorphism, specifically on the two most commonly occurring haplotypes, A and B. Furthermore, I will investigate the relative ages and the demographic history of each of the inversion haplotypes. The remainder of this thesis is structured as follows:

Chapter 2: Defining the Z chromosome karyotype of the zebra finch reference genome

A large (~60-70Mbp) inversion polymorphism found on the Z chromosome of the model organism *Taeniopygia guttata* (the zebra finch), has previously been shown to have profound effects on sperm traits. The recent zebra finch assembly produced by the Vertebrate Genome Project (VGP) has improved the reference genome's quality and completeness, compared to previous iterations. However, the inversion karyotype of the bird used to create

the reference genome is still unknown. I used a set of SNPs which are highly diagnostic of the inversion polymorphism to determine:

- The inversion karyotype of the male zebra finch used to create both the original Sanger reference, and the more recent reference produced by the VGP.
- The inversion karyotype of the female zebra finch that was also recently assembled by the VGP.
- The extent to which the inversion polymorphism on the Z chromosome is correctly assembled both at the phased haplotype assembly level and within the underlying scaffolds.
- The karyotypes of 24 zebra finches originally sequenced in 2013 (Singhal *et al.*, 2015)

Chapter 3: Relaxation of purifying selection within the zebra finch Z chromosome inversion polymorphism

The evolutionary trajectories and implications of supergenes, such as the one found on the Z chromosome of the zebra finch, are determined by a variety of selection pressures. These can include balancing, positive and purifying selection, as well as genetic drift, which can be stronger within inversions than outside of them. These selection pressures may be driving the sequence divergence between inversion haplotypes which determine sperm variation in zebra finches, and potentially other phenotypic variation yet to be discovered. In this Chapter I used sequence data, from birds whose inversion karyotype was determined in Chapter 2, to determine the types of selection that have acted on the inversion polymorphism on the Z chromosome in zebra finches. Focussing on the two most commonly occurring haplotypes, A and B. I used this sequence data to determine:

- The approximate breakpoints of the inversion region
- Evidence of selection which has acted on the sequence within these breakpoints. This was done using a range of summary statistics such as nucleotide diversity, Fixation Index, Tajima's D, Fay and Wu's H and a composite test of D and H known as Zeng's E
- The amount of non-synonymous and synonymous variation within and between supergene haplotypes, as a further tool to infer selection.

Chapter 4: Estimating the age of the inversion polymorphism on the Z chromosome in the zebra finch

Previous work on the zebra finch supergene has identified three distinct inversion haplotypes on the Z chromosome (A, B & C), which are responsible for 67-90% of the additive genetic variance in sperm morphology within male finches. The history of each of these haplotypes was investigated in this chapter. Specifically, using publicly available sequence data (from birds of known karyotype - see Chapter 2) I sought to determine:

- The identity of the ancestral inversion haplotype and the relative order of divergence of the two derived haplotypes.
- The time since divergence of each of these three inversion haplotypes.
- Specific regions of the Z chromosome inversion that best distinguish between specific derived haplotypes.
- The demographic histories of each haplotype.

Chapter 5: General Discussion

Chapter 5 synthesises the findings of Chapters 2-4 and discusses how these results contribute to the understanding of the zebra finch Z chromosome supergene. I also discuss how these findings impact our understanding of zebra finches as species, and the role of supergenes in the evolution of birds and other taxa. I identify some remaining questions that could be investigated in order to build upon the findings of this thesis.

2. Defining the Z chromosome karyotype of the zebra finch reference genome

2.1. Abstract

A large (~60-70Mbp) inversion polymorphism found on the Z chromosome of the model organism *Taeniopygia guttata* (the zebra finch), has previously been shown to have profound effects on sperm traits. However, relatively little is known about differences in gene sequence and gene expression between the different inversion haplotypes. As the zebra finch is a key model organism in numerous fields, it was only the second bird species, and first passeriform to have its genome sequenced. More recently, the Vertebrate Genome Project (VGP), set up by the international genome 10K consortium, has attempted to re-assemble, improve, and phase the zebra finch reference genome. The recent zebra finch assembly produced by VGP has improved the reference genome's quality and completeness. Here I show the male bird used to construct the reference genome is heterozygous for the inversion polymorphism on the Z chromosome, but the assembly pipeline used by VGP was unsuccessful in assembling and accurately phasing the two Z chromosomes into haplotypes.

2.2. Introduction

Whilst a fragmented reference genome is fine for certain analyses, having a complete and high-quality reference genome for any species is an invaluable tool for genomic research (Rhie *et al.*, 2021; Nurk *et al.*, 2022). It is therefore important to produce the best possible version of a reference genome, especially for a model species that is used in a wide variety of biological disciplines. One such model organism is the zebra finch, *Taeniopygia guttata*, an Australian songbird (order Passeriformes, suborder oscine) of the *Estrildidae* family (subfamily *Estrildinae*). Biological research involving zebra finches began around 70 years ago (Morris, 2008), and since then the species has become an important model organism (Zann, 1996). Due to their ability to learn vocalisations, they have been used to understand the neurobiological basis of vocal learning both in birds (Mello, 2014) and in humans (Doupe and Kuhl, 1999). Zebra finches live in large flocks and are highly social, meaning they are widely studied in behavioural ecology (Galoch and Bischof, 2007; Guillette and Healy, 2014). They are also sexually dimorphic, varying in size, weight, feather and beak colouration, and in behaviour; studies of sexual selection in the species are common (Birkhead, Fletcher and Pellatt, 1998; Simons and Verhulst, 2011).

The volume and wide variety of research involving zebra finches resulted in the species being chosen as only the second avian species and first passerine to have its genome sequenced (Warren *et al.*, 2010). This reference genome was produced using the Sanger sequencing method, but in the following decade, the introduction of next generation sequencing (NGS) technologies caused the field of genomics to grow exponentially (Koboldt *et al.*, 2013). Currently, two general forms of NGS technology exist: short-read sequencing and long-read sequencing. Whilst both are still commonly used in biological research, as the complexity of genomes has become apparent, long-read sequences are increasingly used in favour of short-reads for reference genome assembly. At present, the most commonly used long-read sequencing technology is single-molecule real-time sequencing (SMRT) which was developed by Pacific Biosciences (Rhoads and Au, 2015). SMRT can deliver reads with an average length of over 10kb, and a maximum read length of over 60kb (Rhoads and Au, 2015). In comparison, the maximum read length of Illumina HiSeq 2500, a popular short-read sequencing technology, is only 250 bp (Eid *et al.*, 2009; Rhoads and Au, 2015). The greatly increased read length of SMRT (and similar technologies) allows for the resolution of large structural features, which short-read sequences are unable to resolve. Various new technologies for generating scaffolds from sequence reads have also been developed in recent years; these include Linked Reads,

developed by 10X Genomics, which involves partitioning DNA into long fragments and adding a barcode unique to that fragment. These barcodes are then used to guide assembly of short sequence reads into pseudo-long reads since short reads which are in close proximity on a chromosome will have the same unique barcode (Ott *et al.*, 2018). Another recent innovation is Hi-C, a method which investigates the three-dimensional architecture of whole genomes. Sequences that are closely linked on the same chromosome are also closer in physical space; this correlation between genomic and physical distance can be exploited to guide assembly of sequence reads into larger scaffolds (Lieberman-Aiden *et al.*, 2009). A third important technology for genome assembly is long-range optical mapping. Optical mapping has been in use since the 1990s, it involves digesting DNA with restriction enzymes and staining the DNA fragments with fluorescent dye to produce optical maps of each fragment (Chan *et al.*, 2018). These optical maps can then be combined to produce a consensus genomic optical map. However, it was only recently that new technologies have made optical mapping tractable for analysis of large genomes, where it is particularly useful in detecting large structural variation within genomes (Chan *et al.*, 2018).

There are now published reference genomes for thousands of different species, and plans to create references for many more (Rhie *et al.*, 2021). In addition, older references are being improved and completed, using previously unavailable NGS; the most prominent recent example being an improved and completed human genome (Nurk *et al.*, 2022). The source of DNA for the first zebra finch reference genome (TaeGut3.2.4) was a single domesticated male zebra finch, “Black17” (Warren *et al.*, 2010). An updated version of the zebra finch reference genome (bTG1.4) also using DNA from “Black17” was produced in 2021 by the Vertebrate Genomes Project (VGP) (Rhie *et al.*, 2021). This updated zebra finch assembly used NGS technologies and modern assembly techniques, rather than the Sanger sequencing technology used to produce the original reference. The VGP also produced a second reference genome (bTG2) using DNA obtained from a female zebra finch, “Blue55”. Both the updated male reference genome (bTG1.4) and the female reference (bTG2) were produced using the VGP’s assembly pipeline, which is intended to generate high-quality complete reference genomes for all of the roughly 70,000 extant vertebrate species. The pipeline utilises PacBio long-read sequence data and complements it with Linked Reads, Hi-C and long-range optical mapping (Rhie *et al.*, 2021). The references were also phased using an approach known as FALCON-Phase, which uses the inherent phasing information in Hi-C to sort sequence into phased haplotypes (Kronenberg *et al.*, 2021). This updated methodology has resulted in dramatic

improvements to the old reference, including placing 68.5Mb of previously unassigned sequence into chromosomes and assembling seven new chromosomes (Chr 30-36). The assembly also corrects many mis-joins, falsely inferred gene duplications and other substantial errors from the previous iteration.

Reference genomes are vital for identifying and understanding genetic variation within species. Variation is discovered primarily through resequencing, in which sequence reads are assembled using the reference genome as a guide, and any deviation from the reference sequence indicates a new variant. It is therefore important for reference genomes to be as correct and complete as possible, considering the important role the reference will play in future studies of genetic variation. However, it is difficult to resolve some areas of the genome, particularly large scale structural variations such as duplications and inversions (Chaisson, Wilson and Eichler, 2015).

The zebra finch genome is known to contain a number of large inversion polymorphisms (Knief *et al.*, 2016). One of these inversions is found on the Z chromosome (Itoh *et al.*, 2011) and appears to be segregating at high frequencies in all wild and captive zebra finch populations that have been studied to date. This inversion has been found to be responsible for almost all variation in sperm morphology within male zebra finches; this in turn significantly affects sperm motility, and probably male fitness (Kim *et al.*, 2017; Knief *et al.*, 2017). The inversion is characterised by three main haplotypes (A, B & C). If we are to use the reference genome to study the population genetics of the Z chromosome inversion polymorphism, then it is important to know what haplotypes the reference bird(s) carry, and also whether the assembly is correctly ordered. Given there is an inversion, the reference genome could either be correctly oriented for one haplotype and wrong for any others or it could even be assembled in an order that is different from all three haplotypes.

There are still many unanswered questions regarding the origins and maintenance of the zebra finch Z chromosome inversion polymorphism. Identifying the inversion karyotype of the reference bird is important for any future research attempting to answer these questions. Assessing the degree to which this inversion polymorphism is resolved will not only provide an indication of the effectiveness of the VGP assembly pipeline in dealing with large scale inversions, but will also inform further research into the structure, origin, and implications of this inversion polymorphism on the zebra finch Z chromosome.

Here, I aim to identify the Z chromosome inversion karyotype of the male zebra finch used to create both the original Sanger reference, and the more recent reference produced by the VGP. I also aim to identify the inversion karyotype of the female zebra finch which was also recently assembled by the VGP. In addition, I will assess the extent to which the inversion polymorphism on the Z chromosome is correctly assembled both at the phased haplotype assembly level and within the underlying scaffolds.

2.3. Methods and Materials

2.3.1. Reference Genomes

A single domesticated male zebra finch (“Black17”) from the laboratory of Dr. Arthur P. Arnold in the Department of Physiological Science at UCLA, Los Angeles, CA, USA, was the source of DNA for the first zebra finch reference genome (TaeGut3.2.4) (Warren *et al.*, 2010). This reference genome was produced using Sanger sequencing technology (Table 2.1). DNA from this same male was also used to generate an updated and improved version of the zebra finch by the VGP (bTG1.4) (Rhie *et al.*, 2021). This updated version was created using a range of NGS technologies including long-read sequencing, linked reads, optical maps, and Hi-C (Table 2.1). The VGP-produced reference was phased using the FALCON-Unzip algorithm (Chin *et al.*, 2016), producing primary (bTG1_pri) and alternate (bTG_alt) assemblies corresponding to the two haplotypes on each chromosome. The VGP also produced a reference genome based on DNA from a female zebra finch (“Blue55”) (bTG2), applying the same pipeline used to create the reference bTG1.4. Because Blue55 is female, and females are ZW (i.e., they have a single Z chromosome), the Z chromosome of this bird will be correctly phased.

	bTaeGut3.2.4	bTG1.4 (bTG1.4_pri & bTG1.4_alt)	bTG2
Isolate	Black17 (male)	Black17 (male)	Blue55 (female)
Sequencing Technology	Sanger	PacBio continuous long-read sequencing, 10X Genomics Linked Reads, Bionano optical maps, Arima Genomics Hi-C	PacBio continuous long-read sequencing, 10X Genomics Linked Reads, Bionano optical maps, Arima Genomics Hi-C
Total Genome Length (bp)	1,232,118,738	1,056,271,262	1,106,297,559
Total Length of Z chromosome (bp)	72,861,351	75,396,176 (bTG1.4_pri)	79,539,100
Number of Assembled Chromosomes	35	41	33
Unassigned Sequence (bp)	174,341,365	6,298,149	43,388,687
Number of Scaffolds	37,421	199	541
Scaffold N50	8,236,790	70,982,421	71,643,432
Submitter	Washington University Genome Sequencing Centre	Vertebrate Genomes Project	Vertebrate Genomes Project
Date	08/02/2013	04/05/2021	01/04/2020
Genbank Assession Number	GCA_000151805.2	GCA_003957565.4	GCA_009859065.2
URL	https://www.ncbi.nlm.nih.gov/assembly/GCF_000151805.1/	https://www.ncbi.nlm.nih.gov/assembly/GCF_003957565.2/	https://www.ncbi.nlm.nih.gov/assembly/GCA_009859065.2/
Reference	(Warren <i>et al.</i> , 2010)	(Rhie <i>et al.</i> , 2021)	(Rhie <i>et al.</i> , 2021)

Table 2.1: Summary of zebra finch reference genomes.

Details of all currently available reference genomes for the Zebra Finch. Two birds, Black17 and Blue55, were used to create three references, bTaeGut3.2.4, bTG1.4 and bTG2. bTG1.4 was constructed using the same DNA source as bTaeGut3.2.4 (from Black17) but used NGS technology and was phased to produce primary (bTG1.4_pri) and alternate (bTG1.4_alt) assemblies. bTG1.4_pri was assembled to the chromosomal level, whilst bTG1.4_alt was only assembled to the scaffold level. bTG2 was produced using sequence from a female bird, Blue55, and therefore only contains one assembled Z chromosome.

2.3.2. Diagnostic SNPs

A previous study into the zebra finch Z chromosome supergene used PCA analysis to infer inversion karyotypes for 1202 male zebra finches, from a population that was previously maintained at the University of Sheffield (Kim *et al.*, 2017). These birds were used to determine “diagnostic” SNPs which distinguish between the Z chromosome haplotypes. Within this dataset, 497 birds were homokaryotypes at the inversion polymorphism region (200 AA, 223 BB and 54 CC). These homokaryotypic birds were genotyped at 322,720 SNPs, of which 8,989 were Z-linked and in positions spanning most of the chromosome on the bTaeGut3.2.4 reference genome assembly (Kim *et al.*, 2017). SNP genotypes were determined for all birds using an Affymetrix GeneTitan system at Edinburgh Genomics (Kim *et al.*, 2017). Of these SNPs, 3056 showed large allele frequency differences between two of the homokaryotypes and were used in chi-square contingency table tests to find which markers showed the most significant association between genotype and the three haplotypes i.e., to test how well the marker can diagnose the inversion karyotype. These chi-square test statistics were then ranked to determine a subset of SNPs which are very highly diagnostic and could be used to accurately predict the inversion haplotype of other Z chromosome sequence.

Each marker was given a weighted score to indicate its relative diagnostic ability; the score ranges from 0 to 1, with 1 being the highest chi-square test statistic of any SNP found in a given comparison and 0 corresponding to a chi-square test statistic of 0. In the A vs B comparison, the largest chi-square statistic value was 443 at 56 different marker locations; all the remaining SNPs were scored relative to this value with scores ranging from 0 to 1. Therefore, a chi-square statistic of 443 resulted in a weighted chi-square score of 1. Any SNPs with a weighted chi-square score of 0.9 or higher (an A vs B comparison chi-square statistic of 398.7 or higher) were included in the final list of A vs B diagnostic SNPs (Table 2.2, Supplementary Table 1). The maximum chi-square statistics were 443 (A vs B), 277 (B vs C) and 274 (A vs C); in each case these statistics correspond to a perfectly diagnostic SNP. The difference in chi-square statistic between comparisons is due to the different number of homokaryotype birds used in each comparison (Table 2.2). Any SNPs with a weighted chi-square score > 0.9 were considered diagnostic for that comparison. This threshold of 0.9 for the weighted chi-square score was determined through experimentation with various other values. A minimum threshold of 0.9 was chosen as this resulted in a suitably large number of

SNPs whilst ensuring all SNPs were highly diagnostic. A total of 2123 SNPs were found to be diagnostic in the A vs B comparison (Table 2.2, Supplementary Table S2.1), whilst only 1085 were found to be diagnostic in the B vs C comparison (Table 2.2, Supplementary Table S2.2) and just 293 were found to be diagnostic in the A vs C comparison (Table 2.2, Supplementary Table S2.3). Among the 2123 SNPs that are diagnostic between the A and B inversion haplotypes, 919 are also diagnostic between B and C, and 252 are also diagnostic between A and C (Table 2.3). Because of the large number of SNPs diagnostic in multiple comparisons, the total number of SNPs with weighted chi-square scores > 0.9 in at least one comparison was 2330 (Table 2.3).

Inversion haplotype comparison	A vs B	B vs C	A vs C
Number of homokaryotypic birds	423 (200 AA, 223 BB)	277 (223 BB, 54 CC)	254 (200 AA, 54 CC)
Maximum chi-square statistic	443	277	274
90% of Maximum (0.9 weighted chi-square score)	398.7	249.3	246.6
Number of SNPs > 0.9 weighted chi-square score	2123	1085	293

Table 2.2: A summary of inversion karyotype diagnostic SNPs.

Diagnostic SNPs were identified from chi-square test results involving 497 male zebra finches from the University of Sheffield and 3056 SNP positions on the Z chromosome. Each bird has one of three known inversion karyotypes (AA, BB & CC). Pairwise comparisons were made between each of the three inversion karyotypes at all 3056 SNP positions. Weighted chi-square scores were then given to each position for each comparison, where 1 corresponds to the maximum chi-square statistic seen within that comparison, and all other scores are calculated relative to this maximum (all between 0-1). A weighted chi-square of 0.9 (90% of maximum) was chosen as the threshold for diagnostic SNPs.

	Number of SNPs above 0.9 weighted chi-square score		
	A vs B Comparison	B vs C Comparison	A vs C Comparison
>0.9 unique to this comparison	952	166	41
Also >0.9 in A vs B	-	919	252
Also >0.9 in B vs C	919	-	0
Also >0.9 in A vs C	252	0	-
Total	2123	1085	293
Total across all comparisons	2330		

Table 2.3: Counts of inversion karyotype diagnostic SNPs.

A count of SNP positions which meet the threshold of weighted chi-square score > 0.9 in each of three pairwise comparisons between inversion karyotypes (A vs B, B vs C, A vs C). Many SNPs meet the threshold within multiple comparisons, meaning those SNPs are simultaneously diagnostic between both sets of inversion karyotypes. No SNPs are diagnostic (>0.9 weighted chi-square score) in all three pairwise comparisons.

Principal component analysis (PCA) was performed using Plink v1.90 (Purcell *et al.*, 2007) using these 2330 diagnostic SNPs, which showed clear and separate clustering of all inversion haplotypes (Figure 2.1). This indicates that this set of diagnostic SNPs can be used to determine between all three inversion karyotypes, with PC1 differentiating the B haplotype from A and C, and PC2 differentiating C from A and B. The clustering of birds possessing C haplotypes (along the PC2 axis) is less distinct than clustering involving Haplotype B (along the PC1 axis). In an attempt to improve the resolution of the C haplotype detection, PCA using just the 1085 B vs C diagnostic SNPs and just the 293 A vs C diagnostic SNPs were performed. These PCA plots do not show separate clustering of all inversion haplotypes and so these subsets of diagnostic SNPs would be unsuitable for finding the karyotype of unknown birds (Supplementary Figures S2.1 & S2.2). A similar PCA analysis was conducted previously using the same set of birds from the University of Sheffield (Figure 2a of Kim *et al.*, 2017) but with all Z chromosome SNPs. The PCA presented here differs in that a smaller but more conservative group of diagnostic SNPs (>0.9 weighted chi-square score) were included in my analysis, which is the probable cause of the less distinct clustering along the PC2 axis (A/B vs C).

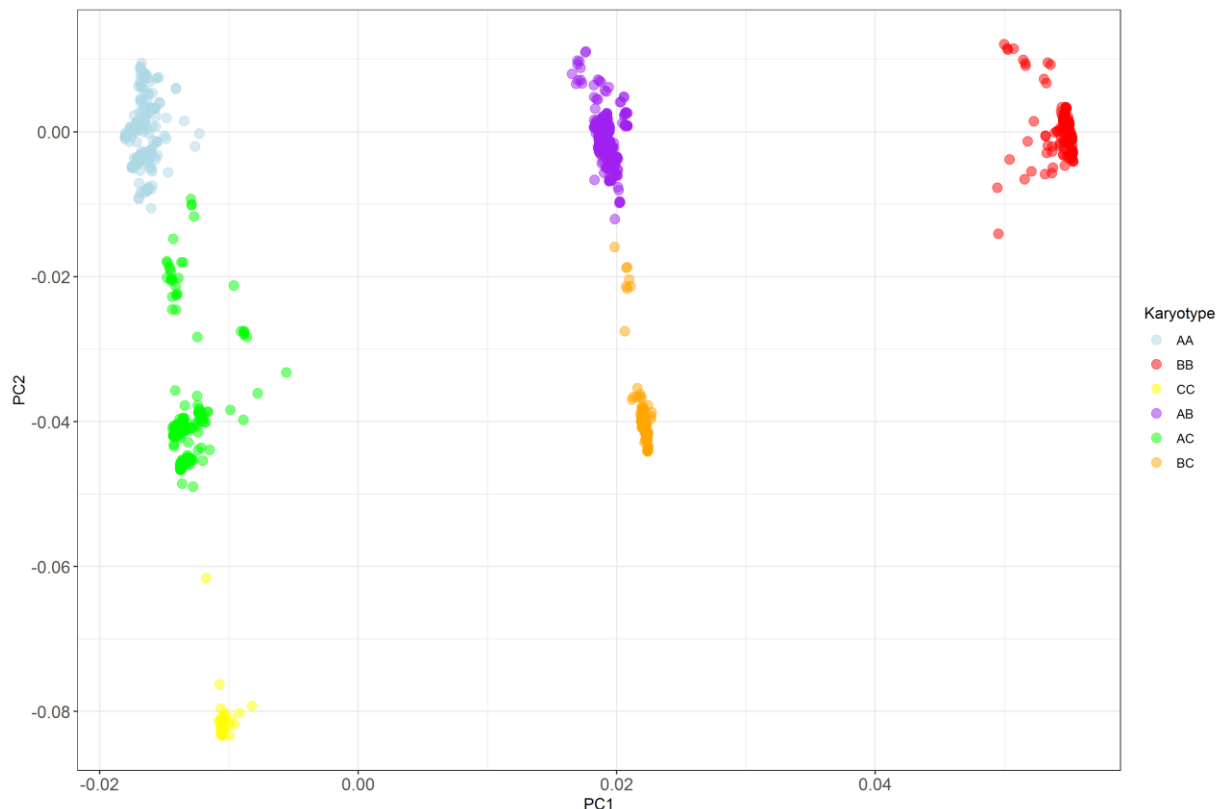


Figure 2.1: Principal component analysis (PCA) of Z chromosome SNP genotypes. Genotypes are of 1202 Principal component analysis (PCA) of 1202 male zebra finches from a population that was previously maintained at the University of Sheffield. The analysis used genotypes from 2330 diagnostic SNP positions with weighted chi-square score >0.9 in at least one of three pairwise comparison between haplotypes (A vs B, A vs C or B vs C). Birds of known inversion karyotypes and are labelled accordingly. The first two principal components explain 93.6% of variance (PC1 = 85.3%, PC2 = 8.3%).

2.3.3. Identification of zebra finch reference genome inversion haplotypes using diagnostic SNPs and Principal component analysis

A principal component analysis was performed, again using the genotypes at the 2330 diagnostic SNP positions from the captive population at the University of Sheffield, but supplementing the data with SNP genotypes from the original reference genome assembly of Black17 (TaeGut3.2.4), phased assemblies of Black17 (bTG1_pri, bTG1_alt) and the assembly of the female finch, Blue 55 (bTG2). Genotypes of the reference zebra finches were found at these diagnostic SNP positions by comparing the flanking sequence (30bp) of diagnostic SNPs with the reference genome sequence using NCBI Blastn v2.8.1 (Camacho *et al.*, 2009). SNPs were sorted into either A- or B-supporting for the 2123 A vs B diagnostic SNPs, B- or C-supporting for the 1085 B vs C diagnostic SNPs and either A- or C-supporting for the 293 A vs C diagnostic SNPs. An overall percentage support for each haplotype in the reference

assemblies was calculated. These percentage supports were the sum of all weighted chi-square scores for SNPs which supported a given haplotype, divided by the sum of weighted chi-square scores for all diagnostic SNPs in the pairwise comparison. For example, the following formula was used to calculate the percentage supports for A and B in the A vs B pairwise comparison.

$$\text{Support for type A} = \frac{\sum \text{weighted chi - square scores for SNPs supporting A}}{\sum \text{weighted chi - square scores for all markers}}$$

$$\text{Support for type B} = \frac{\sum \text{weighted chi - square scores for SNPs supporting B}}{\sum \text{weighted chi - square scores for all markers}}$$

In addition to scoring the whole inversion region of each assembly, the individual scaffolds within the bTG1_alt assembly were scored using the same method but only with the A vs B diagnostic SNPs situated within that scaffold. The bTG1_alt assembly consists of 1,646 scaffolds (N50 = 3.6Mb), but as the assembly is not at chromosome-level, it is not known which of the scaffolds contain Z chromosome sequence. As inversion-wide analysis of the diagnostic sites suggested the bTG1 assembly was of an AB heterozygote bird (for further explanation see results section 2.4.1), SNPs that were diagnostic of Haplotype C were not used. Any scaffolds containing less than two diagnostic markers were removed from the analyses, leaving 137 Z chromosome scaffolds. Scaffolds were typed as either A or B depending on whether the calculated support for either type was greater. This analysis was then repeated after changing the minimum number of diagnostic markers within a scaffold to 3, 4, 5, 6 and 10, to ensure the results were robust to different thresholds.

In addition, any scaffolds from the bTG1_alt assembly containing two or more A vs B diagnostic SNPs and with a mean support for either type >60% were pooled together and SNP identities from within these scaffolds were included in the principal component analysis (bTgu1_alt_A and bTgu1_alt_B) along with other reference sequences. The aim of this analysis was to further confirm the Z-chromosome karyotype of the reference bird Black17.

2.3.4. Using diagnostic SNPs to determine inversion karyotypes of publicly available sequence

The diagnostic SNPs were then used to infer the inversion karyotypes of publicly available sequence reads from 24 Zebra Finches (11 female, 13 male; www.ebi.ac.uk/ena/data/view/PRJEB10586). These birds were originally sequenced using Illumina HiSeq in 2012 and 2013 (Singhal *et al.*, 2015). Of these 24 finches, 19 were wild birds

caught at Fowler's Gap, New South Wales, Australia. The other five birds were bred in captivity at East Carolina University; the group consists of a mother (MP1), father (MP2), and three sons (MP3, MP4 and MP5).

For each of these 24 zebra finches, the sequence reads corresponding to each of the 2330 diagnostic SNP positions (Table 2.2) were determined by comparing reads to the flanking sequence (30bp) of each diagnostic SNP using NCBI Blastn v2.8.1 (Camacho *et al.*, 2009). Reads which perfectly aligned to the flanking sequence of a diagnostic SNP were used to determine the identity of the base at the SNP position for each of the 24 zebra finches. Only one perfectly aligned read was used for each SNP position, meaning birds could not be identified as heterozygous at a given position. However, any birds which were heterokaryomorphic for the inversion polymorphism were still identifiable, as approximately half of the 2330 sites were suggestive of one inversion haplotype, whilst the other half suggested a different inversion haplotype. In order to identify the inversion haplotype of each of the 24 birds, principal component analysis was performed using Plink v1.90 (Purcell *et al.*, 2007) with these SNP identities and data from the 1202 male zebra finches from the University of Sheffield. In addition, the overall percentage support for each zebra finch assembly belonging to each haplotype was calculated, for all three comparisons (A vs B, A vs C, B vs C). This percentage was calculated as the sum of all weighted chi-square scores for SNPs supporting either/or haplotype, divided by the sum of weighted chi-square scores for all diagnostic SNPs in the pairwise comparison (see 2.3.3). An alternative approach, implemented later in the thesis, would use raw reads of each of the 24 birds and perform variant calling at each site. This is more time-consuming, but does not rely on a single consensus sequence for each bird, and is therefore more informative.

2.4. Results

2.4.1. Reference genome inversion haplotypes

A list of 2330 SNPs was produced, all of which were highly diagnostic of the Z chromosome inversion karyotype. They were all located between 6.04Mb and 65.35Mb on the Z chromosome of the original zebra finch reference genome (TaeGut3.2.4). These positions are likely to be the approximate boundaries of the inversion polymorphism, since we expect all of the SNPs that are diagnostic for the inversion karyotype to be located within the inversion, rather than outside of it where recombination between the haplotypes will be possible. These boundaries match closely with the putative breakpoints at 5.7Mb and 60.2-68.8Mb, previously inferred from patterns of heterozygosity of Z chromosome SNPs within male birds of each karyotype (Kim *et al.*, 2017).

Reference	A support vs B (%)	A support vs C (%)	B support vs C (%)	Inversion Type	Sex	Name
TaeGut3.2.4	47.8	43.0	50.7	AB/AC/BC	M	Black17
bTG1_pri	49.0	53.7	54.8	AB/AC/BC	M	Black17
bTG1_alt	48.5	26.2	45.8	AB/AC/BC	M	Black17
bTG2	97.5	95.7	7.5	A	F	Blue55

Table 2.4: Percentage supports for inversion haplotypes in each reference genome.

The percentage of diagnostic markers indicating whether each reference assembly is either A, B or C inversion haplotype. Diagnostic markers (n=2330) are SNPs with a weighted chi-square score > 0.9 in either an A vs B, A vs C or B vs C pairwise comparison. Percentage support was calculated as the sum of all weighted chi-square scores for SNPs supporting either A, B or C, divided by the sum of weighted chi-square scores for all diagnostic SNPs in the pairwise comparison.

This analysis suggests the male used to create the TaeGut3.2.4 and bTG1 references (Black17) is a heterozygote for the inversion polymorphism, as no single haplotype has strong support in either the diploid assembly (TaeGut3.2.4) or the two phased haplotypes (bTG1_pri and bTG1_alt; Table 2.4). As the reference genome does not list variants, any heterozygous positions within the inversion polymorphism will be assembled such that an allele from either of the two inversion haplotypes which make up the heterokaryotype is included. The specific haplotype assembled is random, and therefore across the length of the inversion polymorphism,

we would expect approximately 50% of diagnostic SNP positions to indicate the assembly was of each haplotype. Analysis showed that for the original reference genome (TaeGut3.2.4) 47.8% of A vs B markers are diagnostic of Haplotype A and 52.2% of markers are diagnostic of Haplotype B. When considering diagnostic SNPs versus the C haplotype, there was 43% support for A (A vs C) and 50.7% support for B (B vs C), suggesting the reference bird Black 17 is a heterokaryotype. However, the results are inconclusive about the bird's specific combination of haplotypes (either AB, AC or BC) (Table 2.4). The VGP assembly bTG1.4 used DNA sequence from the same male bird (Black17), but the sequence was phased to produce a chromosome-level primary assembly (bTG1_pri) and a scaffold-level alternate assembly (bTG1_alt). Since these assemblies are phased haplotypes of the heterozygote Black17, we expected one assembly to contain only diagnostic alleles matching one haplotype, and the other assembly to contain only diagnostic alleles matching another haplotype. However, neither the primary assembly (bTG1_pri) nor the alternate assembly (bTG1_alt) showed consistent support for any one haplotype, indicating they still have a mixture of 'A', 'B' and 'C' diagnostic alleles. This lends further support to Black17 being a heterozygote but does not identify which haplotypes it possesses (Table 2.4). The VGP assembly bTG2, which used sequence data from a female bird (Blue55), showed 97.5% and 95.7% support for A against the B and C haplotypes respectively, suggesting the karyotype of Blue55 is A (Table 2.4).

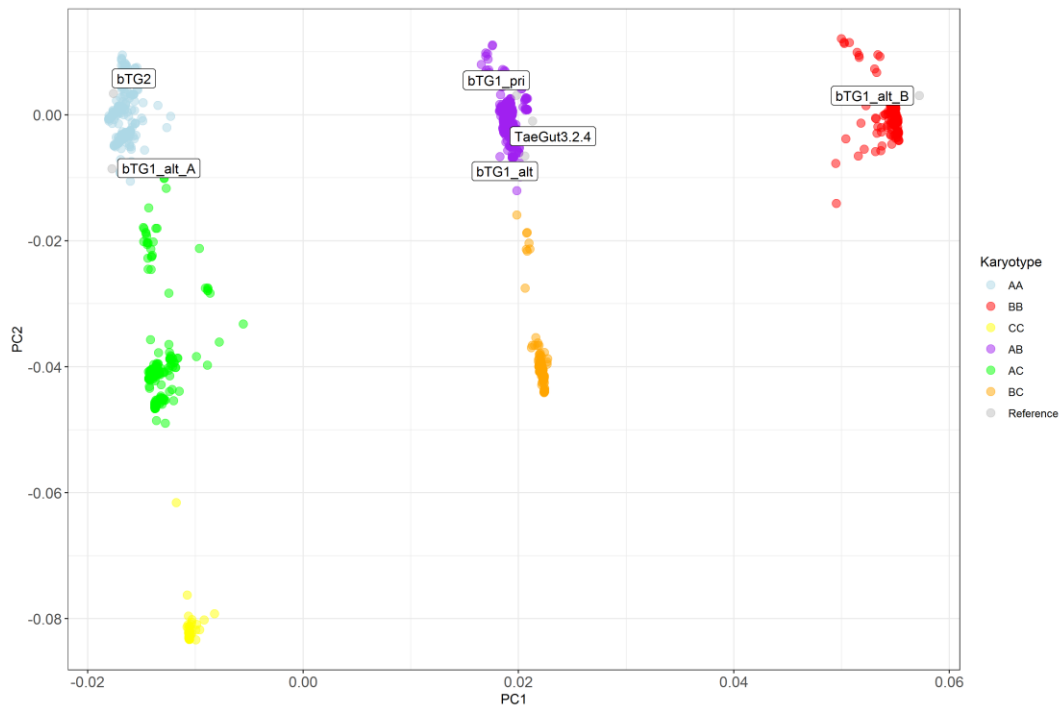


Figure 2.2: PCA including reference assemblies.

Principal Component Analysis of 1202 male zebra finches from the University of Sheffield, along with the reference assemblies. All Sheffield birds are of known inversion karyotypes and so are labelled accordingly. 2330 Z chromosome SNPs, all with a weighted chi-square statistic over 0.9 in comparisons between two of the three haplotypes were used.

A principal component analysis of 2330 highly diagnostic Z chromosome SNPs shows that birds with the same inversion karyotype cluster together (Figure 2.1). The first two principal components explain 93.6% of variance (PC1 = 85.3%, PC2 = 8.3%). Principal component 1 differentiates between the B inversion type and A/C types. Principal component 2 differentiates between C inversion type and A/B types. The various reference assemblies cluster within groups of the genotyped birds, meaning the diagnostic markers can be used on the reference genome birds, and their inversion karyotype can be inferred. The original zebra finch reference (TaeGut3.2.4) was constructed using a male bird (Black17) which was found to be a heterozygote (Table 2.4). On the PCA plot, TaeGut3.2.4 is found clustered with AB-type birds suggesting the karyotype of Black17 is AB (Figure 2.2). The VGP reference constructed from a female bird (bTG2) is found within the cluster of AA-type birds, which is expected since I calculated the female bird's sole Z chromosome to be Haplotype A (Table 2.4). However, the two phased haplotype assemblies produced by VGP (bTG1_pri & bTG1_alt) also cluster alongside the AB birds (Figure 2.2). This is not the result we would expect to see for a correctly phased AB heterozygote male; instead, the A haplotype assembly should cluster alongside the AA males (similar to the female bTG2 assembly) and the B haplotype assembly should cluster alongside the BB males. To further investigate this

unexpected result, the scaffolds which make up the bTG1_alt assembly were individually assessed to find their inversion haplotypes.

2.4.2. Inversion typing of assembly scaffolds

The scaffold-level alternate haplotype assembly (bTG1_alt) of the bird Black17, produced by VGP, contained 137 scaffolds within the inversion each containing more than one diagnostic SNP marker. Using the identities at each marker and the marker's weighted chi-square score, the percentage support for each scaffold being either Haplotype A or Haplotype B sequence was calculated. Of these 137 scaffolds, 76 are consistent with haplotype "A" sequence (>50% support for Haplotype A) and the other 61 are consistent with haplotype "B" sequence (Supplementary Table S2.4). The mean percentage support amongst the 76 scaffolds potentially containing A type sequence was 89.21%, and among the 61 scaffolds containing B type sequence was 93.46% (Table 2.5). The high average support for each scaffold's typing indicates phasing may be correct at the scaffold level, even if phasing at the whole-chromosome level appears to be incorrect (Table 2.4, Figure 2.2).

The average percentage support increases further as the threshold for number of SNPs within a scaffold increase (Table 2.5). At the strictest threshold for number of diagnostic SNPs within a scaffold ($N \geq 10$), there are just 30 scaffolds, but they are all fixed or nearly fixed for either A or B haplotype SNP alleles (Tables 2.5 & 2.6). These results show that a larger number of highly diagnostic SNPs found within a scaffold results in a more reliable assignment to either Haplotype A or Haplotype B. If scaffolds contained inconsistent and erroneous phasing, we would expect at least some scaffolds with high numbers of diagnostic SNPs to have weak support for either haplotype, but that is not the case. There are some scaffolds with weak support for either Haplotype A or B typing, but these all contain relatively few diagnostic SNPs (Supplementary Table S2.4). With this information, we can assume that each scaffold is correctly phased, but the assembly of scaffolds into chromosomal sequence erroneously joins scaffolds from alternative inversion haplotypes and treats them as if they are on the same haplotype. When the bTG1_alt scaffolds of each inferred type (bTG1_alt_A & bTG1_alt_B; number of diagnostic SNPs ≥ 2 , mean average support for typing $> 60\%$) are pooled and included in a PCA, the pooled scaffolds cluster in the positions we would expect from correctly phased haplotype assemblies, with bTG1_alt_A clustering with AA homozygote males and bTG1_alt_B clustering with BB homozygote males (Figure 2.2).

Number of Diagnostic SNPs	Type A			Type B		
	No. Scaffolds	Mean Support (%)	Standard Error	No. Scaffolds	Mean Support (%)	Standard Error
N ≥ 2	76	89.21	2.10	61	93.46	1.96
N ≥ 3	46	91.89	2.06	33	90.89	2.93
N ≥ 4	31	92.25	2.43	28	91.65	3.17
N ≥ 5	25	93.35	2.32	21	98.42	1.35
N ≥ 6	22	93.36	2.53	15	97.79	1.88
N ≥ 10	17	96.96	1.46	13	99.61	0.39

Table 2.5: Support for haplotypes with varying numbers of diagnostic SNPs.

Average percentage support for either the A or B inversion haplotype for scaffolds in the “bTG1_alt” assembly. Scaffolds are only included in the calculation of the mean if they contain greater than or equal to the given threshold of diagnostic SNP markers.

Scaffold	Scaffold Length (bp)	nA	nB	% Score	Consensus haplotype
000058F_015_arrow	1,590,284	210	3	98.6	A
000058F_011_arrow	3,123,494	0	362	100.0	B
000351F_003_arrow	105,831	18	0	100.0	A
000056F_031_arrow	1,307,474	0	133	100.0	B
000056F_027_arrow	510,336	52	0	100.0	A
000056F_016_arrow	1,847,689	193	1	99.5	A
000032F_004_arrow	351,802	0	28	100.0	B
000011F_082_arrow	429,825	21	0	100.0	A
000011F_178_arrow	560,552	0	35	100.0	B
000011F_149_arrow	445,161	20	1	95.4	A
000011F_043_arrow	1,383,358	0	40	100.0	B
000011F_130_arrow	666,447	15	5	75.2	A
000011F_039_arrow	505,353	0	45	100.0	B
000057F_025_arrow	1,026,287	0	31	100.0	B
000143F_010_arrow	441,192	2	37	95.0	B
000056F_003_arrow	806,589	55	0	100.0	A
000056F_021_arrow	82,919	17	0	100.0	A
000157F_007_arrow	872,600	0	35	100.0	B
000011F_008_arrow	844,860	42	1	97.9	A
000011F_059_arrow	548,969	16	1	94.1	A
000011F_033_arrow	540,791	0	11	100.0	B
000011F_152_arrow	730,571	34	1	97.1	A
000165F_002_arrow	165,101	0	11	100.0	B
000056F_011_arrow	484,396	29	1	96.7	A
000011F_066_arrow	737677	14	1	93.8	A
000143F_002_arrow	306,534	0	34	100.0	B
000143F_011_arrow	118,344	23	0	100.0	A
000346F_001_arrow	56,024	11	0	100.0	A
000011F_121_arrow	418,721	14	0	100.0	A
000058F_013_arrow	190,969	0	13	100.0	B

Table 2.6: Inversion haplotypes of *bTG1.4_alt* scaffolds.

The number of diagnostic SNPs and consensus haplotype in each scaffold (with ≥ 10 SNPs) of the alternate assembly of *bTG1.4* (*bTG1.4_alt*) which uses sequence data from the male bird, Black17. A scaffold-level assembly was not available for Black17's primary phased haplotype assembly (*bTG1.4_pri*). *nA* and *nB* are the numbers of diagnostic SNPs with the A and B alleles respectively. % Score is the percentage of SNPs carrying the 'consensus' allele. Note that the % score is weighted by how diagnostic each SNP is. Most scaffolds are fixed (% score = 100.0), or nearly fixed, for either Haplotype A or Haplotype B SNP alleles. The most ambiguous scaffold, 000011F_130_arrow, has 15 SNPs with the A allele and 5 with the B allele and a % score of 75.2%.

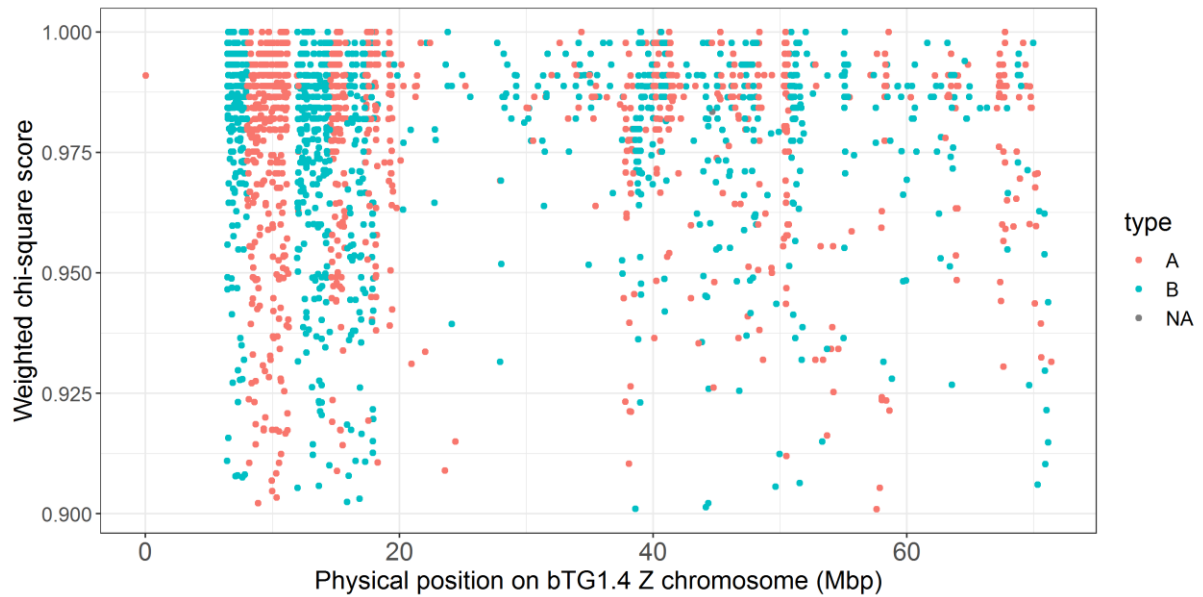


Figure 2.3: Positions of diagnostic SNPs (A vs B) on Z chromosome.

The physical positions of 2123 A vs B diagnostic SNPs on the Z chromosome of the primary phased assembly of the Zebra Finch reference genome bTG1.4 (bTG1.4_pri). Each diagnostic SNP has a weighted chi-square score greater than or equal to 0.9, with a larger chi-square score indicating greater diagnostic power. Points are coloured depending on whether the marker is diagnostic of Haplotype A or Haplotype B in the bTG1.4_pri assembly.

The physical positions of diagnostic SNPs which indicate the assembly bTG1.4_pri inversion type A cluster together in several groups across the length of the Z chromosome; they are interspersed with clustered groups of diagnostic SNPs indicating the assembly is type B (Figure 2.3). Sections of the Z chromosome contain alleles exclusively supporting one inversion karyotype. These sections may correspond to the underlying scaffolds used to construct this whole-chromosome assembly, since the scaffold-level alternate haplotype assembly (bTG1.4_alt) indicates successful phasing of scaffolds in the bTG1.4 reference. Figure 2.3 therefore further indicates that phasing is consistent and correct *within* each scaffold. However, the combination of these correctly phased scaffolds into a continuous chromosomal sequence is incorrect, with A and B type scaffolds frequently being stitched together and treated as being on the same phased haplotype.

2.4.3. Identification of inversion haplotypes for other sequenced zebra finches

To test the application of diagnostic SNPs as a method for identifying the Z chromosome inversion karyotype, publicly available sequence data for 24 zebra finches (11 female, 13 male) sequenced by Singhal *et al.*, (2015) were used. For these birds the genotypes at 2330 diagnostic SNP positions shown to be highly diagnostic between the three inversion karyotypes were called from sequence data (Tables 2.2 & 2.3, Figure 2.1). Principal component analysis of these 2330 diagnostic SNP genotypes and genotypes from 1202 male finches of known inversion karyotypes, showed discrete clusters corresponding to each inversion karyotype (Figure 2.4). As in previous PCAs (Figures 2.1 & 2.2), Principal Component 1 differentiates between the B inversion karyotype and A/C karyotypes, whilst Principal Component 2 differentiates between C inversion karyotype and A/B karyotypes. All 24 zebra finches of unknown karyotype clustered within groups of Sheffield birds whose karyotypes were known (Figure 2.4). This suggests that the diagnostic SNPs are suitable for inferring unknown karyotypes, in different populations, as any birds of an ambiguous karyotype would be separate from any clusters on the PCA plot.

The percentage support for each haplotype in the pairwise comparisons between the three inversion haplotypes enabled assignment of the inversion karyotype for each of 24 zebra finches (Table 2.7). Four female birds were found to be A karyotype, whilst four male birds were AA. In each of these A type birds, the percentage support for A over B and C was between 77.89% and 93.20% (mean = 88.03%), and all birds clustered with AA males on the PCA plot (Figure 2.4). Six female birds were found to be karyotype B, and one male bird was found to be karyotype BB. In each of these cases, the percentage support for B over A and C was between 91.7% and 99.8% (mean = 95.9%), and all were consistent with the results of the PCA (Figure 2.4). One female bird had a high percentage support for C over B (94.22%) and over A (83.30%) suggesting it was karyotype C, although in the PCA plot, this bird (MP1) clustered with AC males (Figure 2.4). An AC karyotype would be impossible for a female bird, and MP1 was assumed to be Haplotype C. The remaining eight birds were all males and were assumed to be heterokaryotypes, with four having >75% support for A over B and >75% support for C over B, suggesting an AC typing, these four birds also clustered with AC heterozygotes in the PCA plot (Figure 2.4). Three of these four AC males (MP3, MP4 & MP5) were part of a captive-bred family; their father (MP2) is most likely an AA homozygote, suggesting the mother (MP1) is likely to possess the C haplotype. The final four birds (26792, 28353, 28404

& 28313) were ambiguous, with percentage support slightly higher for B over A, C over A and B over C, and rarely over 75%. The PCA suggested these four birds were AB heterozygotes (Figure 2.4), and since the percentage support presented here was inconclusive (Table 2.7), the AB heterozygote typing from the PCA was used.

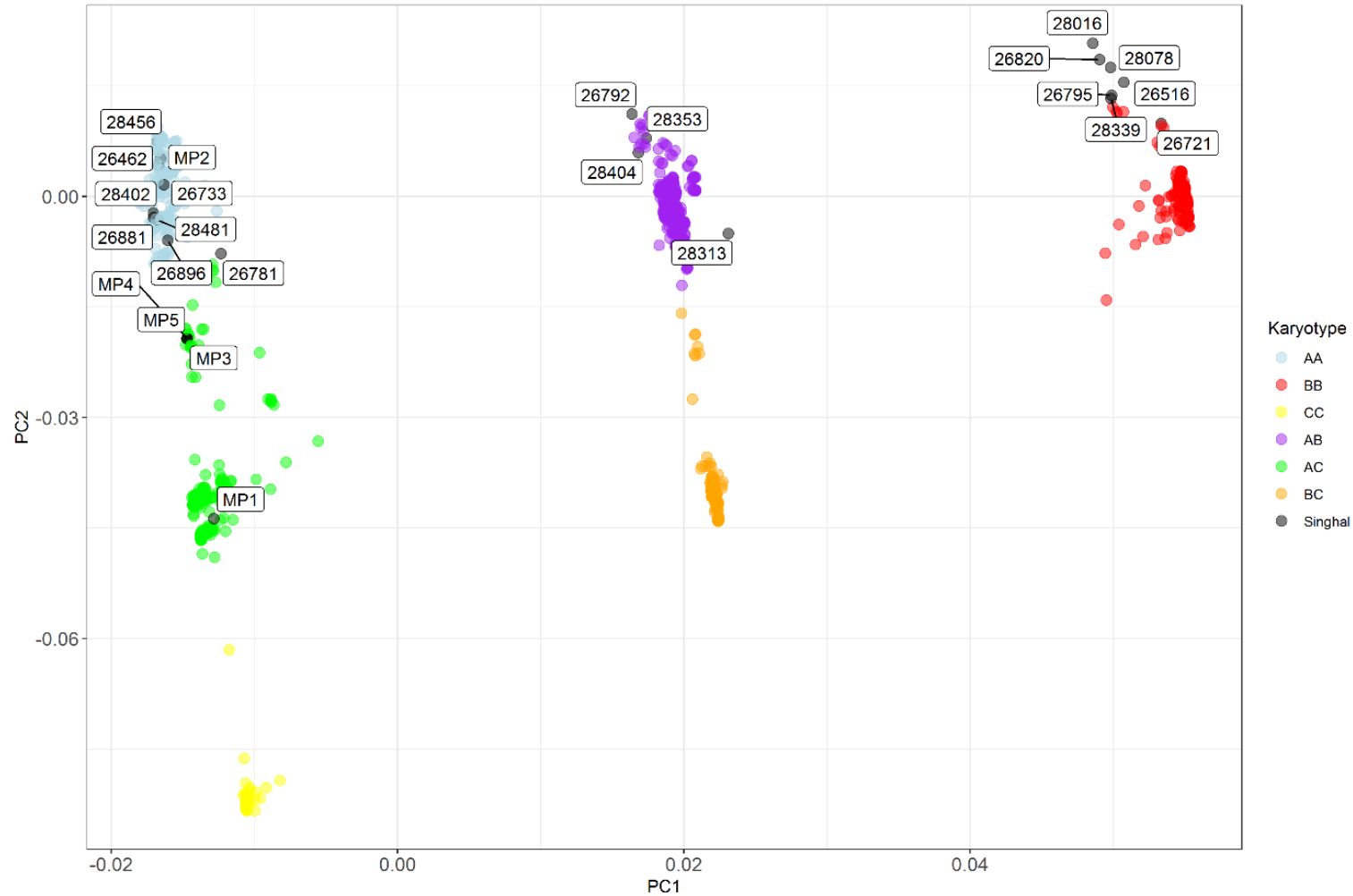


Figure 2.4: PCA including 24 zebra finches from Singhal et al., 2015.

Principal component analysis (PCA) of 1202 male zebra finches from a population that was previously maintained at the University of Sheffield along with 24 zebra finches (11 female, 13 male) sequenced by Singhal et al., and published in 2015. Birds from the University of Sheffield were of known inversion karyotypes and are labelled accordingly. 2330 SNP positions were used, each with weighted chi-square statistics >0.9 in at least one pairwise comparison between haplotypes.

Sample Alias	Sample Accession	Sample Source	Sex	A support vs B (%)	A support vs C (%)	B support vs C (%)	Inferred Karyotype	Mean Coverage
26462	SAMEA3532857	W	F	89.5	93.0	20.0	A	27.9
28456	SAMEA3532861	W	F	90.6	93.2	15.8	A	27.7
26881	SAMEA3532868	W	F	85.1	88.8	10.5	A	50.2
26896	SAMEA3532870	W	F	88.5	92.6	13.8	A	18.7
MP2	SAMEA3532853	D	M	88.3	89.8	17.9	AA	29.8
28402	SAMEA3532862	W	M	85.6	77.9	12.5	AA	16.7
26733	SAMEA3532866	W	M	85.1	88.3	19.5	AA	26.1
28481	SAMEA3532867	W	M	90.3	82.0	13.3	AA	33.4
28404	SAMEA3532864	W	M	37.0	40.6	68.5	AB	20.9
28353	SAMEA3532859	W	M	35.4	44.8	70.2	AB	32.7
26792	SAMEA3532871	W	M	37.7	39.2	66.3	AB	20.7
28313	SAMEA3532875	W	M	30.9	19.3	71.0	AB	28.3
28339	SAMEA3532858	W	F	7.2	8.5	99.1	B	19.9
26721	SAMEA3532860	W	F	6.9	9.5	99.8	B	20.7
26820	SAMEA3532865	W	F	7.4	14.6	99.3	B	29.2
28016	SAMEA3532872	W	F	8.3	15.5	98.8	B	18.7
26795	SAMEA3532873	W	F	7.5	8.9	99.1	B	34.2
28078	SAMEA3532874	W	F	7.7	16.0	99.3	B	25.5
26516	SAMEA3532863	W	M	7.3	11.6	99.4	BB	22.3
MP3	SAMEA3532854	D	M	77.5	43.8	17.0	AC	23.2
MP4	SAMEA3532855	D	M	77.0	44.3	16.4	AC	19.7
MP5	SAMEA3532856	D	M	75.5	41.9	16.5	AC	28.9
26781	SAMEA3532869	W	M	77.4	61.8	26.3	AC	24.2
MP1	SAMEA3532852	D	F	76.5	16.7	5.8	C	27.3

Table 2.7: Inversion haplotypes of birds from Singhal et al., 2015.

Percentage supports for each inversion haplotype (A, B & C) in three pairwise comparisons (A vs B, A vs C & B vs C) for each of 24 zebra finches from Singhal et al., 2015. Percentages were calculated using genotypes at diagnostic SNP positions (2123 A vs B, 1085 B vs C & 293 A vs C positions) across the Z chromosome. Karyotypes were determined by percentage support, with a >75% support for a haplotype in both comparisons in which it appears used to infer the haplotype. E.g., >75% support for A in A vs B and in A vs C comparisons suggests A/AA karyotype. All other male individuals were assumed to be heterokaryotypes. For sample source, D = domesticated, W = wild. Mean coverage as reported by Singhal et al., 2015.

2.5. Discussion

The analyses presented in this chapter suggest the two birds (Black17 & Blue55) used to create the zebra finch reference genomes (bTG1 & bTG2) are of inversion karyotypes AB (Black17) and A (Blue55). The phased haplotypes of the bTG1 reference genome (which was constructed using sequence from Black17), should each exclusively contain sequence of either Haplotype A or Haplotype B. However, although the assembly scaffolds are correctly phased, the assembly of these scaffolds into whole phased chromosome sequences has incorrectly joined scaffolds from different inversion haplotypes. In other words, each assembly is a mixture of correctly phased A and B scaffolds. This explains why, in a PCA, the two bTG1 phased haplotypes do not cluster with the AA and BB karyotype birds from the Sheffield population in the way we expect (Figure 2.2). The clustered physical locations of A-diagnostic or B-diagnostic alleles across the Z chromosome (Figure 2.3), and the consistent typing within assembly scaffolds from the bTG1_alt assembly (Tables 2.4 & 2.5), indicate that the phased haplotype assemblies (bTG1_pri & bTG1_alt) of the reference male are correct at the scaffold-level, even though they contain a mixture of A and B haplotype alleles (Table 2.4).

Although diagnostic SNPs can identify incorrect phasing, I could not use them to correctly phase the entire Z chromosome. Any scaffolds which contain multiple highly diagnostic SNPs can be reliably identified as either the A or B haplotype (Table 2.5), but many areas of the Z chromosome contain few diagnostic SNPs, if any at all (Figure 2.3). Other methods are therefore required to complete phasing of chromosomes, but diagnostic SNPs are nonetheless a useful assessment tool for identifying which haplotypes are within regions of inversion polymorphism.

The Zebra Finch Reference Genome was constructed using the standard VGP assembly pipeline (Rhie *et al.*, 2021). Sequencing data of one male zebra finch (Black17) was used to generate primary and alternate haplotype contigs from sequence reads using FALCON-Unzip (Chin *et al.*, 2016). These contigs were then combined with 10X Genomics Linked Reads to generate scaffolds using Scaff10x (Rhie *et al.*, 2021). Genotypes at diagnostic SNPs indicate that most bTG1 alternate haplotype scaffolds produced by this assembly pipeline are fixed, or nearly fixed, for either Haplotype A or Haplotype B diagnostic SNP alleles (Tables 2.4 & 2.5). This suggests that the erroneous phasing of the haplotype assemblies arises during the generation of chromosome-scale scaffolds from smaller scaffolds. The VGP assembly pipeline achieves chromosome-scale scaffolds using Hi-C technology, which provides information on

contact frequency (a measure of the number of chromatin interactions) between a pair of loci (Lieberman-Aiden *et al.*, 2009). Contact frequency of two loci in a chromosome strongly correlates with the physical distance between them, and so the metric can be used to help guide genome scaffolding. Specifically, a tool known as Salsa2 (Ghurye *et al.*, 2019) was used to integrate Hi-C reads with smaller scaffolds in the VGP assembly pipeline. A large inversion polymorphism would impact the contact frequencies measured by Hi-C since the physical distances between many loci will change depending on the orientation of sequence within the inversion region. In a heterozygous individual such as Black17, there are two different physical distances between any given pair of loci within the inversion, which may make the use of Hi-C technology problematic. The use of Hi-C technology for genome assembly is known to produce ordering errors when the primary sequence is complex, and the scaffolds being combined are short and enriched for segmental duplications and simple repeats (Burton *et al.*, 2013). Sex chromosomes are generally harder to assemble, partially due to their high repeat content (Tomaszkiewicz, Medvedev and Makova, 2017), and VGP report that their zebra finch sex chromosome assemblies were more fragmented than autosomes (Rhie *et al.*, 2021). It is therefore likely that the highly complex nature of the zebra finch Z chromosome and its large polymorphic inversion made Salsa2 ineffective at generating whole-genome scaffolds.

Zebra finches are known to have many other inversion polymorphisms across their genome (Knief *et al.*, 2016; Kim *et al.*, 2017), all of which could cause similar issues with phasing on their respective chromosomes. In addition, many species closely-related to the zebra finch are known to be rich in polymorphic inversions (Christidis, 1986a, 1986b, 1987; Huynh, Maney and Thomas, 2011; Zinzow-Kramer *et al.*, 2015). Therefore, any reference genomes produced for other Passeriformes by the VGP are likely to encounter similar issues if the individuals used are heterokaryomorphic.

The Vertebrate Genomes Project list several areas in which improvement is needed in their assembly pipeline. These include, more accurate and complete haplotype phasing and resolution of long repetitive regions such as telomeres, centromeres, and sex chromosomes (Rhie *et al.*, 2021). It is hoped that identifying haplotype-diagnostic SNPs can help guide the improvement of the project's assembly pipeline in these problem areas, and therefore improve the quality of the zebra finch genome as well as potentially many others.

Nevertheless, the inversion karyotypes of the available zebra finch reference genomes shown through this analysis will inform future studies investigating the zebra finch Z

chromosome. Specifically, any studies focussing on the structure, origin, and implications of the inversion polymorphism found on the Z chromosome will benefit from the findings of this research. For example, diagnostic SNPs are shown to be a suitable method of determining the inversion karyotype of a zebra finch (Figures 2.2 & 2.4). In particular, the diagnostic SNPs found here can identify between the A and B inversion haplotypes but may not be as suitable for identifying the C haplotype (Figure 2.4). This limitation may be because of the relatively small number of SNPs which differentiate between Haplotype C and the other haplotypes. As Haplotype C is not present in any of the reference genomes, and is found at the lowest frequency of all three haplotypes in both wild and captive populations (Kim *et al.*, 2017), this issue is a relatively minor one.

There are many unanswered questions related to this inversion polymorphism, such as the exact breakpoints of the inversion, the age of the polymorphism and the amount and type of selection acting on the inversion. Many of the findings reported here may be invaluable to answering these lingering questions. The two reference genomes recently produced by VGP are likely to be an integral part of many future studies on the zebra finch, and so having known inversion karyotypes for both could be a very useful tool for further research on the Z chromosome. As the zebra finch is a popular model organism, a large quantity of sequence data has been produced (for example in Singhal *et al.*, 2015). In subsequent chapters, these sequence data are repurposed to look at the Z chromosome inversion polymorphism. This is only possible because I could use the diagnostic SNPs as a method to find the inversion haplotypes of the re-sequenced birds (Figure 2.4 & Table 2.7). In summary, the list of 2330 SNPs which are very nearly fixed between different haplotypes, and the alleles which correspond to each haplotype are an invaluable tool for studies looking at sequence evolution and efforts to find out the potential phenotypic impacts of this supergene.

2.6. Supplementary Material

Table S2.1: List of 2123 SNP positions which were found to be highly diagnostic (weighted chi-square score >0.9) between the A and B inversion haplotypes.

Table S2.2: List of 1085 SNP positions which were found to be highly diagnostic (weighted chi-square score >0.9) between the B and C inversion haplotypes.

Table S2.3: List of 293 SNP positions which were found to be highly diagnostic (weighted chi-square score >0.9) between the A and C inversion haplotypes.

Table S2.4: List of 137 scaffolds from the bTGI alt reference assembly. The number of diagnostic SNP positions found on each scaffold are used to determine an overall percentage support for each scaffold being either haplotype A or B sequence.

Figure S2.1: Principal component analysis of 1202 male zebra finches from the University of Sheffield. The analysis used genotypes from 1085 diagnostic SNP positions with weighted chi-square score >0.9 in a comparison between inversion haplotypes B and C.

Figure S2.2: Principal component analysis of 1202 male zebra finches from the University of Sheffield. The analysis used genotypes from 293 diagnostic SNP positions with weighted chi-square score >0.9 in a comparison between inversion haplotypes A and C.

3. Relaxation of purifying selection within the zebra finch Z chromosome inversion polymorphism

3.1. Abstract

An inversion polymorphism, which controls variation in sperm morphology, is found on the Z chromosome of the zebra finch. There are three main haplotypes for this polymorphism (A, B & C), with A and B being the most common. Here, approximate breakpoints for the inversion polymorphism were inferred, allowing an investigation into selection pressures that act on the polymorphism, specifically on the A and B haplotypes. Evidence for a relaxation of purifying selection within the inversion region, particularly in the B haplotype was shown, which has allowed for sequence divergence between the haplotypes to occur. In addition, I found evidence for balancing selection, which maintains the inversion polymorphism.

3.2. Introduction

In Chapter Two, I identified the inversion karyotype of both a male (Black17, AB) and female (Blue55, A) zebra finch used to create the two zebra finch reference genomes (bTG1.4 and bTG2). I also identified the karyotypes of 24 zebra finches originally sequenced in 2013 (Singhal *et al.*, 2015, see Table 2.7). Now that the inversion karyotypes are known, the sequence data from these 24 zebra finches can be used to answer further questions about the supergene. Chromosomal inversions are a type of structural DNA mutation produced by the excision of a genome segment and its subsequent reinsertion in reverse orientation. Large inversions are usually very rare within populations because the fitness effects on heterozygotes are often negative, with gametes produced by these heterozygotes prone to widespread deletions and duplications, as the inversion can often disrupt meiosis and produce unbalanced gametes (Anton *et al.*, 2005). Therefore, this type of structural change is usually highly selected against. In some cases, however, an inversion mutation segregates within a population alongside a non-inverted, ancestral sequence, resulting in an inversion polymorphism.

Due to the reversed orientation of sequence in an inversion polymorphism, recombination is largely prevented between different inversion haplotypes. A large inversion can prevent recombination across a large region encompassing many whole genes, causing elevated linkage disequilibrium between different loci within the inversion. This may facilitate co-adaptation across the entire inverted region because recombination is unlikely to move alleles onto different haplotypes to those in which they arose (Hoffmann and Rieseberg, 2008). Groups of co-adapted genes within an inversion polymorphism are known as supergenes (Turner, 1967; Thompson and Jiggins, 2014). Since an inversion polymorphism can encompass a large region of a chromosome, a supergene can consist of hundreds of linked genes, potentially causing large-scale phenotypic variation whilst retaining the inheritance pattern of a single gene (Thompson and Jiggins, 2014).

Supergenes and their evolutionary implications have been studied for almost a century (Nabours, 1929; Nabours, Larson and Hartwig, 1933; Darlington and Mather, 1950), yet understanding the complexities surrounding their origin and evolution remains a challenge. Advances in genome sequencing technology means large structural variants like supergenes are cheaper and easier to sequence and assemble, whilst long-read sequencing has made it easier to resolve the complex and repetitive sequence that is often found at supergene breakpoints. Improvements in computing power and bioinformatic software have allowed for

genome sequence to be analysed across the length of a supergene (which are often millions of base-pairs in length). All of these advances have allowed for more thorough investigations of these complexities with recent studies suggesting there is considerable variation in the genomic architecture and origin of supergenes across different species (Gutiérrez-Valencia *et al.*, 2021). The continued evolution of a supergene is controlled in large part by the extent and pattern of recombination suppression. This recombination suppression will affect the direction and efficacy of selection acting on the supergene, and play a critical role not only on the evolution of a supergene, but also on its phenotypic consequences within natural populations (Berdan, Flatt, *et al.*, 2022).

Adaptive evolution as a result of positive selection is one example of an evolutionary consequence of supergenes. Alternative adaptive alleles can accumulate and spread between chromosomes of the same inversion type through the influence of positive selection. The lack of recombination between haplotypes prevents adaptive alleles from spreading between opposite haplotypes and therefore allows sequence divergence between inverted and ancestral haplotypes. In this way, supergenes act as a mechanism allowing multiple advantageous trait combinations to persist separately, which would normally be impossible due to the effects of recombination. Suppressed recombination also allows increased co-adaptation of genes within an inversion. Since recombination events that occur between co-adapted alleles will split up advantageous haplotypes, suppression of recombination means favourable alleles are less likely to be removed from the haplotypes in which they arose and therefore favourable combinations of linked alleles can accumulate and be maintained (Hoffmann and Rieseberg, 2008). Concordant with these theories, inversion polymorphisms have been found to control significant aspects of morphology and behaviour in several different organisms where alternative adaptations have arisen within the same populations. Well-documented examples include in fire ants (Wang *et al.*, 2013), ruffs (Küpper *et al.*, 2015; Lamichhaney *et al.*, 2015) and white-throated sparrows (Huynh, Maney and Thomas, 2011; Tuttle *et al.*, 2016). It has been theorised that these alternative adaptations caused by supergenes could facilitate speciation, as the large adaptive differences between haplotypes could lead to impaired fitness of heterozygotes, and homozygous carriers of the inverted haplotype can become an emerging species (Rieseberg, 2001).

Positive selection acting on a supergene could lead to the loss of the polymorphism, as the haplotype with greater fitness is driven to fixation by positive selection. In addition, genetic drift could also cause the frequencies of supergene haplotypes to become unbalanced within a

population. However, many known supergenes are over one million years old (Gutiérrez-Valencia *et al.*, 2021). For supergene polymorphisms to be maintained for this long, the polymorphism needs to be protected by balancing selection; potentially in multiple forms such as overdominance, antagonistic selection and negative frequency-dependent selection (Berdan, Flatt, *et al.*, 2022).

Whilst the reduced rate of recombination between supergene haplotypes can facilitate the acquisition of beneficial alleles, it may also speed up the accumulation of deleterious mutations. Normally, deleterious mutations would be removed from populations through the action of purifying selection. However, suppressed recombination reduces the efficacy of purifying selection, meaning deleterious alleles are able to persist within supergenes (Gutiérrez-Valencia *et al.*, 2021; Berdan, Flatt, *et al.*, 2022). Purifying selection's efficacy is diminished as the reduction in recombination within a supergene region creates a pseudo-population substructure. As opposite supergene haplotypes cannot recombine, the effective population size for each arrangement of the supergene is reduced, thereby reducing the efficacy of purifying selection. (Berdan, Blanckaert, *et al.*, 2022). As deleterious mutations accumulate, they can alter amino acid sequence, disrupt genes through frameshifts or premature stop codons or alter important regulatory regions, resulting in degeneration of one of the haplotypes. It is theorised that this degeneration could destabilise polymorphic supergene systems, and result in loss of the polymorphism entirely, as one haplotype is selected against and removed (Berdan, Blanckaert, *et al.*, 2022; Berdan, Flatt, *et al.*, 2022). It has also been suggested that reduced recombination due to an inversion polymorphism could be the driving force in the evolution of sex chromosomes from autosomes, and degeneration of one sex chromosome could be the result of relaxed purifying selection allowing accumulation of deleterious alleles (Wright *et al.*, 2016). The reduced recombination in supergenes has an impact on the multiple forms of selection that can act on supergenes. Understanding whether positive, balancing, or purifying selection pressures are acting on a given supergene polymorphism is crucial to understanding its potential impact on natural populations, as well as its evolutionary trajectory.

In populations of zebra finches, *Taeniopygia guttata*, a large inversion polymorphism encompassing the majority of the Z chromosome (approximately 60Mbp), was discovered using cytological studies (Itoh *et al.*, 2011) and has subsequently been further characterised by population genetic approaches (Kim *et al.*, 2017; Knief *et al.*, 2017). The zebra finch, an Australian songbird of the *Estrildidae* family, is an important model organism in a variety of biological fields including neurobiology (Mello, 2014), behavioural ecology (Galoch and

Bischof, 2007; Guillette and Healy, 2014) and evolutionary biology, where studies of sexual selection in the species are common (Simons and Verhulst, 2011). Since this supergene is located on a sex chromosome, recessive alleles will always be expressed in females, but less often in males. This could mean that selection against deleterious recessive alleles affecting females will be stronger than deleterious recessive alleles affecting males, and segregating variation might be more likely to affect traits expressed in males. In fact, previous work on the zebra finch supergene has identified three distinct inversion haplotypes on the Z chromosome (A, B & C), which are responsible for 67–90% of the additive genetic variance in sperm morphology within male finches. AA males are characterised by sperm with long overall length, AB and AC males by intermediate length sperm, and BB, BC and CC males by short length sperm (Kim *et al.*, 2017; Knief *et al.*, 2017). The midpiece to tail size ratio is an important factor in explaining variation in sperm motility (Bennison *et al.*, 2016), with an intermediate ratio producing sperm with the fastest swimming speed (in this case males with haplotype AB and AC); the heterozygote advantage due to this increased sperm motility possibly maintains the supergene polymorphism within the zebra finch population (Kim *et al.*, 2017; Knief *et al.*, 2017). The actual sequence variants within the zebra finch inversion polymorphism that result in sperm variation are still undiscovered. In addition, it is unclear whether the inversion causes phenotypic variation in other traits.

Here I aim to determine the types of selection that have acted on the inversion polymorphism on the Z chromosome in zebra finches, specifically on the two most commonly occurring haplotypes, A and B. This will require meeting three objectives: i) to determine the approximate breakpoints of the inversion region, ii) to test for selection which has acted on the sequence within these breakpoints, using a range of summary statistics such as nucleotide diversity, the Fixation Index (F_{ST}), Tajima's D (Tajima, 1989), Fay and Wu's H (Fay and Wu, 2000) and a composite test of D and H known as Zeng's E (Zeng, Shi and Wu, 2007), iii) to determine the amount of non-synonymous and synonymous variation within and between supergene haplotypes as a further indicator of selection. These statistics are all widely used for determining signatures of selection in DNA sequence, in particular allele frequency spectra statistics such as D and H are classic tests which are widely utilised (Nielsen, 2005), including in previous studies of supergenes (Cohen and Privman, 2020; Dagilis *et al.*, 2022). In addition, tests based the amount of non-synonymous and synonymous variation between groups have been suggested to have more power to detect certain types of selection including negative

selection (Zhai, Nielsen and Slatkin, 2009). It is hoped that this work will help to better understand the evolutionary trajectory of a large supergene in an important model organism.

3.3. Methods and Materials

3.3.1. Sequence data

In Chapter 2, male zebra finches (n=1202) from a population that was previously maintained at the University of Sheffield (Kim *et al.*, 2017), were used to determine highly diagnostic SNPs which distinguish between the three haplotypes. These birds were genotyped at 3056 marker positions spanning almost the entirety of the Z chromosome (Kim *et al.*, 2017). These genotypes were used in chi-square tests to determine which markers showed the most significant difference in genotype frequency between the haplotypes and were therefore highly diagnostic of the inversion karyotype. Each marker was given a weighted chi-square score ranging from 0 to 1; any SNPs with a weighted chi-square score greater than 0.9 (90% of maximum score) were included in a list of 2330 “diagnostic SNPs” (see Chapter 2, Supplementary Tables S2.1, S2.2 & S2.3 for the full list).

Publicly available sequence reads (www.ebi.ac.uk/ena/data/view/PRJEB10586) from 24 zebra finches *Taeniopygia guttata* (11 female, 13 male), and one female long-tailed finch *Poephila acuticauda acuticauda* were used in this chapter. All birds were originally sequenced using Illumina HiSeq 2000 paired end sequencing (read length = 100bp) in 2012 and 2013 (Singhal *et al.*, 2015); Table 2.7 lists these 24 birds and their read depths. Of the 24 zebra finches, 19 were wild birds caught at Fowler’s Gap, New South Wales, Australia. The other 5 birds were domesticated zebra finches bred in captivity at East Carolina University; the group consists of a mother (MP1), father (MP2), and three sons (MP3, MP4 & MP5). The inversion karyotypes of all of these 24 zebra finches were inferred in Chapter 2, using the list of 2330 diagnostic SNPs. In total 8 birds were scored as inversion karyotype A or AA (4 females, 4 males), 7 birds were scored as inversion karyotype B or BB (6 females, 1 male), 1 female bird was scored as karyotype C, 4 male birds were scored as AB heterozygotes, and another 4 males were scored as AC heterozygotes (Chapter 2, Figure 2.4 & Table 2.7).

In addition, sequence from a long-tailed finch, *Poephila acuticauda acuticauda*, was used as an outgroup. This long-tailed finch sequence was obtained from a female wild bird caught at Mount House, Western Australia, Australia, and was sequenced using Illumina HiSeq 2000 paired end sequencing (read length = 100bp) in 2012 (Singhal *et al.*, 2015).

3.3.2. Reference genome

A zebra finch reference genome, bTG1.4, assembled by the Vertebrate Genomes Project (Rhie *et al.*, 2021) was used for this analysis. This reference genome was constructed using DNA from a male zebra finch known as “Black17”, which has been shown to be an AB heterokaryotype for the Z chromosome inversion polymorphism (see Chapter 2, Table 2.4 & Figure 2.2). This reference genome was used rather than the reference genome constructed using sequence from the female finch “Blue55” (bTG2), as only bTG1.4 had been fully annotated, meaning further analysis of specific genes was possible.

3.3.3. Variant calling

Sequence reads from 24 zebra finches and one long-tailed finch were aligned to a male zebra finch reference genome (bTG1.4) with Bowtie v2.3.4.3 using its default settings (Langmead and Salzberg, 2012). Variants were called between aligned reads and bTG1.4 using the genome analysis toolkit (GATK) version 4.2.5.0 variant calling pipeline (Poplin *et al.*, 2017). Duplicate reads were removed for each individual, using GATK MarkDuplicates, before HaplotypeCaller was run using aligned reads. Genotypes were then called using GATK GenotypeGVCFs. Indels were discarded whilst called SNPs were retained and then filtered for quality using GATK VariantFiltration. Parameters that were used in variant quality filtering were: variant confidence (QD) <2.0, which is intended to normalize the variant quality in order to avoid inflation caused when there is deep coverage; phred-scaled probability of strand bias (FS) >60.0 & symmetric odds ratio test for strand bias (SOR) >4.0, which both describe whether the alternate allele is seen more or less often on the forward or reverse strand than the reference allele; mapping quality of reads (MQ) <40.0, and the compared mapping qualities of reads supporting the reference and alternate allele (MQRankSum) <-12.5; and a comparison of positions of the reference and alternate alleles within different reads (ReadPosRankSum) < -8.0.

3.3.4. Calculation of summary statistics

SNPs found through variant calling with GATK v4.2.6.1 were used to measure a variety of summary statistics. These statistics include fixation index (F_{ST}) between the 8 zebra finches of inversion karyotype A or AA and the 7 zebra finches of inversion karyotype B or BB. In addition, nucleotide diversity (π), Tajima’s D, Fay & Wu’s H and Zeng’s E (a composite statistic of D and H) were all calculated. Initially these statistics were calculated using SNPs

called using all 24 zebra finches, then using all 8 A or AA karyotype finches, and finally using all 7 B or BB karyotype finches. In each case, the statistic was calculated for sliding windows of 100Kb of sequence, with each window overlapping by 10Kb. As an outgroup is required to calculate Fay & Wu's H and Zeng's E, sequence from a female long-tailed finch was used.

All summary statistics were calculated for sliding windows across the length of the bTG1.4 reference genome Z chromosome, but the three neutrality test statistics (D, H & E) were also calculated for the 30 largest autosomes (Chromosomes 1-30) in the bTG1.4 reference. All summary statistics were calculated using R v4.2.1 (with Rstudio v2022.07.1+554) and the package "PopGenomeR" v2.7.5 (Pfeifer *et al.*, 2014).

Since there were an unequal number of A/AA birds and B/BB birds, nucleotide diversity, Tajima's D, Fay & Wu's H and Zeng's E were also all calculated using just 6 of the 8 A/AA birds (4/4 females, 2/4 males) so that the number of A-type Z chromosome sequences included in the analysis was equal to the number of B-type Z chromosome sequences (n=8). These statistics were also calculated using a subset of 6 birds (2 males, 4 females, total number of Z chromosomes = 8) of a mixture of different inversion karyotypes to check if patterns seen in all 24 birds would also occur with a smaller sample size (karyotypes = A, AA, AB, B, B, C) F_{ST} was also calculated between the 7 B/BB finches (8 Z chromosomes) and the subset of 6 A/AA finches (8 Z chromosomes).

3.3.5. Identifying synonymous and non-synonymous SNPs

Genome annotations from the NCBI *Taeniopygia guttata* annotation release 106 (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Taeniopygia_guttata/106/) were used to determine synonymous and non-synonymous sites within coding regions on the zebra finch (bTG1.4) Z chromosome. Using the R package, PopGenomeR v2.7.5 (Pfeifer *et al.*, 2014), SNPs found through variant calling with GATK v4.2.6.1 were sorted into those falling within and those outside of coding sequence. Then, SNPs within coding regions were sorted into those which were fixed between A and B and at either a non-synonymous ($N_{Dnon-syn}$) or a synonymous (N_{Dsyn}) position, and those which were polymorphic within either the A or B type birds and at either a non-synonymous ($N_{Pnon-syn}$) or synonymous (N_{Psyn}) position. The number of SNPs within each of these categories was then counted, and these counts were used to perform a series of chi-square tests to explore the amount of nonsynonymous and synonymous variation between the A and B haplotypes inside and outside of the inversion region. A list of

genes with fixed differences ($N_{D_{\text{non-syn}}}$ or $N_{D_{\text{syn}}}$) within their coding regions was produced, as within this list could be the genes responsible for variation in sperm morphology between the haplotypes provided the fixed mutations responsible for this morphological variation are within the coding sequence rather than within regulatory regions.

3.4. Results

3.4.1. SNP density

A total of 510,385 Z chromosome SNPs were found through variant calling with GATK using sequence reads from 24 individuals (Singhal *et al.*, 2015), after passing quality thresholds. SNP density was calculated across the length of the bTG1.4 reference genome Z chromosome (Figure 3.1). SNP density is markedly higher at each end of the Z chromosome, specifically between 0 and 6.50Mb and between 70.10 and 75.39Mb, whilst the region between 6.50Mb and 70.10Mb has relatively low SNP density (Figure 3.1). The transitions between low and high SNP density are ‘step-like’ rather than gradual. These physical positions are assumed to be approximate locations of the breakpoints of the inversion polymorphism on the Z chromosome.

When using SNPs called within all 24 zebra finches, the SNP density outside of the predicted breakpoint locations was ~17,800 SNPs/Mb - more than three times the density found within the inversion (Figure 3.1c & Table 3.1). This difference is expected since the reduced recombination within the inversion region makes the effective population size smaller compared to regions outside the inversion, and therefore the likelihood of alleles being lost by genetic drift greater. Therefore, fewer mutations persist and become polymorphisms within the inversion compared to outside the inverted region where recombination is unhindered. This difference in SNP density is less pronounced when only using SNPs found through variant calling with just the 8 zebra finches of inversion karyotype A or AA (Figure 3.1a & Table 3.1). The density of Haplotype A SNPs outside the inversion (~4000 SNPs/Mb) was over six times greater than the density between the predicted inversion breakpoints (~600 SNPs/Mb). However, SNP density inside and outside the inversion differs the most when using SNPs found through variant calling in just the 7 zebra finches of inversion karyotype B or BB. In these birds, the density between the predicted inversion breakpoints was ~200 SNPs/Mb whilst outside the inversion the density was ~13,200 SNPs/Mb - a more than 70-fold increase (Figure 3.1b & Table 3.1).

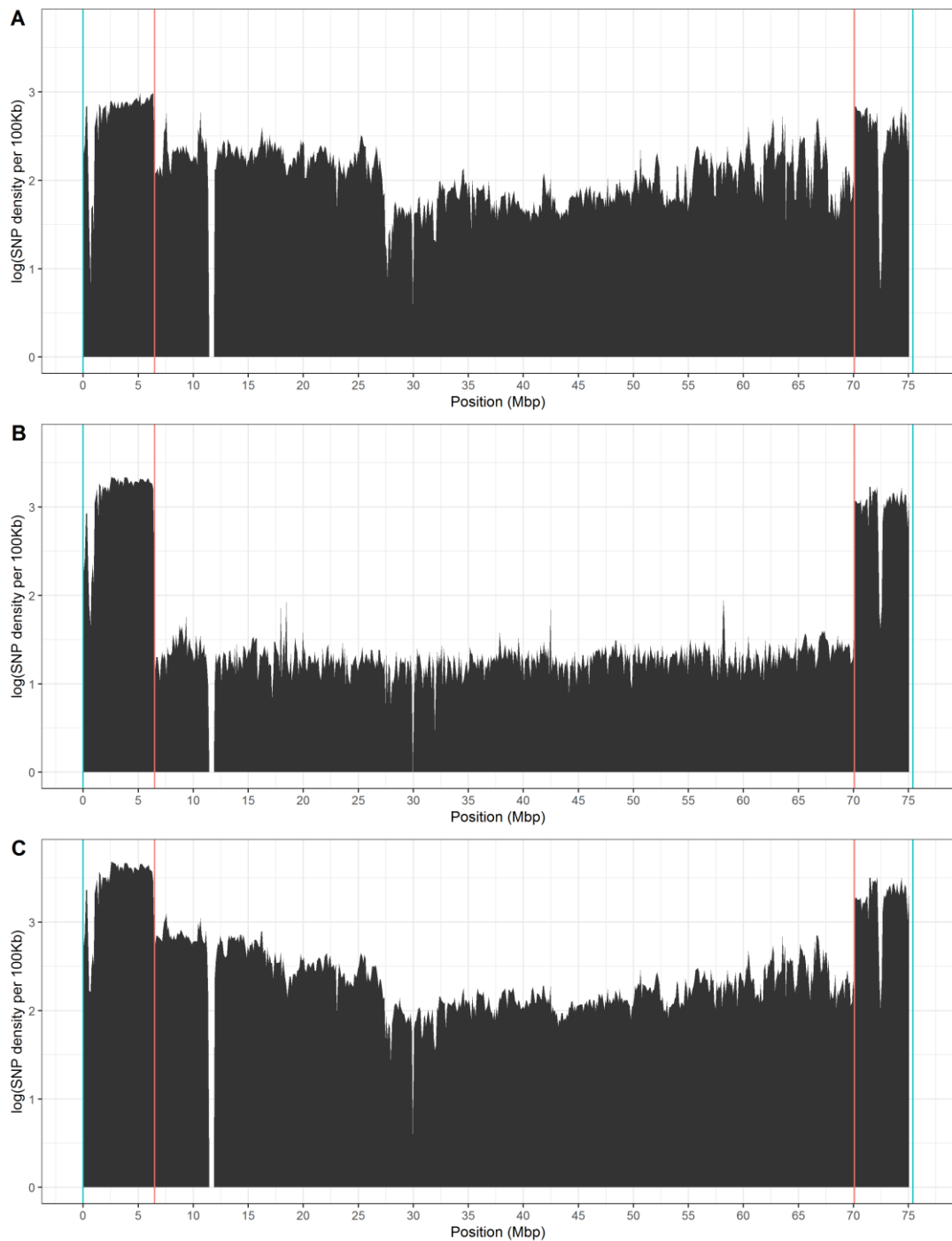


Figure 3.1: SNP density on the Z chromosome.

SNP density was estimated in: **a.** zebra finches scored as inversion karyotype A or AA **b.** zebra finches scored as inversion karyotype B or BB and **c.** all available zebra finch sequence (24 individuals). All sequence reads are from Singhal et al., 2015 and were aligned against the zebra finch reference genome (*bTG1.4*). The number of SNPs located within non-overlapping 100Kb windows were counted and the density was then \log_{10} transformed before plotting. The ends of the *bTG1.4* Z chromosome are indicated with blue lines, whilst inversion breakpoints inferred through a large change in SNP density are indicated with red lines.

	Within Inversion (6.50 – 70.10Mb)	Outside Inversion (0-6.50Mb, 70.10- 75.39Mb)	Whole Z chromosome
Number of SNPs called from all birds	300,379	210,006	510,385
Density of SNPs called from all birds (SNPs/Mb)	~4,700	~17,800	~6,800
Number of SNPs called from type A birds	35,412	47,622	83,034
Density of SNPs called from type A birds (SNPs/Mb)	~600	~4000	~1,100
Number of SNPs called from type B birds	11,598	155,504	167,102
Density of SNPs called from type B birds (SNPs/Mb)	~200	~13,200	~2,200

Table 3.1: SNP density on the Z chromosome.

SNPs were called from zebra finch reads from 24 individuals from Singhal *et al.*, 2015 and the zebra finch reference genome (bTG1.4). The density of SNPs across the length of the bTG1.4 Z chromosome and inside and outside of the approximate breakpoints of the Z chromosome inversion polymorphism were calculated. Breakpoints were inferred through an observed large and sudden change in SNP density. For each region, SNP density within A-type and within B-type birds was calculated.

3.4.2. Pi Diversity Statistic

Nucleotide diversity (π) was considerably higher in regions at either end of the bTG1.4 Z chromosome (0-6.50Mb, 70.10-75.39Mb). The mean per site nucleotide diversity within these regions was 0.0056 (for 0-6.50Mb) and 0.0039 (for 70.10-75.39Mb). The mean average for these two regions combined was 0.0048, more than triple the mean across the intervening region (0.0013; Table 3.2 & Figure 3.2c). The genome wide mean nucleotide diversity was 0.0054, meaning the inversion region on the Z chromosome was less than a quarter of the genome-wide average (Figure 3.2d). This pattern further confirms the approximate boundaries of the inversion polymorphism to be 6.50Mb and 70.10Mb since reduced recombination within the inversion region is expected to reduce genetic diversity. When only SNPs called from either A/AA or B/BB karyotype birds are considered, the pattern of considerably lower diversity

within the inversion region persists, but particularly when only considering B/BB birds (Table 3.2 & Figure 3.2a-b). The disparity in nucleotide diversity between the A and B haplotypes could be explained by the A haplotype being ancestral and therefore retaining genetic diversity which pre-dates the inversion mutation. The lower diversity in the B haplotype could be because it is the derived haplotype and so only contains polymorphism acquired after the inversion mutation. Alternatively, the selection pressures driving the divergence between the A and B haplotypes may be stronger in Haplotype B. A third explanation is that genetic drift was stronger in B than A, as B is much rarer in wild populations (Kim *et al.*, 2017). When accounting for the difference in sample size between A and B, this pattern persists (Supplementary Figure S3.1), suggesting the reduced diversity in B is not due to fewer B-type Z chromosome sequences being present in the dataset.

	Within Inversion (6.50 – 70.10Mb)	Outside Inversion (0-6.50Mb, 70.10- 75.39Mb)	Whole Z chromosome
Mean Pi diversity using SNPs called from all birds	0.00131 (<i>s.e</i> 0.60e-05)	0.00485 (<i>s.e</i> 5.76e-05)	0.00186 (<i>s.e</i> 1.80e-05)
Mean Pi diversity using SNPs called from type A birds	0.00056 (<i>s.e</i> 0.71e-05)	0.00452 (<i>s.e</i> 5.89e-05)	0.00117 (<i>s.e</i> 1.98e-05)
Mean Pi diversity using SNPs called from type B birds	0.00007 (<i>s.e</i> 0.10e-05)	0.00467 (<i>s.e</i> 5.37e-05)	0.00079 (<i>s.e</i> 2.10e-05)

Table 3.2: Nucleotide diversity (π) on the Z chromosome.

Pi was calculated using SNPs called from 24 zebra finches, and then separately using only finches of karyotype A/AA ($n=8$) and then only finches of karyotype B/BB ($n=7$). *Pi* was estimated across the entire Z chromosome as well between approximate inversion breakpoints (6.50-70.10Mb) and outside of these breakpoints (<6.50Mb, >70.10Mb).

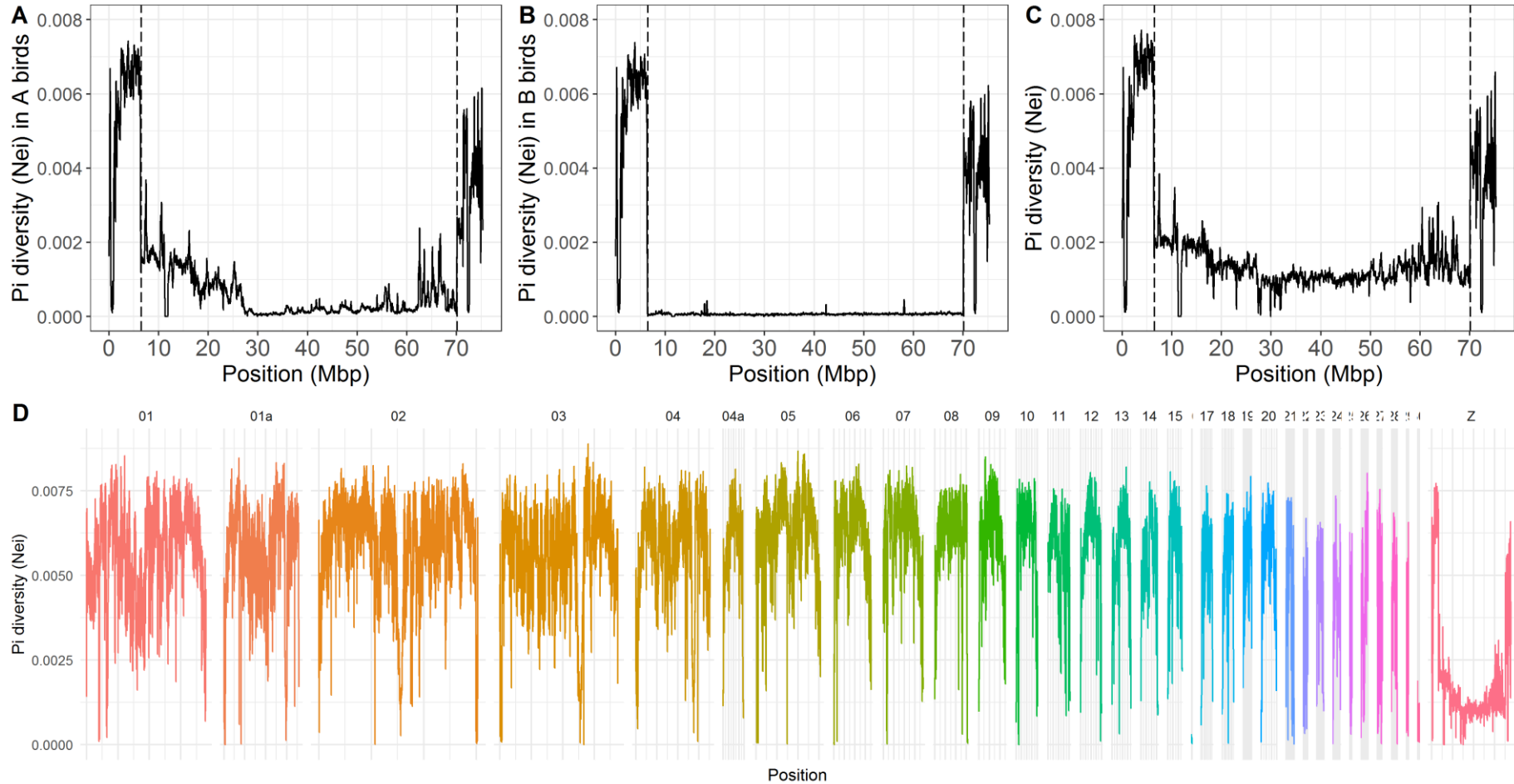


Figure 3.2: Nucleotide diversity (π) on the Z chromosome and across the genome.

Nucleotide diversity (π) plots showing the average number of nucleotide differences per site across 100Kb windows with a 10Kb step. **a.** only using SNPs from A/AA karyotype birds ($n=8$) **b.** only using SNPs from B/BB karyotype birds ($n=7$) **c.** using SNPs from all 24 available zebra finch sequences. Diversity is plotted across the zebra finch reference genome (bTG1.4) Z chromosome, and dashed lines indicate predicted breakpoints of the Z chromosome inversion polymorphism. **d.** Diversity calculated using SNPs from all 24 zebra finches across autosomes 1-30 in addition to the Z chromosome.

3.4.3. Fixation Index (F_{ST})

The fixation index between birds of inversion karyotype A or AA and birds of inversion karyotype B or BB was calculated for each SNP on the bTG1.4 reference genome Z chromosome. F_{ST} was much lower at the ends of the Z chromosome (Figure 3.3b), with a mean of 0.064 (s.e = 0.0021) between 0 and 6.50 Mb and a mean of 0.055 (s.e = 0.0003) between 70.1Mb and the end of the chromosome, compared with a mean of 0.548 (s.e = 0.0016) within the inversion. These results are consistent when only a subset of A/AA karyotype birds are used, so that the number of Z chromosomes of each type are equal (Supplementary Figure S3.2). The pattern of F_{ST} at the ends of the Z chromosome is consistent with the rest of the genome, whilst within the inversion region F_{ST} is much higher (Figure 3.3a). These results indicate a considerable amount of sequence divergence between the two haplotypes has occurred within the inversion region.

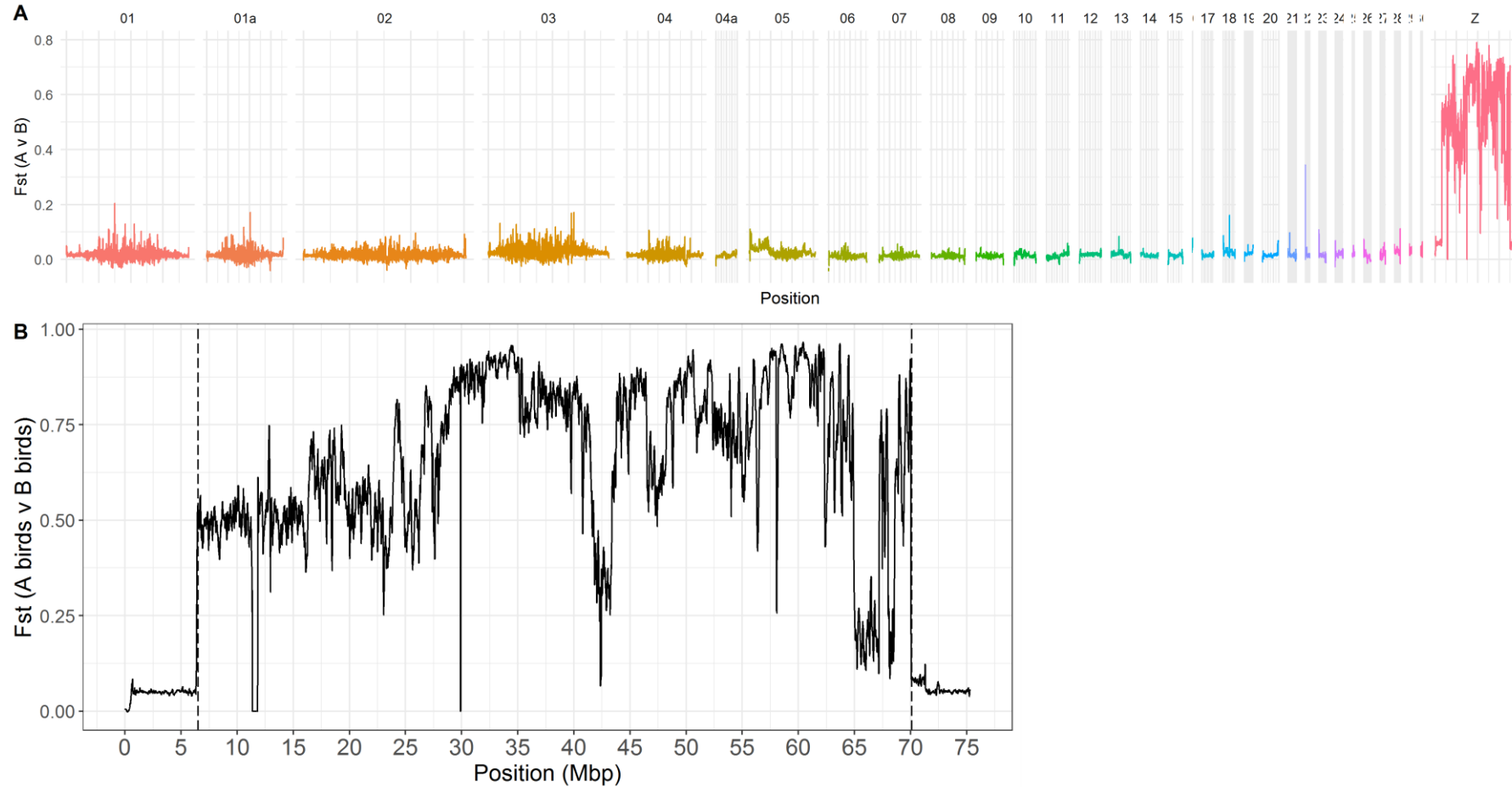


Figure 3.3: Fixation index (F_{ST}) on the Z chromosome and across the genome.

Fixation index (F_{ST}) between zebra finches of inversion karyotype A or AA ($n=8$) and of inversion karyotype B or BB ($n=7$). SNPs were found within all fifteen birds when aligning reads against the zebra finch reference genome (bTG1.4). F_{ST} is averaged for all SNP positions in 100Kb sliding windows of sequence across the **a.** whole bTG1.4 reference genome **b.** bTG1.4 Z chromosome, with each window overlapping by 10Kb. Dashed lines indicate predicted breakpoints of the Z chromosome inversion polymorphism.

3.4.4. Neutrality test statistics

Within the inversion there was little deviation from 0 in either Tajima's D, Fay & Wu's H or Zeng's E when limited to just the A or B haplotype birds (Figures 3.4a, b, d, e, g & h). All three of these metrics are sensitive to positive selection, so a lack of deviation from 0 suggests there is no positive selection acting on either of the inversion haplotypes. When considering all 24 birds, there was little difference in Tajima's D between the inverted and non-inverted regions of the Z chromosome. The mean Tajima's D within the inverted region was -1.83 (s.e = 0.004), whilst outside of the inverted region (0-6.50Mb and 70.10-75.39Mb) the mean average had a similar value of -1.91 (s.e = 0.006; Figure 3.4c). A narrow positive spike in Tajima's D, reaching a maximum value of 1.82 (all birds) and >2 (A birds and B birds), is seen from 58.06Mb to 58.13Mb (Figure 3.4a, b c). This could indicate an area under the influence of balancing selection, but there are no genes of known function within this region. Alternatively, this narrow spike could be a spurious result, possibly due to an assembly error or some reads from a paralogous region being mapped to that region in both sets of birds, resulting in some false SNP calls. However, there is no noticeable difference in SNP density within that region (Figure 3.1b-c). The mean Tajima's D (when considering all 24 finches) within the inversion region on the Z chromosome does not differ greatly to the genome wide average of -1.73 (s.e = 0.001). However, the narrow peak located between 58.06Mb to 58.13Mb, is greater than the genome-wide mean value and in fact only one autosomal location on Chromosome 1a has a larger value (Figure 3.5a).

When considering all birds, there is a far more prominent difference between regions inside and outside of the inversion for both Fay & Wu's H and Zeng's E (Figure 3.4f & i). In these cases, outside the inversion the statistics are very close to 0, suggesting close to neutral evolution, whilst inside the statistics are very negative (H) and very positive (E), indicating an excess of high frequency derived alleles within the inversion region. These values are also very different to the rest of the genome; H has a genome-wide mean of -1.17 (s.e = 0.008; Figure 3.5b), and Zeng's E has a genome-wide mean of -0.68 (s.e = 0.008; Figure 3.5c). These patterns are consistent with balancing selection acting to maintain both haplotypes, which combined with the lack of recombination, results in sequence divergence taking place between the two inversion haplotypes.

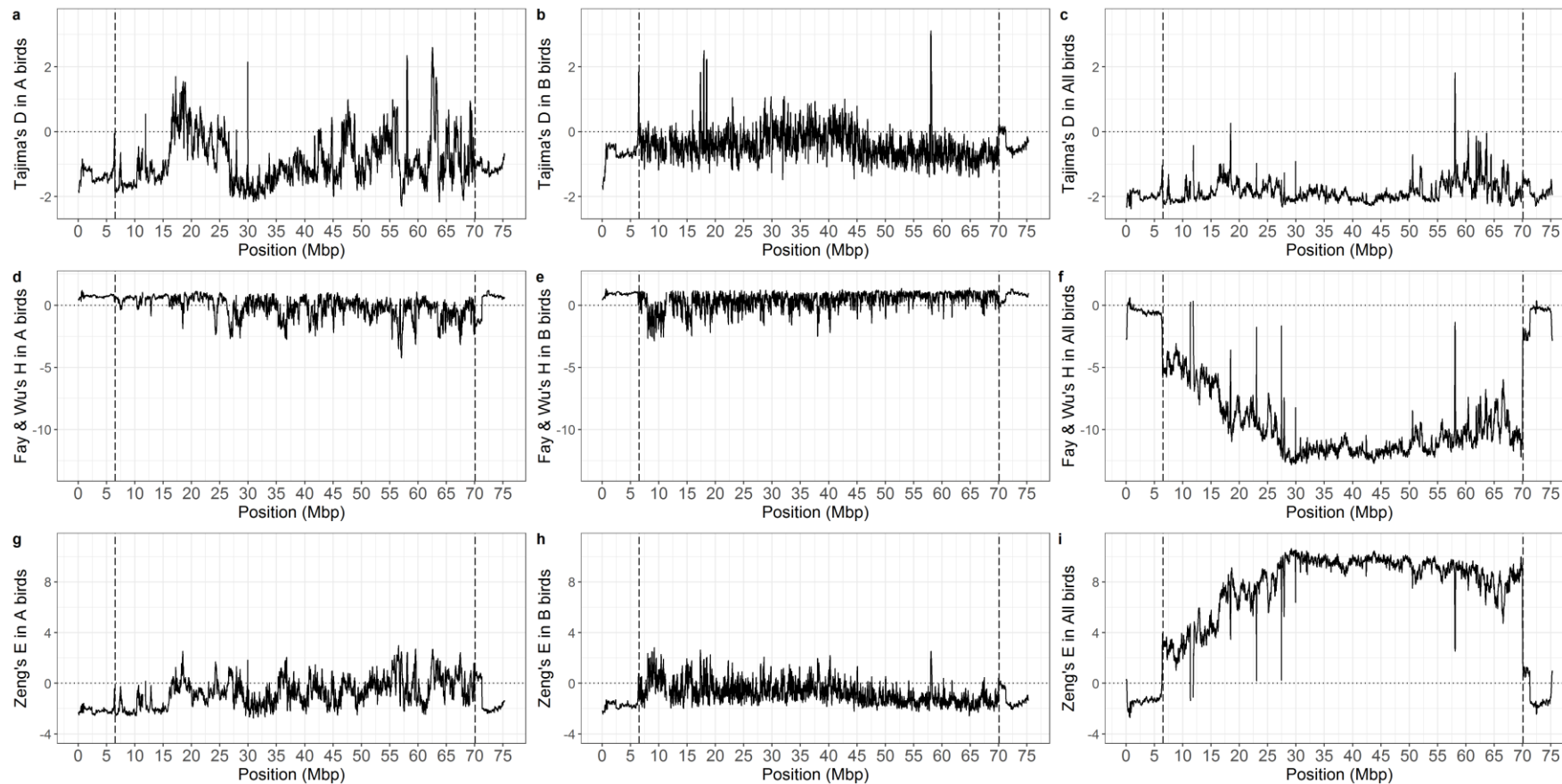


Figure 3.4: Neutrality statistics on the Z chromosome.

Including: Tajima's D calculated using SNPs called using sequence data from **a.** 8 karyotype A/AA finches **b.** 7 B/BB finches and **c.** all 24 available finches, Fay & Wu's H calculated using **d.** 8 karyotype A/AA finches **e.** 7 B/BB finches and **f.** all 24 available finches, and Zeng's E calculated from **g.** 8 karyotype A/AA finches **h.** 7 B/BB finches and **i.** all 24 available finches. Reads were aligned against the reference genome (*bTG1.4*) to give their physical positions. Sequence from a female long-tailed finch was used as an outgroup sequence for calculating H and E . Dashed lines indicate the predicted approximate breakpoints of the Z chromosome inversion region. All statistics were calculated from 100kbp sliding windows across the genome with a 10kbp step between windows.

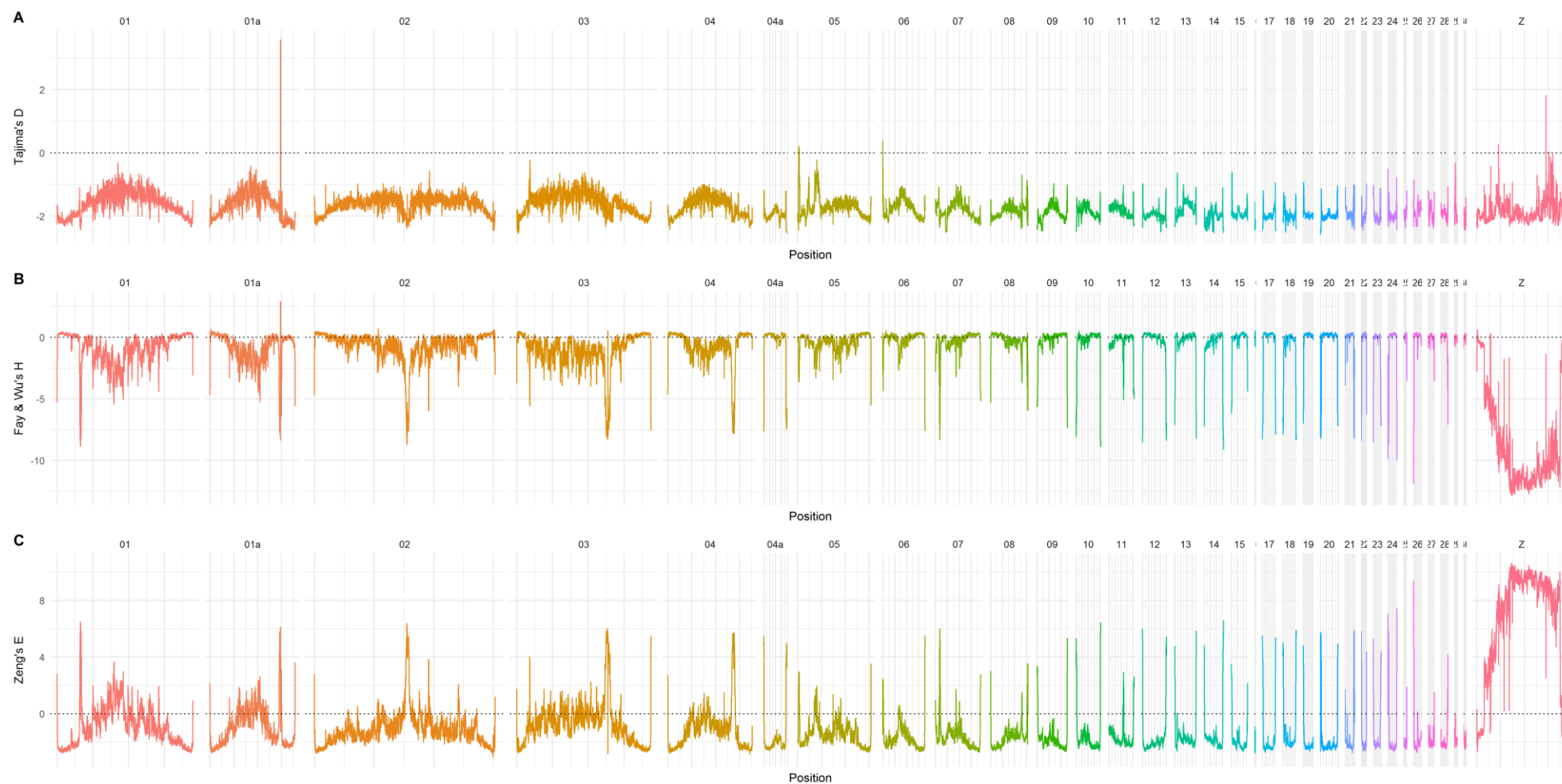


Figure 3.5: Neutrality statistics across the genome.

Genome-wide neutrality statistics calculated using SNPs called using sequence data from 24 zebra finches. Sequence from a female long-tailed finch was used as an outgroup sequence for calculating H and E . **a.** Tajima's D **b.** Fay & Wu's H **c.** Zeng's E . All statistics were calculated from 100kbp sliding windows across the genome with a 10kbp step between windows.

3.4.5. Synonymous and non-synonymous SNPs

A total of 363 fixed differences between A and B inversion haplotypes were found within coding regions on the Z chromosome; all of the fixed coding region differences were in genes within the predicted boundaries of the inversion polymorphism (Table 3.3), further suggesting the approximate boundaries of the inversion region are correct. Within the boundaries of the inversion, there were 2,174 SNPs in A type birds and 781 SNPs segregating within B type birds. In both A and B haplotype birds, nonsynonymous SNPs were more prevalent than synonymous SNPs, with ratios of 1610 to 1399 (1.12:1) in A birds, and 463 to 318 (1.46:1) in B birds respectively (Table 3.3). However, outside of the inversion breakpoints, the ratio of nonsynonymous to synonymous SNPs is very different, with far higher numbers of synonymous SNPs than nonsynonymous SNPs (Table 3.3).

Chi-square tests of independence (degrees of freedom = 1) were performed to examine differences in proportions of nonsynonymous and synonymous polymorphisms: (a) between the two haplotypes, and (b) from the proportion of nonsynonymous and synonymous fixed differences between the two haplotypes. Tests were performed within the inversion, outside of the inversion and across the entire Z chromosome (Table 3.3). These tests are analogous to McDonald-Kreitman tests (McDonald and Kreitman, 1991) except they compare divergence between the A and B haplotypes rather than between two related species.

	Inversion Region (6.50-70.10 Mb)		Outside Inversion (0-6.50Mb, 70.10-75.39Mb)		Whole Z Chromosome		Inversion vs Outside Inversion
	N	S	N	S	N	S	
Polymorphism in A	1150	1024	1584	3351	2734	4375	$\chi^2 = 275.01, p = <0.00001$ $\chi^2 = 217.68, p = <0.00001$
Polymorphism in B	463	318	1220	2663	1683	2981	
Fixed Differences: A vs B	182	181	0	0	182	181	
Polymorphism in A compared to A vs B differences	$\chi^2 = 0.84, p = 0.36$						
Polymorphism in B compared to A vs B differences	$\chi^2 = 8.06, p = 0.0045$						
Polymorphism in A compared to polymorphism in B	$\chi^2 = 9.19, p = 0.0024$		$\chi^2 = 0.43, p = 0.51$		$\chi^2 = 6.67, p = 0.0098$		

Table 3.3: Synonymous and nonsynonymous polymorphism and fixed differences across the Z chromosome.

Comparison of the amount of synonymous and non-synonymous polymorphism within birds of the same inversion haplotype with the number of synonymous and non-synonymous fixed differences between the two inversion haplotypes. Separate comparisons are made using all SNPs on the bTG1.4 reference Z chromosome, just SNPs within the predicted inversion breakpoints, and just SNPs outside of the predicted inversion breakpoints. N = nonsynonymous changes, S = synonymous changes. In both A and B birds there is a significant excess of nonsynonymous polymorphisms within the inversion compared to outside of it. Within the inversion, there is a significant excess of nonsynonymous polymorphisms in B birds, relative to within A birds, and relative to fixed nonsynonymous differences between A and B birds.

There was a striking contrast in the patterns of coding region polymorphism between the inversion regions and the non-inverted regions for both A and B haplotypes ($p = <0.00001$). In particular, for both haplotypes there was a big excess of non-synonymous mutation inside the inversion region (Table 3.3). Non-synonymous mutations are usually selected against by purifying selection, and so this excess could be caused by a relaxation of purifying selection or, if these changes are adaptive, by ongoing positive selection acting on each haplotype. However, there was no significant difference in the proportion of nonsynonymous and synonymous polymorphisms in Haplotype A birds and the proportion of nonsynonymous and synonymous fixed differences between haplotypes A & B. If positive selection were driving divergence of the A haplotype, a greater proportion of nonsynonymous fixed difference relative to nonsynonymous polymorphisms would be expected (McDonald and Kreitman, 1991). There was a significant difference in nonsynonymous:synonymous proportions between Haplotype B SNPs and A vs B fixed differences ($p = 0.0045$). There were relatively more nonsynonymous SNPs than there were nonsynonymous fixed differences. This is the opposite to what would be expected if positive selection was acting on Haplotype B alleles, thus the pattern is more consistent with relaxed purifying selection.

A significant difference in polymorphism patterns was found between haplotypes A and B, both across the whole Z chromosome and within the inversion region, with B birds having relatively more nonsynonymous SNPs. The difference was more significant within the inversion region ($p=0.0024$ compared to $p=0.0098$). When considering just the regions of the Z chromosome outside the inversion region, the difference was not significant ($p = 0.52$), suggesting that the differences in proportions of nonsynonymous to synonymous SNPs seen between the A and B haplotypes is limited to just the inversion region.

3.4.6. Fixed differences in coding regions between inversion haplotypes

A total of 363 fixed differences in coding regions of genes was found between the A and B haplotypes (Table 3.3). Within this list could be the SNPs responsible for variation in sperm morphology between the haplotypes. In particular, nonsynonymous changes in coding regions are more likely to affect phenotype than silent changes. There were 181 synonymous fixed differences between the A and B haplotypes (Supplementary Table S3.3) and 182 nonsynonymous fixed differences (Supplementary Table S3.2). These 363 fixed differences are distributed between the coding regions of 203 genes (Supplementary Table S3.1). Whilst any of these 203 genes may contribute to the variation in sperm traits seen in zebra finches, a

subset of fourteen genes containing at least 3 non-synonymous fixed differences and gene ontology classes associated with them are presented here (Table 3.4). with nine of these genes falling within a ~ 6Mbp region (57.6Mb - 63.7Mb), which suggests a region of particularly suppressed recombination between haplotypes.

Gene Name	Gene Function	CDS Start	CDS End	P(A) non- syn	P(A) syn	P(B) non- syn	P(B) syn	D non- syn	D syn
ANKRD31	Homologous chromosome pairing at meiosis	60,332,675	60,397,704	2	1	1	0	11	4
RIC1	Intracellular protein transport	63,690,131	63,737,954	1	1	0	1	8	2
CMYA5	Regulation of skeletal muscle adaptation	58,398,352	58,444,093	1	0	1	3	7	2
FAM169A	N/A	60,471,809	60,510,269	1	0	0	0	6	4
VPS13A	Flagellated sperm motility	57,618,036	57,727,092	3	0	0	0	5	2
PRUNE2	Apoptotic process	57,830,656	57,934,812	5	0	1	0	5	1
LOC115491269	N/A	51,994,968	52,007,313	3	0	1	0	4	1
FANCC	DNA repair and gamete generation	10,340,132	10,414,964	9	5	1	0	3	0
LOC101233669	N/A	11,911,053	11,929,338	11	6	6	7	3	0
LOC115491137	N/A	17,468,602	17,481,437	8	18	2	0	3	0
SPEF2	Sperm axoneme assembly	44,075,347	44,148,014	1	2	1	0	3	0
LOC116807240	N/A	58,182,748	58,223,849	19	11	24	12	3	1
POLK	DNA repair	60,192,954	60,223,712	1	1	0	0	3	0
LOC105760877	N/A	63,385,678	63,392,309	6	1	1	0	3	3

Table 3.4: Genes containing >3 nonsynonymous fixed differences.

List of zebra finch genes on the Z chromosome with at least three nonsynonymous fixed differences between the A and B inversion haplotypes within their coding regions. Gene function refers to the 'biological process' gene ontology class for homologous genes of the same names in *Mus musculus*. These were used as gene ontology information for *Taeniopygia guttata* was not available from geneontology.org. For the full list of genes containing nonsynonymous fixed differences see Table S3.1, and for positions and further information about the possible effects of the fixed differences see Tables S3.2 & S3.3.

3.5. Discussion

An important first step in investigating the supergene polymorphism on the zebra finch Z chromosome was to determine the breakpoints of the underlying inversion polymorphism. These breakpoints could be approximated using the observed patterns of π nucleotide diversity, SNP density and F_{ST} across the length of the Z chromosome. Within the inversion region, mutations are unable to spread beyond the haplotype they arise in due to recombination suppression (Charlesworth, 2016). Therefore, fewer mutations are able to persist and become polymorphisms within the inversion compared to the rest of the genome. We would expect a pattern of reduced SNP density and π diversity within the inversion compared with outside it. Recombination is known to be suppressed in sub-telomeric regions at the ends of chromosomes and at the centromere (Barton *et al.*, 2008). Consequently, there is low nucleotide diversity seen at the ends of autosomes and at regions in the centre of many autosomes likely corresponding to centromeres (Figure 3.2d). This pattern seen on the autosomes, shows that π diversity is a useful metric for determining areas of reduced recombination. On the Z chromosome, there is a very striking pattern of reduced π diversity and reduced SNP density between 6.50Mb and 70.10Mb (Figures 3.1 & 3.2), suggesting these are the approximate boundaries of the inversion region.

Although new alleles cannot easily recombine onto another inversion haplotype, they can increase in frequency and become fixed within the inversion haplotype they arose in, either due to genetic drift, or because they are under positive selection, or because they are hitchhiking with another linked allele under positive selection. Outside the inversion region this situation is less likely since A and B-associated haplotypes can freely recombine and the effective population size will be larger. This means the number of fixed differences between inversion haplotypes will be far higher within the inversion than outside of it. The inversion region is therefore expected to correspond with a region of high fixation index (F_{ST}) between the A and B haplotypes. The region which fits this expected pattern is also between 6.50Mb and 70.10Mb, whilst the rest of the chromosome shows much lower F_{ST} (Figure 3.3), further suggesting the approximate breakpoints of the inversion region.

As three separate metrics (F_{ST} , SNP density and π diversity) all indicate the same inversion breakpoints for the zebra finch inversion polymorphism (6.50Mb and 70.10Mb on the bTG1.4 reference genome), these positions were used to define the inversion region throughout the remainder of the tests in this chapter. Previous studies used patterns of

heterozygosity of Z chromosome SNPs to predict similarly placed breakpoints at 5.7Mb and between 60.2 and 68.8Mb on the earlier TaeGut3.2.4 reference genome (Kim *et al.*, 2017). Although these predicted breakpoints are different to those indicated here, the disparity is likely due to differences in the assembly of the two reference genomes.

Instead of reduced recombination, an alternative explanation for the low nucleotide diversity in the inversion could be a recent population bottleneck. In this case, polymorphism is lost through genetic drift and insufficient time has passed for new variation to arise through mutation. This explanation is unlikely because it should result in reduced nucleotide diversity across the entire chromosome, not just within the predicted inversion region (Figure 3.1 & 3.2). This demographic explanation can be confidently ruled out because the pattern of reduced diversity is only seen on the Z chromosome, and not across the rest of the genome (Figure 3.2d) as would be expected if a population bottleneck had occurred.

There is a much lower π diversity and density of SNPs within the inversion region of B type birds (Figure 3.1b & 3.2b) than A type birds (Figure 3.1a & 3.2a), suggesting that A is the ancestral haplotype and retains pre-inversion variation, whilst Haplotype B is derived and only has polymorphisms which have arisen after the inversion mutation. This interpretation is consistent with previous work on the zebra finch Z chromosome supergene (Knief *et al.*, 2016; Kim *et al.*, 2017). However, there could be other potential explanations including reduced within-haplotype recombination in B, stronger purifying selection in Haplotype B or a smaller effective population size of Haplotype B birds. Alternatively, the result could be an artifact of the smaller sample size of B birds than A birds in my dataset. There were only 7 B birds, possessing 8 Z chromosomes (1 ZZ male, 6 ZW females) included in the analysis. There were 8 A birds possessing 12 Z chromosomes (4 ZZ males, 4 ZW females), meaning there was less B type sequence and so the full amount of polymorphism within B may not be captured in this analysis i.e. rare alleles will be harder to detect in the B sample. However, when only a subset of A type birds are used, so that the number of Z chromosomes is equal for both haplotypes, the same pattern of reduced diversity in B is seen (Supplementary Figure S3.1a & b), suggesting this result is not a consequence of unequal sample sizes.

The pattern of high fixation (F_{ST}) between A and B haplotypes within the inversion region on the zebra finch Z chromosome is evidence of sequence divergence between the two haplotypes. This pattern of elevated F_{ST} is consistent with divergence caused by chromosomal rearrangements in other species (Berg *et al.*, 2016). When considering all 24 birds Fay and

Wu's H is dramatically lower within the inversion polymorphism than on the rest of the Z, and Zeng's E is higher inside the inversion than outside of it (Figure 3.4f & i). These patterns show an excess of high frequency derived alleles within the inversion region compared with outside the inversion on the Z, which is further evidence of sequence divergence between haplotypes. This pattern of F_{ST} , H and E could also be attributed to demographic effects. However, the patterns are confined to the inversion region on the Z chromosome and are not seen across the rest of the genome (Figures 3.3a & 3.5b-c), meaning demographic effects can be ruled out. The maintenance of this sequence divergence in populations of zebra finches is consistent with balancing selection which is often responsible for maintaining supergene polymorphisms (Berdan, Flatt, *et al.*, 2022; Dagilis *et al.*, 2022; Tafreshi, Otto and Chapuisat, 2022). Previous work has identified heterozygous males as possessing sperm with the highest swimming speed (Kim *et al.*, 2017), which could drive balancing selection in the form of heterozygote advantage. Otherwise, there could be undiscovered phenotypic effects (i.e. on traits other than sperm) which result in overdominance; most likely these phenotypic effects are caused by fixed differences between the haplotypes. There are 203 genes with fixed differences contained within their coding regions (Supplementary Table S3.1), some of which are excellent candidates for having effects on sperm traits. For example, the gene ANKRD31 contains 15 fixed differences within its coding region, 11 of which are nonsynonymous changes (Table 3.4), and has a known role in spermatogenesis in mice (Papanikos *et al.*, 2019; Manfredola *et al.*, 2021). A second strong candidate gene is VPS13A, in which loss-of-function mutations have been shown to cause complete infertility as a result of severely diminished sperm motility in mice (Nagata *et al.*, 2018). These genes could be further investigated in order to determine whether they are responsible for phenotypic variation caused by this supergene. For example, an investigation into the functions of these genes could be performed using gene ontology enrichment analysis (Mi *et al.*, 2019), or a further exploration into the amount of conservative and radical amino acid changes using the CRI index (Sharbrough *et al.*, 2018). An alternative explanation for the observed variation in sperm traits could be variation in adenosine triphosphate production caused by mitochondrial DNA copy number variation which has previously been shown to be linked to this supergene (Knief *et al.*, 2021).

Whilst the difference in sequence observed between the A and B haplotypes is likely to be maintained by balancing selection, this sequence divergence could be driven by either positive selection or by relaxed purifying selection; indeed, the two are not mutually exclusive. Positive selection could be acting on adaptive alleles that have arisen in each haplotype, and

due to the tight linkage across the inversion region, numerous other alleles could be driven to fixation within the haplotype. This genetic hitch-hiking, is a known consequence of supergenes, and it can result in alternate adaptive phenotypes persisting within a single population (Hoffmann, Sgrò and Weeks, 2004; Berdan, Flatt, *et al.*, 2022). Alternatively, purifying selection could be relaxed across the length of the inversion region, resulting in mildly deleterious alleles persisting in the population. As these mutations are prevented from recombining onto the opposite haplotype from which they arose on, each inversion haplotype will accumulate a separate complement of alleles, and so the sequences diverge (Berdan *et al.*, 2021). It is not possible to distinguish between these two explanations using measurements of F_{ST} and nucleotide diversity. However neutrality statistics such as Tajima's D, Fay & Wu's H and Zeng's E can be used to understand the roles of demography and natural selection acting on a genomic region (Achaz, 2009), and so can be used to determine which explanation is most probable.

Tajima's D can be used as an indicator of positive selection (Tajima, 1989; Carlson *et al.*, 2005), where a negative Tajima's D signifies an excess of low frequency polymorphisms which could be due a recent selective sweep. Tajima's D doesn't deviate much from zero in either the A or B haplotype birds across the length of the Z chromosome, either inside or outside of the inversion region (Figure 3.4a-b). If positive selection were acting on either of the two haplotypes to drive sequence divergence, a negative Tajima's D would be expected. Since this is not the case, this indicates there is no broad pattern of positive selection acting on the inversion polymorphism in either the A or B haplotype. There is however a prominent spike of Tajima's D in B-type birds occurring between 58.06Mb and 58.13Mb (Figure 3.4b). This area of positive Tajima's D could indicate a smaller area within the inversion, which is under the influence of balancing selection, but there are no genes of known function within this region, only the predicted gene LOC116807240 which is found nearby (58.18Mb - 58.22Mb).

Fay & Wu's H and Zeng's E are both sensitive to positive selection (Fay and Wu, 2000; Kim and Stephan, 2002; Zeng, Shi and Wu, 2007). In the event of positive selection and a recent selective sweep in one of the haplotypes, an excess of high-frequency derived alleles and a negative Fay and Wu's H would be expected (Fay and Wu, 2000). Zeng's E would also be expected to be negative, since following a selective sweep, once a new variant is established, neutral variation will begin to accumulate. The return to neutral expectations after such an event, however, happens much more rapidly for low frequency variants than high-frequency ones. Therefore, a negative Zeng's E may provide evidence of a recent selective sweep (Zeng,

Shi and Wu, 2007). Since neither of these metrics deviate much from 0 in either the A or B birds (Figure 3.4d, e, g & h) these results further suggest a lack of positive selection.

Given the lack of support for positive selection as the explanation for sequence divergence between inversion haplotypes, an alternate explanation of relaxed purifying selection appears more likely. Under this explanation, reduced recombination leads to a lower effective population size within the inversion region, meaning purifying selection is not able to remove deleterious alleles as effectively as in a larger population (Berdan, Flatt, *et al.*, 2022). As purifying selection is suppressed, mildly deleterious alleles which would otherwise be purged from the population are able to increase in frequency through genetic drift. Since non-synonymous mutations change the structure of proteins, they are usually mildly deleterious and so selected against by purifying selection. We would therefore expect to see relatively more polymorphisms at non-synonymous coding sites within a region of reduced purifying selection. Within both haplotypes, non-synonymous mutations are more prevalent than synonymous mutations (Table 3.3). Outside of the inversion, there are more synonymous than nonsynonymous polymorphisms in birds of both haplotypes. This significant excess of nonsynonymous mutation inside the inversion region is consistent with a reduced efficacy of purifying selection, which allows these nonsynonymous and mildly deleterious mutations to persist. The results suggested that purifying selection is particularly relaxed in Haplotype B, since the ratio of nonsynonymous to synonymous polymorphism within the inversion was greater within the B haplotype than the A haplotype.

An alternative, but less well supported possibility, is that if these nonsynonymous mutations are not in fact mildly deleterious, but instead adaptive, the excess of nonsynonymous polymorphism could still be due to positive selection acting on each haplotype. Positive selection would act to increase the allele frequency of adaptive mutations until they reach fixation within the inversion haplotype they arose in, with many other mutations also reaching fixation through hitch-hiking. In this case, a large amount of non-synonymous fixed differences ($N_{Dnon-syn}$) between the haplotypes would be expected relative to the number of synonymous fixed differences (N_{Dsyn}). The $N_{Dnon-syn} : N_{Dsyn}$ ratio does not reflect this expected pattern. There was no significant difference between Haplotype A polymorphism and fixed differences between A and B birds, but there was a significant excess of non-synonymous Haplotype B polymorphisms compared to fixed differences between A and B birds ($\chi^2 = 8.06$, $p = 0.0045$). This is the opposite to what would be expected if positive selection was driving the divergence between the haplotypes. Therefore, the data suggests positive selection is not the cause of the

excess of non-synonymous polymorphisms within the inversion region. Relaxed purifying selection is the best explanation, since mildly deleterious recessive mutations would be allowed to persist but wouldn't be driven to fixation.

The accumulation of deleterious mutations in this way could have a significant impact on the long-term fate of the zebra finch Z chromosome supergene. Simulation studies have shown the accumulation of deleterious recessive mutations on both inversion haplotypes should lead to overdominance, which maintains the inversion polymorphism, but also the fitness degradation of at least one homokaryotype (Berdan *et al.*, 2021). There is already some evidence for heterozygote advantage with AB males possessing sperm with greater motility than either A or B homozygotes (Kim *et al.*, 2017), heterozygote males showing greater siring success than homozygotes (Knief *et al.*, 2017), and a high value of F_{ST} across the length of the inversion region (Figure 3.3) consistent with balancing selection. More investigation is required to determine if sequence degradation is occurring in either of the inversion haplotypes. However, there is a significant excess of nonsynonymous polymorphisms inside the inversion region within B type birds when compared with polymorphism within A type birds. This could suggest that purifying selection is more strongly reduced, and therefore more deleterious alleles are accumulating, in the B haplotype. In fact, BB homozygote males have especially short and slow-swimming sperm (Knief *et al.*, 2016; Kim *et al.*, 2017), and have the lowest siring success compared to birds with all other inversion karyotypes (Knief *et al.*, 2017).

These results indicate a broad pattern of relaxed purifying selection in both inversion haplotypes which allows mutations to persist, particularly in the B haplotype, resulting in sequence divergence. This sequence divergence is then maintained by balancing selection, possibly through improved fitness of sperm in heterozygote males. To better understand the evolutionary consequences of this broad pattern of selection, it is important to understand the age and demographic history of the supergene. For example, if the supergene is very old, this would further support the idea that balancing selection is maintaining the polymorphism.

3.6. Supplementary Material

Table S3.1: List of zebra finch genes on the Z chromosome with at least one fixed difference between the A and B inversion haplotypes within their coding regions. Genes are ordered according to the number of non-synonymous fixed differences they contain.

Table S3.2: List of nonsynonymous fixed differences between the A and B inversion haplotypes.

Table S3.3: List of synonymous fixed differences between the A and B inversion haplotypes.

Figure S3.1: Nucleotide diversity (π) plots showing the average number of nucleotide differences per site across 100Kb windows with a 10Kb step. **a.** only using SNPs within A/AA karyotype birds (number of individuals = 6, number of Z chromosomes = 8) **b.** only using SNPs within B/BB karyotype birds (number of individuals = 7, number of Z chromosomes = 8) **c.** using SNPs within a subset of birds of a mixture of karyotypes (number of individuals = 6, number of Z chromosomes = 8, karyotypes = A, AA, AB, B, B, C).

Figure S3.2: Fixation index (F_{ST}) between zebra finches of inversion karyotype A or AA (number of individuals = 6, number of Z chromosomes = 8) and of inversion karyotype B or BB (number of individuals = 7, number of Z chromosomes = 8). SNPs were found within all 13 birds when aligning reads against the zebra finch reference genome (bTG1.4). F_{ST} is averaged for all SNP positions in 100Kb sliding windows of sequence across the bTG1.4 Z chromosome, with each window overlapping by 10Kb. Dashed lines indicate predicted breakpoints of the Z chromosome inversion polymorphism.

Figure S3.3: Neutrality statistics calculated using SNPs called using sequence data from 6 karyotype A/AA finches (number of Z chromosomes = 8), 7 B/BB finches (number of Z chromosomes = 8) and 6 finches of mixed karyotypes (number of Z chromosomes = 8, karyotypes = A, AA, AB, B, B, C). Tajima's D calculated using **a.** A/AA finches **b.** B/BB finches **c.** finches of mixed karyotypes. Fay & Wu's H calculated using **d.** A/AA finches **e.** B/BB finches **f.** finches of mixed karyotypes. Zeng's E calculated using **g.** A/AA finches **h.** B/BB finches **i.** finches of mixed karyotypes. Reads were aligned against the reference genome (bTG1.4) to give physical positions. Sequence from a female long-tailed finch was used as an outgroup sequence for calculating H and E. Dashed lines indicate the predicted approximate breakpoints of the Z chromosome inversion region. All statistics were calculated from 100kbp sliding windows across the genome with a 10kbp step between windows.

4. Estimating the age of the zebra finch Z chromosome inversion polymorphism

4.1. Abstract

Phylogenetic inferences of the zebra finch Z chromosome supergene indicate the inversion polymorphism emerged less than 1 million years ago. The ancestral form of this inversion polymorphism is identified as Haplotype A, with two other haplotypes derived from separate inversion mutations. The oldest derived haplotype (C) may have been caused by a relatively small inversion mutation, whilst a second larger mutation may have established Haplotype B. Multiple distinct regions of varying differentiation between inversion haplotypes are identified, which may indicate the locations of each of the mutations that first established the derived haplotypes.

4.2. Introduction

In populations of zebra finches, *Taeniopygia guttata*, a large inversion polymorphism encompassing the majority of the Z chromosome (approximately 60Mbp), was discovered using cytological studies (Itoh et al., 2011) and has subsequently been further characterised by population genetic approaches (Kim et al., 2017; Knief et al., 2017). Previous work on the zebra finch supergene has identified three distinct inversion haplotypes on the Z chromosome (A, B & C), which are responsible for 67–90% of the additive genetic variance in sperm morphology within male finches. AA males are characterised by sperm with long overall length, AB and AC males by intermediate length sperm, and BB, BC and CC males by short length sperm (Kim et al., 2017; Knief et al., 2017). Males which are heterozygous for the inversion polymorphism (AB or AC) have sperm with the fastest swimming speed (Kim *et al.*, 2017; Knief *et al.*, 2017) which may be a source of balancing selection maintaining the supergene polymorphism (see Chapter 3 for further discussion of balancing selection). If the supergene is particularly old, this potential heterozygote advantage would be an explanation for its longevity.

Previous work has suggested that the A inversion haplotype may be the ancestral form of this supergene (Kim *et al.*, 2017), and results outlined in Chapter 3 further support this. However, confirmation of the ancestral haplotype requires further investigation. The aim of this chapter is to use phylogenetic inferences to determine the order of the evolution of the three supergene haplotypes and to estimate the divergence dates between them. Since the approximate boundaries of the inversion polymorphism were estimated in Chapter 3, the sequence within these boundaries can be used to make these phylogenetic inferences. In addition, sub-regions of the inversion polymorphism where recombination is particularly suppressed between pairs of inversion haplotypes may be identifiable. Phylogenetic inferences using just sequence from these sub-regions may produce more accurate predictions of the divergence dates for specific haplotypes (Matschiner *et al.*, 2022).

Supergenes are groups of tightly linked genes with alleles in strong linkage disequilibrium (LD), that are found within regions of reduced recombination, such as an inversion polymorphism (Thompson and Jiggins, 2014). These tightly linked genes can control complex phenotypic variation, which is inherited as if it were a single-gene trait. In addition to sperm traits in zebra finches, other examples of supergenes control mating strategies and plumage morphs in birds such as ruffs, white-throated sparrows and redpoll finches (Huynh,

Maney and Thomas, 2011; Küpper *et al.*, 2015; Lamichhaney *et al.*, 2015; Zinzow-Kramer *et al.*, 2015; Funk *et al.*, 2021), as well as social organisation in fire ants (Wang *et al.*, 2013; Pracana *et al.*, 2017). Supergenes vary in their age; for example the supergene which controls mating strategies in Ruffs was formed approximately 4 million years ago (Lamichhaney *et al.*, 2015), whilst supergenes found in Atlantic cod arose between 0.44 and 1.66 million years ago (Matschiner *et al.*, 2022).

Knowing the age of a supergene is important for understanding its evolutionary context, as a very old supergene implies some form of balancing selection is maintaining the polymorphism. If the supergene is as old as the species it appears in, the supergene could have been a driver of speciation, as supergenes are theorised to facilitate speciation by creating reproductive barriers between carriers of different haplotypes (Rieseberg, 2001).

In this chapter I describe a time-calibrated phylogeny of zebra finch Z chromosome inversion haplotypes, based on sequence in the Z chromosome supergene region, and further phylogenies based on sub-regions of this inversion. These phylogenies provide insight into the origins and age of a very large sex-linked inversion polymorphism that appears to have important effects on male fertility in an important model species, and which has probably been maintained by some form of balancing selection, most likely heterozygous advantage in male birds.

4.3. Methods and Materials

4.3.1. Sequence data

Publicly available (www.ebi.ac.uk/ena/data/view/PRJEB10586) sequence reads for 24 zebra finches (11 female, 13 male) sequenced by Singhal *et al.*, (2015) were used to analyse Z chromosome inversion polymorphism. All birds were sequenced using Illumina HiSeq 2000 paired end sequencing (read length = 100bp) in 2012 and 2013 (Singhal *et al.*, 2015). Of the 24 zebra finches, 19 were wild birds caught at Fowler's Gap, New South Wales, Australia, and 5 domesticated zebra finches were bred in captivity at East Carolina University. The captive group consists of a mother (MP1), father (MP2), and three sons (MP3, MP4 & MP5). In addition, sequence reads from one long-tailed finch, *Poephila acuticauda*, and one double-barred finch, *Stizoptera bichenovii*, which were sequenced in the same study as the zebra finches (Singhal *et al.*, 2015) were used as outgroup sequence. The sequenced long-tailed finch was a wild female bird caught at Mount House, Western Australia, and the sequenced double-barred finch was a male domesticated bird from a private collection. These finches were also sequenced using Illumina HiSeq 2000 paired end sequencing (read length = 100bp) in 2012 and 2013 (Singhal *et al.*, 2015).

Inversion karyotypes for each of the 24 zebra finches were found using a list of SNPs known to be diagnostic of inversion haplotypes (see Chapter 2, Supplementary Tables S2.1, S2.2 & S2.3 for the list of diagnostic SNPs). These diagnostic SNPs were found using 497 male zebra finches from a population previously maintained at the University of Sheffield (Kim *et al.*, 2017), all of known inversion karyotype, and all homokaryotypic. Genotypes of each of the 24 zebra finches at these diagnostic SNPs were used to assign the inversion karyotype of each bird (Chapter 2, Table 2.7 & Figure 2.4). In total 8 birds were scored as inversion karyotype A or AA (4 females, 4 males), 7 birds were scored as inversion karyotype B or BB (6 females, 1 male), 1 female bird was scored as karyotype C, 4 male birds were scored as AB heterozygotes, and another 4 males were scored as AC heterozygotes (Chapter 2, Figure 2.4 & Table 2.7).

In Chapter 3, the sequence reads from these 24 zebra finches and one long-tailed finch were aligned to a male zebra finch reference genome (bTG1.4) using the default settings in Bowtie v2.3.4.3 (Langmead and Salzberg, 2012). Variants were then called between aligned reads and bTG1.4 using the genome analysis toolkit (GATK) version 4.2.5.0 variant calling

pipeline (Poplin *et al.*, 2017). For the analyses in this chapter, the pipeline was repeated with the inclusion of the double-barred finch as a second outgroup species.

4.3.2. Identifying regions within the inversion polymorphism

Approximate breakpoints of the inversion polymorphism were identified in Chapter 3 as 6.50Mb and 70.10Mb (see Chapter 3, Figures 3.1, 3.2 & 3.3). Sequence from between these breakpoints was used to investigate the age of the supergene. However, the fixation index (F_{ST}) across the inversion is variable, with discrete regions of greater or lesser differentiation (Chapter 3, Figure 3.3) being apparent. These regions suggest recombination has not been uniformly suppressed across the inversion. Alternatively, these different regions could have diverged at different times, with greater divergence indicating a longer time to a common ancestor. To investigate possible sub-divisions of the inversion where recombination is likely to have been more suppressed between specific inversion haplotypes, previously described patterns of variation in within-karyotype heterozygosity (Kim *et al.*, 2017) were utilised. In addition, the pattern of F_{ST} between each pairwise combination of the three haplotypes across the inversion region was investigated, as well as the distribution of fixed non-synonymous differences across the inversion.

The following sections outline how these methods were used to determine sub-regions of the inversion where recombination is most likely to have been suppressed between each haplotype, and therefore could be used to measure divergence dates between the haplotypes more accurately.

4.3.3. Regions identified in previous studies

A previous study into the zebra finch Z chromosome inversion polymorphism identified seven regions within the inversion with distinct patterns of heterozygosity within birds of each karyotype (Kim *et al.*, 2017). A companion study found similar patterns (Knief *et al.*, 2017). Patterns of heterozygosity vary across these regions but fall into three broad categories: (i) heterozygosity is roughly the same in all three types of heterokaryotypes; (ii) AB and BC heterokaryotypes show greater heterozygosity than AC types, meaning the B haplotype contains more alternative alleles; (iii) AC and BC heterokaryotypes show greater heterozygosity than AB, meaning the C haplotype contains more alternative alleles (Table 4.1).

The analysis by Kim and colleagues was performed using the zebra finch reference genome released in 2010 (TaeGut3.2.4), which was assembled using Sanger sequence reads from a male finch, “Black17” (Warren *et al.*, 2010). At the time this reference was the only one available. Subsequently an updated reference (bTG1.4) was produced by the Vertebrate Genomes Project (Rhie *et al.*, 2021) using next generation sequencing of DNA from the same male finch. The order of the Z chromosome of the two assemblies has not been formally compared. Therefore, I took sequence (length = 10Kb) from regular intervals of 1Mb across the length of the TaeGut3.2.4 reference Z chromosome and compared it with Z chromosome sequence from the bTG1.4 reference genome. I used NCBI Blastn v2.8.1 (Camacho *et al.*, 2009) to find the equivalent position (blast search result with the lowest e-value) of each 10kbp fragment in the newer reference (Figure 4.1). A change in the order of sequence between the old and new reference assemblies indicated that two regions considered as separate by Kim *et al.*, (regions 2 and 4) were more likely to be adjacent. These two regions have the same broad pattern of heterozygosity (both have high H in AB and in BC birds) and so were treated here as one single region (Figure 4.1).

The seven regions identified by Kim *et al.*, were therefore consolidated into six regions in the updated zebra finch reference (Table 4.2). The number of diagnostic SNPs for each inversion haplotype comparison (A vs B, B vs C and A vs C) varies between these six regions (Table 4.2). Regions 2 and 4 have a higher proportion of AB diagnostic SNPs, whilst 95.59% of diagnostic SNPs in region 3 are B vs C diagnostic, and 71.43% of diagnostic SNPs in region 5 are A vs C diagnostic (Table 4.2). This variability in both heterozygosity and the number of diagnostic SNPs suggests that different regions are responsible for each of the derived inversion haplotypes (B & C). However, establishing exactly which parts of the chromosome are inverted is not easy, especially as the reference genome bird is an AB heterozygote, meaning the assembled chromosome may not fully accurately represent the sequence order of either Haplotype A or Haplotype B.

<u>Region</u>	<u>Position on bTaeGut3.2.4</u>	<u>H_{AB}</u>	<u>H_{AC}</u>	<u>H_{BC}</u>	<u>Haplotype which differs most within region</u>
1	1 - 5,798,614	0.17 (0.19)	0.18 (0.19)	0.19 (0.22)	None
2	5,822,233 - 60,264,951	0.53 (0.45)	0.31 (0.40)	0.64 (0.47)	B
3	60,283,286 - 62,504,794	0.17 (0.18)	0.82 (0.21)	0.97 (0.17)	C
4	62,507,513 - 67,311,640	0.62 (0.44)	0.33 (0.41)	0.62 (0.47)	B
5	67,380,137 - 68,339,423	0.27 (0.30)	0.34 (0.28)	0.3 (0.21)	None
6	68,342,761 - 68,887,401	0.30 (0.31)	0.54 (0.44)	0.46 (0.35)	C
7	68,898,688 - 72,852,975	0.17 (0.17)	0.19 (0.18)	0.17 (0.17)	None

Table 4.1: Regions of the Z chromosome inferred by Kim et al., 2017.

Adapted from Kim et al., 2017. Regions of the Z chromosome inferred by varying levels of mean (s.d.) observed heterozygosity (H). In each region, the haplotype which is most different from the other two is indicated. Regions 1 and 7 are outside of the inversion polymorphism.

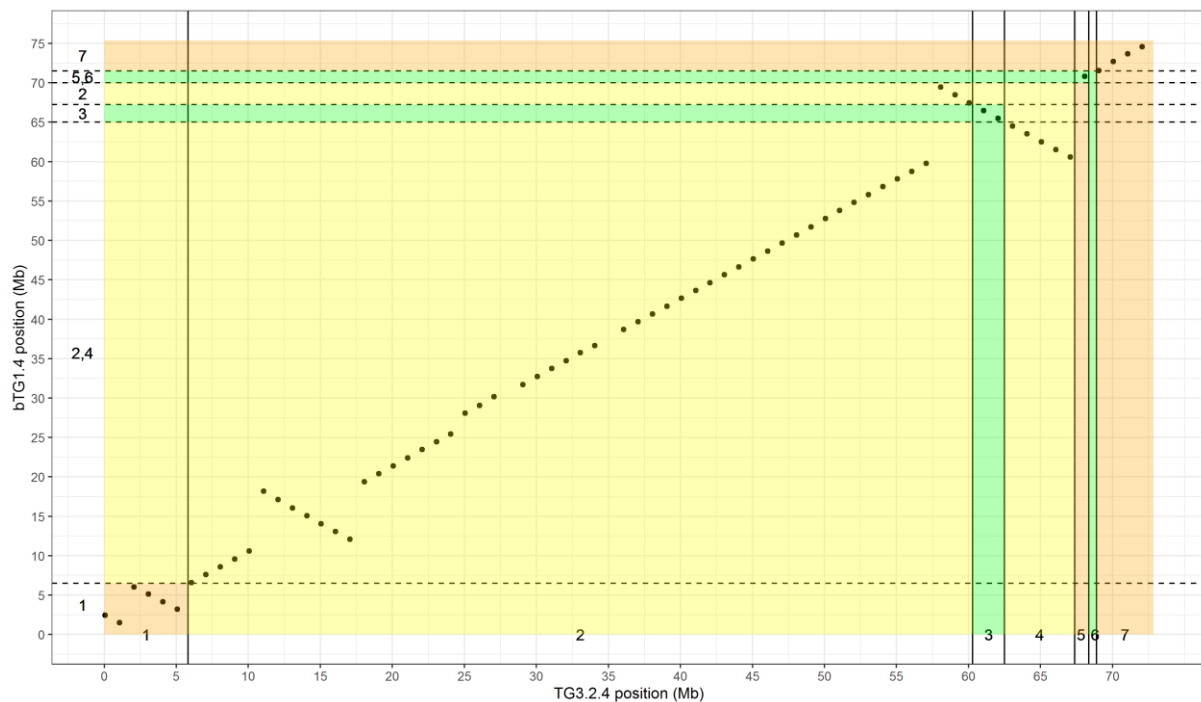


Figure 4.1: Corresponding *bTG1.4* sequence for regions identified by Kim et al., Comparison of physical positions of sequence on the zebra finch Z chromosome in the reference genomes *TaeGut3.2.4* and *bTG1.4*. Sequence (length = 10Kb) taken from regular intervals of 1Mb across the length of the *TaeGut3.2.4* reference Z chromosome was compared with Z chromosome sequence from the *bTG1.4* reference genome using NCBI Blastn v2.8.1 (min word size = 15, min e-value = $1e-10$) to find the equivalent position of each fragment (blast search result with the lowest e-value). Solid vertical lines indicate boundaries of regions of varying heterozygosity found by Kim et al., 2017. Dashed lines indicate equivalent positions of these boundaries in *bTG1.4*. Shading of regions indicates the pattern of heterozygosity found by Kim et al., regions of high H in AB and BC birds are yellow, high H in AC and BC birds are green, and equal H across all heterozygotes are orange.

<u>Position on bTG1.4 (Mb)</u>	<u>Corresponding regions from Kim et al., 2017</u>	<u>Number of Diagnostic SNPs</u>	<u>% AB</u>	<u>% AC</u>	<u>%BC</u>
1.0 - 6.5	1	25	56.0	4.0	40.0
6.5 - 65.0	2, 4	3060	62.9	7.0	30.1
65.0 - 67.3	3	68	4.4	0.0	95.6
67.3 - 70.0	2	87	62.1	0.0	37.9
70.0 - 71.5	5,6	63	27.0	71.4	1.6
71.5 - 75.4	7	0	-	-	-

Table 4.2: Corresponding bTG1.4 sequence for regions identified by Kim et al., Discrete regions of the bTG1.4 Z chromosome and a comparison to the equivalent regions inferred by Kim et al., (on the bTaeGut3.2.4 reference). The number of diagnostic SNPs and the haplotypes they best diagnose varies between these regions.

4.3.4. Regions identified using F_{ST} and using distribution of fixed differences between A and B haplotypes

Using the 24 zebra finches whose inversion karyotypes were identified in Chapter 2 (see Chapter 2, Figure 2.4 & Table 2.7), F_{ST} was calculated between each of the three inversion haplotypes. For the A vs B comparison, all 8 A or AA karyotype finches (4 males, 4 females) and all 7 B or BB karyotype finches (1 male, 6 females) were used. For comparisons involving the C haplotype, since only a single C karyotype female finch was available; it was subsequently used along with 4 AC karyotype males. Average F_{ST} across 200Kb windows with a 100Kb step between windows was calculated along the length of the Z chromosome. All summary statistics were calculated using R v4.2.1 (with Rstudio v2022.07.1+554) and the package “PopGenomeR” v2.7.5 (Pfeifer *et al.*, 2014). F_{ST} values of the three pairwise combinations were compared across the length of the inversion, in order to infer distinct sub-regions, where divergence between one pair of haplotypes was greater than between the other two pairs of haplotypes.

A list of 182 non-synonymous fixed differences between the A and B haplotypes was produced in Chapter 3 (see Chapter 3 Supplementary Table S3.2). In order to further investigate potential sub-regions of the Z chromosome inversion where recombination may be suppressed between the A and B haplotypes, the distribution of these fixed differences was investigated. Any areas with an increased density of non-synonymous changes are likely to have experienced

particularly relaxed purifying selection since those changes are likely to be deleterious and would therefore usually be selected against.

4.3.5. Phylogenetic inference

To construct phylogenetic trees investigating the relative order of divergence of the three inversion haplotypes, as well as the approximate dates of divergence, SNPs called between the 24 zebra finches, 1 long-tailed finch and 1 double-barred finch were used (see Chapter 3 for full methods). Phylogenetic trees were inferred by using the Maximum Likelihood method and Tamura-Nei model (Tamura and Nei, 1993) using the molecular evolutionary genetics analysis (MEGA) software version 11 (Tamura, Stecher and Kumar, 2021). Initial tree(s) for the heuristic search were obtained automatically by applying Neighbour-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Tamura-Nei model, and then selecting the topology with the superior log likelihood value. The reliability of these phylogenetic trees was tested using 100 iterations of bootstrap resampling in MEGA11.

Multiple phylogenetic trees were inferred using this method, using either the entire inversion region, or sub-regions identified through combining F_{ST} , the distribution of non-synonymous fixed differences and data from previous studies. These sub-regions are where recombination is likely to have been most suppressed between specific inversion haplotypes, and so using just these regions for phylogenetic inference may provide more accurate measures of likely divergence dates as these regions are likely to be the oldest divergent strata for each pair of haplotypes. In total 7 phylogenetic trees were inferred.

- (i) Using all SNPs in the entire inversion region (6.5 – 71.5Mb), and all 24 zebra finches
- (ii) Using all SNPs in the entire inversion region, and only female zebra finches and males with homozygous inversion karyotypes (AA or BB as no CC birds were available)
- (iii-vii) Using only SNPs within each identified sub-region of the inversion, and only females and homozygous males.

In each of these phylogenetic trees, the long-tailed finch and the double-barred finch were included as outgroup species. An evolutionary divergence time between the long-tailed finch and the zebra finch of 6.4 million years ago (Hooper and Price, 2015, 2017) was used to calibrate the maximum likelihood tree and produce a timetree (Tamura, Tao and Kumar, 2018). This divergence time was found using the TimeTree 5 database which provides an average of

all published estimates of the divergence time (Kumar *et al.*, 2022). A timetree was inferred using this calibration constraint and by applying the RelTime method with MEGA v11 to estimate divergence times between individuals (Tamura, Tao and Kumar, 2018).

4.3.6. PSMC analysis

To investigate the demographic history of each of the inversion haplotypes, pairwise sequentially Markovian coalescent (PSMC) analysis (Li and Durbin, 2011) was performed. PSMC uses the diploid genome sequence of a single individual to estimate the effective population size (N_e) over a time span of thousands of generations.

It is expected that when the inversion formed, the effective population size would be reduced as the suppressed recombination within the inversion region splits the population into two. The post-inversion effective population sizes of each haplotype, as interpreted by PSMC could then be used to determine if the effect of genetic drift is likely to be the same or different in each haplotype. A low N_e for a given haplotype over a sustained period of time would suggest the sequence of that haplotype would be impacted by the effects of genetic drift. In addition, the derived haplotypes B and C would not have any variation older than the inversion mutation which created them. So, the population size estimates produced by PSMC using a BB or CC homozygote male should only extend backwards in time to the point at which the inversion occurred. Since a BB homozygote male is present in the dataset, PSMC was used to estimate the age of the B haplotype. However no CC male bird was available, so an estimation of the age of the C haplotype was not possible using this method.

PSMC analysis was performed using sequences across the entire Z chromosome inversion region (6.5 – 71.5Mb). Since the method requires diploid sequence, only male birds (n=13) were used, and separate predictions of demographic history of the inversion polymorphism were inferred from each individual. Average read depth across the inversion region was calculated for each individual, and sequence where read depth was less than one third of or more than twice this average was excluded, to minimise the impact of incorrectly called genotypes (i.e. false heterozygote sites) in the data, as recommended for this analysis (Li and Durbin, 2011). Options for the PSMC analysis were set to match those in a previous study of 38 bird species (Nadachowska-Brzyska *et al.*, 2015) which found the parameters (-t5 -r5 -p 4+30*2+4+6+10) to be optimal for all the bird species included. These parameters refer to the upper limit of the time to the most recent common ancestor (-t), the initial θ/ρ value (-r),

the number of atomic time intervals that N_e was inferred across and the number of free interval parameters (both defined by $-p$). In this case the number of atomic time intervals was 84 and the number of free interval parameters was 34, with the first parameter spanning the first 4 atomic time intervals, each of the next 30 parameters spanning 2 intervals, and the 32nd, 33rd and 34th parameters spanning 4, 6 and 10 atomic time intervals respectively ($4+30*2+4+6+10$). The results of the PSMC analysis were then scaled to real time using the per-site per-generation mutation rate (μ), which is estimated as $7e-10$ for zebra finches (Singhal *et al.*, 2015) and the generation interval (g), which was estimated as 2 years.

PSMC determines the history of effective population size change over time, using the distribution of the time since the most recent common ancestor (MRCA) between two alleles in an individual (Li and Durbin, 2011). If the MRCA is recent then the genomic background of both alleles will be similar, but if the MRCA is more distant, more nearby mutations or recombination will usually have occurred (Patterson, 2005), meaning the background around the two focal alleles will be less similar. MRCA can be calculated for a given allele by looking for nearby mutations. Since the Z chromosome inversion occurred and a supergene formed, purifying selection has been reduced (see Chapter 3), enabling more low frequency mutations to accumulate and/or persist longer within the inversion region. These new mutations have very recent MRCAs (i.e. post the inversion event), and because there is an excess of new mutations, the PSMC model can infer a recent, massive population boom, rather than an inversion polymorphism causing relaxed purifying selection and allowing more mutations to persist. PSMC is therefore unsuitable for inferring the effective population sizes of each haplotype over time. However, the timing of the spurious population boom indicates when purifying selection was suppressed within the population i.e. the likely emergence date of that haplotype. Therefore, the use of PSMC can be viewed as a complementary analysis to the phylogenetic analysis described in 4.3.6.

4.4. Results

4.4.1. Regions identified using F_{ST}

By comparing the F_{ST} values of the three pairwise comparisons between inversion haplotypes (A vs B, A vs C/AC & B vs C/AC) across the length of the inversion, distinct sub-regions of the inversion were revealed, where fixation between one pair of haplotypes was greater than between the other two pairs of haplotypes (Figure 4.2). A total of five distinct sub-regions were identified (I, II, III, IV & V; Figure 4.2 & Table 4.3). Two of these sub-regions correspond with regions identified by Kim *et al.*, in 2017; Region IV (Region 3 in Kim *et al.*, 65.00-67.25Mb) and Region V (Region 5/6 from Kim *et al.*, 70.00-71.50Mb; Tables 4.2 & 4.3). Region IV has high AC and BC heterozygosity (Table 4.1), as well as an abundance of B vs C diagnostic SNPs (Table 4.2). Within this same region (65.00-67.25Mb), F_{ST} between B and C/AC birds is greater than F_{ST} between other haplotypes (Region IV in Figure 4.2), suggesting recombination between the B and C haplotypes is particularly suppressed within this region. Similarly, Region V has a high proportion of A vs C diagnostic SNPs (Table 4.2), and F_{ST} between A and C haplotypes is greater within this region than F_{ST} between other haplotypes (Region V in Figure 4.2). It is likely that the actual level of fixation between A and C haplotypes is even greater in this region and across the entire chromosome since the F_{ST} statistic reported here is calculated by grouping together a C-haplotype female and several AC heterokaryotypic males.

Another region of interest is found between 6.50Mb and approximately 17Mb, where B vs C and A vs B F_{ST} is noticeably elevated relative to A vs C F_{ST} (Region I in Figure 4.2). This suggests divergence of the B haplotype from the A and C haplotypes within this region whilst sequence on the A and C haplotypes remains similar. In addition, whilst F_{ST} between A and B birds is greater than F_{ST} for other comparisons across most of the inversion, in some regions this pattern is more noticeable. For example, between 28.25 and 37.55Mb (Region II, Figure 4.2) and between 57.25Mb and 64.50Mb (Region III, Figure 4.2).

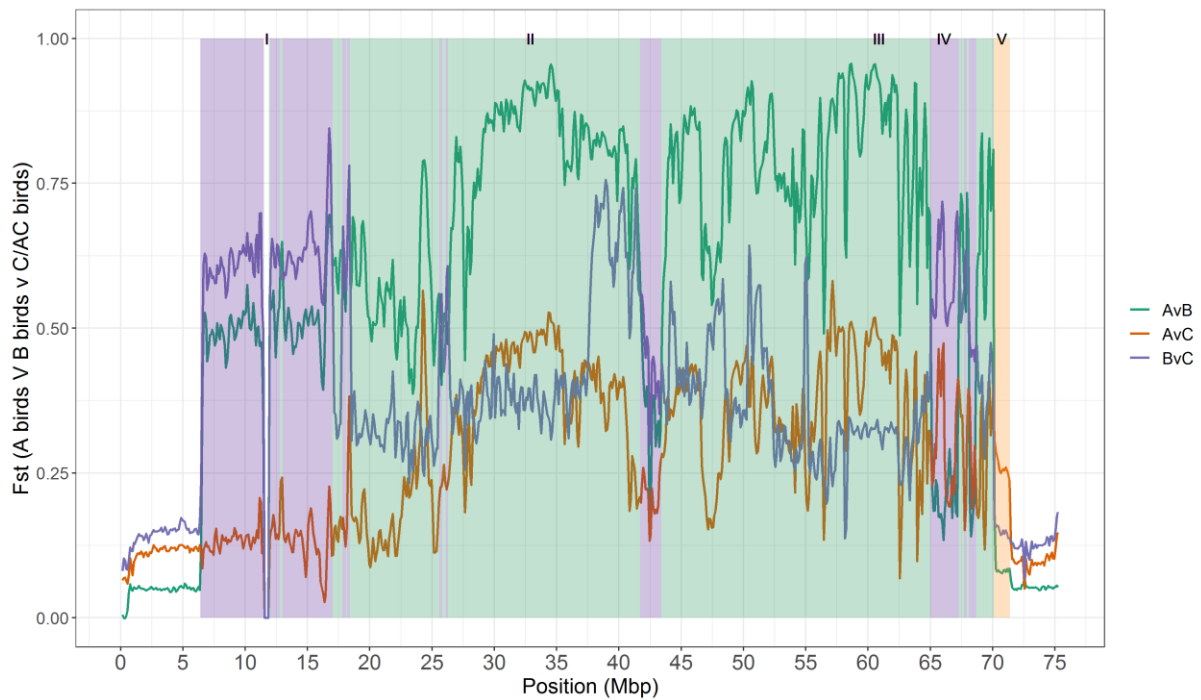


Figure 4.2: F_{ST} and inferred sub-regions of the inversion.

F_{ST} between A/AA ($n=8$) and B/BB ($n=7$) birds, between A/AA and C/AC ($n=5$) birds, and between B/BB and C/AC birds. Average F_{ST} was calculated across 200Kb windows with a 100Kb step between windows, across the length of the Z chromosome. The plot is shaded in the colour of whichever comparison (A vs B, A vs C or B vs C) has the greatest F_{ST} at that position, whenever any F_{ST} score surpasses 0.2.

Region	Position on bTG1.4 (Mb)	Comparison with highest mean F_{ST}	Number of SNPs	Additional evidence
I	6.50- 17.00	B vs C (& A vs B)	225,595	-
II	28.25- 37.55	A vs B	105,403	-
III	57.25- 64.50	A vs B	106,793	Abundance of non-synonymous fixed differences between A and B haplotypes.
IV	65.00- 67.25	B vs C	33,557	Identified in Kim et al., 2017 (Region 3, Table 4.2). Abundance of B vs C diagnostic SNPs.
V	70.00- 71.50	A vs C	54,524	Identified in Kim et al., 2017 (Region 5/6, Table 4.2). Abundance of A vs C diagnostic SNPs.

Table 4.3: Sub-regions of the Z chromosome inversion.

Regions of the bTG1.4 Z chromosome found through comparison of F_{ST} between each of the three haplotypes. Five regions of interest were identified where suppression of recombination between specific haplotypes may be increased (I, II, III, IV & V). Different regions show evidence of particularly suppressed recombination between the A and B haplotypes (regions I,

II and III), between the B and C haplotypes (regions I and IV) and between the A and C haplotypes (region V).

4.4.2. Regions identified using distribution of fixed differences between A and B haplotypes

A list of 182 non-synonymous fixed differences between the A and B haplotypes was produced in Chapter 3 (see Chapter 3 Supplementary Table S3.2). When the distribution of these fixed differences was investigated (Figure 4.3). An abundance of non-synonymous fixed differences was identified in the coding regions of genes between 57.25 and 64.50Mb (Figure 4.3), suggesting recombination is particularly suppressed between the A and B haplotypes within this region. The region corresponds to an area of high F_{ST} between the A and B haplotypes (Region III in Figure 4.2 & Table 4.3).

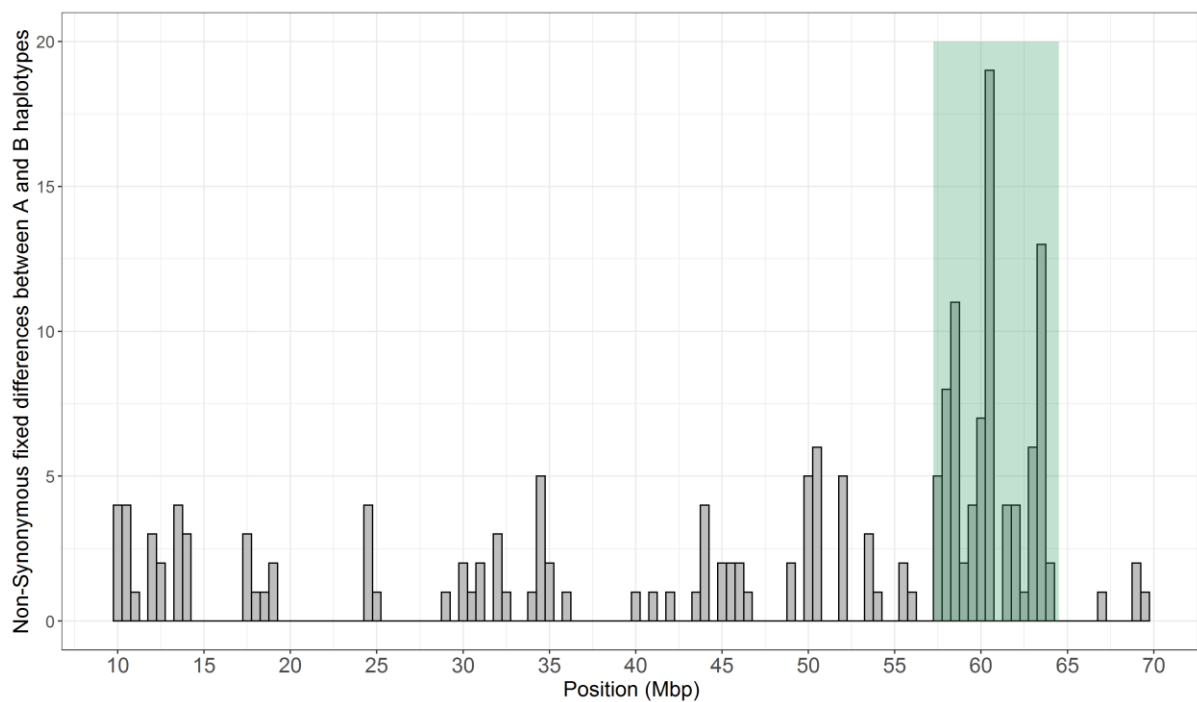


Figure 4.3: A vs B nonsynonymous fixed differences across the Z chromosome. Distribution of non-synonymous fixed differences ($N_{Dnon-syn}$) between birds of the A ($n=8$) and B ($n=7$) inversion haplotypes. $N_{Dnon-syn}$ within 0.5Mb bins across the length of the Z chromosome were counted.

4.4.3. Phylogenetic inference

Multiple phylogenetic trees were inferred, using either the entire inversion region, or sub-regions identified through combining F_{ST} , the distribution of non-synonymous fixed differences and data from previous studies (Table 4.3). In total 7 phylogenetic trees were inferred from: (i) all SNPs in the entire inversion region (6.5 – 71.5Mb), and genotypes at all 24 zebra finches (Figure 4.4); (ii) all SNPs in the entire inversion region, and only female zebra finches and males with homozygous inversion karyotypes (AA or BB as no CC birds were available; Figure 4.5), (iii-vii) only SNPs within each of the five identified sub-regions of the inversion (I - V), and only females and homozygous males (Figure 4.6). When considering the phylogenetic inference made using all SNPs across the entire inversion region and all 24 available birds, the earliest point of divergence between individuals was 0.30 million years ago (95% confidence interval = 0.15 - 0.63 mya) which separates birds into two distinct clades (Figure 4.4b). The first clade contains 8 birds all possessing either A or AA karyotypes, whilst the other clade contains 16 birds all possessing either AC, C, AB, B or BB karyotypes. The birds clustered in this way in 100% of the 100 bootstrap tests which were performed (Figure 4.4a), suggesting this tree topology is reliable. The A haplotype is predicted to be the ancestral haplotype of the inversion polymorphism (Kim *et al.*, 2017), so the oldest node being between A haplotypes and all birds possessing derived haplotypes is expected. Heterozygous individuals are included here to show that birds that are homozygous for the ancestral A haplotype form a distinct clade from all other karyotypes; however, they cannot be used to infer the timings of divergence between haplotypes, and so were not included in any of the subsequent phylogenetic trees.

When heterokaryotypic birds were removed from the analysis, the inferred timings of divergence between Haplotype A birds and birds of the two derived haplotypes is 0.51mya (95% confidence interval = 0.32 - 0.81 mya; Figure 4.5b), and the one Haplotype C bird in the analysis is inferred to have diverged from the Haplotype B birds approximately 0.31mya (95% confidence interval = 0.15 - 0.63 mya; Figure 4.5b). At the nodes relevant to these divergence date predictions, the tree topology was supported across all 100 bootstraps (Figure 4.5a).

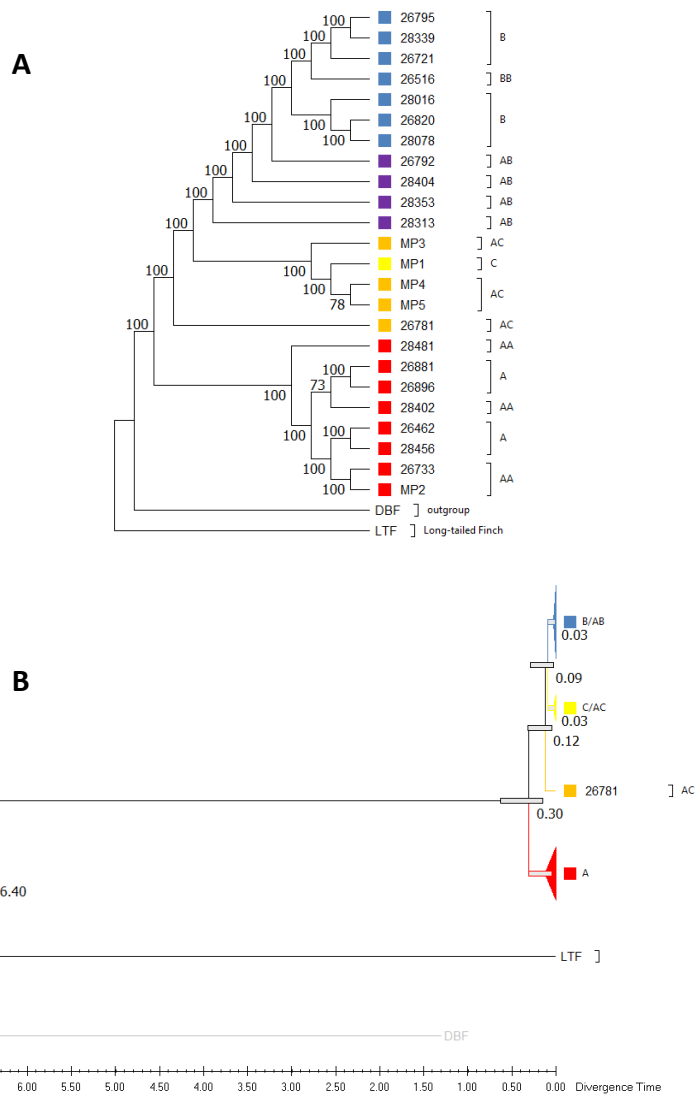


Figure 4.4: Phylogenetic inference using all 24 zebra finches and all SNPs within the inversion region (6.5 – 71.5Mb).

The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model. 2 outgroup species were included. **A.** Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Tamura-Nei model, and then selecting the topology with superior log likelihood value. The tree with the highest log likelihood (-5112552.14) is shown and the percentage of trees (out of 100 bootstrap tests) in which the associated individuals clustered together is shown at each node. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 26 nucleotide sequences and a total of 996,251 positions. **B.** A timetree inferred by applying the RelTime method in MEGA11 to the Maximum Likelihood (ML) phylogenetic tree, which shows divergence times between individuals in real time. The timetree was computed using the divergence time between zebra finches and long-tailed finches (6.4mya) as a calibration constraint. Bars around each node represent 95% confidence intervals.

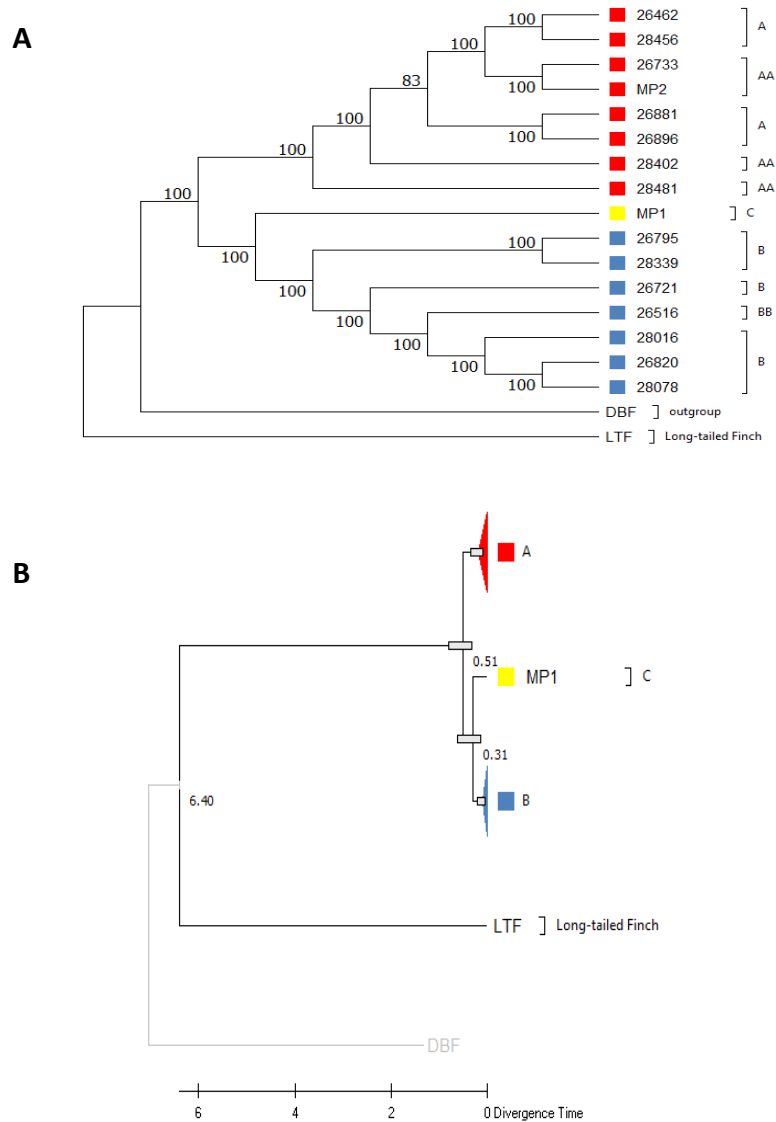


Figure 4.5: Phylogenetic inference excluding heterozygous birds.

*Phylogenetic inference using only female zebra finches, male zebra finches homozygous for the inversion polymorphism ($n=16$) and 2 outgroup species. All SNPs within the inversion region (6.5 - 71.5Mb) were used ($n=996,251$). The evolutionary history was inferred by using the ML method and Tamura-Nei model. **A.** Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Tamura-Nei model, and then selecting the topology with superior log likelihood value (-4922036.96). The percentage of trees (out of 100 bootstrap tests) in which the associated individuals clustered together is shown at each node. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. **B.** A timetree inferred by applying the RelTime method in MEGA11 to the Maximum Likelihood (ML) phylogenetic tree, which shows divergence times between individuals in real time. The timetree was computed using the divergence time between zebra finches and long-tailed finches (6.4mya) as a calibration constraint. Bars around each node represent 95% confidence intervals. Sections of the tree showing divergence between A type individuals and between B type individuals are collapsed.*

When restricting the SNPs to only those within sub-regions of the inversion (see Table 4.3 for sub-regions) the topology of the inferred phylogenetic trees varies somewhat (Figure 4.6, Table 4.4). Trees inferred using regions II (Supplementary Figure S4.2) and III (Supplementary Figure S4.3) suggest a common ancestor of the B and C haplotype birds diverged first, followed by divergence between the B and C haplotypes. Trees inferred using regions I (Supplementary Figure S4.1) and IV (Supplementary Figure 4.4) suggest the B and C haplotypes are both independently derived from Haplotype A in separate divergence events, with the C haplotype emerging first (see Table 4.4 for summary). Assuming the correct topology is inferred when using the entire inversion region (and also when using just regions II & III), these results may suggest the C haplotype arose from an inversion mutation in an ancestral (A haplotype) bird, and the B haplotype is the result of a subsequent mutation event in a Haplotype C bird, relatively soon afterwards.

The estimates for timings of the divergence events between haplotypes change dramatically when only considering SNPs within different sub-regions of the inversion (see Table 4.3 for sub-regions). The inferred time of divergence between the A and C haplotypes was 0.51mya when using region I (95% confidence range = 0.24 - 0.78, Supplementary Figure S4.1), 0.26mya when using region II (95% confidence range = 0.17 - 0.38, Supplementary Figure S4.2), 0.69mya when using region III (95% confidence range = 0.47 - 1.02, Supplementary Figure S4.3), 0.71mya when using region IV (95% confidence range = 0.38 - 1.33, Supplementary Figure S4.4) and 3.44mya when using region V (95% confidence range = 2.25 - 5.26, Supplementary Figure S4.5). Note the confidence intervals for these dates do not all overlap. Whilst the predicted divergence timings from regions I – IV are mostly consistent with those inferred using the whole inversion region, the predicted divergence timing of 3.44mya when using region V is an obvious outlier. This is likely due to the very small length of this region (Region V = 1.5Mb) compared with the inversion polymorphism as a whole (65Mb). In addition to being the smallest sub-regions, regions IV and V also have fewer SNPs within them than the other 3 sub-regions (Table 4.3), which may limit the accuracy of predictions of dates of divergence using these two sub-regions. Alternatively, region V could be an older C-defining inversion which arose before the rest of the inversion region, which would explain the abundance of A vs C diagnostic SNPs found there (Table 4.2) and the relatively high A vs C F_{ST} within this region (Figure 4.2). A third explanation could be selection acting between haplotypes on this region.

The timing of the emergence of the B haplotype (either derived from Haplotype A or C) is more recent, with some inferred timings (using regions II, III and IV) less than 0.1 million years ago (Supplementary Figures S4.1-4 & Table 4.4). When using only SNPs in region V, the A and B haplotype birds do not cluster into separate nodes, suggesting the inversion which defines the B haplotype does not cover this region of the chromosome (Supplementary Figure 4.5).

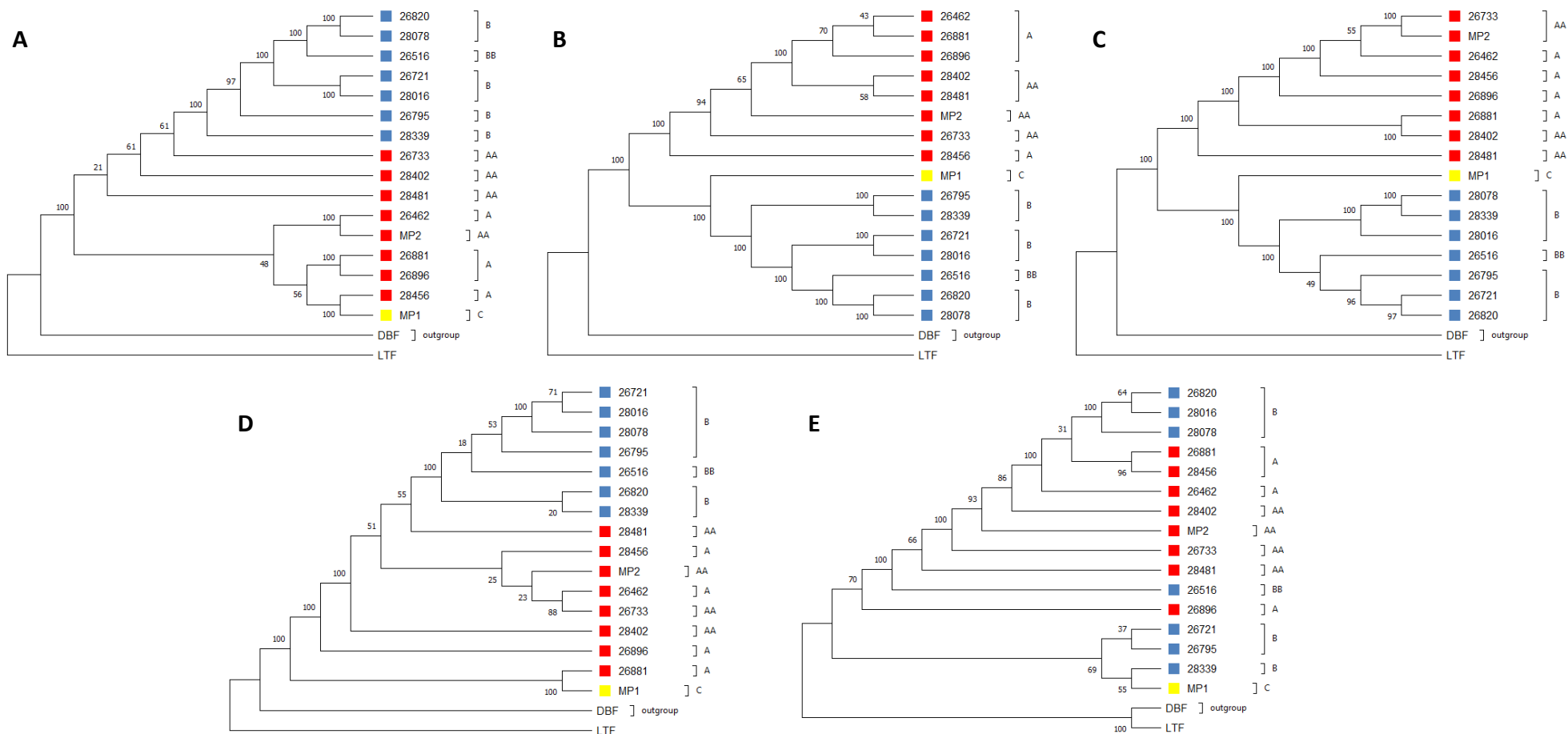


Figure 4.6: Phylogenetic inference using SNPs within sub-regions of the inversion region.

Phylogenetic inference using only female zebra finches, male zebra finches homozygous for the inversion polymorphism ($n=16$) and 2 outgroup species. The percentage support for nodes (from 100 bootstrap tests) is shown at each node. Maximum Likelihood trees were constructed using all SNPs within sub-regions of the inversion polymorphism. **A.** Region I (6.50 - 17.00Mb, 225,595 SNPs, log likelihood = -1056779.76). **B.** Region II (28.25 - 37.55Mb, 105,403 SNPs, log likelihood = -459785.41). **C.** Region III (57.25 - 64.50Mb, 106,793 SNPs, log likelihood = -488352.56). **D.** Region IV (65.00 - 67.25Mb, 33,557 SNPs, log likelihood = -170380.48). **E.** Region V (70.00 - 71.50Mb, 54,525 SNPs, log likelihood = -310170.29).

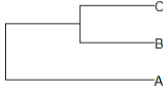
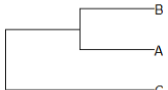
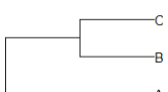
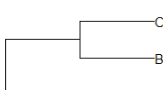
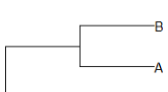
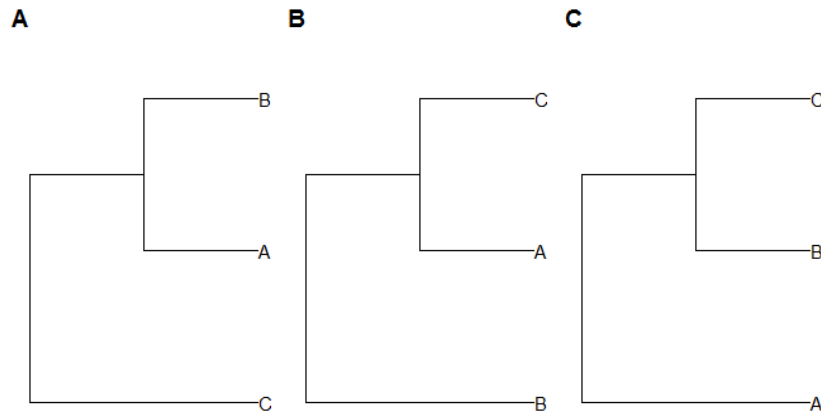
Region Used	Inferred timing of divergence between...			Topology
	A & B	A & C	B & C	
Whole Inversion (6.50 - 71.50Mb, Figure 4.5b)	0.51 (0.32 - 0.81)	0.51 (0.32 - 0.81)	0.31 (0.15 - 0.63)	
Region I (6.50 - 17.00Mb, Supplementary Figure S4.1)	0.41 (0.18 - 0.78)	0.51 (0.24 - 0.78)	0.58 (0.44 - 0.78)	
Region II (28.25 - 37.55Mb, Supplementary Figure S4.2)	0.26 (0.17 - 0.38)	0.26 (0.17 - 0.38)	0.07 (0.04 - 0.11)	
Region III (57.25 - 64.50Mb, Supplementary Figure S4.3)	0.69 (0.47 - 1.02)	0.69 (0.47 - 1.02)	0.05 (0.03 - 0.08)	
Region IV (65.00 - 67.25Mb, Supplementary Figure S4.4)	0.08 (0.04 - 0.18)	0.71 (0.38 - 1.33)	1.00 (0.72 - 1.37)	
Region V (70.00 - 71.50Mb, Supplementary Figure S4.5)	-	3.44 (2.25 - 5.26)	3.16 (1.21 - 5.26)	-

Table 4.4: Summary of phylogenetic inferences.

Phylogenetic inferences were made using different sub-regions of the zebra finch inversion polymorphism. Times of divergence between each of the three inversion haplotypes are shown, with 95% confidence intervals. The tree topologies inferred using each region are shown. When using region V, the Haplotype A and B birds did not form distinct clusters making inference of the order in which the haplotypes arose impossible.

Assuming that Haplotype A is ancestral and Haplotypes B and C are derived, there are just three possible tree topologies that could explain the relationship between the three haplotypes. **(A)** C is first derived from A, then B is later derived from A. **(B)** B is first derived from A, then C is later derived from A. **(C)** either B is derived from A then C is derived from B, or C is derived from A then B is derived from C.



The tree topology inferred using all SNPs from the entire inversion region is topology C, with 100% bootstrap support (Figures 4.4a & 4.5a). The same tree topology is inferred when considering only SNPs from either regions II or III, also with 100% bootstrap support (Figures 4.6b-c). However, an alternative topology (topology A) is inferred when considering only SNPs from within regions I or IV, but with less than 100% bootstrap support (Figures 4.6a & d). The most likely tree topology, is therefore topology C. However, it is also possible that there is no phylogenetic topology which correctly describes the whole supergene region. Multiple inversions or recombination between the inversions would mean that the sub-regions identified here could each have independent phylogenetic topologies.

4.4.4. PSMC analysis

Population size estimates inferred from each of the 13 male birds using PSMC reach maximums of $\sim 1 \times 10^{11}$ (Figure 4.7), which is an impossibly high figure - it's highly unlikely zebra finch populations have ever been this size and certainly not this effective population size. This implausible population boom is likely an artefact due to the relaxation of purifying selection (see Chapter 3), which enables low frequency mutations to accumulate and/or persist longer within the inversion region. As there is an excess of new mutations (arising more recently than the inversion event), the PSMC model infers a recent, massive population increase, rather than the inversion polymorphism causing relaxed purifying selection and allowing more mutations to persist.

The timing of this spurious population boom therefore indicates when purifying selection was suppressed within the population. Whilst the timing varies between birds, the rapid increase in predicted population size always occurs in the last 0.15 million years, suggesting that the purifying selection has only been suppressed on the Z chromosome relatively recently in the history of the zebra finch as a species.

The derived haplotypes B and C would not have any variation older than the inversion mutation which created them. So, the population estimates produced by PSMC using a BB or CC homozygote male should only extend backwards in time to the point at which the inversion occurred. Since a BB homozygote male is present in the dataset, PSMC was used to estimate the age of the B haplotype as approximately 0.25 million years ago (Figure 4.7). No CC male bird was available. These estimates are slightly more recent than those from the phylogenetic analysis of the inversion region, where B and C were estimated to have emerged 0.15 - 0.63 mya and 0.32 - 0.81 mya respectively (Table 4.4).

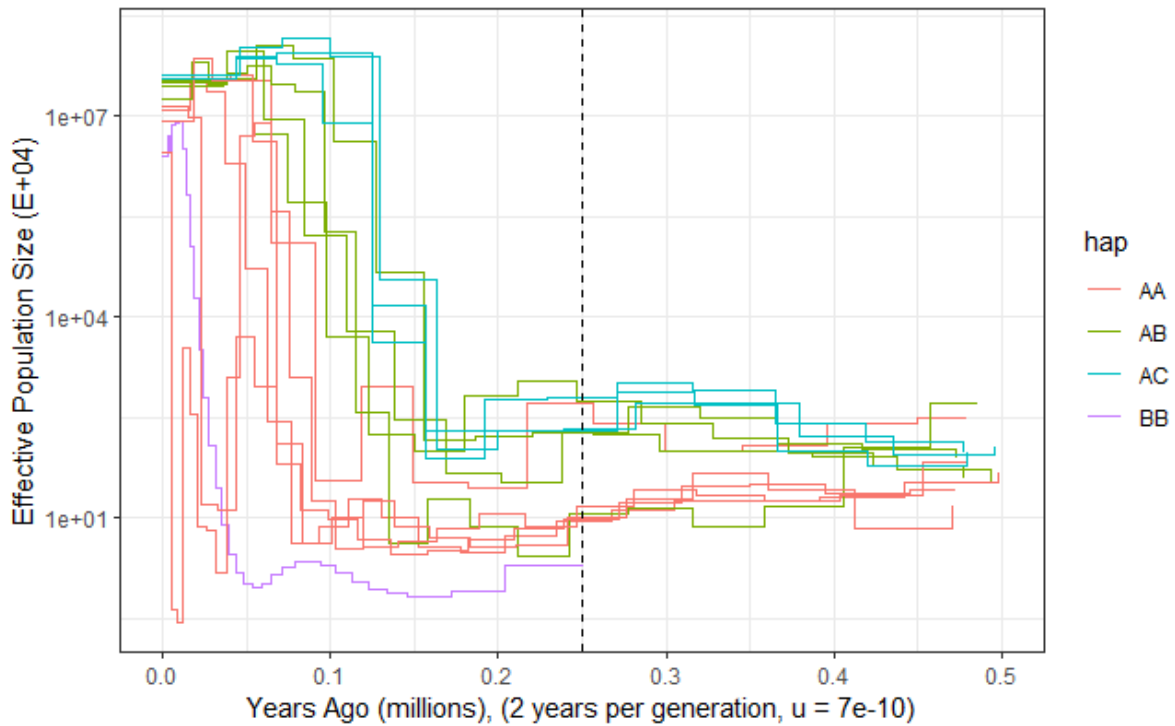


Figure 4.7: N_e of inversion haplotypes inferred by PSMC.

Effective population size of different inversion karyotypes over time as inferred by PSMC analysis using male zebra finches ($n=13$; $n_{AA} = 4$, $n_{AB} = 4$, $n_{AC} = 4$, $n_{BB} = 1$). Reduced purifying selection within the inversion region allows many low frequency mutations to accumulate, which PSMC analysis interprets as an increase in population size. The timing of the sudden population increase indicates when purifying selection was first suppressed. In addition, population estimates for karyotypes consisting of only derived alleles (e.g. BB) will only extend back to the point at which the haplotype first arises (indicated by dashed line).

4.5. Discussion

Consistent with previous studies (Kim *et al.*, 2017; Knief *et al.*, 2017), and with findings from Chapter 3, the phylogenetic inferences made here suggest inversion Haplotype A is the ancestral form of the inversion polymorphism. The phylogenetic inference using the whole inversion region suggests a divergence event between 0.32 and 0.81 million years ago (Figure 4.5b) between the A haplotype and B/C haplotypes. In addition, the divergence times between different Haplotype A individuals are generally older than the divergence times between Haplotype B individuals (Figure 4.4b & 4.5b). Furthermore, when restricting the analysis to SNPs within specific sub-regions of the inversion, divergence between A haplotype individuals is often older than divergence between B haplotype individuals (Supplementary Figures S4.1, S4.3 & S4.4).

Additional evidence for Haplotype A being ancestral was found through PSMC analysis (Figure 4.7), where estimates of effective populations size produced using the BB homozygote male birds do not extend back further than 0.25 million years ago. This suggests the B haplotype is approximately 0.25 million years old since the BB homozygote contains no heterozygous positions predicted to be older than this. Some caution here is required as PSMC analysis is not designed for this purpose (Liu and Hansen, 2017), and the estimated age of these haplotypes is sensitive to user-supplied mutation rates and generation times. By contrast, when PSMC was performed on male birds possessing at least one A haplotype, estimates of effective population size extended much further into the past (Figure 4.7), suggesting the A haplotype contains older mutations and is therefore ancestral.

PSMC analysis was unsuitable for investigating the demographic history of each of the inversion haplotypes, since purifying selection has been relaxed (see Chapter 3) by the Z chromosome inversion, enabling more low frequency mutations to accumulate and/or persist longer within the inversion region. This excess of new mutations, is inferred by the PSMC model as a spurious massive population boom, rather than an inversion polymorphism causing relaxed purifying selection and allowing more mutations to persist. Although PSMC analysis was not successful in producing estimates of effective population size after the inversion mutation, the timing of unlikely rapid increases in reported population sizes may give some insight into when purifying selection was relaxed. In all 13 male birds, PSMC analysis shows the population rapidly rising in the last 0.13 million years, more recently than the emergence of the B haplotype according to the same analysis. This may be evidence of a delay between

the original occurrence of the inversion mutations which produce the derived haplotypes and the haplotype increasing in frequency enough for purifying selection to be suppressed.

Phylogenetic inference suggests that the C haplotype is derived from the A haplotype and emerged 0.15 - 0.63 mya, whilst Haplotype B is derived from Haplotype C and emerged 0.32 - 0.81 mya (Table 4.4). Unfortunately, since only one bird of Haplotype C was available, it is not possible to compare the divergence between C haplotypes with the divergence between B haplotypes. However, the divergence between B haplotypes is always remarkably recent (<0.05mya, Supplementary Figures S4.1-4.5), which is consistent with the B haplotype being the most recently derived. If more birds possessing C haplotypes were available, and the divergence times between them were further in the past than times between B birds, it would confirm the C haplotype is older than B.

It is probable that at least two separate inversion mutations are responsible for creating the three inversion haplotypes, with the first establishing the C haplotype, and the second causing the B haplotype to arise. The sub-regions identified here (Table 4.3), may indicate where these smaller (possibly overlapping) inversion mutations occurred within the larger supergene region. Although, since the reference of the zebra finch Z chromosome is not correctly assembled (see Chapter 2) the order of these sub-regions may not be correct.

Two sub-regions appear to define the C haplotype in particular (regions IV and V), as both have an abundance of SNPs which are diagnostic of the C haplotype (Table 4.2). Phylogenetic trees inferred using these sub-regions indicate much older divergence timings for the C haplotype when compared with trees produced using other sub-regions (Supplementary Figures S4.1-4.5). These regions could be the location of the C-defining inversion, and as they are only separated by ~3Mb of sequence, they may in fact be adjacent but misassembled. Region V is particularly noteworthy as it sits outside of the inversion region identified in Chapter 3, where only the A and B inversion haplotypes were compared, and the boundaries of the inversion were found to be 6.50Mb and 70.10Mb based on F_{ST} (A vs B) and nucleotide diversity (Chapter 3, Figures 3.2 & 3.3). When a phylogenetic tree was constructed using just SNPs in region V, the A and B haplotype birds did not cluster separately, suggesting the B haplotype is not defined by this region. A possible explanation could be that regions IV & V collectively are the locations of the C-defining inversion polymorphism, which overlaps with the B-defining inversion mutation in region IV, but not in region V. Both regions IV and V are small (2.25Mb and 1.50Mb) compared with the inversion region as a whole (~65Mb)

suggesting the older C-defining inversion mutation is much smaller than the later B-defining inversion.

Some sex chromosomes contain regions known as evolutionary strata, where recombination stopped at distinct times (Charlesworth, Charlesworth and Marais, 2005; Bachtrog, 2013), suggesting sex chromosome evolution occurs in multiple successive steps rather than all at once. Evolutionary strata have been attributed to the accumulation of multiple successive chromosomal inversions, where each inversion locally suppresses recombination between proto-sex chromosomes (Bachtrog, 2013). Each inversion allows the proto-sex chromosomes to diverge in sequence within the inversion, with the oldest inversions having accumulated the most mutations, which results in evolutionary strata. An alternative explanation could be that distinct regions of an inversion could result in varied levels of recombination suppression, with regions of greater recombination suppression resulting in more sequence divergence relative to regions of lesser recombination suppression. Evolutionary strata have also been investigated in supergenes, for example in the genus of fungi, *Microbotryum*, where multiple independent strata were discovered in supergenes in several species (Branco *et al.*, 2018). The sub-regions I have identified in the zebra finch Z chromosome supergene, could be evidence of multiple successive inversions. However, the variation in F_{ST} is likely not large enough, or step-like enough to constitute obvious strata (Figure 4.2). This is similar to findings in the fire ant supergene, where evidence of multiple inversions are suggested but there is no evidence of evolutionary strata (Huang *et al.*, 2018). One of the suggested explanations for the lack of strata in fire ants is that the multiple inversions which comprise the supergene happened in such quick succession that any strata would not be detectable (Huang *et al.*, 2018). This explanation is supported by some predictions of the age of the fire ant inversion, which estimate that it arose approximately 390,000 years ago (Wang *et al.*, 2013). According to phylogenetic inferences here, the derived haplotypes of the zebra finch supergene are likely to have occurred within the last ~0.5 million years (Figure 4.5), with the B haplotype in particular likely to have arisen ~0.3 million years ago (Figure 4.5). These young age estimates would suggest that if multiple inversions contributed to the zebra finch supergene, they must have occurred recently and in quick succession, meaning any strata are less likely to be visible.

Further analysis of the phylogenetic relationship between these inversion haplotypes is required, particularly analysis involving multiple C-haplotype individuals. However, these results do give a rough estimate for the age of the zebra finch inversion polymorphism (Figure

4.4 & 4.5). These estimates suggest the Zebra finch inversion is much younger than other well-known examples of supergenes such as the ruff, which is approximately 4 million years old (Lamichhaney *et al.*, 2015). This could explain why there is limited phenotypic difference between individuals of different karyotypes in zebra finches when compared with other examples of supergenes. In zebra finches, the phenotypic differences caused by the supergene which have been reported are mostly limited to sperm traits in males (Kim *et al.*, 2017; Knief *et al.*, 2017). This is a relatively minor phenotypic change compared with the dramatic differences in plumage and behaviour seen in ruffs (Küpper *et al.*, 2015; Lamichhaney *et al.*, 2015). It is possible that in relatively young supergenes such as in the zebra finch, there has been insufficient time for more complex phenotypic variation to evolve. Another example of dramatic phenotype variation caused by a supergene is found in fire ants, where a supergene-containing social chromosome controls social organisation of colonies of ants (Wang *et al.*, 2013; Purcell *et al.*, 2014; Pracana *et al.*, 2017). Although the phenotypic effect of the supergene in fire ants is more dramatic than in zebra finches, some estimates of the ages of the fire ant supergene (0.39 mya; Wang *et al.*, 2013) are very similar to the estimated age of the zebra finch supergene presented here (Figure 4.5). However, another estimate of the age of the fire ant supergene suggests it is much older, at ~1.1 million years ago (Cohen and Privman, 2020). In addition, the fire ant supergene has greatly reduced recombination as the homozygous form of one of the inversion haplotypes is lethal, and this haplotype can therefore only exist in heterozygotes (Wang *et al.*, 2013). This reduced recombination has been shown to result in reduced efficacy of purifying selection (Pracana *et al.*, 2017). If purifying selection were more relaxed in the fire ant supergene than in the zebra finch supergene, a greater amount of sequence divergence may accumulate between haplotypes during the same timeframe. In summary, the complexity of the phenotypic effect of a supergene is likely to be a product of both the age of the supergene, and the amount of recombination suppression.

4.6. Supplementary Material

Table S4.1: A timetree inferred by applying the RelTime method in MEGA11 to the ML phylogenetic tree produced using all SNPs in Region I (shown in Figure 4.6a)

Table S4.2: A timetree inferred by applying the RelTime method in MEGA11 to the ML phylogenetic tree produced using all SNPs in Region II (shown in Figure 4.6b)

Table S4.3: A timetree inferred by applying the RelTime method in MEGA11 to the ML phylogenetic tree produced using all SNPs in Region III (shown in Figure 4.6c)

Table S4.4: A timetree inferred by applying the RelTime method in MEGA11 to the ML phylogenetic tree produced using all SNPs in Region IV (shown in Figure 4.6d)

Table S4.5: A timetree inferred by applying the RelTime method in MEGA11 to the ML phylogenetic tree produced using all SNPs in Region V (shown in Figure 4.6e)

5. General discussion

5.1. Key Findings

1. I produced a list of 2330 diagnostic SNPs, which can be used to determine the Z chromosome inversion polymorphism karyotype of individual zebra finches, based on their Z chromosome DNA sequence. This list is a tool that will be useful for future research involving this supergene.
2. I used my list of highly diagnostic SNPs to determine that the bird used to create the zebra finch reference genome (bTG1) was an AB heterozygote for the inversion polymorphism, and that the female bird used to create an alternate reference genome (bTG2) is karyotype A. I then determined that the reference genome (bTG1) is not correctly phased on the Z chromosome, and that it contains a mixture of both Haplotype A and Haplotype B sequences. However, the reference is correctly phased at the scaffold level. This finding may be useful for future research and to guide future phasing of the zebra reference genome Z chromosome.
3. I also used my list of highly diagnostic SNPs to determine the inversion karyotypes of 24 zebra finches whose genomes were sequenced as part of a previous study (Singhal *et al.*, 2015). Variants were then called within these 24 zebra finch sequences and the resulting variant calls were used to estimate nucleotide diversity across the Z chromosome. This estimate of nucleotide diversity was used to predict the boundaries of the Z inversion polymorphism region as 6.50Mb and 70.1Mb. This is the first estimate of the Z chromosome inversion region boundaries on the updated reference genome (bTG1).
4. Estimates of fixation index and neutrality statistics were used to show that there is considerable sequence divergence between the A and B haplotypes within these inversion region boundaries. I show evidence of balancing selection maintaining this sequence divergence between haplotypes.
5. Based on neutrality statistics and the level of nonsynonymous and synonymous polymorphism/fixed differences between and within haplotypes A and B, I conclude that purifying selection has been relaxed in both A and B haplotypes (but particularly in Haplotype B). This relaxed purifying selection allows more mutations to persist within haplotypes and increase in frequency through genetic drift, which drives sequence divergence between the haplotypes.

6. I produced a list of fixed differences between the A and B haplotypes that are found within coding regions. These fixed differences, especially those which are non-synonymous, are likely to cause some of the observed phenotypic variation in sperm morphology seen in zebra finch males, as well as any other unreported phenotypic variation. This list of fixed differences and the genes they are within could be used to further investigate the phenotypic impact of this supergene.
7. Phylogenetic inferences using SNPs within the inversion region indicate the inversion polymorphism emerged between 0.32 and 0.81 million years ago. Haplotype A is the ancestral form of the polymorphism, Haplotype C was the first to diverge, and Haplotype B was later derived from Haplotype C. Haplotype A was further confirmed as the ancestral haplotype as it has greater nucleotide diversity within the inversion region than Haplotype B does. Much of this increased diversity likely pre-dates the inversion mutation.
8. Five sub-regions within the inversion region are identified, which may be the locations of multiple distinct inversions which have contributed to the supergene region. These sub-regions may include a smaller inversion mutation which established Haplotype C, a much larger mutation which later established Haplotype B, and regions where these two inversions overlap. Alternatively, these sub-regions could be caused by an inversion which established Haplotype B occurring within an earlier, large inversion which established Haplotype C.

5.2. Divergence between supergene haplotypes

Supergene haplotypes contain a tightly linked set of loci which affect a complex phenotype, with sequence divergence between haplotypes resulting in potentially dramatic phenotypic variation such as mating strategies in ruffs (Küpper *et al.*, 2015; Lamichhaney *et al.*, 2015) and social organisation in fire ants (Wang *et al.*, 2013; Pracana *et al.*, 2017). There are two main theories for why this sequence divergence between haplotypes can accumulate; (i) positive selection acting on separate adaptive traits (Hoffmann, Sgrò and Weeks, 2004; Charlesworth, 2016; Berdan, Flatt, *et al.*, 2022), (ii) a relaxation of purifying selection which allows mildly deleterious mutations to accumulate on each haplotype (Berdan *et al.*, 2021).

My study does not find evidence of positive selection in either inversion haplotype. Instead, evidence is presented of relaxed purifying selection across the inversion region which allows mildly deleterious mutations to persist where they would otherwise be purged; a similar

process has been reported in fire ants (Pracana *et al.*, 2017). This accumulation of mutation in each haplotype is a phenomenon that has been previously described (Berdan *et al.*, 2021; Berdan, Flatt, *et al.*, 2022) and modelled (Berdan, Blanckaert, *et al.*, 2022). These models predict that the accumulation of deleterious mutations would reduce fitness of one of the haplotypes (Berdan *et al.*, 2021; Berdan, Blanckaert, *et al.*, 2022), potentially disrupting the balance of the supergene polymorphism. It is also theorised that the accumulation of mutations in each haplotype would lead to heterozygote advantage, since the effects of a deleterious recessive alleles would be mitigated if only present in one chromosome (Tafreshi, Otto and Chapuisat, 2022). This heterozygote advantage could then help to maintain the supergene polymorphism through balancing selection (Dagilis *et al.*, 2022; Tafreshi, Otto and Chapuisat, 2022).

In zebra finches, there is evidence suggesting that heterozygote advantage due to increased sperm motility of heterozygotes (AB and AC) may maintain the supergene polymorphism within the zebra finch population (Kim *et al.*, 2017; Knief *et al.*, 2017). The midpiece to tail size ratio is an important factor in explaining variation in sperm motility (Bennison *et al.*, 2016), with an intermediate ratio producing sperm with the fastest swimming speed or motility (in this case, males with haplotype AB and AC). The findings in this thesis are consistent with relaxed purifying selection allowing mildly deleterious recessive mutations (possibly affecting sperm morphology) to accumulate on each haplotype. Consequently, individuals which are heterozygous at the supergene polymorphism are at an advantage as they will only possess one copy of each deleterious allele.

5.3. Limitations and future directions

In Chapter 2, I showed the zebra finch reference genome is correctly phased at the scaffold-level, but these phased scaffolds are erroneously assembled into an incorrect whole chromosome haplotype. Correction of this phasing in the reference genome could be useful for future research into the Z chromosome inversion polymorphism as it could provide an entire assembled Z chromosome of both the A and the B haplotype (since the bird used to create the reference is an AB heterozygote). Therefore, a future project building on this work could involve correction of the phasing for the Z chromosome of the reference genome. Attempts at correcting this phasing could be checked by repeating the methods outlined in Chapter 2 on any new iterations of the reference genome.

In Chapter 3, only approximate positions of the inversion region were found, rather than exact inversion breakpoints. This means some sequence may be erroneously included/excluded from this analysis which may alter certain results. For example, the number of polymorphic sites inside the inversion region which is presented in Chapter 3 could vary noticeably if the inversion region boundaries were moved, since the level of polymorphism is much greater outside the assumed inversion boundaries. In order to find the exact inversion breakpoints, it is important for the reference genome to be assembled as correctly as possible, which is one reason to correct the incorrect phasing highlighted in Chapter 2. I attempted to find more accurate estimates of the inversion breakpoints using “split-read analysis” where sequence reads were aligned against the reference genome to look for “split” reads which straddle the inversion breakpoint and so partially aligned to two separate locations on the chromosome. For this approach to work, sequence reads would need to be long enough to straddle the inversion breakpoint so that enough sequence for a reliable alignment is located either side of the breakpoint. Genomes of two female zebra finches of inversion karyotype A and B respectively, were sequenced using the 10X Chromium Linked-Reads library preparation system (Ott *et al.*, 2018) and three runs on Illumina HiSeq4000 and HiSeq2500 at the Centre for Genomic Research, University of Liverpool. The Linked Reads system uses molecular barcoding of DNA to generate pseudo-long read sequence data from Illumina short read sequences. The 10X Genomics de novo assembly software, Supernova v2.1.1 (Weisenfeld *et al.*, 2017), was used to assemble each sequenced genome to produce an assembly of both the A and B haplotypes. Sequence reads were then aligned to the assembled Z chromosomes in an attempt to find split reads which would indicate the locations of the inversion breakpoints.

However, this approach was unsuccessful, possibly because the pseudo-long reads were unsuitable for this analysis, and true long read sequence reads may be required instead. This analysis may have formed part of this research project were it not for the impact of the COVID-19 pandemic inhibiting lab-based research and hence the sequencing of an A and B inversion haplotype Z chromosome using further long-read sequencing technology. A future project could re-attempt split-read analysis in order to find the exact locations of breakpoints. In addition, if a Haplotype C bird were included, the locations of the C- and B- defining inversions, which are hypothesised in Chapter 4, could each be determined individually.

The lack of C haplotype sequence means many analyses could only be performed focussing on haplotypes A and B. Since these are the two most common haplotypes, this analysis is still valuable, but a fuller picture of the supergene could be obtained if more

Haplotype C sequence were acquired. For example, whilst Chapter 3 identified relaxed purifying selection in the A and B haplotypes, the lack of Haplotype C birds meant no conclusions could be drawn regarding the selection pressures acting on the C haplotype. A future study could repeat the analysis presented in Chapter 3, but compare Haplotype C to both Haplotypes A and B. The results of this analysis would indicate the level of recombination involving the C haplotype which would paint a fuller picture of the evolutionary context of the supergene. Any regions of reduced recombination between the C haplotype and A/B that are identified could help indicate the location of the C-defining inversion region.

Similarly, including more Haplotype C birds in phylogenetic inferences such as those presented in Chapter 4 would help to confirm the conclusions presented here, including which of the derived haplotypes (B or C) emerged first. As divergence between Haplotype B birds has only occurred very recently, if divergence between Haplotype C birds began earlier, then the C haplotype must have emerged first. A repeat of the analysis in Chapter 4 which utilises multiple Haplotype C birds would therefore be useful for confirming the phylogenetic relationship of the C haplotype to the other two inversion haplotypes.

The list of genes with fixed coding region differences between the A and B haplotypes that is presented in Chapter 3 could include the genes responsible for sperm variation in zebra finches. This is especially true of the genes which contain non-synonymous fixed differences. The top gene on the list, ANKRD31, contains 15 fixed differences within its coding region, 11 of which are nonsynonymous changes (Table 3.4). It also has a known role in spermatogenesis in mice (Papanikos *et al.*, 2019; Manfrevola *et al.*, 2021), specifically mice with a knockout mutation in ANKRD31 produced sperm with abnormal morphological features and which were immotile (Manfrevola *et al.*, 2021). ANKRD31 is therefore an excellent candidate gene for causing the variation in sperm motility and morphology in zebra finches. A second strong candidate gene is VPS13A, in which loss-of-function mutations have been shown to cause complete infertility as a result of severely diminished sperm motility in mice (Nagata *et al.*, 2018). Further studies into the full list of genes could discover more potential candidates. Focusing on the strongest candidate presented here, ANKRD31, further research could utilise gene editing technology such as CRISPR/Cas9 to change the sequence of ANKRD31 in zebra finches or of a homolog in another species, and then observe the impact of this change on sperm morphology and motility.

An ongoing project at the University of Sheffield aims to determine any instances of biased expression of genes between Haplotype A or B. For example, in a heterozygous bird, is either the A or B allele expressed more. If any of the genes presented in Chapter 3 of this study which contain fixed differences in their coding regions are also found to have biased expression between haplotypes, this would be evidence that the sequence divergence between haplotypes has some functional impact and may suggest that these genes are relevant to phenotypic variation between finches. Alternatively, the variation in sperm traits could be caused by non-coding SNPs altering the expression of certain genes. These genes would not have been identified in Chapter 3 but should be identified in this study into gene expression within the supergene. Other ongoing work at the University of Sheffield aims to determine if the supergene has any phenotypic impact in female birds. Currently the only known phenotypic impact of the supergene is on sperm traits, but a study of female phenotypes and of reproductive success of breeding pairs of known karyotypes, could identify other traits affected by the supergene. The list of genes presented here could then be investigated for potential links to these newly discovered phenotypes.

5.4. Concluding Remarks

The work presented in this thesis expands on previous studies of the zebra finch Z chromosome supergene (Knief *et al.*, 2016, 2017; Kim *et al.*, 2017) and provides valuable insight into the structure and evolutionary context of the inversion polymorphism. This adds to the existing literature on supergenes, which are of increasing interest due to their dramatic impact on phenotypes and evolution in a wide variety of species. Specifically, this supergene may be of widespread importance in biological research as it occurs in a model organism used in a variety of fields. Any studies investigating the genetic or phenotypic variation within zebra finches should consider this inversion polymorphism, as well as segregating autosomal inversions.

6. References

- Achaz, G. (2009) 'Frequency spectrum neutrality tests: One for all and all for one', *Genetics*, 183(1), pp. 249–258. Available at: <https://doi.org/10.1534/genetics.109.104042>.
- Anton, E. *et al.* (2005) 'Sperm studies in heterozygote inversion carriers: A review', *Cytogenetic and Genome Research*, 111(3–4), pp. 297–304. Available at: <https://doi.org/10.1159/000086903>.
- Bachtrog, D. (2013) 'Y-chromosome evolution: Emerging insights into processes of Y-chromosome degeneration', *Nature Reviews Genetics*, 14(2), pp. 113–124. Available at: <https://doi.org/10.1038/nrg3366>.
- Barton, A.B. *et al.* (2008) 'Meiotic recombination at the ends of chromosomes in *Saccharomyces cerevisiae*', *Genetics*, 179(3), pp. 1221–1235. Available at: <https://doi.org/10.1534/genetics.107.083493>.
- Bennison, C. *et al.* (2016) 'Sperm morphology, adenosine triphosphate (ATP) concentration and swimming velocity: Unexpected relationships in a passerine bird', *Proceedings of the Royal Society B: Biological Sciences*, 283(1837), p. 20161558. Available at: <https://doi.org/10.1098/rspb.2016.1558>.
- Berdan, E.L. *et al.* (2021) 'Deleterious mutation accumulation and the long-term fate of chromosomal inversions', *PLoS Genetics*, 17(3). Available at: <https://doi.org/10.1371/journal.pgen.1009411>.
- Berdan, E.L., Flatt, T., *et al.* (2022) 'Genomic architecture of supergenes: Connecting form and function', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1856). Available at: <https://doi.org/10.1098/rstb.2021.0192>.
- Berdan, E.L., Blanckaert, A., *et al.* (2022) 'Mutation accumulation opposes polymorphism: supergenes and the curious case of balanced lethals', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1856). Available at: <https://doi.org/10.1098/rstb.2021.0199>.
- Berg, P.R. *et al.* (2016) 'Three chromosomal rearrangements promote genomic divergence between migratory and stationary ecotypes of Atlantic cod', *Scientific Reports*, 6. Available at: <https://doi.org/10.1038/srep23246>.
- Berg, P.R. *et al.* (2017) 'Trans-oceanic genomic divergence of Atlantic cod ecotypes is associated with large inversions', *Heredity*, 119(6), pp. 418–428. Available at: <https://doi.org/10.1038/hdy.2017.54>.
- Birkhead, T.R., Fletcher, F. and Pellatt, E.J. (1998) 'Sexual selection in the zebra finch *Taeniopygia guttata*: Condition, sex traits and immune capacity', *Behavioral Ecology and Sociobiology*, 44(3), pp. 179–191. Available at: <https://doi.org/10.1007/s002650050530>.
- Bourke, A.F.G. and Mank, J.E. (2013) 'Genetics: A social rearrangement', *Nature*, 493(7434), pp. 612–613. Available at: <https://doi.org/10.1038/nature11854>.
- Bradbury, I.R. *et al.* (2014) 'Long distance linkage disequilibrium and limited hybridization suggest cryptic speciation in Atlantic cod', *PLoS ONE*, 9(9), p. 106380. Available at: <https://doi.org/10.1371/journal.pone.0106380>.
- Branco, S. *et al.* (2018) 'Multiple convergent supergene evolution events in mating-type

chromosomes', *Nature Communications*, 9(1). Available at: <https://doi.org/10.1038/s41467-018-04380-9>.

Burton, J.N. *et al.* (2013) 'Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions', *Nature Biotechnology*, 31(12), pp. 1119–1125. Available at: <https://doi.org/10.1038/nbt.2727>.

Camacho, C. *et al.* (2009) 'BLAST+: Architecture and applications', *BMC Bioinformatics*, 10. Available at: <https://doi.org/10.1186/1471-2105-10-421>.

Carlson, C.S. *et al.* (2005) 'Genomic regions exhibiting positive selection identified from dense genotype data', *Genome Research*, 15(11), pp. 1553–1565. Available at: <https://doi.org/10.1101/gr.4326505>.

Chaisson, M.J.P., Wilson, R.K. and Eichler, E.E. (2015) 'Genetic variation and the de novo assembly of human genomes', *Nature Reviews Genetics*, 16(11), pp. 627–640. Available at: <https://doi.org/10.1038/nrg3933>.

Chan, S. *et al.* (2018) 'Structural variation detection and analysis using bionano optical mapping', in D.M. Bickhart (ed.) *Methods in Molecular Biology*. New York, NY: Springer New York, pp. 193–203. Available at: https://doi.org/10.1007/978-1-4939-8666-8_16.

Charlesworth, D. (2016) 'The status of supergenes in the 21st century: Recombination suppression in Batesian mimicry and sex chromosomes and other complex adaptations', *Evolutionary Applications*, 9(1), pp. 74–90. Available at: <https://doi.org/10.1111/eva.12291>.

Charlesworth, D., Charlesworth, B. and Marais, G. (2005) 'Steps in the evolution of heteromorphic sex chromosomes', *Heredity*, 95(2), pp. 118–128. Available at: <https://doi.org/10.1038/sj.hdy.6800697>.

Chin, C.S. *et al.* (2016) 'Phased diploid genome assembly with single-molecule real-time sequencing', *Nature Methods*, 13(12), pp. 1050–1054. Available at: <https://doi.org/10.1038/nmeth.4035>.

Christidis, L. (1986a) 'Chromosomal evolution within the family Estrildidae (Aves) I. The Poephilae', *Genetica*, 71(2), pp. 81–97. Available at: <https://doi.org/10.1007/BF00058691>.

Christidis, L. (1986b) 'Chromosomal evolution within the family Estrildidae (Aves) II. The Lonchurae', *Genetica*, 71(2), pp. 99–113. Available at: <https://doi.org/10.1007/BF00058692>.

Christidis, L. (1987) 'Chromosomal evolution within the family Estrildidae (Aves) III. The Estrildae (waxbill finches)', *Genetica*, 72(2), pp. 93–100. Available at: <https://doi.org/10.1007/BF00123167>.

Cohen, P. and Privman, E. (2020) 'The social supergene dates back to the speciation time of two *Solenopsis* fire ant species', *Scientific Reports*, 10(1), pp. 1–9. Available at: <https://doi.org/10.1038/s41598-020-67999-z>.

Dagilis, A.J. *et al.* (2022) 'Searching for signatures of sexually antagonistic selection on stickleback sex chromosomes', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1856). Available at: <https://doi.org/10.1098/rstb.2021.0205>.

Darlington, C.D. and Mather, K. (1950) *The Elements of Genetics, Population (French Edition)*. Allen & Unwin, London. Available at: <https://doi.org/10.2307/1523397>.

Doupe, A.J. and Kuhl, P.K. (1999) 'Birdsong and human speech: Common themes and

- mechanisms', *Annual Review of Neuroscience*, 22(1), pp. 567–631. Available at: <https://doi.org/10.1146/annurev.neuro.22.1.567>.
- Eid, J. *et al.* (2009) 'Real-time DNA sequencing from single polymerase molecules', *Science*, 323(5910), pp. 133–138. Available at: <https://doi.org/10.1126/science.1162986>.
- Ellegren, H. (2010) 'Evolutionary stasis: the stable chromosomes of birds', *Trends in Ecology and Evolution*, 25(5), pp. 283–291. Available at: <https://doi.org/10.1016/j.tree.2009.12.004>.
- Ellegren, H. (2013) 'The Evolutionary Genomics of Birds', *Annual Review of Ecology, Evolution, and Systematics*, 44(1), pp. 239–259. Available at: <https://doi.org/10.1146/annurev-ecolsys-110411-160327>.
- Fay, J.C. and Wu, C.I. (2000) 'Hitchhiking under positive Darwinian selection', *Genetics*, 155(3), pp. 1405–1413. Available at: <https://doi.org/10.1093/genetics/155.3.1405>.
- Funk, E.R. *et al.* (2021) 'A supergene underlies linked variation in color and morphology in a Holarctic songbird', *Nature Communications* 2021 12:1, 12(1), pp. 1–11. Available at: <https://doi.org/10.1038/s41467-021-27173-z>.
- Galoch, Z. and Bischof, H.J. (2007) 'Behavioural responses to video playbacks by zebra finch males', *Behavioural Processes*, 74(1), pp. 21–26. Available at: <https://doi.org/10.1016/j.beproc.2006.09.002>.
- Ghurye, J. *et al.* (2019) 'Integrating Hi-C links with assembly graphs for chromosome-scale assembly', *PLoS Computational Biology*, 15(8). Available at: <https://doi.org/10.1371/journal.pcbi.1007273>.
- 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), pp. 333–351. Available at: <https://doi.org/10.1038/nrg.2016.49>.
- Guillette, L.M. and Healy, S.D. (2014) 'Mechanisms of copying behaviour in zebra finches', *Behavioural Processes*, 108, pp. 177–182. Available at: <https://doi.org/10.1016/j.beproc.2014.10.011>.
- Gutiérrez-Valencia, J. *et al.* (2021) 'The Genomic Architecture and Evolutionary Fates of Supergenes', *Genome Biology and Evolution*, 13(5). Available at: <https://doi.org/10.1093/gbe/evab057>.
- 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(1), pp. 21–42. Available at: <https://doi.org/10.1146/annurev.ecolsys.39.110707.173532>.
- Hoffmann, A.A., Sgrò, C.M. and Weeks, A.R. (2004) 'Chromosomal inversion polymorphisms and adaptation', *Trends in Ecology and Evolution*, 19(9), pp. 482–488. Available at: <https://doi.org/10.1016/j.tree.2004.06.013>.
- Hooper, D.M. and Price, T.D. (2015) 'Rates of karyotypic evolution in Estrildid finches differ between island and continental clades', *Evolution*, 69(4), pp. 890–903. Available at: <https://doi.org/10.1111/evo.12633>.
- Hooper, D.M. and Price, T.D. (2017) 'Chromosomal inversion differences correlate with range overlap in passerine birds', *Nature Ecology and Evolution*, 1(10), pp. 1526–1534.

Available at: <https://doi.org/10.1038/s41559-017-0284-6>.

Huang, K. and Rieseberg, L.H. (2020) 'Frequency, Origins, and Evolutionary Role of Chromosomal Inversions in Plants', *Frontiers in Plant Science*, 11, p. 296. Available at: <https://doi.org/10.3389/fpls.2020.00296>.

Huang, Y.C. *et al.* (2018) 'Multiple large inversions and breakpoint rewiring of gene expression in the evolution of the fire ant social supergene', *Proceedings of the Royal Society B: Biological Sciences*, 285(1878). Available at: <https://doi.org/10.1098/rspb.2018.0221>.

Huynh, L.Y., Maney, D.L. and Thomas, J.W. (2011) 'Chromosome-wide linkage disequilibrium caused by an inversion polymorphism in the white-throated sparrow (*Zonotrichia albicollis*)', *Heredity*, 106(4), pp. 537–546. Available at: <https://doi.org/10.1038/hdy.2010.85>.

Itoh, Y. *et al.* (2011) 'Karyotypic polymorphism of the zebra finch Z chromosome', *Chromosoma*, 120(3), pp. 255–264. Available at: <https://doi.org/10.1007/s00412-010-0308-3>.

Kim, K.W. *et al.* (2017) 'A sex-linked supergene controls sperm morphology and swimming speed in a songbird', *Nature Ecology and Evolution*, 1(8), pp. 1168–1176. Available at: <https://doi.org/10.1038/s41559-017-0235-2>.

Kim, Y. and Stephan, W. (2002) 'Detecting a local signature of genetic hitchhiking along a recombining chromosome', *Genetics*, 160(2), pp. 765–777. Available at: <https://doi.org/10.1093/genetics/160.2.765>.

Knief, U. *et al.* (2016) 'Fitness consequences of polymorphic inversions in the zebra finch genome', *Genome Biology*, 17(1), pp. 1–22. Available at: <https://doi.org/10.1186/s13059-016-1056-3>.

Knief, U. *et al.* (2017) 'A sex-chromosome inversion causes strong overdominance for sperm traits that affect siring success', *Nature Ecology and Evolution*, 1(8), pp. 1177–1184. Available at: <https://doi.org/10.1038/s41559-017-0236-1>.

Knief, U. *et al.* (2021) 'A sex chromosome inversion is associated with copy number variation of mitochondrial DNA in zebra finch sperm', *Royal Society Open Science*, 8(9). Available at: <https://doi.org/10.1098/RSOS.211025>.

Koboldt, D.C. *et al.* (2013) 'The next-generation sequencing revolution and its impact on genomics', *Cell*, 155(1), p. 27. Available at: <https://doi.org/10.1016/j.cell.2013.09.006>.

Kronenberg, Z.N. *et al.* (2021) 'Extended haplotype-phasing of long-read de novo genome assemblies using Hi-C', *Nature Communications*, 12(1), pp. 1–10. Available at: <https://doi.org/10.1038/s41467-020-20536-y>.

Kumar, S. *et al.* (2022) 'TimeTree 5: An Expanded Resource for Species Divergence Times', *Molecular Biology and Evolution*, 39(8). Available at: <https://doi.org/10.1093/molbev/msac174>.

Küpper, C. *et al.* (2015) 'A supergene determines highly divergent male reproductive morphs in the ruff', *Nature Genetics*, 48(1), pp. 79–83. Available at: <https://doi.org/10.1038/ng.3443>.

Lamichhaney, S. *et al.* (2015) 'Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*)', *Nature Genetics*, 48(1), pp. 84–88. Available at: <https://doi.org/10.1038/ng.3430>.

- Langmead, B. and Salzberg, S.L. (2012) ‘Fast gapped-read alignment with Bowtie 2’, *Nature Methods*, 9(4), pp. 357–359. Available at: <https://doi.org/10.1038/nmeth.1923>.
- Li, H. and Durbin, R. (2011) ‘Inference of human population history from individual whole-genome sequences’, *Nature*, 475(7357), pp. 493–496. Available at: <https://doi.org/10.1038/nature10231>.
- Lieberman-Aiden, E. *et al.* (2009) ‘Comprehensive mapping of long-range interactions reveals folding principles of the human genome’, *Science*, 326(5950), pp. 289–293. Available at: <https://doi.org/10.1126/science.1181369>.
- Liu, S. and Hansen, M.M. (2017) ‘PSMC (pairwise sequentially Markovian coalescent) analysis of RAD (restriction site associated DNA) sequencing data’, *Molecular Ecology Resources*, 17(4), pp. 631–641. Available at: <https://doi.org/10.1111/1755-0998.12606>.
- Lowry, D.B. and Willis, J.H. (2010) ‘A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation’, *PLoS Biology*, 8(9). Available at: <https://doi.org/10.1371/journal.pbio.1000500>.
- Manfrevola, F. *et al.* (2021) ‘Ankrd31 in Sperm and Epididymal Integrity’, *Frontiers in Cell and Developmental Biology*, 9. Available at: <https://doi.org/10.3389/fcell.2021.741975>.
- Martínez-Fundichely, A. *et al.* (2014) ‘InvFEST, a database integrating information of polymorphic inversions in the human genome’, *Nucleic Acids Research*, 42(D1), p. D1027. Available at: <https://doi.org/10.1093/nar/gkt1122>.
- Matschiner, M. *et al.* (2022) ‘Supergene origin and maintenance in Atlantic cod’, *Nature Ecology and Evolution*, 6(4), pp. 469–481. Available at: <https://doi.org/10.1038/s41559-022-01661-x>.
- McDonald, J.H. and Kreitman, M. (1991) ‘Adaptive protein evolution at the Adh locus in *Drosophila*’, *Nature*, 351(6328), pp. 652–654. Available at: <https://doi.org/10.1038/351652a0>.
- Mello, C. V. (2014) ‘The Zebra finch, *taeniopygia guttata*: An avian model for investigating the neurobiological basis of vocal learning’, *Cold Spring Harbor Protocols*, 2014(12), pp. 1237–1242. Available at: <https://doi.org/10.1101/pdb.emo084574>.
- Mi, H. *et al.* (2019) ‘PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools’, *Nucleic Acids Research*, 47(D1), pp. D419–D426. Available at: <https://doi.org/10.1093/NAR/GKY1038>.
- Miller, D.E. (2020) ‘The interchromosomal effect: Different meanings for different organisms’, *Genetics*, 216(3), pp. 621–631. Available at: <https://doi.org/10.1534/genetics.120.303656>.
- Morel, F. *et al.* (2004) ‘Meiotic segregation of translocations during male gametogenesis’, *International Journal of Andrology*, 27(4), pp. 200–212. Available at: <https://doi.org/10.1111/j.1365-2605.2004.00490.x>.
- Morris, D. (2008) *The Reproductive Behaviour of the Zebra Finch (Poephila Guttata), With Special Reference To Pseudofemale Behaviour and Displacement Activities, Behaviour*. Available at: <https://doi.org/10.1163/156853954x00130>.
- Nabours, R.K. (1929) ‘The Genetics of the Tettigidae (Grouse Locusts)’, *The Genetics of the Tettigidae (Grouse Locusts)*, pp. 27–100. Available at: <https://doi.org/10.1007/978-94-011->

9487-7_1.

Nabours, R.K., Larson, I. and Hartwig, N. (1933) ‘Inheritance of Color Patterns in the Grouse Locust *Acrydium Arenosum* Burmeister (Tettigidae)’, *Genetics*, 18(2), pp. 159–171. Available at: <https://doi.org/10.1093/genetics/18.2.159>.

Nadachowska-Brzyska, K. *et al.* (2015) ‘Temporal dynamics of avian populations during pleistocene revealed by whole-genome sequences’, *Current Biology*, 25(10), pp. 1375–1380. Available at: <https://doi.org/10.1016/j.cub.2015.03.047>.

Nagata, O. *et al.* (2018) ‘Mouse model of chorea-acanthocytosis exhibits male infertility caused by impaired sperm motility as a result of ultrastructural morphological abnormalities in the mitochondrial sheath in the sperm midpiece’, *Biochemical and Biophysical Research Communications*, 503(2), pp. 915–920. Available at: <https://doi.org/10.1016/j.bbrc.2018.06.096>.

Nielsen, R. (2005) ‘Molecular Signatures of Natural Selection’, <https://doi.org/10.1146/annurev.genet.39.073003.112420>, 39, pp. 197–218. Available at: <https://doi.org/10.1146/ANNUREV.GENET.39.073003.112420>.

Nurk, S. *et al.* (2022) ‘The complete sequence of a human genome’, *Science*, 376(6588), pp. 44–53. Available at: <https://doi.org/10.1126/science.abj6987>.

Olsson, U. and Alström, P. (2020) ‘A comprehensive phylogeny and taxonomic evaluation of the waxbills (Aves: Estrildidae)’, *Molecular Phylogenetics and Evolution*, 146, p. 106757. Available at: <https://doi.org/10.1016/j.ympev.2020.106757>.

Ott, A. *et al.* (2018) ‘Linked read technology for assembling large complex and polyploid genomes’, *BMC Genomics*, 19(1), p. 651. Available at: <https://doi.org/10.1186/s12864-018-5040-z>.

Papanikos, F. *et al.* (2019) ‘Mouse ANKRD31 Regulates Spatiotemporal Patterning of Meiotic Recombination Initiation and Ensures Recombination between X and Y Sex Chromosomes’, *Molecular Cell*, 74(5), pp. 1069-1085.e11. Available at: <https://doi.org/10.1016/j.molcel.2019.03.022>.

Patterson, N.J. (2005) ‘How old is the most recent ancestor of two copies of an allele?’, *Genetics*, 169(2), pp. 1093–1104. Available at: <https://doi.org/10.1534/genetics.103.015768>.

Pfeifer, B. *et al.* (2014) ‘PopGenome: An efficient swiss army knife for population genomic analyses in R’, *Molecular Biology and Evolution*, 31(7), pp. 1929–1936. Available at: <https://doi.org/10.1093/molbev/msu136>.

Poplin, R. *et al.* (2017) ‘Scaling accurate genetic variant discovery to tens of thousands of samples’, *bioRxiv*, p. 201178. Available at: <https://doi.org/10.1101/201178>.

Pracana, R. *et al.* (2017) ‘The fire ant social chromosome supergene variant Sb shows low diversity but high divergence from SB’, *Molecular Ecology*, 26(11), pp. 2864–2879. Available at: <https://doi.org/10.1111/mec.14054>.

Puig, M. *et al.* (2015) ‘Human inversions and their functional consequences’, *Briefings in Functional Genomics*, 14(5), pp. 369–379. Available at: <https://doi.org/10.1093/bfgp/elv020>.

Purcell, J. *et al.* (2014) ‘Convergent genetic architecture underlies social organization in ants’, *Current Biology*, 24(22), pp. 2728–2732. Available at: <https://doi.org/10.1016/j.cub.2014.09.071>.

- Purcell, S. *et al.* (2007) ‘PLINK: A tool set for whole-genome association and population-based linkage analyses’, *American Journal of Human Genetics*, 81(3), pp. 559–575. Available at: <https://doi.org/10.1086/519795>.
- Rhie, A. *et al.* (2021) ‘Towards complete and error-free genome assemblies of all vertebrate species’, *Nature*, 592(7856), pp. 737–746. Available at: <https://doi.org/10.1038/s41586-021-03451-0>.
- Rhoads, A. and Au, K.F. (2015) ‘PacBio Sequencing and Its Applications’, *Genomics, Proteomics and Bioinformatics*, 13(5), pp. 278–289. Available at: <https://doi.org/10.1016/j.gpb.2015.08.002>.
- Rieseberg, L.H. (2001) ‘Chromosomal rearrangements and speciation’, *Trends in Ecology and Evolution*, 16(7), pp. 351–358. Available at: [https://doi.org/10.1016/S0169-5347\(01\)02187-5](https://doi.org/10.1016/S0169-5347(01)02187-5).
- Ross, K.G. and Keller, L. (1995) ‘Ecology and evolution of social organization: Insights from fire ants and other highly eusocial insects’, *Annual Review of Ecology and Systematics*, 26, pp. 631–656. Available at: <https://doi.org/10.1146/annurev.es.26.110195.003215>.
- Santos, M.D.S. Dos *et al.* (2017) ‘Comparative Cytogenetics between Two Important Songbird, Models: The Zebra Finch and the Canary’, *PLOS ONE*, 12(1), p. e0170997. Available at: <https://doi.org/10.1371/JOURNAL.PONE.0170997>.
- Sharbrough, J. *et al.* (2018) ‘Radical amino acid mutations persist longer in the absence of sex’, *Evolution*, 72(4), pp. 808–824. Available at: <https://doi.org/10.1111/EVO.13465>.
- 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. Available at: <https://doi.org/10.1023/A:1009278914829>.
- Simons, M.J.P. and Verhulst, S. (2011) ‘Zebra finch females prefer males with redder bills independent of song rate—a meta-analysis’, *Behavioral Ecology*, 22(4), pp. 755–762. Available at: <https://doi.org/10.1093/beheco/arr043>.
- Singhal, S. *et al.* (2015) ‘Stable recombination hotspots in birds’, *Science*, 350(6263), pp. 928–932. Available at: <https://doi.org/10.1126/science.aad0843>.
- Stapley, J. *et al.* (2008) ‘A linkage map of the zebra finch *Taeniopygia guttata* provides new insights into avian genome evolution’, *Genetics*, 179(1), pp. 651–667. Available at: <https://doi.org/10.1534/genetics.107.086264>.
- Tafreshi, A.G., Otto, S.P. and Chapuisat, M. (2022) ‘Unbalanced selection: The challenge of maintaining a social polymorphism when a supergene is selfish’, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1856). Available at: <https://doi.org/10.1098/rstb.2021.0197>.
- Tajima, F. (1989) ‘Statistical method for testing the neutral mutation hypothesis by DNA polymorphism’, *Genetics*, 123(3), pp. 585–595. Available at: <https://doi.org/10.1093/genetics/123.3.585>.
- Talukdar, D. (2010) ‘Reciprocal translocations in grass pea (*Lathyrus sativus* L.): Pattern of transmission, detection of multiple interchanges and their independence’, *Journal of Heredity*, 101(2), pp. 169–176. Available at: <https://doi.org/10.1093/jhered/esp106>.
- Tamura, K. and Nei, M. (1993) ‘Estimation of the number of nucleotide substitutions in the

- control region of mitochondrial DNA in humans and chimpanzees', *Molecular Biology and Evolution*, 10(3), pp. 512–526. Available at: <https://doi.org/10.1093/oxfordjournals.molbev.a040023>.
- Tamura, K., Stecher, G. and Kumar, S. (2021) 'MEGA11: Molecular Evolutionary Genetics Analysis Version 11', *Molecular Biology and Evolution*, 38(7), pp. 3022–3027. Available at: <https://doi.org/10.1093/molbev/msab120>.
- Tamura, K., Tao, Q. and Kumar, S. (2018) 'Theoretical foundation of the reltime method for estimating divergence times from variable evolutionary rates', *Molecular Biology and Evolution*, 35(7), pp. 1770–1782. Available at: <https://doi.org/10.1093/molbev/msy044>.
- Thompson, M.J. and Jiggins, C.D. (2014) 'Supergenes and their role in evolution', *Heredity*, 113(1), pp. 1–8. Available at: <https://doi.org/10.1038/hdy.2014.20>.
- Tomaszkiewicz, M., Medvedev, P. and Makova, K.D. (2017) 'Y and W Chromosome Assemblies: Approaches and Discoveries', *Trends in Genetics*, 33(4), pp. 266–282. Available at: <https://doi.org/10.1016/j.tig.2017.01.008>.
- Turner, J.R.G. (1967) 'On Supergenes. I. The Evolution of Supergenes', *The American Naturalist*, 101(919), pp. 195–221. Available at: <https://doi.org/10.1086/282485>.
- Tuttle, E.M. *et al.* (2016) 'Divergence and functional degradation of a sex chromosome-like supergene', *Current Biology*, 26(3), pp. 344–350. Available at: <https://doi.org/10.1016/j.cub.2015.11.069>.
- Wang, J. *et al.* (2013) 'A Y-like social chromosome causes alternative colony organization in fire ants', *Nature*, 493(7434), pp. 664–668. Available at: <https://doi.org/10.1038/nature11832>.
- Warren, W.C. *et al.* (2010) 'The genome of a songbird', *Nature*, 464(7289), pp. 757–762. Available at: <https://doi.org/10.1038/nature08819>.
- Weisenfeld, N.I. *et al.* (2017) 'Direct determination of diploid genome sequences', *Genome Research*, 27(5), pp. 757–767. Available at: <https://doi.org/10.1101/gr.214874.116>.
- Winsor, E.J.T. *et al.* (1978) 'Meiotic analysis of a pericentric inversion, inv(7) (p22q32), in the father of a child with a duplication-deletion of chromosome 7', *Cytogenetic and Genome Research*, 20(1–6), pp. 169–184. Available at: <https://doi.org/10.1159/000130849>.
- Wright, A.E. *et al.* (2016) 'How to make a sex chromosome', *Nature Communications*, 7. Available at: <https://doi.org/10.1038/ncomms12087>.
- Yapan, C. *et al.* (2013) 'P-57 The largest paracentric inversion, the highest rate of recombinant spermatozoa in a 46, XY, inv(2)(q21.2q37.3) carrier', *Reproductive BioMedicine Online*, 26, pp. S48–S49. Available at: [https://doi.org/10.1016/S1472-6483\(13\)60120-0](https://doi.org/10.1016/S1472-6483(13)60120-0).
- Zann, R.A. (1996) *The zebra finch: a synthesis of field and laboratory studies (Vol. 5)*. Oxford University Press.
- Zeng, K., Shi, S. and Wu, C.I. (2007) 'Compound tests for the detection of hitchhiking under positive selection', *Molecular Biology and Evolution*, 24(8), pp. 1898–1908. Available at: <https://doi.org/10.1093/molbev/msm119>.
- Zhai, W., Nielsen, R. and Slatkin, M. (2009) 'An Investigation of the Statistical Power of Neutrality Tests Based on Comparative and Population Genetic Data', *Molecular Biology and Evolution*, 26(2), pp. 273–283. Available at: <https://doi.org/10.1093/MOLBEV/MSN231>.

Zhang, G. *et al.* (2014) ‘Comparative genomics reveals insights into avian genome evolution and adaptation’, *Science*, 346(6215), pp. 1311–1320. Available at: <https://doi.org/10.1126/science.1251385>.

Zinzow-Kramer, W.M. *et al.* (2015) ‘Genes located in a chromosomal inversion are correlated with territorial song in white-throated sparrows’, *Genes, Brain and Behavior*, 14(8), pp. 641–654. Available at: <https://doi.org/10.1111/gbb.12252>.

7. Supporting Material

Table S2.1:

List of 2123 SNP positions which were found to be highly diagnostic (weighted chi-square score >0.9) between the A and B inversion haplotypes. SNPs are sorted by chi-square test statistic and then by position on the Z chromosome.

Physical Position	chi-square statistic	p-value	Weighted chi-square	Physical Position	chi-square statistic	p-value	Weighted chi-square	Physical Position	chi-square statistic	p-value	Weighted chi-square
5,987,059	443.00	6.36E-97	1.000	11,366,826	443.00	6.36E-97	1.000	36,382,866	443.00	6.36E-97	1.000
6,019,307	443.00	6.36E-97	1.000	11,402,084	443.00	6.36E-97	1.000	36,391,362	443.00	6.36E-97	1.000
6,110,634	443.00	6.36E-97	1.000	11,417,108	443.00	6.36E-97	1.000	36,433,250	443.00	6.36E-97	1.000
6,176,914	443.00	6.36E-97	1.000	11,439,964	443.00	6.36E-97	1.000	37,972,587	443.00	6.36E-97	1.000
6,282,469	443.00	6.36E-97	1.000	11,504,926	443.00	6.36E-97	1.000	38,040,290	443.00	6.36E-97	1.000
6,410,624	443.00	6.36E-97	1.000	11,740,993	443.00	6.36E-97	1.000	38,660,542	443.00	6.36E-97	1.000
7,042,446	443.00	6.36E-97	1.000	11,803,650	443.00	6.36E-97	1.000	42,730,338	443.00	6.36E-97	1.000
7,181,872	443.00	6.36E-97	1.000	11,833,469	443.00	6.36E-97	1.000	45,774,277	443.00	6.36E-97	1.000
7,343,825	443.00	6.36E-97	1.000	13,208,795	443.00	6.36E-97	1.000	47,816,666	443.00	6.36E-97	1.000
7,399,220	443.00	6.36E-97	1.000	13,770,625	443.00	6.36E-97	1.000	48,253,340	443.00	6.36E-97	1.000
7,771,946	443.00	6.36E-97	1.000	14,019,360	443.00	6.36E-97	1.000	49,407,383	443.00	6.36E-97	1.000
8,508,892	443.00	6.36E-97	1.000	14,744,843	443.00	6.36E-97	1.000	52,352,191	443.00	6.36E-97	1.000
8,648,270	443.00	6.36E-97	1.000	16,009,509	443.00	6.36E-97	1.000	55,809,486	443.00	6.36E-97	1.000
9,172,182	443.00	6.36E-97	1.000	16,331,942	443.00	6.36E-97	1.000	59,787,552	443.00	6.36E-97	1.000
9,188,355	443.00	6.36E-97	1.000	16,525,366	443.00	6.36E-97	1.000	5,915,079	442.00	1.05E-96	0.998
10,130,660	443.00	6.36E-97	1.000	16,657,708	443.00	6.36E-97	1.000	6,033,335	442.00	1.05E-96	0.998
10,187,185	443.00	6.36E-97	1.000	16,662,209	443.00	6.36E-97	1.000	6,420,011	442.00	1.05E-96	0.998
10,458,579	443.00	6.36E-97	1.000	17,142,459	443.00	6.36E-97	1.000	6,422,602	442.00	1.05E-96	0.998
10,571,541	443.00	6.36E-97	1.000	17,873,122	443.00	6.36E-97	1.000	6,531,036	442.00	1.05E-96	0.998
11,099,340	443.00	6.36E-97	1.000	22,427,407	443.00	6.36E-97	1.000	6,732,492	442.00	1.05E-96	0.998
11,102,627	443.00	6.36E-97	1.000	31,678,886	443.00	6.36E-97	1.000	7,222,157	442.00	1.05E-96	0.998

Physical Position	chi-square statistic	p-value	Weighted chi-square
7,284,575	442.00	1.05E-96	0.998
7,429,575	442.00	1.05E-96	0.998
7,518,957	442.00	1.05E-96	0.998
7,527,849	442.00	1.05E-96	0.998
7,535,523	442.00	1.05E-96	0.998
7,620,061	442.00	1.05E-96	0.998
7,719,533	442.00	1.05E-96	0.998
7,730,156	442.00	1.05E-96	0.998
8,549,396	442.00	1.05E-96	0.998
8,984,561	442.00	1.05E-96	0.998
9,188,713	442.00	1.05E-96	0.998
9,193,369	442.00	1.05E-96	0.998
9,376,933	442.00	1.05E-96	0.998
9,462,187	442.00	1.05E-96	0.998
9,571,565	442.00	1.05E-96	0.998
9,586,756	442.00	1.05E-96	0.998
9,859,623	442.00	1.05E-96	0.998
10,167,657	442.00	1.05E-96	0.998
10,247,786	442.00	1.05E-96	0.998
10,410,507	442.00	1.05E-96	0.998
10,439,530	442.00	1.05E-96	0.998
10,444,696	442.00	1.05E-96	0.998
10,620,323	442.00	1.05E-96	0.998
10,678,178	442.00	1.05E-96	0.998
10,955,656	442.00	1.05E-96	0.998
10,979,773	442.00	1.05E-96	0.998
11,073,692	442.00	1.05E-96	0.998
11,100,639	442.00	1.05E-96	0.998

Physical Position	chi-square statistic	p-value	Weighted chi-square
11,113,911	442.00	1.05E-96	0.998
11,367,099	442.00	1.05E-96	0.998
11,620,348	442.00	1.05E-96	0.998
11,639,247	442.00	1.05E-96	0.998
11,777,583	442.00	1.05E-96	0.998
12,043,419	442.00	1.05E-96	0.998
12,616,766	442.00	1.05E-96	0.998
13,209,083	442.00	1.05E-96	0.998
13,345,145	442.00	1.05E-96	0.998
13,827,779	442.00	1.05E-96	0.998
13,840,861	442.00	1.05E-96	0.998
14,069,822	442.00	1.05E-96	0.998
14,194,625	442.00	1.05E-96	0.998
14,296,298	442.00	1.05E-96	0.998
14,566,076	442.00	1.05E-96	0.998
14,569,213	442.00	1.05E-96	0.998
14,579,310	442.00	1.05E-96	0.998
14,716,790	442.00	1.05E-96	0.998
14,843,467	442.00	1.05E-96	0.998
14,921,232	442.00	1.05E-96	0.998
14,922,472	442.00	1.05E-96	0.998
15,049,969	442.00	1.05E-96	0.998
15,318,326	442.00	1.05E-96	0.998
15,631,562	442.00	1.05E-96	0.998
15,835,085	442.00	1.05E-96	0.998
16,588,278	442.00	1.05E-96	0.998
16,600,582	442.00	1.05E-96	0.998
16,750,326	442.00	1.05E-96	0.998

Physical Position	chi-square statistic	p-value	Weighted chi-square
17,899,164	442.00	1.05E-96	0.998
20,345,625	442.00	1.05E-96	0.998
20,787,356	442.00	1.05E-96	0.998
21,097,349	442.00	1.05E-96	0.998
24,216,827	442.00	1.05E-96	0.998
24,725,909	442.00	1.05E-96	0.998
29,963,041	442.00	1.05E-96	0.998
31,446,386	442.00	1.05E-96	0.998
33,403,615	442.00	1.05E-96	0.998
33,946,345	442.00	1.05E-96	0.998
35,234,205	442.00	1.05E-96	0.998
36,282,054	442.00	1.05E-96	0.998
36,420,199	442.00	1.05E-96	0.998
36,543,831	442.00	1.05E-96	0.998
36,872,834	442.00	1.05E-96	0.998
37,392,576	442.00	1.05E-96	0.998
37,404,821	442.00	1.05E-96	0.998
37,520,412	442.00	1.05E-96	0.998
37,709,759	442.00	1.05E-96	0.998
37,813,719	442.00	1.05E-96	0.998
37,840,295	442.00	1.05E-96	0.998
37,924,088	442.00	1.05E-96	0.998
38,133,363	442.00	1.05E-96	0.998
38,369,891	442.00	1.05E-96	0.998
38,605,463	442.00	1.05E-96	0.998
38,720,173	442.00	1.05E-96	0.998
38,780,367	442.00	1.05E-96	0.998
41,359,173	442.00	1.05E-96	0.998

Physical Position	chi-square statistic	p-value	Weighted chi-square
41,712,855	442.00	1.05E-96	0.998
41,805,944	442.00	1.05E-96	0.998
42,066,523	442.00	1.05E-96	0.998
42,225,311	442.00	1.05E-96	0.998
42,743,499	442.00	1.05E-96	0.998
42,822,786	442.00	1.05E-96	0.998
43,621,324	442.00	1.05E-96	0.998
44,693,179	442.00	1.05E-96	0.998
45,276,026	442.00	1.05E-96	0.998
45,507,668	442.00	1.05E-96	0.998
45,564,524	442.00	1.05E-96	0.998
45,746,251	442.00	1.05E-96	0.998
47,769,567	442.00	1.05E-96	0.998
47,826,755	442.00	1.05E-96	0.998
47,845,399	442.00	1.05E-96	0.998
47,867,070	442.00	1.05E-96	0.998
48,654,397	442.00	1.05E-96	0.998
48,884,209	442.00	1.05E-96	0.998
57,527,380	442.00	1.05E-96	0.998
58,334,552	442.00	1.05E-96	0.998
59,790,796	442.00	1.05E-96	0.998
64,378,809	442.00	1.05E-96	0.998
65,968,913	442.00	1.05E-96	0.998
5,926,566	441.00	1.73E-96	0.995
6,031,737	441.00	1.73E-96	0.995
6,180,171	441.00	1.73E-96	0.995
6,294,965	441.00	1.73E-96	0.995
6,355,256	441.00	1.73E-96	0.995

Physical Position	chi-square statistic	p-value	Weighted chi-square
6,370,390	441.00	1.73E-96	0.995
6,472,139	441.00	1.73E-96	0.995
6,691,034	441.00	1.73E-96	0.995
6,781,941	441.00	1.73E-96	0.995
6,860,176	441.00	1.73E-96	0.995
6,876,899	441.00	1.73E-96	0.995
7,107,163	441.00	1.73E-96	0.995
7,117,358	441.00	1.73E-96	0.995
7,344,659	441.00	1.73E-96	0.995
7,758,457	441.00	1.73E-96	0.995
8,412,172	441.00	1.73E-96	0.995
8,463,284	441.00	1.73E-96	0.995
8,468,435	441.00	1.73E-96	0.995
8,553,777	441.00	1.73E-96	0.995
8,633,788	441.00	1.73E-96	0.995
8,677,084	441.00	1.73E-96	0.995
8,721,890	441.00	1.73E-96	0.995
8,784,247	441.00	1.73E-96	0.995
9,004,844	441.00	1.73E-96	0.995
9,184,636	441.00	1.73E-96	0.995
9,184,727	441.00	1.73E-96	0.995
9,388,617	441.00	1.73E-96	0.995
9,521,589	441.00	1.73E-96	0.995
9,560,154	441.00	1.73E-96	0.995
9,779,623	441.00	1.73E-96	0.995
9,894,090	441.00	1.73E-96	0.995
9,952,904	441.00	1.73E-96	0.995
9,996,940	441.00	1.73E-96	0.995

Physical Position	chi-square statistic	p-value	Weighted chi-square
10,166,850	441.00	1.73E-96	0.995
10,599,990	441.00	1.73E-96	0.995
10,608,879	441.00	1.73E-96	0.995
10,658,432	441.00	1.73E-96	0.995
10,936,950	441.00	1.73E-96	0.995
10,963,076	441.00	1.73E-96	0.995
11,134,217	441.00	1.73E-96	0.995
11,409,951	441.00	1.73E-96	0.995
11,412,456	441.00	1.73E-96	0.995
11,486,514	441.00	1.73E-96	0.995
11,524,034	441.00	1.73E-96	0.995
11,542,643	441.00	1.73E-96	0.995
11,583,883	441.00	1.73E-96	0.995
11,909,169	441.00	1.73E-96	0.995
11,966,810	441.00	1.73E-96	0.995
12,052,304	441.00	1.73E-96	0.995
12,813,075	441.00	1.73E-96	0.995
13,142,968	441.00	1.73E-96	0.995
13,150,395	441.00	1.73E-96	0.995
13,351,060	441.00	1.73E-96	0.995
13,813,142	441.00	1.73E-96	0.995
13,927,960	441.00	1.73E-96	0.995
14,044,893	441.00	1.73E-96	0.995
14,163,906	441.00	1.73E-96	0.995
14,409,861	441.00	1.73E-96	0.995
14,410,800	441.00	1.73E-96	0.995
14,745,061	441.00	1.73E-96	0.995
14,843,761	441.00	1.73E-96	0.995

Physical Position	chi-square statistic	p-value	Weighted chi-square
15,038,294	441.00	1.73E-96	0.995
15,436,358	441.00	1.73E-96	0.995
15,849,371	441.00	1.73E-96	0.995
15,882,467	441.00	1.73E-96	0.995
16,441,156	441.00	1.73E-96	0.995
16,494,379	441.00	1.73E-96	0.995
16,587,033	441.00	1.73E-96	0.995
16,721,291	441.00	1.73E-96	0.995
16,842,755	441.00	1.73E-96	0.995
16,923,279	441.00	1.73E-96	0.995
17,493,272	441.00	1.73E-96	0.995
17,704,236	441.00	1.73E-96	0.995
17,847,415	441.00	1.73E-96	0.995
17,893,386	441.00	1.73E-96	0.995
17,975,605	441.00	1.73E-96	0.995
25,211,613	441.00	1.73E-96	0.995
29,048,233	441.00	1.73E-96	0.995
30,421,704	441.00	1.73E-96	0.995
32,408,293	441.00	1.73E-96	0.995
35,496,045	441.00	1.73E-96	0.995
37,008,928	441.00	1.73E-96	0.995
37,859,965	441.00	1.73E-96	0.995
38,350,363	441.00	1.73E-96	0.995
38,361,582	441.00	1.73E-96	0.995
38,605,826	441.00	1.73E-96	0.995
42,721,543	441.00	1.73E-96	0.995
43,727,945	441.00	1.73E-96	0.995
47,766,627	441.00	1.73E-96	0.995

Physical Position	chi-square statistic	p-value	Weighted chi-square
48,021,560	441.00	1.73E-96	0.995
48,156,696	441.00	1.73E-96	0.995
48,567,181	441.00	1.73E-96	0.995
48,661,297	441.00	1.73E-96	0.995
49,493,715	441.00	1.73E-96	0.995
52,294,743	441.00	1.73E-96	0.995
54,885,492	441.00	1.73E-96	0.995
55,533,253	441.00	1.73E-96	0.995
58,702,208	441.00	1.73E-96	0.995
60,207,080	441.00	1.73E-96	0.995
64,090,486	441.00	1.73E-96	0.995
7,007,236	441.00	1.73E-96	0.995
7,210,599	441.00	1.73E-96	0.995
9,231,187	441.00	1.73E-96	0.995
12,549,818	441.00	1.73E-96	0.995
13,794,068	441.00	1.73E-96	0.995
38,744,226	441.00	1.73E-96	0.995
62,986,871	440.33	2.41E-96	0.994
5,948,230	440.00	2.85E-96	0.993
6,014,404	440.00	2.85E-96	0.993
6,070,177	440.00	2.85E-96	0.993
6,120,236	440.00	2.85E-96	0.993
6,282,858	440.00	2.85E-96	0.993
6,325,701	440.00	2.85E-96	0.993
6,385,353	440.00	2.85E-96	0.993
6,389,865	440.00	2.85E-96	0.993
6,581,416	440.00	2.85E-96	0.993
6,720,009	440.00	2.85E-96	0.993

Physical Position	chi-square statistic	p-value	Weighted chi-square
6,774,678	440.00	2.85E-96	0.993
7,094,254	440.00	2.85E-96	0.993
7,421,269	440.00	2.85E-96	0.993
7,568,851	440.00	2.85E-96	0.993
7,640,443	440.00	2.85E-96	0.993
7,645,116	440.00	2.85E-96	0.993
7,731,180	440.00	2.85E-96	0.993
7,765,427	440.00	2.85E-96	0.993
7,893,191	440.00	2.85E-96	0.993
7,947,883	440.00	2.85E-96	0.993
8,147,703	440.00	2.85E-96	0.993
8,174,126	440.00	2.85E-96	0.993
8,524,507	440.00	2.85E-96	0.993
8,532,836	440.00	2.85E-96	0.993
8,738,807	440.00	2.85E-96	0.993
8,793,186	440.00	2.85E-96	0.993
9,010,360	440.00	2.85E-96	0.993
9,032,147	440.00	2.85E-96	0.993
9,400,138	440.00	2.85E-96	0.993
9,411,569	440.00	2.85E-96	0.993
9,433,289	440.00	2.85E-96	0.993
9,501,032	440.00	2.85E-96	0.993
9,816,269	440.00	2.85E-96	0.993
9,920,639	440.00	2.85E-96	0.993
10,136,858	440.00	2.85E-96	0.993
10,336,155	440.00	2.85E-96	0.993
10,394,854	440.00	2.85E-96	0.993
10,481,970	440.00	2.85E-96	0.993

Physical Position	chi-square statistic	p-value	Weighted chi-square
10,906,159	440.00	2.85E-96	0.993
10,965,942	440.00	2.85E-96	0.993
10,979,550	440.00	2.85E-96	0.993
11,111,369	440.00	2.85E-96	0.993
11,114,750	440.00	2.85E-96	0.993
11,486,065	440.00	2.85E-96	0.993
11,634,635	440.00	2.85E-96	0.993
11,946,483	440.00	2.85E-96	0.993
11,957,145	440.00	2.85E-96	0.993
12,019,563	440.00	2.85E-96	0.993
12,505,315	440.00	2.85E-96	0.993
12,813,876	440.00	2.85E-96	0.993
13,351,090	440.00	2.85E-96	0.993
13,833,678	440.00	2.85E-96	0.993
14,016,414	440.00	2.85E-96	0.993
14,241,561	440.00	2.85E-96	0.993
14,320,384	440.00	2.85E-96	0.993
14,350,137	440.00	2.85E-96	0.993
14,458,119	440.00	2.85E-96	0.993
14,526,659	440.00	2.85E-96	0.993
14,529,278	440.00	2.85E-96	0.993
14,760,129	440.00	2.85E-96	0.993
15,259,246	440.00	2.85E-96	0.993
16,146,237	440.00	2.85E-96	0.993
16,496,947	440.00	2.85E-96	0.993
16,710,479	440.00	2.85E-96	0.993
16,796,968	440.00	2.85E-96	0.993
16,907,889	440.00	2.85E-96	0.993

Physical Position	chi-square statistic	p-value	Weighted chi-square
17,028,567	440.00	2.85E-96	0.993
17,153,979	440.00	2.85E-96	0.993
17,877,506	440.00	2.85E-96	0.993
18,268,600	440.00	2.85E-96	0.993
18,811,280	440.00	2.85E-96	0.993
22,272,644	440.00	2.85E-96	0.993
25,488,154	440.00	2.85E-96	0.993
31,123,103	440.00	2.85E-96	0.993
31,548,520	440.00	2.85E-96	0.993
34,116,199	440.00	2.85E-96	0.993
37,306,855	440.00	2.85E-96	0.993
37,398,228	440.00	2.85E-96	0.993
37,564,329	440.00	2.85E-96	0.993
37,637,937	440.00	2.85E-96	0.993
37,687,131	440.00	2.85E-96	0.993
37,719,009	440.00	2.85E-96	0.993
38,791,001	440.00	2.85E-96	0.993
40,041,454	440.00	2.85E-96	0.993
43,468,605	440.00	2.85E-96	0.993
43,831,905	440.00	2.85E-96	0.993
44,540,068	440.00	2.85E-96	0.993
44,799,062	440.00	2.85E-96	0.993
44,958,308	440.00	2.85E-96	0.993
45,117,540	440.00	2.85E-96	0.993
45,243,264	440.00	2.85E-96	0.993
45,716,957	440.00	2.85E-96	0.993
47,776,994	440.00	2.85E-96	0.993
47,791,866	440.00	2.85E-96	0.993

Physical Position	chi-square statistic	p-value	Weighted chi-square
47,970,279	440.00	2.85E-96	0.993
48,614,188	440.00	2.85E-96	0.993
49,934,784	440.00	2.85E-96	0.993
50,072,375	440.00	2.85E-96	0.993
52,322,633	440.00	2.85E-96	0.993
52,414,459	440.00	2.85E-96	0.993
52,613,763	440.00	2.85E-96	0.993
55,535,495	440.00	2.85E-96	0.993
58,140,169	440.00	2.85E-96	0.993
58,744,598	440.00	2.85E-96	0.993
59,928,521	440.00	2.85E-96	0.993
60,233,428	440.00	2.85E-96	0.993
62,738,836	440.00	2.85E-96	0.993
66,991,610	440.00	2.85E-96	0.993
14,624,590	440.00	2.85E-96	0.993
15,271,413	440.00	2.85E-96	0.993
15,027,879	440.00	2.85E-96	0.993
44,425,529	440.00	2.85E-96	0.993
55,667,298	440.00	2.85E-96	0.993
13,230,358	439.67	3.37E-96	0.992
7,399,021	439.33	3.98E-96	0.992
52,333,397	439.33	3.98E-96	0.992
32,344,591	439.09	4.50E-96	0.991
7,029,235	439.08	4.51E-96	0.991
47,797,375	439.02	4.66E-96	0.991
5,975,168	439.02	4.66E-96	0.991
7,065,847	439.02	4.66E-96	0.991
7,444,461	439.02	4.66E-96	0.991

Physical Position	chi-square statistic	p-value	Weighted chi-square
8,458,865	439.02	4.66E-96	0.991
8,893,122	439.02	4.66E-96	0.991
14,050,553	439.02	4.66E-96	0.991
15,603,157	439.02	4.66E-96	0.991
31,914,960	439.02	4.66E-96	0.991
6,044,206	439.01	1.78E-97	0.991
6,385,701	439.01	1.78E-97	0.991
6,539,029	439.01	1.78E-97	0.991
6,582,283	439.01	1.78E-97	0.991
6,629,178	439.01	1.78E-97	0.991
6,645,262	439.01	1.78E-97	0.991
6,855,264	439.01	1.78E-97	0.991
6,889,699	439.01	1.78E-97	0.991
7,134,876	439.01	1.78E-97	0.991
7,222,611	439.01	1.78E-97	0.991
7,278,496	439.01	1.78E-97	0.991
7,344,160	439.01	1.78E-97	0.991
7,374,326	439.01	1.78E-97	0.991
7,396,272	439.01	1.78E-97	0.991
7,431,237	439.01	1.78E-97	0.991
7,532,375	439.01	1.78E-97	0.991
7,996,694	439.01	1.78E-97	0.991
8,332,299	439.01	1.78E-97	0.991
8,561,933	439.01	1.78E-97	0.991
8,624,136	439.01	1.78E-97	0.991
8,645,232	439.01	1.78E-97	0.991
8,841,688	439.01	1.78E-97	0.991
8,907,339	439.01	1.78E-97	0.991

Physical Position	chi-square statistic	p-value	Weighted chi-square
9,148,846	439.01	1.78E-97	0.991
9,246,308	439.01	1.78E-97	0.991
9,274,448	439.01	1.78E-97	0.991
9,293,367	439.01	1.78E-97	0.991
9,384,456	439.01	1.78E-97	0.991
9,576,268	439.01	1.78E-97	0.991
9,609,333	439.01	1.78E-97	0.991
9,706,463	439.01	1.78E-97	0.991
9,804,434	439.01	1.78E-97	0.991
9,944,789	439.01	1.78E-97	0.991
10,013,287	439.01	1.78E-97	0.991
10,449,184	439.01	1.78E-97	0.991
10,451,508	439.01	1.78E-97	0.991
10,578,246	439.01	1.78E-97	0.991
10,596,480	439.01	1.78E-97	0.991
10,611,893	439.01	1.78E-97	0.991
10,620,066	439.01	1.78E-97	0.991
10,620,871	439.01	1.78E-97	0.991
10,717,668	439.01	1.78E-97	0.991
11,729,810	439.01	1.78E-97	0.991
11,751,848	439.01	1.78E-97	0.991
11,857,148	439.01	1.78E-97	0.991
11,876,024	439.01	1.78E-97	0.991
13,228,196	439.01	1.78E-97	0.991
13,385,149	439.01	1.78E-97	0.991
14,270,068	439.01	1.78E-97	0.991
14,389,882	439.01	1.78E-97	0.991
14,880,654	439.01	1.78E-97	0.991

Physical Position	chi-square statistic	p-value	Weighted chi-square
15,880,123	439.01	1.78E-97	0.991
16,393,083	439.01	1.78E-97	0.991
16,597,960	439.01	1.78E-97	0.991
16,960,260	439.01	1.78E-97	0.991
16,994,254	439.01	1.78E-97	0.991
17,904,749	439.01	1.78E-97	0.991
17,952,800	439.01	1.78E-97	0.991
23,571,862	439.01	1.78E-97	0.991
26,090,126	439.01	1.78E-97	0.991
29,202,254	439.01	1.78E-97	0.991
29,347,341	439.01	1.78E-97	0.991
31,287,417	439.01	1.78E-97	0.991
32,307,508	439.01	1.78E-97	0.991
32,921,570	439.01	1.78E-97	0.991
36,689,727	439.01	1.78E-97	0.991
37,039,577	439.01	1.78E-97	0.991
37,234,356	439.01	1.78E-97	0.991
37,283,287	439.01	1.78E-97	0.991
37,403,706	439.01	1.78E-97	0.991
37,489,505	439.01	1.78E-97	0.991
37,492,262	439.01	1.78E-97	0.991
37,769,549	439.01	1.78E-97	0.991
38,141,288	439.01	1.78E-97	0.991
38,416,744	439.01	1.78E-97	0.991
38,687,865	439.01	1.78E-97	0.991
38,825,751	439.01	1.78E-97	0.991
38,842,925	439.01	1.78E-97	0.991
38,869,094	439.01	1.78E-97	0.991

Physical Position	chi-square statistic	p-value	Weighted chi-square
38,883,690	439.01	1.78E-97	0.991
38,899,511	439.01	1.78E-97	0.991
38,929,473	439.01	1.78E-97	0.991
40,875,768	439.01	1.78E-97	0.991
41,255,029	439.01	1.78E-97	0.991
42,471,351	439.01	1.78E-97	0.991
42,643,911	439.01	1.78E-97	0.991
43,699,104	439.01	1.78E-97	0.991
44,745,845	439.01	1.78E-97	0.991
45,650,419	439.01	1.78E-97	0.991
46,160,298	439.01	1.78E-97	0.991
46,768,665	439.01	1.78E-97	0.991
47,684,447	439.01	1.78E-97	0.991
47,853,986	439.01	1.78E-97	0.991
48,213,229	439.01	1.78E-97	0.991
48,509,170	439.01	1.78E-97	0.991
48,558,898	439.01	1.78E-97	0.991
48,667,534	439.01	1.78E-97	0.991
48,695,985	439.01	1.78E-97	0.991
54,411,607	439.01	1.78E-97	0.991
54,640,705	439.01	1.78E-97	0.991
58,554,385	439.01	1.78E-97	0.991
58,753,412	439.01	1.78E-97	0.991
59,846,869	439.01	1.78E-97	0.991
63,833,556	439.01	1.78E-97	0.991
63,969,433	439.01	1.78E-97	0.991
64,075,579	439.01	1.78E-97	0.991
64,289,610	439.01	1.78E-97	0.991

Physical Position	chi-square statistic	p-value	Weighted chi-square
6,226,449	439.00	4.70E-96	0.991
6,227,588	439.00	4.70E-96	0.991
6,352,763	439.00	4.70E-96	0.991
6,894,756	439.00	4.70E-96	0.991
6,905,963	439.00	4.70E-96	0.991
6,943,400	439.00	4.70E-96	0.991
7,147,626	439.00	4.70E-96	0.991
7,221,367	439.00	4.70E-96	0.991
7,427,834	439.00	4.70E-96	0.991
7,516,495	439.00	4.70E-96	0.991
7,719,710	439.00	4.70E-96	0.991
7,759,837	439.00	4.70E-96	0.991
7,812,562	439.00	4.70E-96	0.991
7,970,654	439.00	4.70E-96	0.991
8,059,391	439.00	4.70E-96	0.991
8,341,953	439.00	4.70E-96	0.991
8,709,642	439.00	4.70E-96	0.991
9,169,064	439.00	4.70E-96	0.991
9,422,685	439.00	4.70E-96	0.991
10,021,195	439.00	4.70E-96	0.991
10,428,766	439.00	4.70E-96	0.991
10,434,445	439.00	4.70E-96	0.991
10,469,740	439.00	4.70E-96	0.991
10,588,313	439.00	4.70E-96	0.991
11,702,004	439.00	4.70E-96	0.991
11,706,217	439.00	4.70E-96	0.991
12,001,215	439.00	4.70E-96	0.991
12,047,381	439.00	4.70E-96	0.991

Physical Position	chi-square statistic	p-value	Weighted chi-square
12,636,480	439.00	4.70E-96	0.991
12,636,690	439.00	4.70E-96	0.991
12,929,437	439.00	4.70E-96	0.991
13,149,175	439.00	4.70E-96	0.991
13,150,695	439.00	4.70E-96	0.991
13,538,506	439.00	4.70E-96	0.991
14,139,065	439.00	4.70E-96	0.991
14,718,796	439.00	4.70E-96	0.991
14,804,087	439.00	4.70E-96	0.991
15,422,754	439.00	4.70E-96	0.991
15,653,968	439.00	4.70E-96	0.991
15,717,450	439.00	4.70E-96	0.991
15,911,661	439.00	4.70E-96	0.991
15,957,118	439.00	4.70E-96	0.991
16,404,222	439.00	4.70E-96	0.991
16,521,021	439.00	4.70E-96	0.991
16,537,052	439.00	4.70E-96	0.991
16,570,879	439.00	4.70E-96	0.991
17,185,501	439.00	4.70E-96	0.991
18,515,673	439.00	4.70E-96	0.991
20,077,133	439.00	4.70E-96	0.991
26,234,684	439.00	4.70E-96	0.991
28,658,990	439.00	4.70E-96	0.991
28,787,835	439.00	4.70E-96	0.991
29,767,034	439.00	4.70E-96	0.991
31,602,576	439.00	4.70E-96	0.991
35,606,300	439.00	4.70E-96	0.991
36,254,438	439.00	4.70E-96	0.991

Physical Position	chi-square statistic	p-value	Weighted chi-square
37,769,223	439.00	4.70E-96	0.991
38,172,076	439.00	4.70E-96	0.991
39,281,712	439.00	4.70E-96	0.991
40,193,048	439.00	4.70E-96	0.991
41,355,181	439.00	4.70E-96	0.991
42,770,137	439.00	4.70E-96	0.991
44,325,177	439.00	4.70E-96	0.991
47,794,660	439.00	4.70E-96	0.991
47,844,802	439.00	4.70E-96	0.991
51,010,922	439.00	4.70E-96	0.991
52,135,504	439.00	4.70E-96	0.991
52,320,191	439.00	4.70E-96	0.991
55,613,969	439.00	4.70E-96	0.991
60,207,640	439.00	4.70E-96	0.991
60,250,851	439.00	4.70E-96	0.991
63,519,758	439.00	4.70E-96	0.991
65,348,726	439.00	4.70E-96	0.991
68,884,033	439.00	4.70E-96	0.991
8,266,959	439.00	4.70E-96	0.991
14,865,809	439.00	4.70E-96	0.991
16,230,766	439.00	4.70E-96	0.991
48,372,323	439.00	4.70E-96	0.991
37,764,646	439.00	4.70E-96	0.991
9,438,868	439.00	4.70E-96	0.991
6,283,074	438.80	5.20E-96	0.991
7,082,658	438.67	5.56E-96	0.990
7,953,100	438.67	5.56E-96	0.990
12,926,943	438.67	5.56E-96	0.990

Physical Position	chi-square statistic	p-value	Weighted chi-square
14,539,688	438.57	5.83E-96	0.990
15,366,630	438.57	5.83E-96	0.990
60,187,515	438.36	6.46E-96	0.990
10,217,493	438.33	6.56E-96	0.989
14,635,480	438.33	6.56E-96	0.989
14,874,029	438.33	6.56E-96	0.989
14,901,418	438.33	6.56E-96	0.989
9,823,644	438.33	6.56E-96	0.989
11,152,731	438.33	6.56E-96	0.989
14,793,260	438.33	6.56E-96	0.989
14,853,848	438.33	6.56E-96	0.989
32,377,583	438.33	6.56E-96	0.989
7,099,723	438.02	7.68E-96	0.989
9,213,019	438.02	7.68E-96	0.989
13,534,647	438.02	7.68E-96	0.989
13,963,399	438.02	7.68E-96	0.989
14,050,014	438.02	7.68E-96	0.989
15,086,213	438.02	7.68E-96	0.989
15,534,459	438.02	7.68E-96	0.989
15,553,299	438.02	7.68E-96	0.989
59,944,570	438.02	7.68E-96	0.989
10,635,608	438.02	7.69E-96	0.989
5,927,066	438.01	2.94E-97	0.989
6,134,943	438.01	2.94E-97	0.989
6,585,073	438.01	2.94E-97	0.989
6,626,315	438.01	2.94E-97	0.989
6,872,106	438.01	2.94E-97	0.989
7,091,869	438.01	2.94E-97	0.989

Physical Position	chi-square statistic	p-value	Weighted chi-square
7,133,035	438.01	2.94E-97	0.989
7,209,189	438.01	2.94E-97	0.989
7,219,413	438.01	2.94E-97	0.989
7,286,930	438.01	2.94E-97	0.989
7,409,149	438.01	2.94E-97	0.989
7,433,603	438.01	2.94E-97	0.989
7,477,462	438.01	2.94E-97	0.989
7,503,174	438.01	2.94E-97	0.989
7,660,094	438.01	2.94E-97	0.989
7,943,551	438.01	2.94E-97	0.989
8,630,987	438.01	2.94E-97	0.989
8,785,762	438.01	2.94E-97	0.989
8,959,132	438.01	2.94E-97	0.989
8,990,468	438.01	2.94E-97	0.989
9,360,723	438.01	2.94E-97	0.989
9,612,116	438.01	2.94E-97	0.989
9,681,582	438.01	2.94E-97	0.989
10,028,107	438.01	2.94E-97	0.989
10,398,392	438.01	2.94E-97	0.989
10,730,674	438.01	2.94E-97	0.989
11,737,647	438.01	2.94E-97	0.989
13,058,984	438.01	2.94E-97	0.989
13,322,374	438.01	2.94E-97	0.989
13,412,121	438.01	2.94E-97	0.989
13,515,877	438.01	2.94E-97	0.989
13,926,610	438.01	2.94E-97	0.989
14,140,491	438.01	2.94E-97	0.989
14,446,149	438.01	2.94E-97	0.989

Physical Position	chi-square statistic	p-value	Weighted chi-square
14,647,779	438.01	2.94E-97	0.989
14,941,392	438.01	2.94E-97	0.989
15,203,932	438.01	2.94E-97	0.989
15,684,180	438.01	2.94E-97	0.989
15,779,448	438.01	2.94E-97	0.989
16,333,925	438.01	2.94E-97	0.989
16,785,232	438.01	2.94E-97	0.989
16,865,992	438.01	2.94E-97	0.989
16,943,072	438.01	2.94E-97	0.989
17,013,119	438.01	2.94E-97	0.989
17,097,179	438.01	2.94E-97	0.989
18,028,954	438.01	2.94E-97	0.989
18,082,877	438.01	2.94E-97	0.989
18,102,640	438.01	2.94E-97	0.989
20,032,394	438.01	2.94E-97	0.989
22,734,306	438.01	2.94E-97	0.989
23,676,159	438.01	2.94E-97	0.989
23,801,524	438.01	2.94E-97	0.989
26,208,051	438.01	2.94E-97	0.989
28,310,320	438.01	2.94E-97	0.989
30,263,329	438.01	2.94E-97	0.989
31,203,476	438.01	2.94E-97	0.989
35,368,634	438.01	2.94E-97	0.989
35,773,135	438.01	2.94E-97	0.989
36,213,320	438.01	2.94E-97	0.989
36,504,428	438.01	2.94E-97	0.989
37,263,912	438.01	2.94E-97	0.989
37,308,227	438.01	2.94E-97	0.989

Physical Position	chi-square statistic	p-value	Weighted chi-square
37,453,890	438.01	2.94E-97	0.989
37,555,904	438.01	2.94E-97	0.989
37,616,567	438.01	2.94E-97	0.989
37,968,351	438.01	2.94E-97	0.989
38,036,197	438.01	2.94E-97	0.989
38,187,815	438.01	2.94E-97	0.989
38,194,429	438.01	2.94E-97	0.989
38,403,903	438.01	2.94E-97	0.989
38,652,657	438.01	2.94E-97	0.989
38,761,652	438.01	2.94E-97	0.989
38,893,654	438.01	2.94E-97	0.989
38,960,816	438.01	2.94E-97	0.989
39,447,176	438.01	2.94E-97	0.989
39,554,314	438.01	2.94E-97	0.989
40,747,268	438.01	2.94E-97	0.989
41,621,743	438.01	2.94E-97	0.989
43,602,383	438.01	2.94E-97	0.989
43,864,979	438.01	2.94E-97	0.989
44,614,188	438.01	2.94E-97	0.989
45,734,561	438.01	2.94E-97	0.989
47,978,339	438.01	2.94E-97	0.989
48,430,131	438.01	2.94E-97	0.989
50,137,492	438.01	2.94E-97	0.989
52,333,760	438.01	2.94E-97	0.989
52,356,137	438.01	2.94E-97	0.989
52,363,718	438.01	2.94E-97	0.989
55,482,632	438.01	2.94E-97	0.989
55,619,362	438.01	2.94E-97	0.989

Physical Position	chi-square statistic	p-value	Weighted chi-square
57,383,326	438.01	2.94E-97	0.989
58,489,086	438.01	2.94E-97	0.989
60,159,298	438.01	2.94E-97	0.989
63,501,397	438.01	2.94E-97	0.989
63,619,150	438.01	2.94E-97	0.989
64,072,402	438.01	2.94E-97	0.989
64,860,354	438.01	2.94E-97	0.989
64,880,594	438.01	2.94E-97	0.989
67,216,608	438.01	2.94E-97	0.989
6,730,056	438.01	2.94E-97	0.989
6,753,792	438.01	2.94E-97	0.989
8,144,128	438.01	2.94E-97	0.989
8,233,061	438.01	2.94E-97	0.989
9,625,112	438.01	2.94E-97	0.989
10,071,364	438.01	2.94E-97	0.989
11,649,218	438.01	2.94E-97	0.989
11,831,493	438.01	2.94E-97	0.989
12,009,714	438.01	2.94E-97	0.989
12,880,149	438.01	2.94E-97	0.989
13,030,127	438.01	2.94E-97	0.989
14,442,338	438.01	2.94E-97	0.989
15,765,035	438.01	2.94E-97	0.989
16,511,325	438.01	2.94E-97	0.989
16,588,478	438.01	2.94E-97	0.989
16,726,366	438.01	2.94E-97	0.989
16,770,729	438.01	2.94E-97	0.989
17,955,491	438.01	2.94E-97	0.989
18,971,443	438.01	2.94E-97	0.989

Physical Position	chi-square statistic	p-value	Weighted chi-square
22,426,270	438.01	2.94E-97	0.989
29,225,896	438.01	2.94E-97	0.989
31,211,207	438.01	2.94E-97	0.989
32,816,216	438.01	2.94E-97	0.989
33,983,274	438.01	2.94E-97	0.989
38,861,143	438.01	2.94E-97	0.989
42,464,308	438.01	2.94E-97	0.989
43,746,519	438.01	2.94E-97	0.989
44,620,411	438.01	2.94E-97	0.989
44,934,695	438.01	2.94E-97	0.989
45,286,512	438.01	2.94E-97	0.989
47,709,158	438.01	2.94E-97	0.989
47,711,942	438.01	2.94E-97	0.989
48,190,987	438.01	2.94E-97	0.989
55,609,854	438.01	2.94E-97	0.989
55,627,501	438.01	2.94E-97	0.989
56,610,878	438.01	2.94E-97	0.989
58,126,172	438.01	2.94E-97	0.989
59,667,896	438.01	2.94E-97	0.989
59,855,255	438.01	2.94E-97	0.989
60,264,951	438.01	2.94E-97	0.989
65,252,835	438.01	2.94E-97	0.989
5,949,790	438.00	7.75E-96	0.989
6,197,412	438.00	7.75E-96	0.989
6,254,538	438.00	7.75E-96	0.989
7,090,331	438.00	7.75E-96	0.989
7,181,437	438.00	7.75E-96	0.989
7,266,073	438.00	7.75E-96	0.989

Physical Position	chi-square statistic	p-value	Weighted chi-square
7,325,542	438.00	7.75E-96	0.989
7,543,460	438.00	7.75E-96	0.989
7,624,289	438.00	7.75E-96	0.989
8,053,221	438.00	7.75E-96	0.989
8,191,091	438.00	7.75E-96	0.989
9,618,647	438.00	7.75E-96	0.989
9,743,381	438.00	7.75E-96	0.989
9,879,867	438.00	7.75E-96	0.989
10,015,144	438.00	7.75E-96	0.989
10,113,548	438.00	7.75E-96	0.989
10,260,025	438.00	7.75E-96	0.989
11,126,297	438.00	7.75E-96	0.989
11,368,251	438.00	7.75E-96	0.989
11,595,909	438.00	7.75E-96	0.989
11,620,838	438.00	7.75E-96	0.989
12,044,761	438.00	7.75E-96	0.989
12,634,688	438.00	7.75E-96	0.989
13,400,729	438.00	7.75E-96	0.989
13,888,375	438.00	7.75E-96	0.989
14,044,640	438.00	7.75E-96	0.989
14,490,678	438.00	7.75E-96	0.989
14,666,028	438.00	7.75E-96	0.989
15,263,953	438.00	7.75E-96	0.989
15,365,819	438.00	7.75E-96	0.989
15,888,476	438.00	7.75E-96	0.989
16,056,447	438.00	7.75E-96	0.989
16,636,143	438.00	7.75E-96	0.989
16,705,635	438.00	7.75E-96	0.989

Physical Position	chi-square statistic	p-value	Weighted chi-square
16,886,609	438.00	7.75E-96	0.989
17,130,283	438.00	7.75E-96	0.989
29,250,516	438.00	7.75E-96	0.989
36,802,855	438.00	7.75E-96	0.989
36,939,887	438.00	7.75E-96	0.989
37,971,896	438.00	7.75E-96	0.989
38,639,590	438.00	7.75E-96	0.989
39,995,001	438.00	7.75E-96	0.989
42,739,943	438.00	7.75E-96	0.989
42,831,075	438.00	7.75E-96	0.989
43,680,565	438.00	7.75E-96	0.989
47,779,113	438.00	7.75E-96	0.989
47,830,101	438.00	7.75E-96	0.989
48,592,570	438.00	7.75E-96	0.989
48,908,956	438.00	7.75E-96	0.989
56,565,291	438.00	7.75E-96	0.989
57,791,301	438.00	7.75E-96	0.989
7,209,691	438.00	7.75E-96	0.989
14,664,339	438.00	7.75E-96	0.989
16,386,368	438.00	7.75E-96	0.989
16,697,514	438.00	7.75E-96	0.989
7,557,110	438.00	7.75E-96	0.989
10,618,411	438.00	7.75E-96	0.989
11,945,817	438.00	7.75E-96	0.989
18,037,707	438.00	7.75E-96	0.989
42,911,109	438.00	7.75E-96	0.989
62,607,913	438.00	7.75E-96	0.989
11,098,889	438.00	7.75E-96	0.989

Physical Position	chi-square statistic	p-value	Weighted chi-square
19,889,432	438.00	7.76E-96	0.989
15,664,248	437.80	8.57E-96	0.988
16,938,557	437.80	8.57E-96	0.988
16,643,395	437.67	9.16E-96	0.988
14,942,989	437.67	9.16E-96	0.988
10,415,394	437.57	9.61E-96	0.988
7,389,120	437.50	9.96E-96	0.988
12,949,843	437.33	1.08E-95	0.987
47,870,401	437.33	1.08E-95	0.987
8,412,437	437.33	1.08E-95	0.987
13,557,466	437.33	1.08E-95	0.987
16,573,851	437.33	1.08E-95	0.987
17,837,069	437.33	1.08E-95	0.987
18,065,373	437.33	1.08E-95	0.987
25,387,972	437.33	1.08E-95	0.987
47,827,412	437.33	1.08E-95	0.987
6,254,014	437.33	1.08E-95	0.987
6,719,067	437.33	1.08E-95	0.987
8,761,809	437.33	1.08E-95	0.987
16,489,920	437.33	1.08E-95	0.987
28,951,894	437.33	1.08E-95	0.987
42,497,974	437.33	1.08E-95	0.987
49,890,397	437.33	1.08E-95	0.987
60,239,459	437.31	1.10E-95	0.987
10,708,091	437.31	1.10E-95	0.987
18,066,380	437.20	1.16E-95	0.987
64,269,483	437.20	1.16E-95	0.987
12,432,300	437.09	1.22E-95	0.987

Physical Position	chi-square statistic	p-value	Weighted chi-square
9,186,413	437.08	1.23E-95	0.987
6,650,612	437.02	1.27E-95	0.986
8,893,348	437.02	1.27E-95	0.986
15,524,850	437.02	1.27E-95	0.986
15,559,201	437.02	1.27E-95	0.986
9,596,960	437.02	1.27E-95	0.986
13,961,628	437.02	1.27E-95	0.986
15,536,759	437.02	1.27E-95	0.986
6,012,273	437.01	4.85E-97	0.986
6,096,870	437.01	4.85E-97	0.986
6,176,084	437.01	4.85E-97	0.986
6,757,559	437.01	4.85E-97	0.986
7,136,488	437.01	4.85E-97	0.986
7,323,654	437.01	4.85E-97	0.986
7,698,427	437.01	4.85E-97	0.986
9,775,244	437.01	4.85E-97	0.986
10,041,713	437.01	4.85E-97	0.986
10,375,912	437.01	4.85E-97	0.986
10,573,835	437.01	4.85E-97	0.986
13,376,466	437.01	4.85E-97	0.986
13,475,630	437.01	4.85E-97	0.986
13,906,475	437.01	4.85E-97	0.986
14,647,750	437.01	4.85E-97	0.986
14,718,224	437.01	4.85E-97	0.986
15,023,297	437.01	4.85E-97	0.986
16,303,165	437.01	4.85E-97	0.986
16,619,329	437.01	4.85E-97	0.986
16,732,066	437.01	4.85E-97	0.986

Physical Position	chi-square statistic	p-value	Weighted chi-square
31,465,133	437.01	4.85E-97	0.986
32,500,667	437.01	4.85E-97	0.986
32,844,906	437.01	4.85E-97	0.986
33,052,779	437.01	4.85E-97	0.986
35,029,065	437.01	4.85E-97	0.986
35,088,506	437.01	4.85E-97	0.986
36,710,230	437.01	4.85E-97	0.986
37,297,483	437.01	4.85E-97	0.986
37,389,330	437.01	4.85E-97	0.986
37,391,525	437.01	4.85E-97	0.986
37,636,437	437.01	4.85E-97	0.986
38,124,600	437.01	4.85E-97	0.986
38,322,864	437.01	4.85E-97	0.986
38,756,769	437.01	4.85E-97	0.986
38,757,200	437.01	4.85E-97	0.986
38,879,120	437.01	4.85E-97	0.986
39,798,953	437.01	4.85E-97	0.986
41,579,313	437.01	4.85E-97	0.986
43,560,195	437.01	4.85E-97	0.986
45,489,629	437.01	4.85E-97	0.986
48,025,408	437.01	4.85E-97	0.986
48,220,918	437.01	4.85E-97	0.986
48,246,184	437.01	4.85E-97	0.986
48,320,612	437.01	4.85E-97	0.986
48,505,144	437.01	4.85E-97	0.986
49,313,696	437.01	4.85E-97	0.986
55,482,435	437.01	4.85E-97	0.986
56,864,720	437.01	4.85E-97	0.986

Physical Position	chi-square statistic	p-value	Weighted chi-square
57,714,415	437.01	4.85E-97	0.986
60,103,826	437.01	4.85E-97	0.986
60,263,483	437.01	4.85E-97	0.986
64,235,680	437.01	4.85E-97	0.986
6,028,045	437.01	4.85E-97	0.986
6,032,726	437.01	4.85E-97	0.986
7,189,083	437.01	4.85E-97	0.986
7,268,976	437.01	4.85E-97	0.986
7,617,236	437.01	4.85E-97	0.986
7,625,870	437.01	4.85E-97	0.986
8,984,270	437.01	4.85E-97	0.986
9,653,103	437.01	4.85E-97	0.986
13,427,369	437.01	4.85E-97	0.986
14,914,647	437.01	4.85E-97	0.986
15,627,993	437.01	4.85E-97	0.986
15,663,262	437.01	4.85E-97	0.986
15,944,561	437.01	4.85E-97	0.986
16,621,178	437.01	4.85E-97	0.986
16,692,488	437.01	4.85E-97	0.986
19,949,068	437.01	4.85E-97	0.986
24,194,621	437.01	4.85E-97	0.986
25,109,299	437.01	4.85E-97	0.986
25,278,250	437.01	4.85E-97	0.986
27,548,784	437.01	4.85E-97	0.986
28,871,157	437.01	4.85E-97	0.986
32,759,321	437.01	4.85E-97	0.986
36,465,934	437.01	4.85E-97	0.986
38,107,627	437.01	4.85E-97	0.986

Physical Position	chi-square statistic	p-value	Weighted chi-square
38,293,781	437.01	4.85E-97	0.986
39,025,009	437.01	4.85E-97	0.986
42,662,841	437.01	4.85E-97	0.986
43,736,348	437.01	4.85E-97	0.986
48,439,590	437.01	4.85E-97	0.986
48,953,533	437.01	4.85E-97	0.986
52,316,759	437.01	4.85E-97	0.986
52,613,793	437.01	4.85E-97	0.986
55,685,837	437.01	4.85E-97	0.986
64,377,192	437.01	4.85E-97	0.986
65,688,896	437.01	4.85E-97	0.986
6,555,954	437.01	4.85E-97	0.986
6,909,045	437.01	4.85E-97	0.986
7,096,363	437.01	4.85E-97	0.986
8,606,223	437.01	4.85E-97	0.986
9,249,328	437.01	4.85E-97	0.986
9,310,848	437.01	4.85E-97	0.986
9,592,257	437.01	4.85E-97	0.986
14,477,651	437.01	4.85E-97	0.986
15,025,568	437.01	4.85E-97	0.986
15,730,041	437.01	4.85E-97	0.986
17,055,645	437.01	4.85E-97	0.986
26,203,597	437.01	4.85E-97	0.986
28,958,928	437.01	4.85E-97	0.986
36,636,753	437.01	4.85E-97	0.986
38,067,843	437.01	4.85E-97	0.986
44,249,267	437.01	4.85E-97	0.986
52,344,785	437.01	4.85E-97	0.986

Physical Position	chi-square statistic	p-value	Weighted chi-square
62,620,775	437.01	4.85E-97	0.986
64,274,081	437.01	4.85E-97	0.986
65,435,760	437.01	4.85E-97	0.986
5,930,658	437.00	1.28E-95	0.986
5,951,611	437.00	1.28E-95	0.986
6,068,065	437.00	1.28E-95	0.986
6,554,297	437.00	1.28E-95	0.986
6,861,413	437.00	1.28E-95	0.986
8,928,941	437.00	1.28E-95	0.986
9,310,504	437.00	1.28E-95	0.986
9,713,078	437.00	1.28E-95	0.986
10,532,464	437.00	1.28E-95	0.986
11,376,028	437.00	1.28E-95	0.986
11,419,425	437.00	1.28E-95	0.986
11,440,167	437.00	1.28E-95	0.986
11,658,703	437.00	1.28E-95	0.986
13,491,351	437.00	1.28E-95	0.986
14,190,552	437.00	1.28E-95	0.986
14,911,455	437.00	1.28E-95	0.986
15,324,851	437.00	1.28E-95	0.986
32,526,887	437.00	1.28E-95	0.986
37,465,211	437.00	1.28E-95	0.986
37,610,409	437.00	1.28E-95	0.986
38,720,544	437.00	1.28E-95	0.986
43,312,571	437.00	1.28E-95	0.986
44,513,098	437.00	1.28E-95	0.986
45,008,398	437.00	1.28E-95	0.986
45,730,520	437.00	1.28E-95	0.986

Physical Position	chi-square statistic	p-value	Weighted chi-square
48,045,396	437.00	1.28E-95	0.986
48,323,820	437.00	1.28E-95	0.986
58,657,946	437.00	1.28E-95	0.986
59,836,727	437.00	1.28E-95	0.986
59,997,810	437.00	1.28E-95	0.986
63,830,539	437.00	1.28E-95	0.986
67,311,640	437.00	1.28E-95	0.986
9,318,621	437.00	1.28E-95	0.986
11,788,078	437.00	1.28E-95	0.986
36,533,645	437.00	1.28E-95	0.986
8,071,603	437.00	1.28E-95	0.986
45,133,862	437.00	1.28E-95	0.986
14,273,661	437.00	1.28E-95	0.986
33,673,260	437.00	1.28E-95	0.986
52,749,793	437.00	1.28E-95	0.986
42,726,604	436.80	1.41E-95	0.986
12,026,269	436.80	1.41E-95	0.986
16,432,390	436.80	1.41E-95	0.986
38,918,130	436.67	1.51E-95	0.986
8,790,850	436.67	1.51E-95	0.986
6,571,622	436.57	1.58E-95	0.985
33,653,047	436.50	1.64E-95	0.985
6,916,873	436.40	1.73E-95	0.985
6,471,262	436.33	1.78E-95	0.985
6,536,360	436.33	1.78E-95	0.985
8,163,215	436.33	1.78E-95	0.985
8,386,745	436.33	1.78E-95	0.985
10,920,423	436.33	1.78E-95	0.985

Physical Position	chi-square statistic	p-value	Weighted chi-square
11,789,513	436.33	1.78E-95	0.985
34,779,202	436.33	1.78E-95	0.985
48,981,258	436.33	1.78E-95	0.985
11,113,690	436.20	1.91E-95	0.985
16,184,893	436.08	2.03E-95	0.984
6,819,968	436.02	2.09E-95	0.984
13,991,640	436.02	2.09E-95	0.984
16,603,012	436.02	2.09E-95	0.984
42,468,910	436.02	2.09E-95	0.984
13,994,104	436.02	2.09E-95	0.984
15,458,552	436.02	2.09E-95	0.984
64,012,306	436.02	2.09E-95	0.984
5,965,116	436.01	8.00E-97	0.984
6,029,164	436.01	8.00E-97	0.984
6,150,331	436.01	8.00E-97	0.984
6,410,421	436.01	8.00E-97	0.984
7,057,813	436.01	8.00E-97	0.984
9,650,408	436.01	8.00E-97	0.984
9,916,868	436.01	8.00E-97	0.984
10,249,702	436.01	8.00E-97	0.984
10,615,913	436.01	8.00E-97	0.984
10,697,005	436.01	8.00E-97	0.984
12,620,886	436.01	8.00E-97	0.984
13,338,347	436.01	8.00E-97	0.984
14,246,795	436.01	8.00E-97	0.984
15,580,942	436.01	8.00E-97	0.984
15,838,010	436.01	8.00E-97	0.984
16,624,858	436.01	8.00E-97	0.984

Physical Position	chi-square statistic	p-value	Weighted chi-square
17,045,082	436.01	8.00E-97	0.984
26,931,464	436.01	8.00E-97	0.984
32,427,006	436.01	8.00E-97	0.984
36,480,776	436.01	8.00E-97	0.984
37,027,550	436.01	8.00E-97	0.984
37,969,430	436.01	8.00E-97	0.984
38,660,726	436.01	8.00E-97	0.984
38,712,145	436.01	8.00E-97	0.984
40,705,366	436.01	8.00E-97	0.984
45,344,845	436.01	8.00E-97	0.984
45,733,372	436.01	8.00E-97	0.984
45,935,124	436.01	8.00E-97	0.984
48,425,665	436.01	8.00E-97	0.984
54,797,352	436.01	8.00E-97	0.984
57,658,355	436.01	8.00E-97	0.984
57,718,072	436.01	8.00E-97	0.984
59,846,569	436.01	8.00E-97	0.984
5,925,283	436.01	8.00E-97	0.984
6,379,052	436.01	8.00E-97	0.984
8,718,029	436.01	8.00E-97	0.984
8,935,264	436.01	8.00E-97	0.984
11,887,578	436.01	8.00E-97	0.984
14,691,709	436.01	8.00E-97	0.984
25,257,105	436.01	8.00E-97	0.984
31,438,594	436.01	8.00E-97	0.984
38,650,522	436.01	8.00E-97	0.984
43,619,506	436.01	8.00E-97	0.984
43,894,444	436.01	8.00E-97	0.984

Physical Position	chi-square statistic	p-value	Weighted chi-square
48,650,302	436.01	8.00E-97	0.984
48,726,070	436.01	8.00E-97	0.984
49,056,461	436.01	8.00E-97	0.984
65,878,389	436.01	8.00E-97	0.984
66,996,844	436.01	8.00E-97	0.984
7,178,255	436.01	8.00E-97	0.984
7,889,312	436.01	8.00E-97	0.984
8,940,445	436.01	8.00E-97	0.984
10,248,857	436.01	8.00E-97	0.984
14,907,907	436.01	8.00E-97	0.984
17,091,332	436.01	8.00E-97	0.984
38,736,901	436.01	8.00E-97	0.984
45,364,072	436.01	8.00E-97	0.984
52,330,306	436.01	8.00E-97	0.984
61,212,120	436.01	8.00E-97	0.984
7,178,427	436.01	8.00E-97	0.984
17,054,004	436.01	8.00E-97	0.984
5,996,734	436.00	2.11E-95	0.984
7,479,286	436.00	2.11E-95	0.984
8,022,949	436.00	2.11E-95	0.984
8,126,707	436.00	2.11E-95	0.984
8,465,159	436.00	2.11E-95	0.984
9,634,281	436.00	2.11E-95	0.984
10,223,857	436.00	2.11E-95	0.984
10,371,149	436.00	2.11E-95	0.984
11,020,890	436.00	2.11E-95	0.984
13,350,136	436.00	2.11E-95	0.984
13,393,654	436.00	2.11E-95	0.984

Physical Position	chi-square statistic	p-value	Weighted chi-square
13,812,707	436.00	2.11E-95	0.984
14,008,068	436.00	2.11E-95	0.984
14,192,472	436.00	2.11E-95	0.984
14,329,403	436.00	2.11E-95	0.984
14,396,949	436.00	2.11E-95	0.984
14,847,837	436.00	2.11E-95	0.984
15,727,910	436.00	2.11E-95	0.984
15,745,570	436.00	2.11E-95	0.984
16,494,578	436.00	2.11E-95	0.984
17,909,534	436.00	2.11E-95	0.984
27,071,226	436.00	2.11E-95	0.984
27,133,048	436.00	2.11E-95	0.984
35,632,292	436.00	2.11E-95	0.984
37,287,170	436.00	2.11E-95	0.984
38,709,977	436.00	2.11E-95	0.984
41,868,367	436.00	2.11E-95	0.984
43,602,169	436.00	2.11E-95	0.984
59,607,549	436.00	2.11E-95	0.984
60,418,377	436.00	2.11E-95	0.984
61,760,034	436.00	2.11E-95	0.984
63,720,754	436.00	2.11E-95	0.984
9,569,217	436.00	2.11E-95	0.984
9,636,356	436.00	2.11E-95	0.984
11,795,120	436.00	2.11E-95	0.984
15,270,087	436.00	2.11E-95	0.984
35,155,676	436.00	2.11E-95	0.984
14,747,110	436.00	2.11E-95	0.984
11,803,939	436.00	2.11E-95	0.984

Physical Position	chi-square statistic	p-value	Weighted chi-square
8,658,013	436.00	2.11E-95	0.984
7,068,602	435.80	2.33E-95	0.984
10,529,949	435.80	2.33E-95	0.984
11,078,809	435.80	2.33E-95	0.984
16,802,441	435.80	2.33E-95	0.984
48,024,757	435.78	2.36E-95	0.984
9,444,599	435.67	2.49E-95	0.983
15,845,170	435.67	2.49E-95	0.983
42,079,649	435.67	2.49E-95	0.983
13,917,568	435.36	2.90E-95	0.983
16,594,645	435.35	2.92E-95	0.983
5,968,025	435.33	2.94E-95	0.983
8,575,370	435.33	2.94E-95	0.983
14,734,016	435.33	2.94E-95	0.983
29,204,457	435.33	2.94E-95	0.983
44,163,699	435.33	2.94E-95	0.983
16,337,181	435.33	2.94E-95	0.983
12,560,202	435.33	2.94E-95	0.983
9,377,155	435.08	3.34E-95	0.982
6,563,231	435.07	3.35E-95	0.982
10,140,965	435.07	3.35E-95	0.982
11,070,388	435.07	3.35E-95	0.982
11,207,440	435.07	3.35E-95	0.982
13,947,503	435.07	3.35E-95	0.982
14,953,874	435.07	3.35E-95	0.982
15,387,433	435.07	3.35E-95	0.982
16,544,647	435.07	3.35E-95	0.982
16,547,718	435.07	3.35E-95	0.982

Physical Position	chi-square statistic	p-value	Weighted chi-square
18,124,083	435.07	3.35E-95	0.982
36,065,878	435.07	3.35E-95	0.982
36,396,676	435.07	3.35E-95	0.982
37,399,866	435.07	3.35E-95	0.982
14,245,380	435.02	3.44E-95	0.982
7,451,908	435.02	3.44E-95	0.982
44,197,838	435.02	3.44E-95	0.982
10,048,702	435.02	3.44E-95	0.982
15,504,005	435.02	3.44E-95	0.982
14,247,128	435.02	3.44E-95	0.982
13,297,409	435.02	3.44E-95	0.982
14,267,144	435.02	3.45E-95	0.982
6,200,910	435.01	1.32E-96	0.982
6,990,850	435.01	1.32E-96	0.982
7,003,990	435.01	1.32E-96	0.982
7,099,123	435.01	1.32E-96	0.982
7,408,959	435.01	1.32E-96	0.982
8,725,767	435.01	1.32E-96	0.982
9,422,983	435.01	1.32E-96	0.982
9,536,516	435.01	1.32E-96	0.982
9,618,097	435.01	1.32E-96	0.982
12,878,328	435.01	1.32E-96	0.982
13,560,399	435.01	1.32E-96	0.982
13,572,413	435.01	1.32E-96	0.982
15,020,952	435.01	1.32E-96	0.982
15,030,618	435.01	1.32E-96	0.982
15,781,275	435.01	1.32E-96	0.982
16,535,733	435.01	1.32E-96	0.982

Physical Position	chi-square statistic	p-value	Weighted chi-square
16,737,120	435.01	1.32E-96	0.982
16,822,040	435.01	1.32E-96	0.982
16,870,913	435.01	1.32E-96	0.982
25,807,192	435.01	1.32E-96	0.982
27,056,805	435.01	1.32E-96	0.982
30,628,026	435.01	1.32E-96	0.982
31,221,781	435.01	1.32E-96	0.982
32,623,136	435.01	1.32E-96	0.982
32,627,935	435.01	1.32E-96	0.982
37,660,710	435.01	1.32E-96	0.982
37,744,791	435.01	1.32E-96	0.982
38,363,296	435.01	1.32E-96	0.982
38,523,154	435.01	1.32E-96	0.982
39,592,137	435.01	1.32E-96	0.982
41,582,810	435.01	1.32E-96	0.982
44,628,276	435.01	1.32E-96	0.982
45,661,162	435.01	1.32E-96	0.982
48,473,938	435.01	1.32E-96	0.982
48,582,162	435.01	1.32E-96	0.982
62,875,513	435.01	1.32E-96	0.982
67,221,906	435.01	1.32E-96	0.982
5,950,163	435.01	1.32E-96	0.982
6,984,681	435.01	1.32E-96	0.982
8,247,673	435.01	1.32E-96	0.982
8,469,394	435.01	1.32E-96	0.982
9,330,892	435.01	1.32E-96	0.982
15,574,560	435.01	1.32E-96	0.982
15,656,479	435.01	1.32E-96	0.982

Physical Position	chi-square statistic	p-value	Weighted chi-square
15,979,400	435.01	1.32E-96	0.982
38,747,448	435.01	1.32E-96	0.982
38,860,333	435.01	1.32E-96	0.982
67,104,453	435.01	1.32E-96	0.982
8,466,892	435.01	1.32E-96	0.982
9,497,105	435.01	1.32E-96	0.982
12,549,597	435.01	1.32E-96	0.982
14,406,533	435.01	1.32E-96	0.982
15,716,768	435.01	1.32E-96	0.982
42,498,723	435.01	1.32E-96	0.982
62,846,491	435.01	1.32E-96	0.982
10,507,066	435.01	1.32E-96	0.982
11,621,631	435.01	1.32E-96	0.982
17,052,920	435.01	1.32E-96	0.982
6,231,085	435.00	3.47E-95	0.982
7,091,537	435.00	3.47E-95	0.982
7,174,255	435.00	3.47E-95	0.982
7,418,713	435.00	3.47E-95	0.982
7,432,504	435.00	3.47E-95	0.982
8,043,725	435.00	3.47E-95	0.982
8,163,465	435.00	3.47E-95	0.982
9,536,971	435.00	3.47E-95	0.982
9,739,745	435.00	3.47E-95	0.982
9,967,629	435.00	3.47E-95	0.982
10,963,235	435.00	3.47E-95	0.982
11,873,991	435.00	3.47E-95	0.982
11,945,664	435.00	3.47E-95	0.982
11,986,275	435.00	3.47E-95	0.982

Physical Position	chi-square statistic	p-value	Weighted chi-square
12,313,884	435.00	3.47E-95	0.982
14,027,136	435.00	3.47E-95	0.982
16,683,181	435.00	3.47E-95	0.982
17,114,761	435.00	3.47E-95	0.982
17,832,633	435.00	3.47E-95	0.982
35,286,019	435.00	3.47E-95	0.982
41,769,131	435.00	3.47E-95	0.982
44,885,099	435.00	3.47E-95	0.982
48,388,786	435.00	3.47E-95	0.982
48,537,255	435.00	3.47E-95	0.982
8,917,271	435.00	3.47E-95	0.982
14,325,390	435.00	3.48E-95	0.982
11,413,299	435.00	3.48E-95	0.982
37,991,798	435.00	3.48E-95	0.982
7,669,305	434.82	3.81E-95	0.982
9,300,481	434.80	3.84E-95	0.981
38,413,985	434.80	3.84E-95	0.981
6,782,815	434.67	4.11E-95	0.981
12,818,919	434.67	4.11E-95	0.981
26,844,231	434.67	4.11E-95	0.981
8,874,034	434.67	4.11E-95	0.981
6,699,671	434.67	4.11E-95	0.981
13,189,749	434.57	4.31E-95	0.981
49,103,516	434.57	4.31E-95	0.981
14,527,069	434.45	4.56E-95	0.981
6,253,853	434.33	4.85E-95	0.980
13,829,489	434.33	4.85E-95	0.980
14,630,687	434.33	4.85E-95	0.980

Physical Position	chi-square statistic	p-value	Weighted chi-square
47,715,130	434.33	4.85E-95	0.980
60,037,161	434.33	4.85E-95	0.980
7,481,733	434.33	4.85E-95	0.980
16,440,041	434.33	4.85E-95	0.980
41,775,137	434.33	4.85E-95	0.980
47,879,755	434.33	4.85E-95	0.980
11,806,679	434.33	4.85E-95	0.980
10,602,815	434.29	4.97E-95	0.980
14,849,423	434.20	5.18E-95	0.980
8,488,105	434.20	5.18E-95	0.980
6,043,786	434.10	5.45E-95	0.980
11,055,654	434.07	5.53E-95	0.980
12,288,069	434.07	5.53E-95	0.980
14,833,773	434.07	5.53E-95	0.980
35,250,448	434.07	5.53E-95	0.980
35,443,105	434.07	5.53E-95	0.980
35,744,495	434.07	5.53E-95	0.980
35,998,161	434.07	5.53E-95	0.980
36,036,040	434.07	5.53E-95	0.980
36,091,539	434.07	5.53E-95	0.980
36,206,662	434.07	5.53E-95	0.980
13,947,803	434.07	5.53E-95	0.980
13,948,224	434.07	5.53E-95	0.980
16,444,584	434.07	5.53E-95	0.980
35,845,363	434.07	5.53E-95	0.980
36,313,239	434.07	5.53E-95	0.980
15,531,626	434.02	5.68E-95	0.980
6,886,992	434.02	5.68E-95	0.980

Physical Position	chi-square statistic	p-value	Weighted chi-square
13,867,617	434.01	2.18E-96	0.980
16,888,038	434.01	2.18E-96	0.980
37,887,091	434.01	2.18E-96	0.980
6,387,047	434.01	2.18E-96	0.980
7,430,296	434.01	2.18E-96	0.980
7,485,818	434.01	2.18E-96	0.980
8,092,324	434.01	2.18E-96	0.980
8,772,608	434.01	2.18E-96	0.980
9,236,003	434.01	2.18E-96	0.980
9,602,996	434.01	2.18E-96	0.980
9,621,699	434.01	2.18E-96	0.980
10,119,428	434.01	2.18E-96	0.980
10,283,831	434.01	2.18E-96	0.980
10,569,469	434.01	2.18E-96	0.980
13,789,843	434.01	2.18E-96	0.980
14,133,770	434.01	2.18E-96	0.980
15,643,879	434.01	2.18E-96	0.980
21,469,544	434.01	2.18E-96	0.980
37,591,078	434.01	2.18E-96	0.980
38,650,749	434.01	2.18E-96	0.980
44,663,063	434.01	2.18E-96	0.980
44,776,578	434.01	2.18E-96	0.980
47,846,279	434.01	2.18E-96	0.980
6,192,397	434.01	2.18E-96	0.980
7,753,887	434.01	2.18E-96	0.980
15,660,604	434.01	2.18E-96	0.980
37,715,316	434.01	2.18E-96	0.980
47,905,329	434.01	2.18E-96	0.980

Physical Position	chi-square statistic	p-value	Weighted chi-square
7,150,615	434.00	5.73E-95	0.980
9,711,359	434.00	5.73E-95	0.980
10,113,864	434.00	5.73E-95	0.980
10,544,484	434.00	5.73E-95	0.980
10,573,181	434.00	5.73E-95	0.980
10,992,288	434.00	5.73E-95	0.980
15,669,866	434.00	5.73E-95	0.980
19,546,473	434.00	5.73E-95	0.980
36,821,322	434.00	5.73E-95	0.980
42,896,042	434.00	5.73E-95	0.980
45,289,403	434.00	5.73E-95	0.980
47,770,812	434.00	5.73E-95	0.980
8,266,303	434.00	5.73E-95	0.980
9,297,720	434.00	5.73E-95	0.980
12,388,984	434.00	5.73E-95	0.980
16,863,605	434.00	5.73E-95	0.980
14,364,628	434.00	5.73E-95	0.980
11,594,377	434.00	5.73E-95	0.980
14,595,479	434.00	5.73E-95	0.980
15,812,490	434.00	5.73E-95	0.980
16,986,706	434.00	5.73E-95	0.980
10,593,229	434.00	5.73E-95	0.980
47,281,733	433.80	6.33E-95	0.979
6,196,407	433.80	6.33E-95	0.979
14,303,867	433.80	6.33E-95	0.979
15,776,848	433.80	6.33E-95	0.979
47,900,041	433.80	6.33E-95	0.979
14,247,615	433.68	6.71E-95	0.979

Physical Position	chi-square statistic	p-value	Weighted chi-square
49,237,024	433.67	6.77E-95	0.979
13,774,182	433.67	6.77E-95	0.979
16,868,190	433.67	6.77E-95	0.979
46,696,893	433.67	6.77E-95	0.979
11,743,387	433.67	6.77E-95	0.979
7,842,417	433.67	6.77E-95	0.979
14,869,740	433.57	7.10E-95	0.979
7,709,508	433.57	7.10E-95	0.979
6,896,220	433.33	8.00E-95	0.978
29,226,605	433.33	8.00E-95	0.978
8,262,531	433.33	8.00E-95	0.978
10,082,499	433.33	8.00E-95	0.978
64,862,034	433.33	8.00E-95	0.978
7,812,946	433.29	8.19E-95	0.978
64,523,663	433.27	8.27E-95	0.978
6,925,714	433.20	8.55E-95	0.978
14,814,654	433.14	8.80E-95	0.978
15,950,131	433.14	8.80E-95	0.978
14,698,872	433.08	9.05E-95	0.978
6,690,476	433.07	9.11E-95	0.978
11,067,555	433.07	9.11E-95	0.978
15,537,028	433.07	9.11E-95	0.978
36,319,402	433.07	9.11E-95	0.978
11,073,447	433.07	9.12E-95	0.978
34,966,242	433.07	9.12E-95	0.978
35,449,941	433.07	9.12E-95	0.978
35,701,874	433.07	9.12E-95	0.978
35,800,447	433.07	9.12E-95	0.978

Physical Position	chi-square statistic	p-value	Weighted chi-square
35,981,382	433.07	9.12E-95	0.978
14,212,230	433.07	9.12E-95	0.978
21,528,012	433.07	9.12E-95	0.978
12,544,430	433.02	9.36E-95	0.977
6,119,618	433.01	3.60E-96	0.977
8,257,236	433.01	3.60E-96	0.977
8,307,240	433.01	3.60E-96	0.977
27,286,086	433.01	3.60E-96	0.977
37,986,326	433.01	3.60E-96	0.977
55,474,934	433.01	3.60E-96	0.977
8,173,768	433.01	3.60E-96	0.977
13,570,981	433.01	3.60E-96	0.977
48,243,710	433.01	3.60E-96	0.977
55,523,457	433.01	3.60E-96	0.977
11,893,916	433.01	3.60E-96	0.977
11,881,390	433.00	9.45E-95	0.977
11,883,745	433.00	9.45E-95	0.977
11,971,762	433.00	9.45E-95	0.977
14,416,367	433.00	9.45E-95	0.977
14,455,952	433.00	9.45E-95	0.977
16,582,461	433.00	9.45E-95	0.977
17,957,484	433.00	9.45E-95	0.977
27,458,328	433.00	9.45E-95	0.977
28,594,673	433.00	9.45E-95	0.977
35,415,613	433.00	9.45E-95	0.977
45,333,577	433.00	9.45E-95	0.977
48,410,061	433.00	9.45E-95	0.977
56,026,995	433.00	9.45E-95	0.977

Physical Position	chi-square statistic	p-value	Weighted chi-square
8,913,408	433.00	9.45E-95	0.977
16,382,093	433.00	9.45E-95	0.977
45,768,065	433.00	9.45E-95	0.977
40,620,259	433.00	9.45E-95	0.977
43,183,011	433.00	9.45E-95	0.977
6,904,161	433.00	9.45E-95	0.977
8,747,269	433.00	9.45E-95	0.977
13,891,958	433.00	9.45E-95	0.977
48,937,447	433.00	9.45E-95	0.977
8,186,435	433.00	9.45E-95	0.977
18,990,798	432.80	1.04E-94	0.977
6,854,906	432.80	1.04E-94	0.977
54,797,458	432.80	1.04E-94	0.977
8,163,975	432.78	1.06E-94	0.977
16,418,221	432.68	1.11E-94	0.977
37,359,381	432.67	1.12E-94	0.977
12,632,073	432.67	1.12E-94	0.977
43,388,370	432.50	1.21E-94	0.976
60,257,023	432.44	1.25E-94	0.976
63,907,926	432.36	1.30E-94	0.976
15,471,499	432.35	1.31E-94	0.976
17,033,969	432.33	1.32E-94	0.976
7,124,435	432.33	1.32E-94	0.976
16,442,082	432.29	1.35E-94	0.976
35,723,443	432.21	1.40E-94	0.976
6,993,255	432.20	1.41E-94	0.976
6,965,383	432.08	1.49E-94	0.975
35,320,417	432.07	1.50E-94	0.975

Physical Position	chi-square statistic	p-value	Weighted chi-square
35,401,276	432.07	1.50E-94	0.975
35,976,850	432.07	1.50E-94	0.975
17,098,797	432.07	1.50E-94	0.975
35,846,569	432.07	1.50E-94	0.975
36,208,820	432.07	1.50E-94	0.975
36,031,422	432.07	1.50E-94	0.975
36,319,218	432.07	1.50E-94	0.975
36,402,175	432.07	1.50E-94	0.975
18,096,255	432.07	1.50E-94	0.975
14,818,800	432.02	1.54E-94	0.975
56,882,151	432.02	1.55E-94	0.975
59,923,897	432.01	5.94E-96	0.975
65,210,688	432.01	5.94E-96	0.975
8,234,333	432.01	5.94E-96	0.975
14,439,018	432.01	5.94E-96	0.975
17,097,418	432.01	5.94E-96	0.975
28,823,093	432.01	5.94E-96	0.975
48,058,932	432.01	5.94E-96	0.975
6,361,940	432.01	5.94E-96	0.975
37,977,951	432.01	5.94E-96	0.975
38,102,420	432.01	5.94E-96	0.975
47,798,767	432.01	5.94E-96	0.975
58,846,140	432.01	5.94E-96	0.975
14,307,648	432.01	5.94E-96	0.975
10,496,773	432.01	5.94E-96	0.975
6,201,763	432.00	1.56E-94	0.975
7,254,355	432.00	1.56E-94	0.975
7,571,199	432.00	1.56E-94	0.975

Physical Position	chi-square statistic	p-value	Weighted chi-square
7,906,557	432.00	1.56E-94	0.975
8,057,654	432.00	1.56E-94	0.975
11,630,549	432.00	1.56E-94	0.975
11,933,574	432.00	1.56E-94	0.975
14,078,286	432.00	1.56E-94	0.975
14,651,966	432.00	1.56E-94	0.975
16,560,765	432.00	1.56E-94	0.975
30,690,870	432.00	1.56E-94	0.975
37,953,444	432.00	1.56E-94	0.975
52,372,178	432.00	1.56E-94	0.975
66,869,612	432.00	1.56E-94	0.975
11,987,145	432.00	1.56E-94	0.975
15,597,959	431.82	1.71E-94	0.975
14,772,000	431.82	1.71E-94	0.975
7,550,779	431.80	1.72E-94	0.975
10,182,254	431.78	1.74E-94	0.975
15,599,955	431.68	1.82E-94	0.974
9,572,934	431.67	1.84E-94	0.974
37,282,367	431.67	1.84E-94	0.974
9,388,897	431.67	1.84E-94	0.974
53,080,853	431.67	1.84E-94	0.974
8,501,245	431.67	1.84E-94	0.974
10,164,646	431.67	1.84E-94	0.974
64,116,363	431.50	2.00E-94	0.974
48,279,027	431.44	2.06E-94	0.974
16,772,338	431.44	2.06E-94	0.974
42,433,173	431.40	2.10E-94	0.974
15,599,306	431.35	2.16E-94	0.974

Physical Position	chi-square statistic	p-value	Weighted chi-square
15,402,293	431.33	2.17E-94	0.974
14,309,093	431.33	2.17E-94	0.974
6,108,121	431.33	2.17E-94	0.974
16,301,165	431.30	2.21E-94	0.974
9,842,817	431.20	2.32E-94	0.973
7,966,909	431.20	2.32E-94	0.973
14,633,077	431.19	2.33E-94	0.973
41,743,408	431.16	2.37E-94	0.973
42,551,076	431.16	2.37E-94	0.973
43,176,057	431.16	2.37E-94	0.973
18,753,176	431.16	2.37E-94	0.973
7,485,572	431.14	2.39E-94	0.973
7,202,566	431.14	2.39E-94	0.973
7,026,725	431.08	2.46E-94	0.973
35,210,715	431.07	2.48E-94	0.973
36,093,564	431.07	2.48E-94	0.973
36,202,311	431.07	2.48E-94	0.973
14,952,602	431.07	2.48E-94	0.973
16,938,867	431.01	9.80E-96	0.973
7,146,525	431.01	9.80E-96	0.973
10,282,501	431.01	9.80E-96	0.973
17,509,058	431.01	9.80E-96	0.973
9,779,358	431.01	9.81E-96	0.973
11,925,039	431.01	9.81E-96	0.973
17,932,672	431.01	9.81E-96	0.973
6,390,095	431.00	2.57E-94	0.973
8,071,159	431.00	2.57E-94	0.973
8,144,398	431.00	2.57E-94	0.973

Physical Position	chi-square statistic	p-value	Weighted chi-square
9,331,149	431.00	2.57E-94	0.973
9,954,171	431.00	2.57E-94	0.973
11,862,995	431.00	2.57E-94	0.973
15,955,894	431.00	2.57E-94	0.973
16,696,605	431.00	2.57E-94	0.973
59,797,478	431.00	2.57E-94	0.973
16,431,204	431.00	2.57E-94	0.973
7,219,853	431.00	2.57E-94	0.973
58,739,038	431.00	2.57E-94	0.973
16,684,898	431.00	2.57E-94	0.973
10,971,674	430.67	3.03E-94	0.972
16,454,544	430.64	3.07E-94	0.972
16,470,023	430.50	3.30E-94	0.972
63,944,494	430.45	3.37E-94	0.972
37,684,805	430.40	3.47E-94	0.972
14,934,644	430.29	3.66E-94	0.971
15,072,753	430.29	3.67E-94	0.971
57,975,561	430.29	3.67E-94	0.971
36,051,651	430.27	3.70E-94	0.971
36,311,855	430.26	3.71E-94	0.971
45,192,047	430.23	3.77E-94	0.971
48,937,250	430.20	3.83E-94	0.971
12,843,335	430.20	3.83E-94	0.971
41,378,326	430.16	3.91E-94	0.971
43,490,640	430.16	3.91E-94	0.971
44,636,462	430.16	3.91E-94	0.971
6,925,463	430.16	3.91E-94	0.971
12,251,996	430.16	3.91E-94	0.971

Physical Position	chi-square statistic	p-value	Weighted chi-square
14,627,441	430.16	3.91E-94	0.971
11,973,583	430.16	3.91E-94	0.971
12,154,125	430.16	3.91E-94	0.971
17,101,936	430.07	4.08E-94	0.971
36,304,234	430.07	4.08E-94	0.971
35,229,364	430.07	4.09E-94	0.971
9,216,374	430.02	4.20E-94	0.971
7,254,793	430.01	1.62E-95	0.971
39,388,362	430.01	1.62E-95	0.971
52,393,018	430.01	1.62E-95	0.971
9,701,284	430.01	1.62E-95	0.971
6,601,278	430.00	4.23E-94	0.971
7,556,012	430.00	4.23E-94	0.971
9,938,127	430.00	4.23E-94	0.971
10,249,206	430.00	4.23E-94	0.971
13,470,351	430.00	4.23E-94	0.971
14,853,281	430.00	4.23E-94	0.971
15,940,094	430.00	4.23E-94	0.971
17,996,376	430.00	4.23E-94	0.971
29,239,014	430.00	4.23E-94	0.971
35,851,793	430.00	4.23E-94	0.971
36,499,552	430.00	4.23E-94	0.971
37,703,041	430.00	4.23E-94	0.971
68,620,839	430.00	4.24E-94	0.971
18,065,611	430.00	4.24E-94	0.971
68,809,341	429.91	4.44E-94	0.970
15,070,717	429.80	4.68E-94	0.970
43,442,615	429.76	4.77E-94	0.970

Physical Position	chi-square statistic	p-value	Weighted chi-square
6,504,651	429.66	5.01E-94	0.970
37,052,151	429.57	5.25E-94	0.970
15,418,187	429.50	5.44E-94	0.970
12,043,166	429.50	5.44E-94	0.970
58,605,338	429.50	5.44E-94	0.970
57,306,739	429.42	5.67E-94	0.969
8,818,634	429.34	5.89E-94	0.969
38,447,275	429.33	5.91E-94	0.969
47,828,359	429.33	5.91E-94	0.969
24,963,204	429.33	5.93E-94	0.969
16,907,366	429.31	5.98E-94	0.969
17,845,089	429.28	6.06E-94	0.969
8,754,882	429.25	6.16E-94	0.969
13,224,277	429.20	6.32E-94	0.969
42,357,254	429.16	6.44E-94	0.969
12,208,743	429.16	6.44E-94	0.969
12,298,829	429.16	6.44E-94	0.969
17,108,293	429.16	6.44E-94	0.969
9,479,865	429.16	6.44E-94	0.969
12,228,097	429.16	6.44E-94	0.969
12,142,630	429.16	6.45E-94	0.969
35,571,261	429.07	6.73E-94	0.969
36,187,701	429.07	6.73E-94	0.969
5,997,365	429.07	6.74E-94	0.969
6,857,701	429.07	6.75E-94	0.969
7,113,542	429.07	6.75E-94	0.969
14,119,107	429.02	6.92E-94	0.968
29,253,948	429.02	6.92E-94	0.968

Physical Position	chi-square statistic	p-value	Weighted chi-square
15,935,112	429.01	2.67E-95	0.968
7,335,337	429.01	2.67E-95	0.968
16,925,095	429.01	2.67E-95	0.968
48,677,808	429.01	2.67E-95	0.968
8,239,187	429.00	6.98E-94	0.968
8,633,955	429.00	6.98E-94	0.968
10,601,767	429.00	6.98E-94	0.968
13,381,447	429.00	6.98E-94	0.968
15,623,751	429.00	6.98E-94	0.968
37,367,707	429.00	6.98E-94	0.968
45,774,499	429.00	6.98E-94	0.968
13,964,200	428.96	7.12E-94	0.968
18,108,206	428.87	7.45E-94	0.968
68,809,576	428.71	8.05E-94	0.968
10,092,023	428.67	8.25E-94	0.968
16,714,770	428.67	8.25E-94	0.968
9,831,146	428.67	8.25E-94	0.968
6,350,425	428.60	8.53E-94	0.967
13,171,153	428.50	8.96E-94	0.967
6,487,519	428.40	9.42E-94	0.967
16,697,979	428.33	9.74E-94	0.967
16,915,558	428.33	9.74E-94	0.967
18,166,932	428.33	9.74E-94	0.967
7,691,002	428.31	9.87E-94	0.967
7,668,527	428.27	1.00E-93	0.967
36,203,387	428.27	1.00E-93	0.967
11,939,714	428.23	1.03E-93	0.967
6,496,713	428.22	1.03E-93	0.967

Physical Position	chi-square statistic	p-value	Weighted chi-square
38,000,902	428.20	1.04E-93	0.967
34,231,786	428.19	1.04E-93	0.967
42,914,256	428.16	1.06E-93	0.967
12,206,082	428.16	1.06E-93	0.967
12,138,943	428.16	1.06E-93	0.967
59,555,318	428.16	1.06E-93	0.967
35,662,538	428.16	1.06E-93	0.966
11,615,362	428.16	1.06E-93	0.966
11,619,727	428.14	1.07E-93	0.966
7,444,677	428.14	1.07E-93	0.966
36,422,835	428.07	1.11E-93	0.966
15,473,434	428.02	1.14E-93	0.966
57,004,208	428.02	1.14E-93	0.966
46,426,586	428.00	1.15E-93	0.966
47,926,415	428.00	1.15E-93	0.966
8,179,143	428.00	1.15E-93	0.966
15,389,315	428.00	1.15E-93	0.966
16,489,033	428.00	1.15E-93	0.966
44,576,044	428.00	1.15E-93	0.966
8,004,212	427.82	1.26E-93	0.966
58,894,748	427.66	1.36E-93	0.965
7,491,917	427.60	1.40E-93	0.965
59,628,963	427.50	1.48E-93	0.965
16,717,727	427.33	1.61E-93	0.965
6,020,207	427.29	1.64E-93	0.965
13,813,777	427.29	1.64E-93	0.965
47,861,226	427.29	1.64E-93	0.965
48,383,380	427.28	1.65E-93	0.965

Physical Position	chi-square statistic	p-value	Weighted chi-square
6,331,411	427.28	1.65E-93	0.965
7,943,139	427.28	1.65E-93	0.965
10,358,484	427.28	1.65E-93	0.965
11,235,692	427.28	1.65E-93	0.965
11,239,619	427.28	1.65E-93	0.965
11,265,403	427.28	1.65E-93	0.965
11,273,222	427.28	1.65E-93	0.965
11,299,452	427.28	1.65E-93	0.965
17,069,960	427.28	1.65E-93	0.965
17,136,056	427.28	1.65E-93	0.965
21,444,381	427.28	1.65E-93	0.965
8,937,749	427.27	1.66E-93	0.964
14,589,111	427.23	1.69E-93	0.964
48,428,161	427.20	1.72E-93	0.964
14,118,211	427.20	1.72E-93	0.964
44,306,787	427.16	1.75E-93	0.964
43,859,224	427.16	1.75E-93	0.964
36,260,120	427.07	1.83E-93	0.964
8,942,648	427.07	1.83E-93	0.964
18,099,317	427.07	1.84E-93	0.964
32,787,110	427.01	7.28E-95	0.964
10,005,140	427.00	1.90E-93	0.964
28,740,349	427.00	1.90E-93	0.964
7,536,982	427.00	1.90E-93	0.964
11,548,152	427.00	1.90E-93	0.964
17,109,058	426.83	2.07E-93	0.963
15,744,254	426.80	2.10E-93	0.963
18,454,559	426.80	2.10E-93	0.963

Physical Position	chi-square statistic	p-value	Weighted chi-square
13,031,922	426.80	2.10E-93	0.963
11,071,392	426.80	2.10E-93	0.963
63,565,015	426.78	2.12E-93	0.963
63,696,911	426.78	2.12E-93	0.963
48,845,358	426.73	2.17E-93	0.963
18,944,902	426.66	2.25E-93	0.963
68,480,136	426.52	2.41E-93	0.963
17,101,011	426.52	2.42E-93	0.963
13,445,439	426.50	2.44E-93	0.963
55,285,669	426.50	2.44E-93	0.963
6,614,103	426.50	2.44E-93	0.963
16,332,094	426.40	2.57E-93	0.963
10,253,083	426.36	2.61E-93	0.962
12,960,819	426.33	2.65E-93	0.962
68,042,593	426.31	2.68E-93	0.962
35,306,022	426.30	2.70E-93	0.962
46,257,782	426.29	2.71E-93	0.962
9,701,529	426.28	2.72E-93	0.962
11,288,491	426.28	2.72E-93	0.962
11,333,676	426.28	2.72E-93	0.962
40,859,963	426.28	2.72E-93	0.962
11,307,134	426.28	2.72E-93	0.962
17,081,171	426.28	2.72E-93	0.962
65,047,460	426.28	2.72E-93	0.962
9,184,860	426.28	2.72E-93	0.962
9,190,177	426.23	2.79E-93	0.962
14,169,703	426.22	2.81E-93	0.962
45,456,818	426.20	2.83E-93	0.962

Physical Position	chi-square statistic	p-value	Weighted chi-square
13,391,090	426.00	3.13E-93	0.962
14,184,305	426.00	3.13E-93	0.962
45,353,693	426.00	3.13E-93	0.962
45,412,517	426.00	3.13E-93	0.962
7,404,023	426.00	3.13E-93	0.962
35,302,920	425.91	3.27E-93	0.961
13,585,317	425.52	3.99E-93	0.961
41,938,871	425.40	4.22E-93	0.960
57,820,295	425.33	4.37E-93	0.960
15,876,763	425.33	4.37E-93	0.960
15,517,160	425.30	4.43E-93	0.960
47,986,754	425.29	4.47E-93	0.960
45,539,782	425.29	4.47E-93	0.960
45,677,876	425.29	4.47E-93	0.960
48,069,745	425.29	4.47E-93	0.960
60,003,756	425.29	4.47E-93	0.960
9,483,123	425.28	4.47E-93	0.960
12,000,022	425.28	4.47E-93	0.960
15,300,533	425.28	4.47E-93	0.960
16,288,654	425.28	4.47E-93	0.960
41,106,740	425.28	4.47E-93	0.960
10,140,086	425.28	4.48E-93	0.960
11,326,471	425.28	4.48E-93	0.960
13,823,970	425.27	4.50E-93	0.960
6,700,854	425.23	4.59E-93	0.960
40,406,948	425.21	4.64E-93	0.960
6,310,639	425.16	4.76E-93	0.960
13,670,952	425.16	4.76E-93	0.960

Physical Position	chi-square statistic	p-value	Weighted chi-square
68,737,847	425.14	4.80E-93	0.960
13,486,295	425.12	1.88E-94	0.960
59,104,105	425.09	4.93E-93	0.960
15,505,695	425.04	1.95E-94	0.959
9,538,764	425.00	5.16E-93	0.959
55,266,633	425.00	5.16E-93	0.959
8,506,220	425.00	5.16E-93	0.959
8,504,614	425.00	5.16E-93	0.959
59,752,894	424.89	5.45E-93	0.959
9,146,251	424.82	5.63E-93	0.959
44,556,418	424.78	5.76E-93	0.959
14,755,879	424.67	6.09E-93	0.959
52,903,772	424.66	6.10E-93	0.959
14,592,523	424.40	6.96E-93	0.958
8,481,229	424.33	7.20E-93	0.958
10,990,953	424.33	7.20E-93	0.958
14,391,014	424.33	7.21E-93	0.958
47,870,643	424.29	7.37E-93	0.958
14,930,572	424.29	7.37E-93	0.958
47,987,178	424.29	7.37E-93	0.958
12,919,019	424.28	7.38E-93	0.958
11,234,483	424.28	7.39E-93	0.958
11,284,418	424.28	7.39E-93	0.958
6,381,357	424.23	7.57E-93	0.958
10,634,561	424.00	8.50E-93	0.957
47,697,763	424.00	8.50E-93	0.957
13,989,498	423.82	9.33E-93	0.957
48,584,559	423.78	9.50E-93	0.957

Physical Position	chi-square statistic	p-value	Weighted chi-square
59,896,094	423.78	9.51E-93	0.957
8,069,606	423.67	1.00E-92	0.956
11,848,011	423.66	1.01E-92	0.956
14,865,618	423.57	1.05E-92	0.956
16,663,113	423.50	1.09E-92	0.956
5,929,595	423.44	1.13E-92	0.956
11,088,360	423.44	1.13E-92	0.956
12,741,790	423.44	1.13E-92	0.956
12,953,521	423.35	1.18E-92	0.956
14,841,823	423.33	1.19E-92	0.956
50,481,373	423.29	1.21E-92	0.956
51,434,445	423.29	1.21E-92	0.956
47,634,945	423.29	1.22E-92	0.955
7,656,479	423.28	1.22E-92	0.955
10,331,093	423.27	1.23E-92	0.955
7,852,903	423.23	1.25E-92	0.955
42,613,702	423.16	1.29E-92	0.955
43,501,983	423.16	1.29E-92	0.955
13,402,632	423.16	1.30E-92	0.955
12,194,150	423.15	1.30E-92	0.955
52,330,534	423.14	1.31E-92	0.955
13,205,024	423.09	1.34E-92	0.955
17,115,076	423.02	1.39E-92	0.955
17,043,628	423.01	5.41E-94	0.955
6,784,806	423.00	1.40E-92	0.955
59,589,886	423.00	1.40E-92	0.955
48,432,937	423.00	1.40E-92	0.955
7,202,976	423.00	1.40E-92	0.955

Physical Position	chi-square statistic	p-value	Weighted chi-square
14,505,469	423.00	1.40E-92	0.955
42,415,703	423.00	1.40E-92	0.955
68,752,564	422.91	1.47E-92	0.955
38,661,822	422.67	1.66E-92	0.954
68,055,465	422.56	1.75E-92	0.954
10,475,714	422.45	1.84E-92	0.954
10,394,679	422.44	1.85E-92	0.954
12,119,148	422.44	1.85E-92	0.954
12,648,233	422.44	1.85E-92	0.954
13,887,552	422.44	1.85E-92	0.954
63,704,825	422.44	1.85E-92	0.954
10,340,893	422.35	1.95E-92	0.953
38,522,006	422.34	1.95E-92	0.953
7,810,315	422.33	1.96E-92	0.953
9,528,711	422.33	1.96E-92	0.953
12,205,659	422.29	2.00E-92	0.953
11,329,152	422.29	2.00E-92	0.953
36,388,346	422.28	2.01E-92	0.953
12,982,570	422.28	2.01E-92	0.953
17,162,683	422.23	2.06E-92	0.953
64,899,174	422.20	2.09E-92	0.953
11,688,162	422.06	2.24E-92	0.953
9,978,686	422.00	2.31E-92	0.953
12,817,125	422.00	2.31E-92	0.953
34,963,331	422.00	2.31E-92	0.953
9,859,046	421.89	2.44E-92	0.952
14,391,597	421.86	2.48E-92	0.952
25,022,288	421.66	2.74E-92	0.952

Physical Position	chi-square statistic	p-value	Weighted chi-square
32,242,940	421.60	2.82E-92	0.952
48,103,228	421.47	3.01E-92	0.951
14,824,267	421.44	3.05E-92	0.951
11,097,873	421.44	3.06E-92	0.951
64,161,441	421.44	3.06E-92	0.951
44,927,588	421.39	3.13E-92	0.951
7,940,241	421.35	3.20E-92	0.951
10,332,278	421.29	3.30E-92	0.951
46,728,234	421.28	3.30E-92	0.951
10,587,957	421.15	3.53E-92	0.951
45,678,123	421.11	3.61E-92	0.951
17,985,725	421.07	3.69E-92	0.950
17,120,579	420.95	3.91E-92	0.950
13,941,552	420.87	4.07E-92	0.950
46,767,617	420.86	4.09E-92	0.950
14,926,110	420.79	4.24E-92	0.950
34,997,604	420.77	4.28E-92	0.950
10,740,382	420.77	4.28E-92	0.950
15,534,888	420.74	4.35E-92	0.950
9,807,664	420.69	4.44E-92	0.950
11,055,421	420.64	4.57E-92	0.950
41,394,858	420.57	4.73E-92	0.949
5,929,020	420.44	5.03E-92	0.949
14,078,870	420.44	5.03E-92	0.949
14,480,380	420.44	5.03E-92	0.949
15,768,988	420.44	5.03E-92	0.949
12,115,477	420.44	5.04E-92	0.949
7,931,319	420.44	5.04E-92	0.949

Physical Position	chi-square statistic	p-value	Weighted chi-square
10,096,058	420.44	5.04E-92	0.949
13,660,455	420.44	5.04E-92	0.949
15,223,103	420.44	5.04E-92	0.949
11,091,894	420.43	5.05E-92	0.949
15,961,072	420.40	5.14E-92	0.949
7,972,905	420.40	5.15E-92	0.949
14,834,844	420.40	5.15E-92	0.949
45,256,812	420.40	5.15E-92	0.949
10,340,142	420.35	5.27E-92	0.949
11,237,840	420.35	5.28E-92	0.949
6,220,943	420.33	5.33E-92	0.949
38,299,381	420.28	5.47E-92	0.949
8,281,423	420.27	5.48E-92	0.949
10,006,576	420.25	5.53E-92	0.949
63,627,150	420.20	5.68E-92	0.949
8,148,516	420.19	5.71E-92	0.949
15,759,966	420.16	5.81E-92	0.948
45,064,601	420.14	5.86E-92	0.948
57,237,769	420.13	5.89E-92	0.948
37,680,141	420.09	6.00E-92	0.948
57,006,529	420.07	6.06E-92	0.948
44,032,304	420.07	6.06E-92	0.948
60,179,391	420.01	6.27E-92	0.948
42,229,158	420.00	6.28E-92	0.948
14,062,240	419.91	6.56E-92	0.948
36,427,712	419.83	6.83E-92	0.948
14,301,369	419.63	7.57E-92	0.947
15,216,502	419.63	7.57E-92	0.947

Physical Position	chi-square statistic	p-value	Weighted chi-square
13,786,668	419.60	7.67E-92	0.947
68,884,283	419.60	7.67E-92	0.947
6,258,252	419.47	8.19E-92	0.947
47,873,053	419.44	8.30E-92	0.947
6,677,072	419.44	8.32E-92	0.947
8,366,682	419.43	8.34E-92	0.947
15,406,015	419.41	8.43E-92	0.947
11,320,737	419.35	8.68E-92	0.947
13,636,540	419.35	8.70E-92	0.947
5,928,358	419.34	8.74E-92	0.947
12,029,176	419.33	8.78E-92	0.947
11,921,505	419.00	1.04E-91	0.946
35,597,876	418.92	1.08E-91	0.946
35,903,201	418.87	1.10E-91	0.946
36,423,629	418.86	1.11E-91	0.945
41,553,519	418.73	1.18E-91	0.945
41,660,449	418.64	1.24E-91	0.945
11,406,544	418.63	1.25E-91	0.945
11,077,483	418.63	1.25E-91	0.945
35,075,971	418.53	1.31E-91	0.945
14,433,980	418.53	1.31E-91	0.945
15,448,290	418.53	1.31E-91	0.945
9,545,020	418.51	1.32E-91	0.945
40,407,428	418.50	1.33E-91	0.945
7,925,391	418.47	1.35E-91	0.945
16,575,473	418.45	1.37E-91	0.945
15,818,384	418.43	1.38E-91	0.945
12,581,560	418.36	1.43E-91	0.944

Physical Position	chi-square statistic	p-value	Weighted chi-square
12,986,426	418.35	1.43E-91	0.944
14,083,676	418.33	1.45E-91	0.944
60,105,036	418.26	1.50E-91	0.944
14,081,174	418.22	1.53E-91	0.944
67,758,932	418.16	1.58E-91	0.944
14,885,254	418.16	1.58E-91	0.944
68,837,118	418.02	1.69E-91	0.944
47,116,302	418.00	1.71E-91	0.944
47,863,272	418.00	1.71E-91	0.944
7,857,638	418.00	1.71E-91	0.944
9,958,345	417.99	1.71E-91	0.944
8,461,524	417.80	1.89E-91	0.943
10,139,595	417.63	2.06E-91	0.943
8,276,877	417.53	2.16E-91	0.943
18,124,371	417.47	2.22E-91	0.942
13,160,429	417.44	2.26E-91	0.942
11,239,004	417.35	2.36E-91	0.942
9,558,886	417.33	2.38E-91	0.942
38,310,771	417.28	2.45E-91	0.942
12,246,919	417.22	2.52E-91	0.942
16,449,189	417.22	2.52E-91	0.942
45,072,924	417.14	2.62E-91	0.942
6,266,400	417.02	2.79E-91	0.941
48,403,033	417.00	2.82E-91	0.941
44,864,853	416.86	3.02E-91	0.941
16,448,622	416.74	3.21E-91	0.941
9,999,997	416.62	3.40E-91	0.940
12,110,490	416.57	3.50E-91	0.940

Physical Position	chi-square statistic	p-value	Weighted chi-square
11,344,833	416.55	3.53E-91	0.940
10,751,974	416.54	3.55E-91	0.940
15,217,074	416.53	3.57E-91	0.940
10,142,107	416.47	3.67E-91	0.940
11,325,980	416.34	3.91E-91	0.940
35,527,622	416.25	4.10E-91	0.940
68,349,390	416.17	4.27E-91	0.939
7,767,937	416.14	4.32E-91	0.939
22,745,790	416.14	4.33E-91	0.939
17,913,620	416.00	4.64E-91	0.939
10,626,066	416.00	4.65E-91	0.939
49,148,785	415.86	4.97E-91	0.939
51,363,940	415.86	4.97E-91	0.939
10,754,725	415.85	5.00E-91	0.939
13,710,550	415.85	5.00E-91	0.939
16,715,835	415.85	5.00E-91	0.939
16,274,698	415.63	5.59E-91	0.938
45,767,518	415.58	5.73E-91	0.938
11,075,618	415.56	5.79E-91	0.938
12,741,419	415.52	5.89E-91	0.938
12,609,369	415.52	5.89E-91	0.938
14,604,026	415.42	6.21E-91	0.938
16,502,754	415.25	6.76E-91	0.937
12,899,906	415.14	7.13E-91	0.937
44,772,575	415.08	7.37E-91	0.937
9,855,433	415.00	7.66E-91	0.937
15,658,446	415.00	7.66E-91	0.937
9,538,994	415.00	7.66E-91	0.937

Physical Position	chi-square statistic	p-value	Weighted chi-square
37,494,748	414.86	8.20E-91	0.936
52,245,627	414.86	8.20E-91	0.936
48,569,818	414.86	8.20E-91	0.936
44,162,731	414.86	8.21E-91	0.936
6,937,144	414.86	8.23E-91	0.936
9,593,977	414.85	8.25E-91	0.936
15,238,581	414.77	8.59E-91	0.936
18,128,562	414.77	8.59E-91	0.936
36,214,039	414.74	8.72E-91	0.936
13,055,372	414.68	9.00E-91	0.936
41,261,097	414.50	9.85E-91	0.936
12,109,048	414.45	1.01E-90	0.936
41,011,559	414.36	1.05E-90	0.935
41,569,215	414.33	1.07E-90	0.935
7,040,093	414.20	1.14E-90	0.935
14,950,374	414.02	1.25E-90	0.935
51,290,230	413.87	1.35E-90	0.934
50,974,145	413.87	1.35E-90	0.934
51,847,184	413.86	1.35E-90	0.934
10,059,070	413.85	1.36E-90	0.934
13,556,056	413.72	1.45E-90	0.934
20,695,658	413.58	1.55E-90	0.934
16,457,563	413.51	1.62E-90	0.933
7,928,161	413.32	1.77E-90	0.933
9,298,912	413.21	1.87E-90	0.933
8,302,157	413.11	1.97E-90	0.933
68,312,638	413.08	2.00E-90	0.932
49,071,024	412.87	2.22E-90	0.932

Physical Position	chi-square statistic	p-value	Weighted chi-square
50,079,354	412.87	2.22E-90	0.932
7,230,999	412.86	2.23E-90	0.932
50,679,566	412.86	2.23E-90	0.932
46,067,505	412.86	2.23E-90	0.932
9,340,362	412.86	2.23E-90	0.932
9,998,867	412.85	2.24E-90	0.932
12,758,854	412.79	2.32E-90	0.932
17,119,503	412.72	2.39E-90	0.932
68,887,401	412.66	2.47E-90	0.932
19,632,177	412.47	2.72E-90	0.931
8,723,193	412.35	2.88E-90	0.931
59,919,835	412.22	3.08E-90	0.931
6,745,718	411.90	3.60E-90	0.930
68,027,024	411.84	3.71E-90	0.930
8,898,122	411.82	3.76E-90	0.930
9,187,560	411.26	4.96E-90	0.928
11,971,606	411.20	5.12E-90	0.928
56,160,520	411.12	5.32E-90	0.928
10,141,705	411.11	5.37E-90	0.928
7,085,708	411.07	5.45E-90	0.928
6,910,419	411.00	5.66E-90	0.928
15,434,219	410.91	5.91E-90	0.928
10,149,840	410.91	5.92E-90	0.928
8,156,071	410.91	5.93E-90	0.928
10,053,533	410.86	6.07E-90	0.927
6,320,821	410.72	6.49E-90	0.927
7,831,157	410.69	6.59E-90	0.927
64,012,430	410.56	7.04E-90	0.927

Physical Position	chi-square statistic	p-value	Weighted chi-square
57,855,459	410.54	7.13E-90	0.927
15,222,412	410.53	7.14E-90	0.927
35,635,361	410.41	7.60E-90	0.926
16,003,314	410.35	7.84E-90	0.926
12,210,784	410.30	8.03E-90	0.926
42,168,778	410.28	8.09E-90	0.926
45,661,389	410.27	8.15E-90	0.926
41,785,819	410.20	8.44E-90	0.926
44,207,137	410.00	9.32E-90	0.926
10,413,980	409.96	9.51E-90	0.925
51,458,033	409.87	9.93E-90	0.925
55,274,215	409.40	1.26E-89	0.924
7,602,577	409.23	1.37E-89	0.924
14,379,166	409.14	1.43E-89	0.924
55,276,081	409.13	1.44E-89	0.924
55,602,063	409.12	1.45E-89	0.924
15,443,856	409.05	1.50E-89	0.923
35,200,680	408.98	1.55E-89	0.923
8,040,796	408.98	1.55E-89	0.923
6,734,950	408.98	1.55E-89	0.923
36,361,166	408.91	1.61E-89	0.923
15,247,016	408.91	1.61E-89	0.923
15,418,457	408.90	1.62E-89	0.923
11,260,098	408.29	2.19E-89	0.922
67,879,328	408.22	2.27E-89	0.921
55,868,110	408.18	2.32E-89	0.921
15,351,450	408.13	2.38E-89	0.921
35,560,034	408.12	2.38E-89	0.921

Physical Position	chi-square statistic	p-value	Weighted chi-square
15,239,761	408.10	2.41E-89	0.921
35,666,059	408.07	2.44E-89	0.921
10,682,064	407.92	2.64E-89	0.921
15,226,647	407.77	2.84E-89	0.920
8,899,285	407.56	3.16E-89	0.920
11,240,398	407.39	3.43E-89	0.920
11,630,734	407.27	3.66E-89	0.919
14,396,028	407.17	3.84E-89	0.919
8,152,674	406.92	4.36E-89	0.919
13,214,148	406.88	4.43E-89	0.918
8,769,379	406.52	5.31E-89	0.918
13,766,257	406.41	5.62E-89	0.917
9,457,109	406.40	5.64E-89	0.917
9,768,991	406.40	5.64E-89	0.917
12,910,877	406.40	5.64E-89	0.917
8,772,293	406.39	5.66E-89	0.917
10,674,584	406.39	5.66E-89	0.917
13,738,457	406.39	5.67E-89	0.917
10,005,310	406.28	6.00E-89	0.917
10,515,005	406.07	6.65E-89	0.917
12,088,274	406.05	6.72E-89	0.917
50,985,681	405.88	7.31E-89	0.916
5,989,747	405.66	8.17E-89	0.916
11,265,906	405.39	9.35E-89	0.915
23,024,514	405.33	9.62E-89	0.915
50,614,445	405.32	9.68E-89	0.915
67,757,827	405.27	9.91E-89	0.915
8,150,701	405.08	1.09E-88	0.914

Physical Position	chi-square statistic	p-value	Weighted chi-square
15,948,628	405.07	1.10E-88	0.914
13,584,244	405.01	1.13E-88	0.914
5,822,233	404.90	1.19E-88	0.914
11,265,057	404.29	1.62E-88	0.913
47,330,762	404.19	1.70E-88	0.912
10,150,085	404.18	1.71E-88	0.912
15,949,979	404.13	1.76E-88	0.912
47,898,685	404.00	1.87E-88	0.912
5,823,156	403.55	2.34E-88	0.911
12,585,603	403.54	2.35E-88	0.911
11,922,551	403.48	2.42E-88	0.911
10,942,801	403.41	2.52E-88	0.911
7,633,294	403.39	2.54E-88	0.911
10,002,317	403.37	2.57E-88	0.911
35,489,200	403.31	2.64E-88	0.910

Physical Position	chi-square statistic	p-value	Weighted chi-square
67,992,771	403.26	2.71E-88	0.910
14,632,216	403.16	2.85E-88	0.910
22,201,164	402.69	3.61E-88	0.909
14,008,682	402.63	3.72E-88	0.909
9,768,565	402.40	4.16E-88	0.908
10,094,645	402.39	4.18E-88	0.908
7,268,756	402.29	4.41E-88	0.908
6,719,796	402.21	4.58E-88	0.908
13,085,499	402.18	4.64E-88	0.908
14,421,137	402.16	4.70E-88	0.908
6,563,460	402.15	4.72E-88	0.908
7,105,109	402.04	4.99E-88	0.908
9,440,612	401.73	5.82E-88	0.907
1,860,778	401.67	6.01E-88	0.907
48,930,847	401.52	6.46E-88	0.906

Physical Position	chi-square statistic	p-value	Weighted chi-square
68,621,027	401.38	6.94E-88	0.906
15,442,865	401.26	7.38E-88	0.906
47,075,784	401.19	7.65E-88	0.906
55,138,030	401.09	8.01E-88	0.905
17,146,769	401.07	8.12E-88	0.905
9,463,569	400.78	9.37E-88	0.905
9,814,554	400.20	1.25E-87	0.903
12,258,745	400.08	1.33E-87	0.903
13,214,896	399.77	1.55E-87	0.902
8,344,440	399.68	1.62E-87	0.902
41,777,505	399.68	1.63E-87	0.902
41,562,413	399.29	1.97E-87	0.901
35,986,521	399.15	2.12E-87	0.901
54,873,004	399.10	2.17E-87	0.901

Table S2.2:

List of 1085 SNP positions which were found to be highly diagnostic (weighted chi-square score >0.9) between the B and C inversion haplotypes.

Physical Position	Chi-square statistic	p-value	Weighted chi-square	Physical Position	Chi-square statistic	p-value	Weighted chi-square	Physical Position	Chi-square statistic	p-value	Weighted chi-square
7,759,837	277.00	7.08E-61	1.000	13,230,358	277.00	7.08E-61	1.000	7,893,191	276.00	1.17E-60	0.996
9,516,702	277.00	7.08E-61	1.000	15,027,879	277.00	7.08E-61	1.000	9,779,623	276.00	1.17E-60	0.996
9,781,388	277.00	7.08E-61	1.000	15,366,630	277.00	7.08E-61	1.000	10,166,850	276.00	1.17E-60	0.996
10,595,748	277.00	7.08E-61	1.000	16,331,942	277.00	7.08E-61	1.000	10,167,179	276.00	1.17E-60	0.996
11,091,894	277.00	7.08E-61	1.000	35,234,205	277.00	7.08E-61	1.000	10,195,772	276.00	1.17E-60	0.996
11,486,065	277.00	7.08E-61	1.000	38,040,290	277.00	7.08E-61	1.000	10,247,786	276.00	1.17E-60	0.996
14,301,369	277.00	7.08E-61	1.000	38,660,542	277.00	7.08E-61	1.000	10,326,420	276.00	1.17E-60	0.996
15,851,462	277.00	7.08E-61	1.000	41,695,755	277.00	7.08E-61	1.000	10,444,696	276.00	1.17E-60	0.996
38,709,977	277.00	7.08E-61	1.000	44,693,179	277.00	7.08E-61	1.000	10,596,216	276.00	1.17E-60	0.996
60,854,601	277.00	7.08E-61	1.000	59,928,521	277.00	7.08E-61	1.000	12,077,358	276.00	1.17E-60	0.996
64,274,081	277.00	7.08E-61	1.000	60,728,238	277.00	7.08E-61	1.000	12,179,694	276.00	1.17E-60	0.996
6,410,624	277.00	7.08E-61	1.000	60,864,931	277.00	7.08E-61	1.000	12,616,766	276.00	1.17E-60	0.996
6,472,139	277.00	7.08E-61	1.000	11,077,483	276.00	1.17E-60	0.996	12,926,943	276.00	1.17E-60	0.996
6,886,992	277.00	7.08E-61	1.000	6,254,014	276.00	1.17E-60	0.996	13,149,175	276.00	1.17E-60	0.996
7,007,236	277.00	7.08E-61	1.000	6,420,011	276.00	1.17E-60	0.996	14,008,068	276.00	1.17E-60	0.996
7,343,825	277.00	7.08E-61	1.000	6,531,036	276.00	1.17E-60	0.996	14,016,414	276.00	1.17E-60	0.996
7,944,116	277.00	7.08E-61	1.000	6,781,941	276.00	1.17E-60	0.996	14,247,128	276.00	1.17E-60	0.996
9,146,251	277.00	7.08E-61	1.000	7,082,658	276.00	1.17E-60	0.996	14,391,014	276.00	1.17E-60	0.996
9,190,177	277.00	7.08E-61	1.000	7,151,423	276.00	1.17E-60	0.996	14,793,260	276.00	1.17E-60	0.996
9,462,187	277.00	7.08E-61	1.000	7,284,575	276.00	1.17E-60	0.996	14,843,761	276.00	1.17E-60	0.996
10,340,893	277.00	7.08E-61	1.000	7,363,635	276.00	1.17E-60	0.996	14,853,848	276.00	1.17E-60	0.996
10,571,541	277.00	7.08E-61	1.000	7,389,120	276.00	1.17E-60	0.996	14,922,472	276.00	1.17E-60	0.996
12,266,218	277.00	7.08E-61	1.000	7,399,021	276.00	1.17E-60	0.996	15,238,581	276.00	1.17E-60	0.996
12,953,521	277.00	7.08E-61	1.000	7,730,156	276.00	1.17E-60	0.996	15,271,413	276.00	1.17E-60	0.996

Physical Position	Chi-square statistic	p-value	Weighted chi-square
15,434,219	276.00	1.17E-60	0.996
15,448,290	276.00	1.17E-60	0.996
16,404,060	276.00	1.17E-60	0.996
16,432,390	276.00	1.17E-60	0.996
16,454,544	276.00	1.17E-60	0.996
16,588,278	276.00	1.17E-60	0.996
16,932,544	276.00	1.17E-60	0.996
17,109,058	276.00	1.17E-60	0.996
17,150,322	276.00	1.17E-60	0.996
21,097,349	276.00	1.17E-60	0.996
23,676,159	276.00	1.17E-60	0.996
33,673,260	276.00	1.17E-60	0.996
33,946,345	276.00	1.17E-60	0.996
35,302,920	276.00	1.17E-60	0.996
36,311,855	276.00	1.17E-60	0.996
37,392,576	276.00	1.17E-60	0.996
37,404,821	276.00	1.17E-60	0.996
37,709,759	276.00	1.17E-60	0.996
37,813,719	276.00	1.17E-60	0.996
37,840,295	276.00	1.17E-60	0.996
38,133,363	276.00	1.17E-60	0.996
38,369,891	276.00	1.17E-60	0.996
38,605,463	276.00	1.17E-60	0.996
38,780,367	276.00	1.17E-60	0.996
38,869,094	276.00	1.17E-60	0.996
45,564,524	276.00	1.17E-60	0.996
48,654,397	276.00	1.17E-60	0.996
48,884,209	276.00	1.17E-60	0.996

Physical Position	Chi-square statistic	p-value	Weighted chi-square
50,072,375	276.00	1.17E-60	0.996
60,187,515	276.00	1.17E-60	0.996
60,709,636	276.00	1.17E-60	0.996
60,815,065	276.00	1.17E-60	0.996
60,857,136	276.00	1.17E-60	0.996
61,266,823	276.00	1.17E-60	0.996
64,377,192	276.00	1.17E-60	0.996
6,375,355	276.00	1.17E-60	0.996
12,549,818	276.00	1.17E-60	0.996
16,594,645	276.00	1.17E-60	0.996
38,522,006	276.00	1.17E-60	0.996
42,730,338	276.00	1.17E-60	0.996
6,120,236	275.00	1.93E-60	0.993
6,227,588	275.00	1.93E-60	0.993
7,812,562	275.00	1.93E-60	0.993
5,925,283	275.00	1.93E-60	0.993
6,070,177	275.00	1.93E-60	0.993
6,180,171	275.00	1.93E-60	0.993
7,209,691	275.00	1.93E-60	0.993
7,516,495	275.00	1.93E-60	0.993
7,831,157	275.00	1.93E-60	0.993
8,266,959	275.00	1.93E-60	0.993
8,463,284	275.00	1.93E-60	0.993
8,818,634	275.00	1.93E-60	0.993
9,216,374	275.00	1.93E-60	0.993
9,440,612	275.00	1.93E-60	0.993
9,593,977	275.00	1.93E-60	0.993
9,927,467	275.00	1.93E-60	0.993

Physical Position	Chi-square statistic	p-value	Weighted chi-square
9,952,904	275.00	1.93E-60	0.993
10,059,070	275.00	1.93E-60	0.993
10,126,414	275.00	1.93E-60	0.993
10,217,493	275.00	1.93E-60	0.993
10,415,394	275.00	1.93E-60	0.993
10,658,432	275.00	1.93E-60	0.993
10,687,040	275.00	1.93E-60	0.993
11,113,690	275.00	1.93E-60	0.993
11,413,299	275.00	1.93E-60	0.993
11,945,817	275.00	1.93E-60	0.993
12,553,853	275.00	1.93E-60	0.993
12,563,053	275.00	1.93E-60	0.993
12,636,690	275.00	1.93E-60	0.993
12,758,854	275.00	1.93E-60	0.993
13,297,409	275.00	1.93E-60	0.993
13,823,970	275.00	1.93E-60	0.993
14,192,472	275.00	1.93E-60	0.993
14,247,615	275.00	1.93E-60	0.993
14,410,800	275.00	1.93E-60	0.993
14,527,069	275.00	1.93E-60	0.993
14,747,110	275.00	1.93E-60	0.993
14,865,809	275.00	1.93E-60	0.993
14,901,418	275.00	1.93E-60	0.993
15,217,074	275.00	1.93E-60	0.993
15,471,499	275.00	1.93E-60	0.993
16,230,766	275.00	1.93E-60	0.993
16,643,395	275.00	1.93E-60	0.993
16,705,635	275.00	1.93E-60	0.993

Physical Position	Chi-square statistic	p-value	Weighted chi-square
17,098,797	275.00	1.93E-60	0.993
17,185,501	275.00	1.93E-60	0.993
17,704,236	275.00	1.93E-60	0.993
37,008,928	275.00	1.93E-60	0.993
37,564,329	275.00	1.93E-60	0.993
37,859,965	275.00	1.93E-60	0.993
41,355,181	275.00	1.93E-60	0.993
42,911,109	275.00	1.93E-60	0.993
43,727,945	275.00	1.93E-60	0.993
43,746,519	275.00	1.93E-60	0.993
45,117,540	275.00	1.93E-60	0.993
48,323,820	275.00	1.93E-60	0.993
49,493,715	275.00	1.93E-60	0.993
61,613,362	275.00	1.93E-60	0.993
62,515,014	275.00	1.93E-60	0.993
64,090,486	275.00	1.93E-60	0.993
66,991,610	275.00	1.93E-60	0.993
5,997,365	275.00	1.93E-60	0.993
7,953,100	275.00	1.93E-60	0.993
8,984,561	275.00	1.93E-60	0.993
10,708,091	275.00	1.93E-60	0.993
14,433,980	275.00	1.93E-60	0.993
16,489,920	275.00	1.93E-60	0.993
17,120,579	275.00	1.93E-60	0.993
42,743,499	275.00	1.93E-60	0.993
48,021,560	275.00	1.93E-60	0.993
60,728,448	275.00	1.93E-60	0.993
62,387,321	275.00	1.93E-60	0.993

Physical Position	Chi-square statistic	p-value	Weighted chi-square
7,124,435	274.00	3.17E-60	0.989
10,697,005	274.00	3.17E-60	0.989
14,772,000	274.00	3.17E-60	0.989
15,818,384	274.00	3.17E-60	0.989
30,421,704	274.00	3.17E-60	0.989
47,830,101	274.00	3.17E-60	0.989
6,043,786	274.00	3.17E-60	0.989
6,774,678	274.00	3.17E-60	0.989
7,117,358	274.00	3.17E-60	0.989
7,230,999	274.00	3.17E-60	0.989
8,262,531	274.00	3.17E-60	0.989
8,412,437	274.00	3.17E-60	0.989
9,010,360	274.00	3.17E-60	0.989
9,377,155	274.00	3.17E-60	0.989
10,155,682	274.00	3.17E-60	0.989
10,394,854	274.00	3.17E-60	0.989
10,618,411	274.00	3.17E-60	0.989
11,075,618	274.00	3.17E-60	0.989
11,078,809	274.00	3.17E-60	0.989
12,184,055	274.00	3.17E-60	0.989
13,393,654	274.00	3.17E-60	0.989
14,139,065	274.00	3.17E-60	0.989
14,241,561	274.00	3.17E-60	0.989
14,458,119	274.00	3.17E-60	0.989
14,526,659	274.00	3.17E-60	0.989
14,664,339	274.00	3.17E-60	0.989
14,926,110	274.00	3.17E-60	0.989
15,531,626	274.00	3.17E-60	0.989

Physical Position	Chi-square statistic	p-value	Weighted chi-square
15,534,888	274.00	3.17E-60	0.989
15,597,959	274.00	3.17E-60	0.989
16,274,698	274.00	3.17E-60	0.989
16,431,204	274.00	3.17E-60	0.989
16,537,052	274.00	3.17E-60	0.989
16,796,968	274.00	3.17E-60	0.989
25,488,154	274.00	3.17E-60	0.989
31,548,520	274.00	3.17E-60	0.989
35,606,300	274.00	3.17E-60	0.989
36,533,645	274.00	3.17E-60	0.989
37,719,009	274.00	3.17E-60	0.989
38,639,590	274.00	3.17E-60	0.989
44,799,062	274.00	3.17E-60	0.989
44,864,853	274.00	3.17E-60	0.989
47,879,755	274.00	3.17E-60	0.989
48,592,570	274.00	3.17E-60	0.989
48,981,258	274.00	3.17E-60	0.989
49,237,024	274.00	3.17E-60	0.989
57,791,301	274.00	3.17E-60	0.989
58,744,598	274.00	3.17E-60	0.989
60,207,640	274.00	3.17E-60	0.989
60,808,354	274.00	3.17E-60	0.989
60,889,144	274.00	3.17E-60	0.989
6,745,718	273.81	3.48E-60	0.988
62,986,871	273.81	3.48E-60	0.988
6,310,639	273.00	5.23E-60	0.986
6,896,220	273.00	5.23E-60	0.986
6,905,963	273.00	5.23E-60	0.986

Physical Position	Chi-square statistic	p-value	Weighted chi-square
7,181,437	273.00	5.23E-60	0.986
8,163,215	273.00	5.23E-60	0.986
9,572,934	273.00	5.23E-60	0.986
10,182,254	273.00	5.23E-60	0.986
10,529,949	273.00	5.23E-60	0.986
10,626,066	273.00	5.23E-60	0.986
12,818,919	273.00	5.23E-60	0.986
16,886,609	273.00	5.23E-60	0.986
39,281,712	273.00	5.23E-60	0.986
47,715,130	273.00	5.23E-60	0.986
63,944,494	273.00	5.23E-60	0.986
7,421,269	273.00	5.23E-60	0.986
7,633,294	273.00	5.23E-60	0.986
7,947,883	273.00	5.23E-60	0.986
8,341,953	273.00	5.23E-60	0.986
8,709,642	273.00	5.23E-60	0.986
9,954,171	273.00	5.23E-60	0.986
12,585,603	273.00	5.23E-60	0.986
13,491,351	273.00	5.23E-60	0.986
14,119,107	273.00	5.23E-60	0.986
15,418,457	273.00	5.23E-60	0.986
15,911,661	273.00	5.23E-60	0.986
16,146,237	273.00	5.23E-60	0.986
16,382,093	273.00	5.23E-60	0.986
36,260,120	273.00	5.23E-60	0.986
47,905,329	273.00	5.23E-60	0.986
62,219,076	273.00	5.23E-60	0.986
6,196,407	273.00	5.23E-60	0.986

Physical Position	Chi-square statistic	p-value	Weighted chi-square
6,254,538	273.00	5.23E-60	0.986
6,894,756	273.00	5.23E-60	0.986
6,904,161	273.00	5.23E-60	0.986
7,404,023	273.00	5.23E-60	0.986
8,163,975	273.00	5.23E-60	0.986
10,021,195	273.00	5.23E-60	0.986
15,331,844	273.00	5.23E-60	0.986
16,440,041	273.00	5.23E-60	0.986
17,119,503	273.00	5.23E-60	0.986
36,203,387	273.00	5.23E-60	0.986
38,720,544	273.00	5.23E-60	0.986
10,149,840	272.82	5.72E-60	0.985
13,053,468	272.82	5.72E-60	0.985
14,396,028	272.82	5.72E-60	0.985
16,491,618	272.82	5.72E-60	0.985
60,840,453	272.82	5.72E-60	0.985
64,378,809	272.82	5.72E-60	0.985
7,842,417	272.78	5.85E-60	0.985
17,111,078	272.78	5.85E-60	0.985
7,481,733	272.00	8.63E-60	0.982
48,410,061	272.00	8.63E-60	0.982
48,614,188	272.00	8.63E-60	0.982
6,253,853	272.00	8.63E-60	0.982
6,554,297	272.00	8.63E-60	0.982
7,202,566	272.00	8.63E-60	0.982
8,053,221	272.00	8.63E-60	0.982
8,059,391	272.00	8.63E-60	0.982
8,266,303	272.00	8.63E-60	0.982

Physical Position	Chi-square statistic	p-value	Weighted chi-square
9,297,720	272.00	8.63E-60	0.982
12,388,984	272.00	8.63E-60	0.982
13,829,489	272.00	8.63E-60	0.982
14,698,872	272.00	8.63E-60	0.982
14,849,423	272.00	8.63E-60	0.982
14,869,740	272.00	8.63E-60	0.982
14,911,455	272.00	8.63E-60	0.982
15,473,434	272.00	8.63E-60	0.982
15,669,866	272.00	8.63E-60	0.982
15,776,848	272.00	8.63E-60	0.982
15,876,763	272.00	8.63E-60	0.982
16,863,605	272.00	8.63E-60	0.982
17,101,011	272.00	8.63E-60	0.982
28,594,673	272.00	8.63E-60	0.982
35,306,022	272.00	8.63E-60	0.982
39,388,362	272.00	8.63E-60	0.982
40,010,636	272.00	8.63E-60	0.982
43,442,615	272.00	8.63E-60	0.982
43,680,565	272.00	8.63E-60	0.982
45,768,065	272.00	8.63E-60	0.982
47,779,113	272.00	8.63E-60	0.982
48,537,255	272.00	8.63E-60	0.982
60,875,485	272.00	8.63E-60	0.982
61,160,592	272.00	8.63E-60	0.982
63,830,539	272.00	8.63E-60	0.982
6,980,366	271.83	9.39E-60	0.981
8,071,603	271.83	9.39E-60	0.981
10,323,382	271.83	9.39E-60	0.981

Physical Position	Chi-square statistic	p-value	Weighted chi-square
9,186,413	271.79	9.60E-60	0.981
10,126,686	271.79	9.60E-60	0.981
11,910,071	271.79	9.60E-60	0.981
14,273,661	271.79	9.60E-60	0.981
47,766,627	271.79	9.60E-60	0.981
60,859,380	271.79	9.60E-60	0.981
11,071,392	271.76	9.71E-60	0.981
15,226,647	271.76	9.71E-60	0.981
15,599,955	271.76	9.71E-60	0.981
16,174,333	271.70	1.00E-59	0.981
12,101,494	271.23	1.27E-59	0.979
36,872,834	271.23	1.27E-59	0.979
5,996,734	271.00	1.42E-59	0.978
7,113,542	271.00	1.42E-59	0.978
7,925,391	271.00	1.42E-59	0.978
8,043,725	271.00	1.42E-59	0.978
9,310,504	271.00	1.42E-59	0.978
10,005,310	271.00	1.42E-59	0.978
10,371,149	271.00	1.42E-59	0.978
10,971,674	271.00	1.42E-59	0.978
12,177,593	271.00	1.42E-59	0.978
32,526,887	271.00	1.42E-59	0.978
35,075,971	271.00	1.42E-59	0.978
35,903,201	271.00	1.42E-59	0.978
44,513,098	271.00	1.42E-59	0.978
47,770,812	271.00	1.42E-59	0.978
60,883,736	271.00	1.42E-59	0.978
60,925,404	271.00	1.42E-59	0.978

Physical Position	Chi-square statistic	p-value	Weighted chi-square
8,504,614	271.00	1.42E-59	0.978
9,223,067	271.00	1.42E-59	0.978
9,842,817	271.00	1.42E-59	0.978
10,323,636	271.00	1.42E-59	0.978
41,938,871	271.00	1.42E-59	0.978
44,163,699	271.00	1.42E-59	0.978
60,808,730	271.00	1.42E-59	0.978
62,184,193	271.00	1.42E-59	0.978
6,381,357	270.84	1.54E-59	0.978
12,110,490	270.84	1.54E-59	0.978
60,239,459	270.80	1.58E-59	0.978
60,902,600	270.75	1.61E-59	0.977
47,797,375	270.74	1.62E-59	0.977
58,605,338	270.70	1.65E-59	0.977
5,929,595	270.66	8.13E-61	0.977
5,987,059	270.66	8.13E-61	0.977
6,044,206	270.66	8.13E-61	0.977
6,539,029	270.66	8.13E-61	0.977
6,555,954	270.66	8.13E-61	0.977
6,582,283	270.66	8.13E-61	0.977
6,629,178	270.66	8.13E-61	0.977
6,732,492	270.66	8.13E-61	0.977
6,857,701	270.66	8.13E-61	0.977
6,909,045	270.66	8.13E-61	0.977
7,134,876	270.66	8.13E-61	0.977
7,178,427	270.66	8.13E-61	0.977
7,181,872	270.66	8.13E-61	0.977
7,222,611	270.66	8.13E-61	0.977

Physical Position	Chi-square statistic	p-value	Weighted chi-square
7,278,496	270.66	8.13E-61	0.977
7,325,542	270.66	8.13E-61	0.977
7,344,160	270.66	8.13E-61	0.977
7,396,272	270.66	8.13E-61	0.977
7,431,237	270.66	8.13E-61	0.977
7,501,476	270.66	8.13E-61	0.977
7,527,849	270.66	8.13E-61	0.977
7,532,375	270.66	8.13E-61	0.977
7,571,199	270.66	8.13E-61	0.977
7,620,061	270.66	8.13E-61	0.977
7,656,479	270.66	8.13E-61	0.977
7,731,180	270.66	8.13E-61	0.977
7,943,139	270.66	8.13E-61	0.977
7,996,694	270.66	8.13E-61	0.977
8,332,299	270.66	8.13E-61	0.977
8,508,892	270.66	8.13E-61	0.977
8,549,396	270.66	8.13E-61	0.977
8,645,232	270.66	8.13E-61	0.977
8,648,270	270.66	8.13E-61	0.977
8,772,293	270.66	8.13E-61	0.977
9,193,369	270.66	8.13E-61	0.977
9,249,328	270.66	8.13E-61	0.977
9,310,848	270.66	8.13E-61	0.977
9,376,933	270.66	8.13E-61	0.977
9,560,154	270.66	8.13E-61	0.977
9,706,463	270.66	8.13E-61	0.977
9,804,434	270.66	8.13E-61	0.977
10,071,364	270.66	8.13E-61	0.977

Physical Position	Chi-square statistic	p-value	Weighted chi-square
10,321,437	270.66	8.13E-61	0.977
10,323,789	270.66	8.13E-61	0.977
10,451,508	270.66	8.13E-61	0.977
10,578,246	270.66	8.13E-61	0.977
10,596,480	270.66	8.13E-61	0.977
10,620,871	270.66	8.13E-61	0.977
10,635,608	270.66	8.13E-61	0.977
10,674,584	270.66	8.13E-61	0.977
11,073,692	270.66	8.13E-61	0.977
11,831,493	270.66	8.13E-61	0.977
12,154,125	270.66	8.13E-61	0.977
13,030,127	270.66	8.13E-61	0.977
13,150,395	270.66	8.13E-61	0.977
13,208,795	270.66	8.13E-61	0.977
13,228,196	270.66	8.13E-61	0.977
13,345,145	270.66	8.13E-61	0.977
13,385,149	270.66	8.13E-61	0.977
13,710,550	270.66	8.13E-61	0.977
14,270,068	270.66	8.13E-61	0.977
14,477,651	270.66	8.13E-61	0.977
14,569,213	270.66	8.13E-61	0.977
15,025,568	270.66	8.13E-61	0.977
15,880,123	270.66	8.13E-61	0.977
16,054,453	270.66	8.13E-61	0.977
16,544,647	270.66	8.13E-61	0.977
16,547,718	270.66	8.13E-61	0.977
16,770,729	270.66	8.13E-61	0.977
16,994,254	270.66	8.13E-61	0.977

Physical Position	Chi-square statistic	p-value	Weighted chi-square
17,069,960	270.66	8.13E-61	0.977
17,081,171	270.66	8.13E-61	0.977
17,136,056	270.66	8.13E-61	0.977
17,893,386	270.66	8.13E-61	0.977
26,090,126	270.66	8.13E-61	0.977
26,203,597	270.66	8.13E-61	0.977
29,347,341	270.66	8.13E-61	0.977
31,287,417	270.66	8.13E-61	0.977
32,307,508	270.66	8.13E-61	0.977
32,921,570	270.66	8.13E-61	0.977
36,065,878	270.66	8.13E-61	0.977
36,313,239	270.66	8.13E-61	0.977
36,420,199	270.66	8.13E-61	0.977
36,543,831	270.66	8.13E-61	0.977
36,636,753	270.66	8.13E-61	0.977
36,689,727	270.66	8.13E-61	0.977
37,039,577	270.66	8.13E-61	0.977
37,283,287	270.66	8.13E-61	0.977
37,403,706	270.66	8.13E-61	0.977
37,489,505	270.66	8.13E-61	0.977
38,350,363	270.66	8.13E-61	0.977
38,687,865	270.66	8.13E-61	0.977
38,825,751	270.66	8.13E-61	0.977
38,842,925	270.66	8.13E-61	0.977
38,883,690	270.66	8.13E-61	0.977
38,899,511	270.66	8.13E-61	0.977
39,092,992	270.66	8.13E-61	0.977
39,134,154	270.66	8.13E-61	0.977

Physical Position	Chi-square statistic	p-value	Weighted chi-square
39,145,547	270.66	8.13E-61	0.977
40,089,561	270.66	8.13E-61	0.977
40,250,225	270.66	8.13E-61	0.977
40,303,602	270.66	8.13E-61	0.977
40,321,696	270.66	8.13E-61	0.977
40,875,768	270.66	8.13E-61	0.977
44,249,267	270.66	8.13E-61	0.977
45,286,512	270.66	8.13E-61	0.977
45,650,419	270.66	8.13E-61	0.977
46,160,298	270.66	8.13E-61	0.977
46,768,665	270.66	8.13E-61	0.977
47,709,158	270.66	8.13E-61	0.977
47,816,666	270.66	8.13E-61	0.977
48,190,987	270.66	8.13E-61	0.977
48,213,229	270.66	8.13E-61	0.977
48,509,170	270.66	8.13E-61	0.977
48,558,898	270.66	8.13E-61	0.977
48,567,181	270.66	8.13E-61	0.977
48,667,534	270.66	8.13E-61	0.977
48,677,808	270.66	8.13E-61	0.977
48,695,985	270.66	8.13E-61	0.977
57,527,380	270.66	8.13E-61	0.977
58,554,385	270.66	8.13E-61	0.977
58,753,412	270.66	8.13E-61	0.977
59,787,552	270.66	8.13E-61	0.977
59,846,869	270.66	8.13E-61	0.977
59,855,255	270.66	8.13E-61	0.977
60,264,951	270.66	8.13E-61	0.977

Physical Position	Chi-square statistic	p-value	Weighted chi-square
60,714,087	270.66	8.13E-61	0.977
60,808,124	270.66	8.13E-61	0.977
60,906,499	270.66	8.13E-61	0.977
60,938,401	270.66	8.13E-61	0.977
60,966,490	270.66	8.13E-61	0.977
60,966,977	270.66	8.13E-61	0.977
60,982,601	270.66	8.13E-61	0.977
61,203,492	270.66	8.13E-61	0.977
61,614,317	270.66	8.13E-61	0.977
62,378,638	270.66	8.13E-61	0.977
62,396,437	270.66	8.13E-61	0.977
62,620,775	270.66	8.13E-61	0.977
63,833,556	270.66	8.13E-61	0.977
63,969,433	270.66	8.13E-61	0.977
9,340,362	270.25	2.07E-59	0.976
12,789,291	270.18	2.14E-59	0.975
13,189,749	270.00	2.35E-59	0.975
7,432,504	270.00	2.35E-59	0.975
7,479,286	270.00	2.35E-59	0.975
10,223,857	270.00	2.35E-59	0.975
12,205,659	270.00	2.35E-59	0.975
12,634,688	270.00	2.35E-59	0.975
13,350,136	270.00	2.35E-59	0.975
13,445,439	270.00	2.35E-59	0.975
14,062,240	270.00	2.35E-59	0.975
14,190,552	270.00	2.35E-59	0.975
14,329,403	270.00	2.35E-59	0.975
14,592,523	270.00	2.35E-59	0.975

Physical Position	Chi-square statistic	p-value	Weighted chi-square
15,442,865	270.00	2.35E-59	0.975
15,812,490	270.00	2.35E-59	0.975
15,935,112	270.00	2.35E-59	0.975
15,949,979	270.00	2.35E-59	0.975
16,683,181	270.00	2.35E-59	0.975
16,986,706	270.00	2.35E-59	0.975
37,282,367	270.00	2.35E-59	0.975
47,861,226	270.00	2.35E-59	0.975
48,045,396	270.00	2.35E-59	0.975
48,428,161	270.00	2.35E-59	0.975
59,607,549	270.00	2.35E-59	0.975
59,919,835	270.00	2.35E-59	0.975
60,653,428	270.00	2.35E-59	0.975
61,760,034	270.00	2.35E-59	0.975
62,475,991	270.00	2.35E-59	0.975
10,150,085	269.85	2.53E-59	0.974
14,633,077	269.85	2.53E-59	0.974
6,854,906	269.80	2.59E-59	0.974
10,682,064	269.80	2.59E-59	0.974
12,219,837	269.80	2.59E-59	0.974
8,575,370	269.79	2.61E-59	0.974
14,364,628	269.79	2.61E-59	0.974
8,488,105	269.79	2.61E-59	0.974
60,665,913	269.79	2.61E-59	0.974
7,812,946	269.73	2.69E-59	0.974
6,031,737	269.68	1.33E-60	0.974
6,585,073	269.68	1.33E-60	0.974
6,691,034	269.68	1.33E-60	0.974

Physical Position	Chi-square statistic	p-value	Weighted chi-square
6,876,899	269.68	1.33E-60	0.974
7,268,976	269.68	1.33E-60	0.974
7,477,462	269.68	1.33E-60	0.974
7,503,174	269.68	1.33E-60	0.974
7,617,236	269.68	1.33E-60	0.974
7,660,094	269.68	1.33E-60	0.974
7,943,551	269.68	1.33E-60	0.974
8,532,836	269.68	1.33E-60	0.974
8,630,987	269.68	1.33E-60	0.974
8,677,084	269.68	1.33E-60	0.974
8,738,807	269.68	1.33E-60	0.974
8,784,247	269.68	1.33E-60	0.974
8,984,270	269.68	1.33E-60	0.974
8,990,468	269.68	1.33E-60	0.974
9,004,844	269.68	1.33E-60	0.974
9,184,636	269.68	1.33E-60	0.974
9,188,713	269.68	1.33E-60	0.974
9,278,106	269.68	1.33E-60	0.974
9,360,723	269.68	1.33E-60	0.974
9,388,617	269.68	1.33E-60	0.974
9,586,756	269.68	1.33E-60	0.974
9,596,960	269.68	1.33E-60	0.974
9,612,116	269.68	1.33E-60	0.974
9,701,529	269.68	1.33E-60	0.974
10,028,107	269.68	1.33E-60	0.974
10,094,645	269.68	1.33E-60	0.974
10,156,163	269.68	1.33E-60	0.974
10,167,657	269.68	1.33E-60	0.974

Physical Position	Chi-square statistic	p-value	Weighted chi-square
10,248,857	269.68	1.33E-60	0.974
10,311,323	269.68	1.33E-60	0.974
10,394,679	269.68	1.33E-60	0.974
10,410,507	269.68	1.33E-60	0.974
10,906,159	269.68	1.33E-60	0.974
11,097,873	269.68	1.33E-60	0.974
11,486,514	269.68	1.33E-60	0.974
12,251,996	269.68	1.33E-60	0.974
12,738,690	269.68	1.33E-60	0.974
13,160,429	269.68	1.33E-60	0.974
13,209,083	269.68	1.33E-60	0.974
13,813,142	269.68	1.33E-60	0.974
14,194,625	269.68	1.33E-60	0.974
14,406,533	269.68	1.33E-60	0.974
14,409,861	269.68	1.33E-60	0.974
14,446,149	269.68	1.33E-60	0.974
14,833,773	269.68	1.33E-60	0.974
14,941,392	269.68	1.33E-60	0.974
15,038,294	269.68	1.33E-60	0.974
15,086,213	269.68	1.33E-60	0.974
15,458,552	269.68	1.33E-60	0.974
15,553,299	269.68	1.33E-60	0.974
15,627,993	269.68	1.33E-60	0.974
15,631,562	269.68	1.33E-60	0.974
15,716,768	269.68	1.33E-60	0.974
15,779,448	269.68	1.33E-60	0.974
16,441,156	269.68	1.33E-60	0.974
16,600,582	269.68	1.33E-60	0.974

Physical Position	Chi-square statistic	p-value	Weighted chi-square
16,621,178	269.68	1.33E-60	0.974
16,692,488	269.68	1.33E-60	0.974
16,750,326	269.68	1.33E-60	0.974
16,842,755	269.68	1.33E-60	0.974
16,865,992	269.68	1.33E-60	0.974
16,943,072	269.68	1.33E-60	0.974
17,097,179	269.68	1.33E-60	0.974
18,082,877	269.68	1.33E-60	0.974
20,345,625	269.68	1.33E-60	0.974
25,109,299	269.68	1.33E-60	0.974
28,310,320	269.68	1.33E-60	0.974
32,759,321	269.68	1.33E-60	0.974
33,403,615	269.68	1.33E-60	0.974
35,250,448	269.68	1.33E-60	0.974
35,368,634	269.68	1.33E-60	0.974
35,662,538	269.68	1.33E-60	0.974
35,701,874	269.68	1.33E-60	0.974
35,981,382	269.68	1.33E-60	0.974
35,998,161	269.68	1.33E-60	0.974
36,206,662	269.68	1.33E-60	0.974
36,213,320	269.68	1.33E-60	0.974
36,282,054	269.68	1.33E-60	0.974
36,504,428	269.68	1.33E-60	0.974
37,263,912	269.68	1.33E-60	0.974
37,308,227	269.68	1.33E-60	0.974
37,555,904	269.68	1.33E-60	0.974
38,036,197	269.68	1.33E-60	0.974
38,107,627	269.68	1.33E-60	0.974

Physical Position	Chi-square statistic	p-value	Weighted chi-square
38,194,429	269.68	1.33E-60	0.974
38,293,781	269.68	1.33E-60	0.974
38,403,903	269.68	1.33E-60	0.974
38,605,826	269.68	1.33E-60	0.974
39,158,179	269.68	1.33E-60	0.974
39,189,912	269.68	1.33E-60	0.974
39,554,314	269.68	1.33E-60	0.974
39,620,612	269.68	1.33E-60	0.974
40,747,268	269.68	1.33E-60	0.974
43,621,324	269.68	1.33E-60	0.974
43,736,348	269.68	1.33E-60	0.974
43,864,979	269.68	1.33E-60	0.974
44,486,459	269.68	1.33E-60	0.974
45,507,668	269.68	1.33E-60	0.974
47,826,755	269.68	1.33E-60	0.974
47,845,399	269.68	1.33E-60	0.974
47,978,339	269.68	1.33E-60	0.974
48,953,533	269.68	1.33E-60	0.974
57,383,326	269.68	1.33E-60	0.974
58,489,086	269.68	1.33E-60	0.974
60,159,298	269.68	1.33E-60	0.974
60,717,087	269.68	1.33E-60	0.974
60,928,997	269.68	1.33E-60	0.974
62,451,078	269.68	1.33E-60	0.974
62,520,814	269.68	1.33E-60	0.974
6,330,400	269.67	2.76E-59	0.974
8,186,435	269.66	2.77E-59	0.974
3,027,700	269.65	2.80E-59	0.973

Physical Position	Chi-square statistic	p-value	Weighted chi-square
6,331,411	269.59	1.39E-60	0.973
6,645,262	269.59	1.39E-60	0.973
6,855,264	269.59	1.39E-60	0.973
6,889,699	269.59	1.39E-60	0.973
7,518,957	269.59	1.39E-60	0.973
8,841,688	269.59	1.39E-60	0.973
10,717,668	269.59	1.39E-60	0.973
11,102,627	269.59	1.39E-60	0.973
11,207,440	269.59	1.39E-60	0.973
14,350,137	269.59	1.39E-60	0.973
16,393,083	269.59	1.39E-60	0.973
16,588,478	269.59	1.39E-60	0.973
42,464,308	269.59	1.39E-60	0.973
60,714,501	269.59	1.39E-60	0.973
62,379,375	269.59	1.39E-60	0.973
12,919,019	269.29	3.35E-59	0.972
6,197,412	269.26	3.40E-59	0.972
62,392,137	269.26	3.40E-59	0.972
8,071,159	269.00	3.87E-59	0.971
14,416,367	269.00	3.87E-59	0.971
42,896,042	269.00	3.87E-59	0.971
6,266,400	269.00	3.87E-59	0.971
12,843,335	269.00	3.87E-59	0.971
13,812,707	269.00	3.87E-59	0.971
15,247,302	269.00	3.87E-59	0.971
15,418,187	269.00	3.87E-59	0.971
7,026,725	269.00	3.87E-59	0.971
7,678,310	269.00	3.87E-59	0.971

Physical Position	Chi-square statistic	p-value	Weighted chi-square
8,144,398	269.00	3.87E-59	0.971
10,142,107	269.00	3.87E-59	0.971
36,427,712	269.00	3.87E-59	0.971
14,760,129	268.94	3.98E-59	0.971
5,968,025	268.80	4.28E-59	0.970
6,980,794	268.80	4.28E-59	0.970
8,928,941	268.74	4.41E-59	0.970
12,949,843	268.74	4.41E-59	0.970
36,939,887	268.74	4.41E-59	0.970
14,303,867	268.74	4.41E-59	0.970
5,926,566	268.70	2.18E-60	0.970
6,012,273	268.70	2.18E-60	0.970
6,176,084	268.70	2.18E-60	0.970
6,294,965	268.70	2.18E-60	0.970
6,379,052	268.70	2.18E-60	0.970
6,650,612	268.70	2.18E-60	0.970
6,690,476	268.70	2.18E-60	0.970
6,984,681	268.70	2.18E-60	0.970
7,090,331	268.70	2.18E-60	0.970
7,136,488	268.70	2.18E-60	0.970
7,753,887	268.70	2.18E-60	0.970
9,472,939	268.70	2.18E-60	0.970
9,483,123	268.70	2.18E-60	0.970
9,701,284	268.70	2.18E-60	0.970
9,768,565	268.70	2.18E-60	0.970
9,775,244	268.70	2.18E-60	0.970
9,779,358	268.70	2.18E-60	0.970
10,375,912	268.70	2.18E-60	0.970

Physical Position	Chi-square statistic	p-value	Weighted chi-square
10,936,950	268.70	2.18E-60	0.970
12,208,743	268.70	2.18E-60	0.970
12,298,829	268.70	2.18E-60	0.970
12,813,075	268.70	2.18E-60	0.970
12,813,876	268.70	2.18E-60	0.970
13,376,466	268.70	2.18E-60	0.970
14,163,906	268.70	2.18E-60	0.970
14,647,750	268.70	2.18E-60	0.970
14,691,709	268.70	2.18E-60	0.970
14,718,224	268.70	2.18E-60	0.970
15,023,297	268.70	2.18E-60	0.970
15,504,005	268.70	2.18E-60	0.970
15,524,850	268.70	2.18E-60	0.970
15,559,201	268.70	2.18E-60	0.970
16,288,654	268.70	2.18E-60	0.970
16,303,165	268.70	2.18E-60	0.970
16,404,222	268.70	2.18E-60	0.970
16,603,012	268.70	2.18E-60	0.970
16,619,329	268.70	2.18E-60	0.970
16,807,250	268.70	2.18E-60	0.970
17,108,293	268.70	2.18E-60	0.970
17,153,979	268.70	2.18E-60	0.970
31,123,103	268.70	2.18E-60	0.970
31,465,133	268.70	2.18E-60	0.970
32,500,667	268.70	2.18E-60	0.970
35,029,065	268.70	2.18E-60	0.970
35,088,506	268.70	2.18E-60	0.970
35,846,569	268.70	2.18E-60	0.970

Physical Position	Chi-square statistic	p-value	Weighted chi-square
36,319,402	268.70	2.18E-60	0.970
37,297,483	268.70	2.18E-60	0.970
37,391,525	268.70	2.18E-60	0.970
37,636,437	268.70	2.18E-60	0.970
37,715,316	268.70	2.18E-60	0.970
38,124,600	268.70	2.18E-60	0.970
38,322,864	268.70	2.18E-60	0.970
38,650,522	268.70	2.18E-60	0.970
38,860,333	268.70	2.18E-60	0.970
39,135,381	268.70	2.18E-60	0.970
39,176,232	268.70	2.18E-60	0.970
39,192,144	268.70	2.18E-60	0.970
39,596,740	268.70	2.18E-60	0.970
41,106,740	268.70	2.18E-60	0.970
41,579,313	268.70	2.18E-60	0.970
43,560,195	268.70	2.18E-60	0.970
43,619,506	268.70	2.18E-60	0.970
44,325,177	268.70	2.18E-60	0.970
57,714,415	268.70	2.18E-60	0.970
59,555,318	268.70	2.18E-60	0.970
60,103,826	268.70	2.18E-60	0.970
60,207,080	268.70	2.18E-60	0.970
60,263,483	268.70	2.18E-60	0.970
60,656,790	268.70	2.18E-60	0.970
60,701,280	268.70	2.18E-60	0.970
60,714,862	268.70	2.18E-60	0.970
62,394,800	268.70	2.18E-60	0.970
62,424,725	268.70	2.18E-60	0.970

Physical Position	Chi-square statistic	p-value	Weighted chi-square
65,348,726	268.70	2.18E-60	0.970
48,156,696	268.66	4.58E-59	0.970
6,032,726	268.61	2.28E-60	0.970
6,134,943	268.61	2.28E-60	0.970
6,872,106	268.61	2.28E-60	0.970
7,221,367	268.61	2.28E-60	0.970
7,433,603	268.61	2.28E-60	0.970
7,931,319	268.61	2.28E-60	0.970
9,331,149	268.61	2.28E-60	0.970
10,730,674	268.61	2.28E-60	0.970
12,549,597	268.61	2.28E-60	0.970
13,351,090	268.61	2.28E-60	0.970
15,223,103	268.61	2.28E-60	0.970
15,534,459	268.61	2.28E-60	0.970
15,663,262	268.61	2.28E-60	0.970
16,333,925	268.61	2.28E-60	0.970
36,031,422	268.61	2.28E-60	0.970
36,091,539	268.61	2.28E-60	0.970
36,319,218	268.61	2.28E-60	0.970
36,465,934	268.61	2.28E-60	0.970
37,453,890	268.61	2.28E-60	0.970
62,407,804	268.61	2.28E-60	0.970
62,846,491	268.61	2.28E-60	0.970
10,136,858	268.52	2.39E-60	0.969
11,100,639	268.52	2.39E-60	0.969
13,840,861	268.52	2.39E-60	0.969
60,820,752	268.52	2.39E-60	0.969
60,827,377	268.52	2.39E-60	0.969

Physical Position	Chi-square statistic	p-value	Weighted chi-square
60,857,546	268.52	2.39E-60	0.969
6,916,873	268.31	5.45E-59	0.969
11,630,549	268.14	5.96E-59	0.968
8,754,882	268.13	5.97E-59	0.968
13,224,277	268.00	6.38E-59	0.968
60,910,697	268.00	6.38E-59	0.968
12,109,048	268.00	6.38E-59	0.968
13,738,457	268.00	6.38E-59	0.968
13,813,777	268.00	6.38E-59	0.968
16,696,605	268.00	6.38E-59	0.968
16,697,979	268.00	6.38E-59	0.968
7,852,903	268.00	6.38E-59	0.968
14,755,879	268.00	6.38E-59	0.968
14,837,045	268.00	6.38E-59	0.968
61,042,353	268.00	6.38E-59	0.968
36,214,039	267.96	6.51E-59	0.967
7,242,455	267.90	6.71E-59	0.967
15,222,412	267.82	6.97E-59	0.967
60,257,023	267.82	6.97E-59	0.967
12,088,274	267.81	7.01E-59	0.967
6,536,360	267.75	7.22E-59	0.967
5,929,020	267.72	3.57E-60	0.966
6,029,164	267.72	3.57E-60	0.966
6,150,331	267.72	3.57E-60	0.966
6,990,850	267.72	3.57E-60	0.966
7,057,813	267.72	3.57E-60	0.966
7,408,959	267.72	3.57E-60	0.966
8,524,507	267.72	3.57E-60	0.966

Physical Position	Chi-square statistic	p-value	Weighted chi-square
9,916,868	267.72	3.57E-60	0.966
10,615,913	267.72	3.57E-60	0.966
10,979,550	267.72	3.57E-60	0.966
11,114,750	267.72	3.57E-60	0.966
12,505,315	267.72	3.57E-60	0.966
12,620,886	267.72	3.57E-60	0.966
13,338,347	267.72	3.57E-60	0.966
14,078,870	267.72	3.57E-60	0.966
14,480,380	267.72	3.57E-60	0.966
15,580,942	267.72	3.57E-60	0.966
15,759,966	267.72	3.57E-60	0.966
15,768,988	267.72	3.57E-60	0.966
15,838,010	267.72	3.57E-60	0.966
16,496,947	267.72	3.57E-60	0.966
16,624,858	267.72	3.57E-60	0.966
17,045,082	267.72	3.57E-60	0.966
35,320,417	267.72	3.57E-60	0.966
35,401,276	267.72	3.57E-60	0.966
35,976,850	267.72	3.57E-60	0.966
36,093,564	267.72	3.57E-60	0.966
37,027,550	267.72	3.57E-60	0.966
38,660,726	267.72	3.57E-60	0.966
39,614,004	267.72	3.57E-60	0.966
40,705,366	267.72	3.57E-60	0.966
41,582,810	267.72	3.57E-60	0.966
44,197,838	267.72	3.57E-60	0.966
44,628,276	267.72	3.57E-60	0.966
45,344,845	267.72	3.57E-60	0.966

Physical Position	Chi-square statistic	p-value	Weighted chi-square
45,733,372	267.72	3.57E-60	0.966
45,935,124	267.72	3.57E-60	0.966
48,425,665	267.72	3.57E-60	0.966
57,658,355	267.72	3.57E-60	0.966
59,846,569	267.72	3.57E-60	0.966
60,837,463	267.72	3.57E-60	0.966
6,581,416	267.63	3.73E-60	0.966
10,041,713	267.63	3.73E-60	0.966
14,745,061	267.63	3.73E-60	0.966
33,052,779	267.63	3.73E-60	0.966
60,716,632	267.63	3.73E-60	0.966
10,164,646	267.62	7.69E-59	0.966
64,116,363	267.62	7.69E-59	0.966
60,974,714	267.60	7.79E-59	0.966
5,928,358	267.56	7.96E-59	0.966
6,860,176	267.54	3.91E-60	0.966
12,228,097	267.54	3.91E-60	0.966
9,859,623	267.44	4.10E-60	0.965
7,668,527	267.29	9.11E-59	0.965
10,331,093	267.15	9.75E-59	0.964
17,033,969	267.10	9.98E-59	0.964
10,990,953	267.00	1.05E-58	0.964
7,254,355	267.00	1.05E-58	0.964
14,391,597	267.00	1.05E-58	0.964
15,303,712	267.00	1.05E-58	0.964
27,458,328	267.00	1.05E-58	0.964
45,333,577	267.00	1.05E-58	0.964
6,993,255	266.82	1.15E-58	0.963

Physical Position	Chi-square statistic	p-value	Weighted chi-square
16,442,082	266.76	1.18E-58	0.963
8,465,159	266.76	1.18E-58	0.963
6,200,910	266.73	5.84E-60	0.963
6,352,763	266.73	5.84E-60	0.963
7,003,990	266.73	5.84E-60	0.963
7,099,123	266.73	5.84E-60	0.963
9,422,983	266.73	5.84E-60	0.963
9,618,097	266.73	5.84E-60	0.963
10,469,740	266.73	5.84E-60	0.963
12,223,556	266.73	5.84E-60	0.963
12,878,328	266.73	5.84E-60	0.963
13,560,399	266.73	5.84E-60	0.963
14,804,087	266.73	5.84E-60	0.963
15,030,618	266.73	5.84E-60	0.963
15,351,450	266.73	5.84E-60	0.963
15,957,118	266.73	5.84E-60	0.963
16,056,447	266.73	5.84E-60	0.963
16,822,040	266.73	5.84E-60	0.963
16,870,913	266.73	5.84E-60	0.963
30,628,026	266.73	5.84E-60	0.963
36,304,234	266.73	5.84E-60	0.963
38,363,296	266.73	5.84E-60	0.963
38,523,154	266.73	5.84E-60	0.963
45,661,162	266.73	5.84E-60	0.963
62,875,513	266.73	5.84E-60	0.963
67,221,906	266.73	5.84E-60	0.963
8,163,465	266.70	1.22E-58	0.963
16,457,563	266.70	1.22E-58	0.963

Physical Position	Chi-square statistic	p-value	Weighted chi-square
58,894,748	266.67	1.24E-58	0.963
26,931,464	266.65	6.11E-60	0.963
48,473,938	266.65	6.11E-60	0.963
60,889,486	266.65	6.11E-60	0.963
62,477,244	266.65	6.11E-60	0.963
63,720,754	266.65	6.11E-60	0.963
7,709,508	266.64	1.26E-58	0.963
10,336,155	266.55	6.40E-60	0.962
14,396,949	266.55	6.40E-60	0.962
43,894,444	266.55	6.40E-60	0.962
63,519,758	266.55	6.40E-60	0.962
63,696,911	266.51	1.34E-58	0.962
62,379,129	266.46	6.72E-60	0.962
9,528,711	266.31	1.49E-58	0.961
64,523,663	266.13	1.62E-58	0.961
14,841,823	266.09	1.66E-58	0.961
9,298,912	266.03	1.71E-58	0.960
6,496,713	266.00	1.73E-58	0.960
8,937,749	266.00	1.73E-58	0.960
9,978,686	266.00	1.73E-58	0.960
13,470,351	266.00	1.73E-58	0.960
15,955,894	266.00	1.73E-58	0.960
16,560,765	266.00	1.73E-58	0.960
42,415,703	266.00	1.73E-58	0.960
48,584,559	266.00	1.73E-58	0.960
8,092,324	265.75	9.57E-60	0.959
9,236,003	265.75	9.57E-60	0.959
9,602,996	265.75	9.57E-60	0.959

Physical Position	Chi-square statistic	p-value	Weighted chi-square
10,283,831	265.75	9.57E-60	0.959
14,133,770	265.75	9.57E-60	0.959
15,643,879	265.75	9.57E-60	0.959
21,469,544	265.75	9.57E-60	0.959
37,591,078	265.75	9.57E-60	0.959
44,663,063	265.75	9.57E-60	0.959
44,776,578	265.75	9.57E-60	0.959
47,846,279	265.75	9.57E-60	0.959
55,474,934	265.75	9.57E-60	0.959
59,923,897	265.75	9.57E-60	0.959
60,720,822	265.75	9.57E-60	0.959
17,101,936	265.66	1.00E-59	0.959
61,027,277	265.66	1.00E-59	0.959
6,350,425	265.65	2.07E-58	0.959
62,483,935	265.65	2.07E-58	0.959
63,565,015	265.58	2.14E-58	0.959
9,920,639	265.57	1.05E-59	0.959
59,836,727	265.57	1.05E-59	0.959
58,739,038	265.49	2.23E-58	0.958
7,698,427	265.48	1.10E-59	0.958
9,634,281	265.44	2.30E-58	0.958
9,538,764	265.31	2.44E-58	0.958
38,447,275	265.25	2.52E-58	0.958
6,201,763	265.15	2.65E-58	0.957
35,723,443	265.06	2.77E-58	0.957
6,220,943	265.00	2.86E-58	0.957
8,769,379	265.00	2.86E-58	0.957
8,173,768	264.77	1.57E-59	0.956

Physical Position	Chi-square statistic	p-value	Weighted chi-square
10,332,278	264.77	1.57E-59	0.956
13,486,295	264.77	1.57E-59	0.956
17,114,761	264.77	1.57E-59	0.956
28,823,093	264.77	1.57E-59	0.956
48,243,710	264.77	1.57E-59	0.956
45,010,693	264.57	3.54E-58	0.955
15,365,819	264.52	3.63E-58	0.955
41,868,367	264.49	1.80E-59	0.955
14,598,018	264.39	3.89E-58	0.954
43,388,370	264.37	3.91E-58	0.954
16,915,558	264.32	4.02E-58	0.954
8,481,229	264.28	4.09E-58	0.954
44,032,304	264.11	4.46E-58	0.953
60,662,246	264.04	4.61E-58	0.953
9,187,560	264.00	4.71E-58	0.953
10,320,995	264.00	4.71E-58	0.953
35,851,793	264.00	4.71E-58	0.953
45,353,693	264.00	4.71E-58	0.953
8,633,955	263.80	5.21E-58	0.952
6,361,940	263.79	2.57E-59	0.952
38,102,420	263.79	2.57E-59	0.952
58,846,140	263.79	2.57E-59	0.952
36,187,701	263.70	2.68E-59	0.952
48,937,250	263.66	5.58E-58	0.952
16,575,473	263.61	2.81E-59	0.952
14,505,469	263.58	5.81E-58	0.952
7,202,976	263.50	6.04E-58	0.951
45,192,047	263.50	6.05E-58	0.951

Physical Position	Chi-square statistic	p-value	Weighted chi-square
15,727,910	263.41	3.09E-59	0.951
12,761,358	263.37	6.45E-58	0.951
34,963,331	263.34	6.56E-58	0.951
45,064,601	263.26	6.82E-58	0.950
60,925,578	263.20	7.01E-58	0.950
35,560,034	263.03	7.66E-58	0.950
13,381,447	263.00	7.77E-58	0.949
61,040,395	263.00	7.77E-58	0.949
7,335,337	262.80	4.20E-59	0.949
17,509,058	262.80	4.20E-59	0.949
8,747,269	262.50	9.97E-58	0.948
42,433,173	262.16	1.18E-57	0.946
44,927,588	262.14	1.20E-57	0.946
14,184,305	261.87	1.37E-57	0.945
14,604,026	261.81	1.41E-57	0.945
35,635,361	261.47	1.67E-57	0.944
35,489,200	261.41	1.72E-57	0.944
61,091,117	261.23	1.89E-57	0.943
59,752,894	261.08	2.02E-57	0.943
7,268,756	261.06	2.05E-57	0.942
9,958,345	261.00	2.11E-57	0.942
22,745,790	260.84	2.29E-57	0.942
41,562,413	260.68	2.48E-57	0.941
15,623,751	260.57	1.29E-58	0.941
13,214,896	260.54	2.66E-57	0.941
7,536,982	260.45	2.78E-57	0.940
47,116,302	260.44	2.80E-57	0.940
60,666,358	260.44	2.80E-57	0.940

Physical Position	Chi-square statistic	p-value	Weighted chi-square
7,085,708	260.35	2.92E-57	0.940
15,961,072	260.33	2.96E-57	0.940
7,810,315	260.18	3.19E-57	0.939
62,545,488	260.04	3.41E-57	0.939
9,855,433	260.00	3.48E-57	0.939
8,004,212	260.00	3.49E-57	0.939
12,817,125	259.29	4.98E-57	0.936
1,655,775	259.25	5.05E-57	0.936
57,820,295	259.12	5.42E-57	0.935
43,176,057	258.89	6.06E-57	0.935
6,601,278	258.79	3.15E-58	0.934
1,854,534	258.64	6.87E-57	0.934
35,527,622	258.60	7.00E-57	0.934
60,827,885	258.08	9.10E-57	0.932
7,906,557	258.02	9.36E-57	0.931
25,022,288	258.00	9.46E-57	0.931
43,490,640	257.94	9.75E-57	0.931
44,636,462	257.94	9.75E-57	0.931
6,390,095	257.69	1.11E-56	0.930
8,665,743	257.26	1.37E-56	0.929
6,258,252	257.14	1.45E-56	0.928
59,589,886	257.00	1.56E-56	0.928
42,357,254	256.99	1.57E-56	0.928
48,432,937	256.17	2.36E-56	0.925
5,822,233	256.00	2.57E-56	0.924
44,306,787	255.80	2.84E-56	0.923
14,379,166	255.53	3.25E-56	0.923
13,786,668	255.10	4.03E-56	0.921

Physical Position	Chi-square statistic	p-value	Weighted chi-square
43,859,224	255.09	4.06E-56	0.921
7,644,859	255.00	4.24E-56	0.921
7,602,577	255.00	4.24E-56	0.921
8,148,516	254.94	4.36E-56	0.920
68,443,334	254.66	5.02E-56	0.919
61,194,194	254.61	5.14E-56	0.919
23,024,514	254.44	5.60E-56	0.919
5,943,423	254.09	6.68E-56	0.917
8,276,877	254.04	6.84E-56	0.917
5,823,156	254.00	6.99E-56	0.917
10,716,696	253.49	9.00E-56	0.915
48,383,380	253.27	1.01E-55	0.914
42,613,702	253.19	1.05E-55	0.914
7,230,672	252.91	1.21E-55	0.913
24,026,208	252.66	1.37E-55	0.912
62,183,780	252.63	1.38E-55	0.912
13,649,139	252.35	1.60E-55	0.911
46,257,782	252.34	1.61E-55	0.911
42,168,778	252.00	1.90E-55	0.910
61,046,067	251.92	1.98E-55	0.909
47,330,762	251.79	2.11E-55	0.909
50,481,373	251.40	2.57E-55	0.908
45,539,782	251.40	2.57E-55	0.908
45,677,876	251.40	2.57E-55	0.908
48,069,745	251.40	2.57E-55	0.908
6,563,460	251.00	3.13E-55	0.906
5,989,747	251.00	3.13E-55	0.906
12,092,357	250.77	3.52E-55	0.905

Physical Position	Chi-square statistic	p-value	Weighted chi-square
47,987,178	250.47	4.09E-55	0.904
47,870,643	250.16	4.77E-55	0.903

Physical Position	Chi-square statistic	p-value	Weighted chi-square
35,986,521	250.10	4.92E-55	0.903
47,634,945	249.53	6.54E-55	0.901

Physical Position	Chi-square statistic	p-value	Weighted chi-square
9,814,554	249.37	7.08E-55	0.900

Table S2.3:

List of 293 SNP positions which were found to be highly diagnostic (weighted chi-square score >0.9) between the A and C inversion haplotypes.

Physical Position	Chi-square statistic	p-value	Weighted chi-square	Physical Position	Chi-square statistic	p-value	Weighted chi-square	Physical Position	Chi-square statistic	p-value	Weighted chi-square
20,032,394	274.00	3.17E-60	1.000	12,043,419	273.00	5.23E-60	0.996	42,721,543	272.00	8.63E-60	0.993
51,010,922	274.00	3.17E-60	1.000	18,971,443	273.00	5.23E-60	0.996	52,330,306	272.00	8.63E-60	0.993
52,333,760	274.00	3.17E-60	1.000	19,949,068	273.00	5.23E-60	0.996	52,344,785	272.00	8.63E-60	0.993
11,702,004	274.00	3.17E-60	1.000	20,077,133	273.00	5.23E-60	0.996	11,594,377	271.00	1.42E-59	0.989
11,740,993	274.00	3.17E-60	1.000	52,135,504	273.00	5.23E-60	0.996	11,634,635	271.00	1.42E-59	0.989
11,803,650	274.00	3.17E-60	1.000	52,294,743	273.00	5.23E-60	0.996	11,803,939	271.00	1.42E-59	0.989
11,833,469	274.00	3.17E-60	1.000	52,316,759	273.00	5.23E-60	0.996	11,957,145	271.00	1.42E-59	0.989
11,946,483	274.00	3.17E-60	1.000	54,411,607	273.00	5.23E-60	0.996	12,047,381	271.00	1.42E-59	0.989
12,052,304	274.00	3.17E-60	1.000	11,788,078	273.00	5.23E-60	0.996	55,535,495	271.00	1.42E-59	0.989
17,847,415	274.00	3.17E-60	1.000	11,939,714	273.00	5.23E-60	0.996	17,996,376	271.00	1.42E-59	0.989
17,873,122	274.00	3.17E-60	1.000	26,844,231	273.00	5.23E-60	0.996	11,743,387	270.00	2.35E-59	0.985
17,877,506	274.00	3.17E-60	1.000	11,524,034	272.00	8.63E-60	0.993	11,986,275	270.00	2.35E-59	0.985
17,899,164	274.00	3.17E-60	1.000	11,542,643	272.00	8.63E-60	0.993	11,987,145	270.00	2.35E-59	0.985
31,446,386	274.00	3.17E-60	1.000	11,583,883	272.00	8.63E-60	0.993	19,546,473	270.00	2.35E-59	0.985
32,344,591	274.00	3.17E-60	1.000	11,620,838	272.00	8.63E-60	0.993	31,602,576	270.00	2.35E-59	0.985
55,533,253	274.00	3.17E-60	1.000	11,795,120	272.00	8.63E-60	0.993	68,884,033	270.00	2.35E-59	0.985
11,595,909	273.00	5.23E-60	0.996	11,909,169	272.00	8.63E-60	0.993	16,386,368	269.26	3.40E-59	0.983
11,620,348	273.00	5.23E-60	0.996	11,966,810	272.00	8.63E-60	0.993	11,548,152	269.00	3.87E-59	0.982
17,493,272	273.00	5.23E-60	0.996	12,019,563	272.00	8.63E-60	0.993	17,832,633	269.00	3.87E-59	0.982
17,837,069	273.00	5.23E-60	0.996	18,811,280	272.00	8.63E-60	0.993	14,307,648	269.00	3.87E-59	0.982
18,037,707	273.00	5.23E-60	0.996	25,211,613	272.00	8.63E-60	0.993	17,909,534	269.00	3.87E-59	0.982
18,065,373	273.00	5.23E-60	0.996	25,257,105	272.00	8.63E-60	0.993	9,172,182	268.94	3.98E-59	0.982
29,226,605	273.00	5.23E-60	0.996	28,951,894	272.00	8.63E-60	0.993	14,019,360	268.94	3.98E-59	0.982
55,667,298	273.00	5.23E-60	0.996	29,767,034	272.00	8.63E-60	0.993	16,923,279	268.86	4.15E-59	0.981
10,573,835	273.00	5.23E-60	0.996	32,377,583	272.00	8.63E-60	0.993	52,322,633	268.86	4.15E-59	0.981

Physical Position	Chi-square statistic	p-value	Weighted chi-square
11,417,108	268.47	5.04E-59	0.980
11,402,084	268.38	5.27E-59	0.979
11,439,964	268.38	5.27E-59	0.979
18,066,380	268.21	5.75E-59	0.979
54,797,458	268.21	5.75E-59	0.979
11,806,679	268.00	6.38E-59	0.978
11,945,664	268.00	6.38E-59	0.978
46,696,893	268.00	6.38E-59	0.978
17,957,484	268.00	6.38E-59	0.978
64,862,034	268.00	6.38E-59	0.978
68,605,605	267.96	6.51E-59	0.978
52,749,793	267.87	6.82E-59	0.978
18,268,600	267.81	7.01E-59	0.977
11,366,826	267.80	7.04E-59	0.977
31,914,960	267.80	7.06E-59	0.977
11,729,810	267.72	3.57E-60	0.977
11,737,647	267.72	3.57E-60	0.977
11,751,848	267.72	3.57E-60	0.977
11,857,148	267.72	3.57E-60	0.977
11,876,024	267.72	3.57E-60	0.977
17,904,749	267.72	3.57E-60	0.977
17,952,800	267.72	3.57E-60	0.977
18,028,954	267.72	3.57E-60	0.977
29,202,254	267.72	3.57E-60	0.977
30,263,329	267.72	3.57E-60	0.977
31,203,476	267.72	3.57E-60	0.977
38,893,654	267.72	3.57E-60	0.977
41,255,029	267.72	3.57E-60	0.977

Physical Position	Chi-square statistic	p-value	Weighted chi-square
49,407,383	267.72	3.57E-60	0.977
52,352,191	267.72	3.57E-60	0.977
52,356,137	267.72	3.57E-60	0.977
52,363,718	267.72	3.57E-60	0.977
54,640,705	267.72	3.57E-60	0.977
54,797,352	267.72	3.57E-60	0.977
55,482,435	267.72	3.57E-60	0.977
55,482,632	267.72	3.57E-60	0.977
55,613,969	267.72	3.57E-60	0.977
55,619,362	267.72	3.57E-60	0.977
55,809,486	267.72	3.57E-60	0.977
64,075,579	267.72	3.57E-60	0.977
67,216,608	267.72	3.57E-60	0.977
68,671,224	267.68	7.48E-59	0.977
68,789,853	267.68	7.48E-59	0.977
68,620,839	267.60	7.79E-59	0.977
68,349,390	267.49	8.25E-59	0.976
10,573,181	267.00	1.05E-58	0.974
16,925,095	267.00	1.05E-58	0.974
16,938,557	266.90	1.10E-58	0.974
19,889,432	266.83	1.14E-58	0.974
18,990,798	266.82	1.15E-58	0.974
11,367,099	266.82	1.15E-58	0.974
42,468,910	266.81	1.15E-58	0.974
11,639,247	266.73	5.84E-60	0.973
11,887,578	266.73	5.84E-60	0.973
17,975,605	266.73	5.84E-60	0.973
29,225,896	266.73	5.84E-60	0.973

Physical Position	Chi-square statistic	p-value	Weighted chi-square
29,250,516	266.73	5.84E-60	0.973
52,613,763	266.73	5.84E-60	0.973
52,613,793	266.73	5.84E-60	0.973
55,609,854	266.73	5.84E-60	0.973
55,627,501	266.73	5.84E-60	0.973
55,685,837	266.73	5.84E-60	0.973
56,610,878	266.73	5.84E-60	0.973
65,252,835	266.73	5.84E-60	0.973
65,688,896	266.73	5.84E-60	0.973
68,795,231	266.70	1.22E-58	0.973
9,996,940	266.61	1.28E-58	0.973
68,491,287	266.61	1.28E-58	0.973
68,778,587	266.61	1.28E-58	0.973
33,653,047	266.50	1.35E-58	0.973
52,333,397	266.42	1.41E-58	0.972
9,400,138	266.30	1.49E-58	0.972
18,065,611	266.00	1.73E-58	0.971
56,026,995	266.00	1.73E-58	0.971
68,884,283	266.00	1.73E-58	0.971
11,619,727	265.99	1.75E-58	0.971
18,515,673	265.83	1.88E-58	0.970
64,012,306	265.83	1.89E-58	0.970
11,706,217	265.75	9.57E-60	0.970
11,873,991	265.75	9.57E-60	0.970
28,958,928	265.75	9.57E-60	0.970
43,468,605	265.75	9.57E-60	0.970
65,435,760	265.75	9.57E-60	0.970
68,815,607	265.72	2.00E-58	0.970

Physical Position	Chi-square statistic	p-value	Weighted chi-square
11,649,218	265.66	1.00E-59	0.970
12,009,714	265.66	1.00E-59	0.970
17,955,491	265.66	1.00E-59	0.970
24,216,827	265.66	1.00E-59	0.970
27,056,805	265.66	1.00E-59	0.970
29,204,457	265.66	1.00E-59	0.970
68,393,072	265.57	2.14E-58	0.969
11,412,456	265.44	2.30E-58	0.969
68,716,301	265.14	2.67E-58	0.968
29,253,948	265.00	2.86E-58	0.967
52,320,191	265.00	2.86E-58	0.967
66,869,612	265.00	2.86E-58	0.967
55,285,669	265.00	2.86E-58	0.967
12,026,269	264.86	3.07E-58	0.967
67,104,453	264.68	1.64E-59	0.966
68,418,580	264.60	3.50E-58	0.966
29,048,233	264.59	1.71E-59	0.966
52,414,459	264.59	1.71E-59	0.966
68,472,636	264.46	3.75E-58	0.965
68,404,234	264.24	4.18E-58	0.964
68,678,461	264.24	4.18E-58	0.964
11,419,425	264.23	4.20E-58	0.964
68,259,624	264.17	4.33E-58	0.964
68,667,080	264.17	4.33E-58	0.964
10,544,484	263.80	5.21E-58	0.963
11,621,631	263.79	2.57E-59	0.963
17,130,283	263.18	7.09E-58	0.961
68,121,004	263.15	7.22E-58	0.960

Physical Position	Chi-square statistic	p-value	Weighted chi-square
12,043,166	262.98	7.84E-58	0.960
11,368,251	262.89	8.22E-58	0.959
57,306,739	262.88	8.24E-58	0.959
11,893,916	262.80	4.20E-59	0.959
68,633,881	262.77	8.73E-58	0.959
67,757,827	262.77	8.73E-58	0.959
55,523,457	262.72	4.39E-59	0.959
68,041,779	262.69	9.06E-58	0.959
53,080,853	262.64	9.28E-58	0.959
11,883,745	262.53	4.82E-59	0.958
11,933,574	262.00	1.28E-57	0.956
68,887,401	262.00	1.28E-57	0.956
11,971,606	262.00	1.28E-57	0.956
52,330,534	261.97	1.30E-57	0.956
14,050,014	261.90	1.35E-57	0.956
14,050,553	261.90	1.35E-57	0.956
13,963,399	261.87	1.36E-57	0.956
52,393,018	261.82	6.88E-59	0.956
18,124,083	261.81	1.41E-57	0.956
68,042,593	261.81	1.41E-57	0.956
68,837,118	261.77	1.44E-57	0.955
57,237,769	261.73	1.47E-57	0.955
68,737,847	261.72	1.48E-57	0.955
18,166,932	261.45	8.28E-59	0.954
68,241,396	261.41	1.72E-57	0.954
56,882,151	260.91	2.20E-57	0.952
13,961,628	260.88	2.24E-57	0.952
68,809,341	260.88	2.24E-57	0.952

Physical Position	Chi-square statistic	p-value	Weighted chi-square
68,055,465	260.86	2.26E-57	0.952
68,679,734	260.86	2.26E-57	0.952
34,231,786	260.85	2.28E-57	0.952
13,991,640	260.84	2.28E-57	0.952
68,282,788	260.81	2.32E-57	0.952
68,480,136	260.80	2.33E-57	0.952
68,809,576	260.80	2.33E-57	0.952
68,312,638	260.59	2.59E-57	0.951
68,686,556	260.54	2.66E-57	0.951
11,862,995	260.44	2.80E-57	0.951
68,634,076	260.24	3.08E-57	0.950
46,426,586	260.00	3.48E-57	0.949
13,994,104	259.92	3.63E-57	0.949
18,108,206	259.88	3.70E-57	0.948
21,528,012	259.88	3.70E-57	0.948
11,921,505	259.86	3.73E-57	0.948
17,932,672	259.86	1.85E-58	0.948
18,454,559	259.85	3.75E-57	0.948
67,758,932	259.65	4.15E-57	0.948
13,964,200	259.59	4.26E-57	0.947
11,440,167	259.58	4.29E-57	0.947
68,346,959	259.42	4.65E-57	0.947
68,729,091	259.12	5.40E-57	0.946
57,004,208	258.95	5.89E-57	0.945
18,096,255	258.91	5.99E-57	0.945
14,267,144	258.87	6.14E-57	0.945
68,701,590	258.69	6.71E-57	0.944
68,013,068	258.39	7.80E-57	0.943

Physical Position	Chi-square statistic	p-value	Weighted chi-square
68,407,626	258.25	8.35E-57	0.943
18,944,902	258.03	9.33E-57	0.942
68,752,564	257.98	9.54E-57	0.942
11,409,951	257.80	5.17E-58	0.941
55,266,633	257.44	1.25E-56	0.940
12,029,176	257.17	1.43E-56	0.939
11,688,162	257.00	1.56E-56	0.938
11,881,390	256.91	8.10E-58	0.938
11,848,011	256.79	1.74E-56	0.937
17,985,725	256.50	2.01E-56	0.936
13,947,503	255.92	2.67E-56	0.934
10,140,965	255.73	2.95E-56	0.933
68,298,333	255.70	3.00E-56	0.933
52,903,772	255.62	3.11E-56	0.933
68,827,829	255.61	3.12E-56	0.933
68,305,894	255.57	3.19E-56	0.933
11,973,583	255.09	4.06E-56	0.931

Physical Position	Chi-square statistic	p-value	Weighted chi-square
13,947,803	254.97	4.30E-56	0.931
13,948,224	254.97	4.30E-56	0.931
11,971,762	254.94	2.17E-57	0.930
13,989,498	254.88	4.51E-56	0.930
17,845,089	254.86	2.27E-57	0.930
68,605,046	254.77	2.37E-57	0.930
68,685,388	254.70	4.93E-56	0.930
68,503,852	254.31	5.98E-56	0.928
68,120,842	254.11	6.61E-56	0.927
18,099,317	254.08	6.72E-56	0.927
68,555,502	253.93	7.23E-56	0.927
11,630,734	253.86	7.48E-56	0.927
68,740,328	253.79	7.76E-56	0.926
52,372,178	253.60	4.26E-57	0.926
68,651,898	253.51	8.93E-56	0.925
19,632,177	253.45	9.19E-56	0.925
64,012,430	253.24	1.02E-55	0.924

Physical Position	Chi-square statistic	p-value	Weighted chi-square
11,615,362	253.19	1.05E-55	0.924
13,585,317	252.40	1.56E-55	0.921
68,764,721	251.34	2.64E-55	0.917
8,942,648	250.89	3.31E-55	0.916
18,124,371	250.58	3.87E-55	0.915
57,006,529	250.51	4.00E-55	0.914
12,000,022	250.47	4.09E-55	0.914
21,444,381	250.47	4.09E-55	0.914
18,753,176	249.93	5.35E-55	0.912
55,274,215	249.78	5.76E-55	0.912
65,047,460	249.53	6.54E-55	0.911
29,239,014	248.60	5.25E-56	0.907
67,879,328	248.47	1.11E-54	0.907
68,043,753	248.43	1.13E-54	0.907
67,776,818	247.75	1.59E-54	0.904
67,992,771	246.65	2.75E-54	0.900

Table S2.4:

List of 137 scaffolds from the *bTGI_alt* reference assembly. The number of diagnostic SNP positions found on each scaffold are used to determine an overall percentage support for each scaffold being either Haplotype A or B sequence.

Scaffold	n SNPs (A)	n SNPs (B)	Total SNPs	Weighted % (A)	Weighted % (B)	Inferred Type
000058F_015_arrow	210	3	213	98.62	1.38	A
000056F_016_arrow	193	1	194	99.48	0.52	A
000056F_003_arrow	55	0	55	100.00	0.00	A
000056F_027_arrow	52	0	52	100.00	0.00	A
000011F_008_arrow	42	1	43	97.85	2.15	A
000011F_152_arrow	34	1	35	97.12	2.88	A
000056F_011_arrow	29	1	30	96.69	3.31	A
000143F_011_arrow	23	0	23	100.00	0.00	A
000011F_082_arrow	21	0	21	100.00	0.00	A
000011F_149_arrow	20	1	21	95.42	4.58	A
000011F_130_arrow	15	5	20	75.18	24.82	A
000351F_003_arrow	18	0	18	100.00	0.00	A
000056F_021_arrow	17	0	17	100.00	0.00	A
000011F_059_arrow	16	1	17	94.12	5.88	A
000011F_066_arrow	14	1	15	93.80	6.20	A
000011F_121_arrow	14	0	14	100.00	0.00	A
000346F_001_arrow	11	0	11	100.00	0.00	A
000056F_014_arrow	7	2	9	77.75	22.25	A
000011F_069_arrow	8	0	8	100.00	0.00	A
000011F_011_arrow	4	3	7	57.14	42.86	A
000091F_015_arrow	5	2	7	70.68	29.32	A
000058F_006_arrow	6	0	6	100.00	0.00	A
000011F_045_arrow	5	0	5	100.00	0.00	A
000056F_023_arrow	5	0	5	100.00	0.00	A

Scaffold	n SNPs (A)	n SNPs (B)	Total SNPs	Weighted % (A)	Weighted % (B)	Inferred Type
000058F_019_arrow	4	1	5	79.89	20.11	A
000058F_003_arrow	4	0	4	100.00	0.00	A
000057F_040_arrow	4	0	4	100.00	0.00	A
000011F_024_arrow	2	2	4	50.87	49.13	A
000056F_012_arrow	4	0	4	100.00	0.00	A
000011F_133_arrow	3	1	4	75.03	24.97	A
000011F_140_arrow	4	0	4	100.00	0.00	A
000084F_011_arrow	2	1	3	66.46	33.54	A
000011F_026_arrow	2	1	3	66.36	33.64	A
000057F_024_arrow	3	0	3	100.00	0.00	A
000084F_028_arrow	3	0	3	100.00	0.00	A
000058F_017_arrow	2	1	3	67.82	32.18	A
000087F_022_arrow	2	1	3	66.68	33.32	A
000143F_008_arrow	3	0	3	100.00	0.00	A
000011F_058_arrow	3	0	3	100.00	0.00	A
000157F_012_arrow	3	0	3	100.00	0.00	A
000011F_134_arrow	3	0	3	100.00	0.00	A
000057F_009_arrow	3	0	3	100.00	0.00	A
000056F_025_arrow	3	0	3	100.00	0.00	A
000011F_031_arrow	3	0	3	100.00	0.00	A
000098F_009_arrow	3	0	3	100.00	0.00	A
000142F_008_arrow	3	0	3	100.00	0.00	A
000147F_013_arrow	1	1	2	50.21	49.79	A
000057F_002_arrow	1	1	2	50.11	49.89	A
000032F_028_arrow	2	0	2	100.00	0.00	A
000142F_001_arrow	1	1	2	50.23	49.77	A
000032F_057_arrow	2	0	2	100.00	0.00	A

Scaffold	n SNPs (A)	n SNPs (B)	Total SNPs	Weighted % (A)	Weighted % (B)	Inferred Type
000066F_006_arrow	2	0	2	100.00	0.00	A
000058F_018_arrow	2	0	2	100.00	0.00	A
000057F_005_arrow	2	0	2	100.00	0.00	A
000011F_175_arrow	2	0	2	100.00	0.00	A
000058F_020_arrow	1	1	2	50.23	49.77	A
000058F_009_arrow	2	0	2	100.00	0.00	A
000032F_034_arrow	1	1	2	50.51	49.49	A
000165F_005_arrow	2	0	2	100.00	0.00	A
000147F_005_arrow	1	1	2	50.69	49.31	A
000011F_023_arrow	2	0	2	100.00	0.00	A
000011F_123_arrow	2	0	2	100.00	0.00	A
000157F_009_arrow	2	0	2	100.00	0.00	A
000057F_020_arrow	2	0	2	100.00	0.00	A
000011F_064_arrow	2	0	2	100.00	0.00	A
000058F_002_arrow	2	0	2	100.00	0.00	A
000066F_013_arrow	1	1	2	50.47	49.53	A
000311F_003_arrow	2	0	2	100.00	0.00	A
000311F_002_arrow	2	0	2	100.00	0.00	A
000311F_001_arrow	2	0	2	100.00	0.00	A
000066F_011_arrow	1	1	2	50.27	49.73	A
000032F_067_arrow	1	1	2	50.05	49.95	A
000087F_034_arrow	2	0	2	100.00	0.00	A
000011F_128_arrow	2	0	2	100.00	0.00	A
000098F_012_arrow	2	0	2	100.00	0.00	A
000057F_007_arrow	2	0	2	100.00	0.00	A
000058F_011_arrow	0	362	362	0.00	100.00	B
000056F_031_arrow	0	133	133	0.00	100.00	B

Scaffold	n SNPs (A)	n SNPs (B)	Total SNPs	Weighted % (A)	Weighted % (B)	Inferred Type
000011F_039_arrow	0	45	45	0.00	100.00	B
000011F_043_arrow	0	40	40	0.00	100.00	B
000143F_010_arrow	2	37	39	5.03	94.97	B
000011F_178_arrow	0	35	35	0.00	100.00	B
000157F_007_arrow	0	35	35	0.00	100.00	B
000143F_002_arrow	0	34	34	0.00	100.00	B
000057F_025_arrow	0	31	31	0.00	100.00	B
000032F_004_arrow	0	28	28	0.00	100.00	B
000058F_013_arrow	0	13	13	0.00	100.00	B
000011F_033_arrow	0	11	11	0.00	100.00	B
000165F_002_arrow	0	11	11	0.00	100.00	B
000011F_131_arrow	0	8	8	0.00	100.00	B
000091F_001_arrow	2	5	7	28.06	71.94	B
000057F_041_arrow	0	5	5	0.00	100.00	B
000143F_014_arrow	0	5	5	0.00	100.00	B
000058F_007_arrow	0	5	5	0.00	100.00	B
000056F_026_arrow	0	5	5	0.00	100.00	B
000011F_083_arrow	0	5	5	0.00	100.00	B
000098F_001_arrow	0	5	5	0.00	100.00	B
000011F_151_arrow	0	4	4	0.00	100.00	B
000087F_026_arrow	1	3	4	25.19	74.81	B
000011F_086_arrow	2	2	4	49.91	50.09	B
000011F_075_arrow	2	2	4	49.64	50.36	B
000011F_052_arrow	1	3	4	25.91	74.09	B
000098F_005_arrow	2	2	4	49.95	50.05	B
000058F_021_arrow	0	4	4	0.00	100.00	B
000011F_016_arrow	1	2	3	33.93	66.07	B

Scaffold	n SNPs (A)	n SNPs (B)	Total SNPs	Weighted % (A)	Weighted % (B)	Inferred Type
000087F_015_arrow	0	3	3	0.00	100.00	B
000011F_154_arrow	0	3	3	0.00	100.00	B
000057F_031_arrow	0	3	3	0.00	100.00	B
000351F_001_arrow	1	2	3	32.75	67.25	B
000056F_020_arrow	0	2	2	0.00	100.00	B
000157F_003_arrow	0	2	2	0.00	100.00	B
000057F_014_arrow	0	2	2	0.00	100.00	B
000057F_011_arrow	0	2	2	0.00	100.00	B
000351F_002_arrow	1	1	2	49.72	50.28	B
000098F_013_arrow	0	2	2	0.00	100.00	B
000011F_170_arrow	0	2	2	0.00	100.00	B
000011F_157_arrow	0	2	2	0.00	100.00	B
000011F_107_arrow	0	2	2	0.00	100.00	B
000084F_019_arrow	0	2	2	0.00	100.00	B
000087F_033_arrow	0	2	2	0.00	100.00	B
000032F_054_arrow	1	1	2	48.84	51.16	B
scaffold_52_arrow_ctg1	0	2	2	0.00	100.00	B
000147F_011_arrow	0	2	2	0.00	100.00	B
000098F_017_arrow	0	2	2	0.00	100.00	B
000011F_110_arrow	0	2	2	0.00	100.00	B
000011F_153_arrow	0	2	2	0.00	100.00	B
000011F_150_arrow	0	2	2	0.00	100.00	B
000066F_002_arrow	0	2	2	0.00	100.00	B
000011F_090_arrow	0	2	2	0.00	100.00	B
scaffold_213_arrow_ctg1	0	2	2	0.00	100.00	B
000011F_165_arrow	0	2	2	0.00	100.00	B
000011F_147_arrow	0	2	2	0.00	100.00	B

Scaffold	n SNPs (A)	n SNPs (B)	Total SNPs	Weighted % (A)	Weighted % (B)	Inferred Type
000011F_124_arrow	0	2	2	0.00	100.00	B
000011F_164_arrow	0	2	2	0.00	100.00	B
000057F_013_arrow	0	2	2	0.00	100.00	B
000058F_010_arrow	0	2	2	0.00	100.00	B
000066F_010_arrow	0	2	2	0.00	100.00	B

Figure S2.1:

Principal component analysis of 1202 male zebra finches from the University of Sheffield. The analysis used genotypes from 1085 diagnostic SNP positions with weighted chi-square score >0.9 in a comparison between inversion haplotypes B and C.



Figure S2.2:

Principal component analysis of 1202 male zebra finches from the University of Sheffield. The analysis used genotypes from 293 diagnostic SNP positions with weighted chi-square score >0.9 in a comparison between inversion haplotypes A and C.

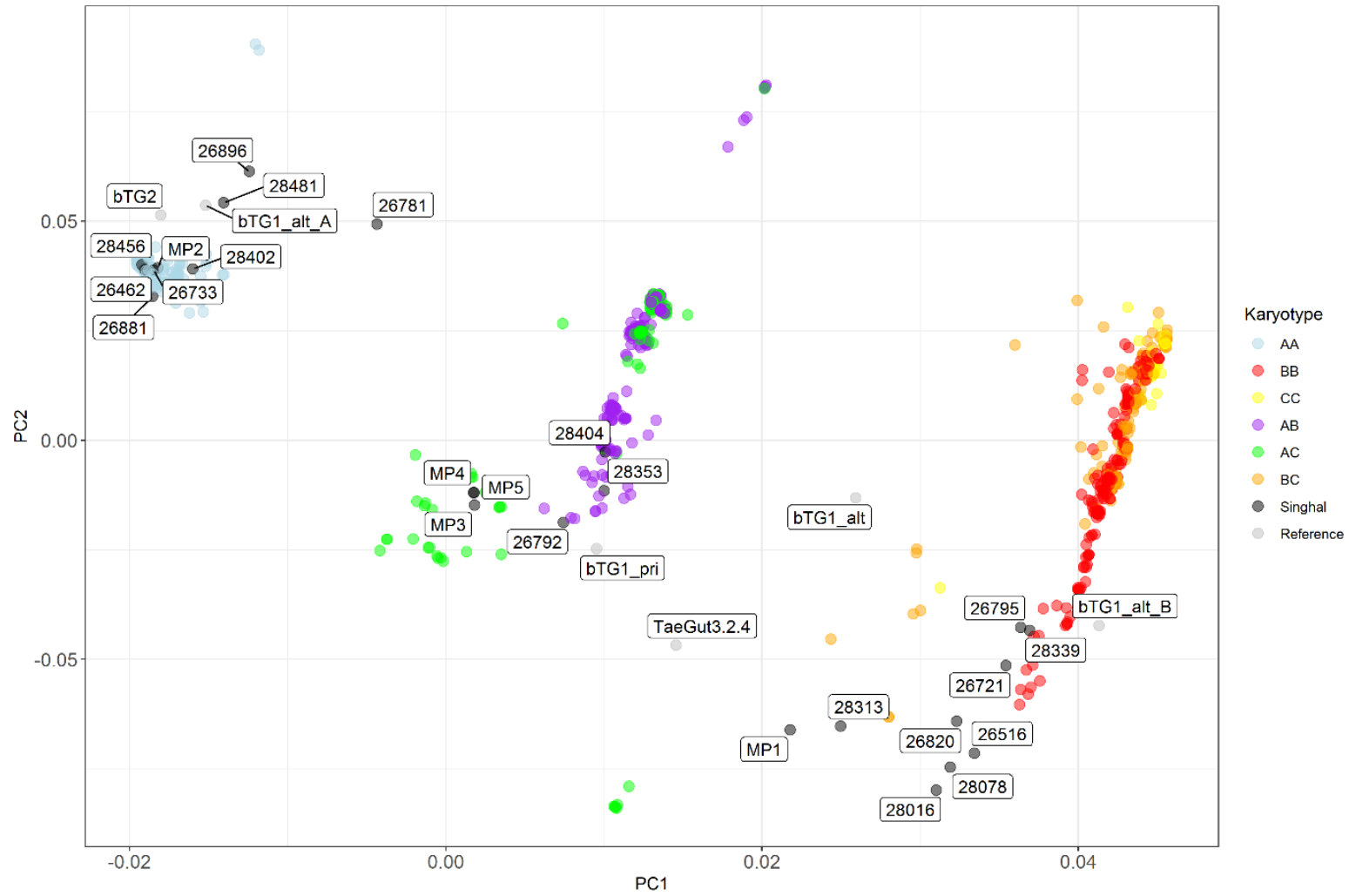


Table S3.1:

List of zebra finch genes on the Z chromosome with at least one fixed difference between the A and B inversion haplotypes within their coding regions. Genes are ordered according to the number of non-synonymous fixed differences they contain.

Gene Name	CDS Start	CDS End	P(A) non- syn	P(A) syn	P(B) non- syn	P(B) syn	D non- syn	D syn
ANKRD31	60,332,675	60,397,704	2	1	1	0	11	4
RIC1	63,690,131	63,737,954	1	1	0	1	8	2
CMYA5	58,398,352	58,444,093	1	0	1	3	7	2
FAM169A	60,471,809	60,510,269	1	0	0	0	6	4
VPS13A	57,618,036	57,727,092	3	0	0	0	5	2
PRUNE2	57,830,656	57,934,812	5	0	1	0	5	1
LOC115491269	51,994,968	52,007,313	3	0	1	0	4	1
FANCC	10,340,132	10,414,964	9	5	1	0	3	0
LOC101233669	11,911,053	11,929,338	11	6	6	7	3	0
LOC115491137	17,468,602	17,481,437	8	18	2	0	3	0
SPEF2	44,075,347	44,148,014	1	2	1	0	3	0
LOC116807240	58,182,748	58,223,849	19	11	24	12	3	1
POLK	60,192,954	60,223,712	1	1	0	0	3	0
LOC105760877	63,385,678	63,392,309	6	1	1	0	3	3
ERCC6L2	10,045,621	10,094,639	13	2	1	0	2	2
CEMIP2	12,585,077	12,635,615	3	3	0	0	2	2
TRPM6	13,594,760	13,674,816	10	17	0	1	2	2
TTC37	14,089,091	14,157,563	10	20	1	0	2	2
CSNK1G3	24,313,954	24,393,225	4	0	0	0	2	1
CEP120	24,407,941	24,449,369	3	2	0	0	2	0
ELAVL2	30,950,329	31,030,898	0	0	0	0	2	0
LOC115491242	32,008,270	32,023,745	10	3	8	4	2	1
CPLANE1	44,725,613	44,784,876	1	1	0	1	2	1
IL31RA	50,129,304	50,163,690	1	0	1	1	2	0
MAP3K1	50,463,251	50,567,501	5	2	0	0	2	3
LOC121468155	50,616,533	50,623,614	2	0	0	0	2	0
UBQLN1	55,535,439	55,571,637	1	0	0	0	2	0
JMY	58,602,658	58,666,117	0	2	0	0	2	1
ARSB	58,763,042	58,833,658	0	0	0	0	2	0
IQGAP2	59,738,189	59,862,699	1	4	0	0	2	2
GFM2	60,519,279	60,538,444	1	0	0	1	2	3
MAP1B	61,608,469	61,681,181	1	1	2	0	2	2
DOCK8	61,927,477	62,018,757	2	1	0	2	2	0
DMRT1	62,175,397	62,238,344	0	0	2	0	2	2
LOC105760879	62,937,789	62,939,378	3	1	0	0	2	1
GLIS3	63,134,435	63,272,713	0	1	0	1	2	1
ADAMTSL1	68,723,619	69,113,720	4	3	2	1	2	1
LOC115491274	9,775,718	9,785,866	2	5	0	0	1	0
SLC35D2	9,918,439	9,941,404	0	3	0	0	1	0

Gene Name	CDS Start	CDS End	P(A) non- syn	P(A) syn	P(B) non- syn	P(B) syn	D non- syn	D syn
LOC100219012	10,643,506	10,664,011	0	4	0	1	1	1
TUT7	11,191,396	11,224,524	8	15	0	0	1	0
CZH9orf40	13,695,217	13,701,371	5	3	0	0	1	0
NMRK1	13,726,040	13,735,545	8	3	0	0	1	1
ARSK	14,055,491	14,089,573	11	16	1	1	1	1
GRIN3A	17,904,339	17,970,973	2	7	0	0	1	0
SLCO4C1	18,711,426	18,738,533	2	0	0	0	1	0
PAM	18,815,649	18,931,899	0	2	0	0	1	0
GIN1	18,938,580	18,945,688	3	0	0	0	1	1
LOC115490958	24,351,806	24,352,726	3	0	0	0	1	1
LOC115491159	24,960,906	24,973,043	2	1	0	0	1	0
SMC2	29,074,169	29,095,147	0	0	0	0	1	0
TOPORS	29,752,478	29,757,793	0	0	0	0	1	0
PGGT1B	30,158,662	30,196,595	1	0	0	0	1	0
LOC115491124	30,719,441	30,722,225	0	0	0	0	1	0
LOC115491125	32,040,080	32,058,967	1	0	0	0	1	1
LINGO2	32,263,451	32,436,573	0	0	0	0	1	0
B4GALT1	34,174,138	34,201,950	0	0	0	0	1	0
TDRD7	34,384,024	34,430,220	0	1	1	0	1	1
DGKQ	34,441,623	34,529,265	1	0	0	0	1	1
ACO1	34,549,677	34,593,009	0	0	0	0	1	0
DDX58	34,585,249	34,609,253	0	0	0	0	1	2
LOC100217918	34,713,314	34,717,900	0	0	0	0	1	0
BDP1	34,936,049	34,988,648	0	0	1	1	1	1
LOC100221697	34,995,141	35,000,474	0	0	0	0	1	0
SETBP1	35,978,762	36,249,191	2	2	0	1	1	0
KIAA1328	40,183,801	40,359,677	1	0	0	0	1	0
DNAI1	40,976,255	41,112,882	0	0	0	0	1	0
LOC115491047	40,986,475	41,025,114	0	0	0	0	1	0
UNC13B	41,870,579	42,078,758	3	5	0	1	1	0
ADAMTS12	43,465,253	43,612,329	2	0	0	1	1	0
PRLR	43,891,870	44,075,303	0	1	0	0	1	0
LIFR	45,247,582	45,302,959	0	1	2	0	1	1
FYB1	45,510,057	45,554,909	2	1	0	0	1	0
PRKAA1	46,127,041	46,146,673	0	1	0	0	1	0
C7	46,171,258	46,195,484	0	0	1	1	1	0
OXCT1	46,412,371	46,491,189	1	1	0	0	1	0
MOCS2	49,047,758	49,056,166	0	0	0	0	1	0
FST	49,198,861	49,205,572	0	0	0	0	1	0
CDC20B	49,841,873	49,861,363	1	0	0	0	1	1
CCNO	49,891,263	49,893,614	1	0	0	0	1	0
IL6ST	50,176,185	50,208,724	2	1	0	0	1	0
SETD9	50,570,566	50,578,060	0	3	0	0	1	1

Gene Name	CDS Start	CDS End	P(A) non- syn	P(A) syn	P(B) non- syn	P(B) syn	D non- syn	D syn
MIER3	50,583,831	50,606,763	1	1	0	0	1	0
ELOVL7	51,941,585	51,973,259	0	0	0	0	1	1
CWC27	53,293,660	53,390,001	1	1	0	0	1	0
TRAPPC13	53,657,297	53,685,764	0	0	0	0	1	0
SGTB	53,681,764	53,705,521	0	0	1	0	1	0
ERBIN	53,780,910	53,898,348	0	3	0	1	1	0
RASEF	55,753,189	55,786,779	0	4	1	0	1	0
THBS4	58,276,212	58,321,353	0	0	0	0	1	1
TENT2	58,444,687	58,491,848	0	0	0	0	1	0
LOC115491249	59,444,335	59,456,787	2	0	0	0	1	0
PDE8B	59,484,131	59,555,060	0	1	0	0	1	0
LOC116806746	59,488,515	59,506,081	0	0	0	0	1	0
AGGF1	59,592,131	59,614,776	0	3	0	1	1	0
F2R	59,726,260	59,734,936	0	0	0	0	1	1
F2RL2	59,768,620	59,773,164	0	2	0	0	1	1
POC5	60,134,705	60,163,319	0	0	1	0	1	0
ANKDD1B	60,163,320	60,192,809	1	1	0	0	1	1
FCHO2	61,242,477	61,324,258	0	0	0	0	1	2
TNPO1	61,333,539	61,409,064	3	1	0	1	1	0
SMARCA2	62,579,399	62,695,750	1	5	0	0	1	1
PUM3	62,894,706	62,918,177	0	0	0	1	1	1
LOC100222616	62,928,264	62,932,371	2	0	0	0	1	0
RCL1	63,457,818	63,491,278	1	0	0	0	1	1
ERMP1	63,741,886	63,762,562	2	4	0	1	1	2
KIAA2026	63,803,187	63,853,533	5	3	0	1	1	0
KDM4C	64,065,189	64,312,746	1	1	1	0	1	0
NFIB	67,059,101	67,229,064	1	0	0	1	1	0
ACER2	69,285,478	69,300,294	3	1	0	0	1	0
SYK	7,198,078	7,247,590	1	6	0	0	0	1
GADD45G	7,643,488	7,645,206	0	3	0	0	0	1
SEMA4D	7,686,393	7,783,343	6	11	1	0	0	1
SPIN1	8,069,935	8,131,270	0	1	0	0	0	1
LRRC2	9,589,261	9,665,621	0	1	0	0	0	1
PTCH1	10,240,866	10,313,961	3	11	1	1	0	1
AOPEP	10,423,689	10,604,075	7	7	0	0	0	1
DAPK1	10,709,668	10,795,646	5	12	0	0	0	3
CENPH	11,295,835	11,303,511	0	0	0	0	0	1
SMC5	12,041,726	12,081,093	2	4	0	0	0	1
TRPM3	12,146,479	12,549,693	2	4	1	0	0	1
RORB	13,440,189	13,577,903	1	4	0	0	0	1
MCTP1	14,196,342	14,410,870	0	6	1	0	0	1
NR2F1	14,983,500	14,996,478	0	1	0	0	0	1
ADGRV1	15,962,038	16,215,632	30	45	3	2	0	1

Gene Name	CDS Start	CDS End	P(A) non- syn	P(A) syn	P(B) non- syn	P(B) syn	D non- syn	D syn
TMEM161B	17,124,388	17,166,779	0	0	0	1	0	1
LOC115491106	18,061,152	18,077,645	0	0	0	0	0	1
ALDOB	18,069,863	18,079,832	0	0	0	0	0	1
MACIR	19,002,876	19,020,665	9	6	4	0	0	1
EFNA5	20,473,380	20,680,712	0	1	0	0	0	1
FBXL17	20,737,720	21,018,550	0	1	0	0	0	1
FER	21,102,415	21,263,204	0	1	0	0	0	1
EPB41L4A	22,292,081	22,409,217	3	4	0	1	0	1
SRP19	22,590,602	22,595,136	0	0	0	0	0	1
MCC	22,648,364	22,829,928	1	1	1	0	0	1
APBA1	22,864,261	22,959,912	0	2	0	0	0	1
KIAA1958	28,536,066	28,602,450	1	2	0	0	0	1
SVEP1	29,166,037	29,283,514	3	0	1	0	0	2
GRAMD2B	29,811,333	29,846,775	1	0	0	0	0	1
FEM1C	30,022,819	30,119,403	0	0	1	0	0	1
CCDC112	30,149,241	30,158,320	0	0	1	0	0	1
LOC115490934	30,157,076	30,158,669	0	0	0	0	0	1
TEK	31,917,924	31,945,291	0	0	0	0	0	1
NCBP1	34,306,938	34,338,210	0	0	0	0	0	2
TMOD1	34,352,055	34,378,848	0	0	0	0	0	1
LOC100217858	35,013,774	35,025,733	0	0	0	0	0	1
CZH18orf25	35,396,436	35,426,145	5	25	7	31	0	1
CELF4	39,445,857	40,160,719	0	0	0	0	0	1
NOL6	40,517,332	40,544,106	1	0	0	1	0	1
RANBP3L	44,306,810	44,342,858	0	0	0	0	0	2
EGFLAM	45,162,455	45,242,328	2	1	0	0	0	1
RICTOR	45,423,023	45,499,159	0	0	0	0	0	1
RPL37	46,153,836	46,156,411	0	0	0	0	0	1
LOC121468186	46,491,084	46,504,450	1	0	0	0	0	1
SELENOP	46,811,419	46,819,749	0	0	0	0	0	1
ITGA2	48,975,196	49,042,119	1	1	0	0	0	1
HSPB3	49,569,584	49,571,811	0	0	0	0	0	1
MCIDAS	49,886,112	49,889,655	0	1	0	0	0	1
DHX29	49,893,739	49,922,731	0	0	1	0	0	1
MTREX	49,922,831	49,966,523	0	0	0	0	0	1
DDX4	50,098,518	50,118,178	1	0	2	1	0	1
ANKRD55	50,236,565	50,282,299	0	0	1	0	0	1
GPBP1	50,649,093	50,687,575	0	0	0	0	0	1
PDE4D	51,358,597	51,709,246	0	0	0	1	0	1
LOC121468141	51,708,861	51,715,147	0	0	0	0	0	1
ERCC8	52,007,424	52,036,012	1	0	0	0	0	1
NDUFAF2	52,036,056	52,088,074	0	0	0	0	0	1
ZSWIM6	52,166,179	52,273,151	1	0	0	0	0	2

Gene Name	CDS Start	CDS End	P(A) non- syn	P(A) syn	P(B) non- syn	P(B) syn	D non- syn	D syn
PPWD1	53,619,729	53,631,083	0	0	0	0	0	1
LOC115491052	53,949,706	53,983,069	1	0	1	0	0	1
MAST4	54,073,124	54,351,248	1	0	1	0	0	1
GOLM1	54,766,563	54,802,374	0	0	0	0	0	1
AGTPBP1	54,838,592	54,899,033	1	2	0	0	0	1
NTRK2	55,072,224	55,265,363	0	0	0	0	0	1
SLC28A3	55,363,030	55,400,924	2	1	2	0	0	1
FRMD3	55,590,688	55,713,496	2	1	1	0	0	2
LOC100221996	56,163,894	56,244,161	0	4	0	0	0	3
GCNT1	57,962,517	57,972,165	0	0	0	0	0	1
SERINC5	58,230,788	58,272,887	0	1	0	0	0	2
HOMER1	58,505,186	58,593,629	0	0	0	0	0	2
DMGDH	58,718,218	58,757,995	2	1	1	1	0	2
LHFPL2	58,843,244	58,964,353	0	1	0	0	0	2
LOC100224500	59,648,239	59,657,104	0	0	0	0	0	1
F2RL1	59,691,456	59,697,268	0	0	0	0	0	1
CERT1	60,223,834	60,304,536	0	1	0	1	0	2
HMGCR	60,312,318	60,332,351	0	0	0	0	0	1
NSA2	60,513,706	60,519,297	1	1	0	0	0	1
HEXB	60,538,578	60,555,471	1	0	0	0	0	1
ARHGEF28	60,877,385	60,955,961	6	2	1	0	0	3
UTP15	61,005,293	61,016,569	1	0	0	2	0	3
ANKRA2	61,016,621	61,025,610	0	0	0	0	0	1
TMEM171	61,216,777	61,228,052	2	0	0	0	0	1
ZNF366	61,478,611	61,514,908	0	0	0	0	0	1
PTCD2	61,542,596	61,557,792	1	1	0	0	0	1
MRPS27	61,557,429	61,599,909	1	0	0	0	0	2
PGM5	61,816,563	61,885,756	1	0	0	0	0	1
CBWD1	61,898,225	61,920,048	0	0	0	0	0	1
DMRT3	62,242,826	62,253,657	0	0	0	0	0	1
JAK2	63,506,797	63,594,739	0	6	0	0	0	1
GLDC	64,003,360	64,042,395	2	2	0	0	0	1
FREM1	67,387,601	67,452,522	8	7	2	1	0	2
CCDC171	67,643,503	67,812,201	7	5	1	0	0	1
FXN	69,911,988	69,923,365	0	1	0	0	0	1

Table S3.2:*List of nonsynonymous fixed differences between the A and B inversion haplotypes.*

Position	Codons (Minor)	Codons (Major)	Protein (Minor)	Protein (Major)	Polarity (Minor)	Polarity (Major)	Gene Name
9,784,437	GTT	GCT	V	A	nonpolar	nonpolar	LOC115491274
9,935,968	TTG	TTT	L	F	nonpolar	nonpolar	SLC35D2
10,063,048	GAA	AAA	E	K	acidic	basic	ERCC6L2
10,088,866	GTT	ATT	V	I	nonpolar	nonpolar	ERCC6L2
10,360,564	TAC	TGC	Y	C	polar	polar	FANCC
10,383,697	AAG	GAG	K	E	basic	acidic	FANCC
10,405,887	ATG	GTG	M	V	nonpolar	nonpolar	FANCC
10,653,792	AAA	AGA	K	R	basic	basic	LOC100219012
11,208,194	GAT	GAA	D	E	acidic	acidic	TUT7
11,914,780	GAA	AAA	E	K	acidic	basic	LOC101233669
11,914,820	TGT	TAT	C	Y	polar	polar	LOC101233669
11,914,844	GCC	GTC	A	V	nonpolar	nonpolar	LOC101233669
12,609,209	CTA	GTA	L	V	nonpolar	nonpolar	CEMIP2
12,617,801	AAC	AGC	N	S	polar	polar	CEMIP2
13,629,337	ACC	AGC	T	S	polar	polar	TRPM6
13,638,236	TTC	ATC	F	I	nonpolar	nonpolar	TRPM6
13,700,970	GGT	GAT	G	D	nonpolar	acidic	CZH9orf40
13,735,022	GTG	GCG	V	A	nonpolar	nonpolar	NMRK1
14,088,949	TCC	GCC	S	A	polar	nonpolar	ARSK
14,091,260	GAC	GAA	D	E	acidic	acidic	TTC37
14,123,138	AAT	GAT	N	D	polar	acidic	TTC37
17,472,146	AGA	AAA	R	K	basic	basic	LOC115491137
17,472,970	TAC	CAC	Y	H	polar	basic	LOC115491137
17,475,451	TAC	CAC	Y	H	polar	basic	LOC115491137
17,966,674	CAT	CAG	H	Q	basic	basic	GRIN3A
18,729,197	CGC	CAC	R	H	basic	basic	SLCO4C1
18,916,164	AGA	AAA	R	K	basic	basic	PAM
18,941,134	ACC	TCC	T	S	polar	polar	GIN1
24,315,852	GTC	ATC	V	I	nonpolar	nonpolar	CSNK1G3
24,351,843	GGA	GAA	G	E	nonpolar	acidic	CSNK1G3
24,423,761	CAG	CGG	Q	R	basic	basic	CEP120
24,429,877	GTG	ATG	V	M	nonpolar	nonpolar	CEP120
24,971,733	AAT	TAT	N	Y	polar	polar	LOC115491159
29,074,716	CAA	CGA	Q	R	basic	basic	SMC2
29,756,216	CAC	TAC	H	Y	basic	polar	TOPORS
30,164,130	GAT	GAA	D	E	acidic	acidic	PGGT1B
30,720,381	TTC	CTC	F	L	nonpolar	nonpolar	LOC115491124
30,958,251	GCA	ACA	A	T	nonpolar	polar	ELAVL2
30,968,830	TTG	TTT	L	F	nonpolar	nonpolar	ELAVL2
32,009,720	AAG	AAT	K	N	basic	polar	LOC115491242
32,023,322	CGC	TGC	R	C	basic	polar	LOC115491242

Position	Codons (Minor)	Codons (Major)	Protein (Minor)	Protein (Major)	Polarity (Minor)	Polarity (Major)	Gene Name
32,050,579	ATG	CTG	M	L	nonpolar	nonpolar	LOC115491125
32,311,250	AGA	AGT	R	S	basic	polar	LINGO2
34,188,622	CGT	CAT	R	H	basic	basic	B4GALT1
34,393,974	GCC	GAC	A	D	nonpolar	acidic	TDRD7
34,471,343	GTT	ATT	V	I	nonpolar	nonpolar	DGKQ
34,574,291	GAC	AAC	D	N	acidic	polar	ACO1
34,603,497	CGC	TGC	R	C	basic	polar	DDX58
34,714,935	GCG	GTG	A	V	nonpolar	nonpolar	LOC100217918
34,966,146	GTA	TTA	V	L	nonpolar	nonpolar	BDP1
34,996,597	AAT	AAA	N	K	polar	basic	LOC100221697
36,060,484	GCT	CCT	A	P	nonpolar	nonpolar	SETBP1
40,246,536	TAT	CAT	Y	H	polar	basic	KIAA1328
40,995,949	CTA	CCA	L	P	nonpolar	nonpolar	DNAI1
41,998,027	GCT	GGT	A	G	nonpolar	nonpolar	UNC13B
43,482,996	GTA	GGA	V	G	nonpolar	nonpolar	ADAMTS12
43,899,266	GCA	ACA	A	T	nonpolar	polar	PRLR
44,117,742	GTG	GCG	V	A	nonpolar	nonpolar	SPEF2
44,129,734	CAT	CGT	H	R	basic	basic	SPEF2
44,135,787	ATT	ACT	I	T	nonpolar	polar	SPEF2
44,761,763	GGT	GAT	G	D	nonpolar	acidic	CPLANE1
44,773,815	AGT	AAT	S	N	polar	polar	CPLANE1
45,279,858	GAA	GGA	E	G	acidic	nonpolar	LIFR
45,548,560	TTA	ATA	L	I	nonpolar	nonpolar	FYB1
46,146,570	TCC	TTC	S	F	polar	nonpolar	PRKAA1
46,174,449	GAA	GAC	E	D	acidic	acidic	C7
46,486,069	ATA	GTA	I	V	nonpolar	nonpolar	OXCT1
49,051,517	GCG	GTG	A	V	nonpolar	nonpolar	MOCS2
49,204,476	GAA	GAT	E	D	acidic	acidic	FST
49,847,228	TGG	TGT	W	C	nonpolar	polar	CDC20B
49,892,121	TCC	TTC	S	F	polar	nonpolar	CCNO
50,144,800	AAA	AGA	K	R	basic	basic	IL31RA
50,144,880	AAA	GAA	K	E	basic	acidic	IL31RA
50,189,211	CGT	CAT	R	H	basic	basic	IL6ST
50,510,317	TGC	TTC	C	F	polar	nonpolar	MAP3K1
50,510,419	GAC	GCC	D	A	acidic	nonpolar	MAP3K1
50,577,482	GAA	GGA	E	G	acidic	nonpolar	SETD9
50,605,920	TGC	CGC	C	R	polar	basic	MIER3
50,622,336	ATG	GTG	M	V	nonpolar	nonpolar	LOC121468155
50,622,411	AAG	GAG	K	E	basic	acidic	LOC121468155
51,942,589	GTC	ATC	V	I	nonpolar	nonpolar	ELOVL7
52,000,986	CTG	CCG	L	P	nonpolar	nonpolar	LOC115491269
52,001,308	GTT	ATT	V	I	nonpolar	nonpolar	LOC115491269
52,001,722	AAC	GAC	N	D	polar	acidic	LOC115491269
52,001,940	ATC	ACC	I	T	nonpolar	polar	LOC115491269

Position	Codons (Minor)	Codons (Major)	Protein (Minor)	Protein (Major)	Polarity (Minor)	Polarity (Major)	Gene Name
53,308,411	ACC	AAC	T	N	polar	polar	CWC27
53,676,242	CAC	CGC	H	R	basic	basic	TRAPPC13
53,705,398	GAC	TAC	D	Y	acidic	polar	SGTB
53,878,945	TAC	CAC	Y	H	polar	basic	ERBIN
55,553,096	CAG	CTG	Q	L	basic	nonpolar	UBQLN1
55,559,101	ATA	CTA	I	L	nonpolar	nonpolar	UBQLN1
55,768,452	CAC	TAC	H	Y	basic	polar	RASEF
57,656,125	GGA	GTA	G	V	nonpolar	nonpolar	VPS13A
57,656,215	GCA	GGA	A	G	nonpolar	nonpolar	VPS13A
57,656,309	TTA	ATA	L	I	nonpolar	nonpolar	VPS13A
57,660,043	GGT	CGT	G	R	nonpolar	basic	VPS13A
57,693,719	CTT	CCT	L	P	nonpolar	nonpolar	VPS13A
57,880,719	CTT	GTT	L	V	nonpolar	nonpolar	PRUNE2
57,881,976	GCT	TCT	A	S	nonpolar	polar	PRUNE2
57,882,946	ATG	ACG	M	T	nonpolar	polar	PRUNE2
57,884,005	ACC	AAC	T	N	polar	polar	PRUNE2
57,884,380	GGA	GAA	G	E	nonpolar	acidic	PRUNE2
58,187,459	CTA	CCA	L	P	nonpolar	nonpolar	LOC116807240
58,189,319	CCT	ACT	P	T	nonpolar	polar	LOC116807240
58,199,489	CCT	ACT	P	T	nonpolar	polar	LOC116807240
58,307,574	CGA	CAA	R	Q	basic	basic	THBS4
58,430,279	GCT	TCT	A	S	nonpolar	polar	CMYA5
58,431,908	ACT	GCT	T	A	polar	nonpolar	CMYA5
58,432,556	GAA	AAA	E	K	acidic	basic	CMYA5
58,432,759	CTG	CCG	L	P	nonpolar	nonpolar	CMYA5
58,433,363	AGT	CGT	S	R	polar	basic	CMYA5
58,433,735	CAA	AAA	Q	K	basic	basic	CMYA5
58,434,020	TCA	ACA	S	T	polar	polar	CMYA5
58,490,064	CTG	CCG	L	P	nonpolar	nonpolar	TENT2
58,613,554	AGA	ATA	R	I	basic	nonpolar	JMY
58,665,384	GCG	GGG	A	G	nonpolar	nonpolar	JMY
58,768,500	CTC	TTC	L	F	nonpolar	nonpolar	ARSB
58,818,222	ATG	ACG	M	T	nonpolar	polar	ARSB
59,444,948	ACA	AGA	T	R	polar	basic	LOC115491249
59,492,912	TCC	TTC	S	F	polar	nonpolar	PDE8B
59,612,936	ATT	ACT	I	T	nonpolar	polar	AGGF1
59,726,633	GGT	AGT	G	S	nonpolar	polar	F2R
59,769,304	CTC	TTC	L	F	nonpolar	nonpolar	IQGAP2
59,780,891	GTT	ATT	V	I	nonpolar	nonpolar	IQGAP2
60,152,285	ACT	AAT	T	N	polar	polar	POC5
60,169,683	ACG	ATG	T	M	polar	nonpolar	ANKDD1B
60,202,771	TTT	TTG	F	L	nonpolar	nonpolar	POLK
60,209,095	CAG	CGG	Q	R	basic	basic	POLK
60,214,842	ACA	GCA	T	A	polar	nonpolar	POLK

Position	Codons (Minor)	Codons (Major)	Protein (Minor)	Protein (Major)	Polarity (Minor)	Polarity (Major)	Gene Name
60,345,704	ATG	ACG	M	T	nonpolar	polar	ANKRD31
60,345,719	ACA	ATA	T	I	polar	nonpolar	ANKRD31
60,345,929	TGT	TTT	C	F	polar	nonpolar	ANKRD31
60,346,120	CAA	GAA	Q	E	basic	acidic	ANKRD31
60,346,145	CAA	CGA	Q	R	basic	basic	ANKRD31
60,349,336	GGT	TGT	G	C	nonpolar	polar	ANKRD31
60,349,445	ATA	ACA	I	T	nonpolar	polar	ANKRD31
60,349,456	TCA	ACA	S	T	polar	polar	ANKRD31
60,355,847	CAT	CAA	H	Q	basic	basic	ANKRD31
60,366,327	TCA	ACA	S	T	polar	polar	ANKRD31
60,370,043	GGA	GAA	G	E	nonpolar	acidic	ANKRD31
60,493,220	GTT	CTT	V	L	nonpolar	nonpolar	FAM169A
60,499,349	GCG	GTG	A	V	nonpolar	nonpolar	FAM169A
60,502,215	CTC	CCC	L	P	nonpolar	nonpolar	FAM169A
60,504,278	TCT	TTT	S	F	polar	nonpolar	FAM169A
60,504,386	CAT	CGT	H	R	basic	basic	FAM169A
60,506,209	GCA	GTA	A	V	nonpolar	nonpolar	FAM169A
60,519,447	GGA	GCA	G	A	nonpolar	nonpolar	GFM2
60,521,095	ATA	ATG	I	M	nonpolar	nonpolar	GFM2
61,300,588	CAT	AAT	H	N	basic	polar	FCHO2
61,392,830	GCC	ACC	A	T	nonpolar	polar	TNPO1
61,619,453	GCA	ACA	A	T	nonpolar	polar	MAP1B
61,620,151	TTT	TCT	F	S	nonpolar	polar	MAP1B
61,945,726	GTC	GGC	V	G	nonpolar	nonpolar	DOCK8
61,959,504	TGT	CGT	C	R	polar	basic	DOCK8
62,175,899	GTC	ATC	V	I	nonpolar	nonpolar	DMRT1
62,208,735	GGC	GAC	G	D	nonpolar	acidic	DMRT1
62,610,175	AGT	GGT	S	G	polar	nonpolar	SMARCA2
62,917,798	CAC	TAC	H	Y	basic	polar	PUM3
62,928,792	AGA	GGA	R	G	basic	nonpolar	LOC100222616
62,938,935	GAG	AAG	E	K	acidic	basic	LOC105760879
62,938,938	GAG	AAG	E	K	acidic	basic	LOC105760879
63,156,177	TCT	TGT	S	C	polar	polar	GLIS3
63,239,995	CTG	CAG	L	Q	nonpolar	basic	GLIS3
63,385,943	AAT	AAG	N	K	polar	basic	LOC105760877
63,386,928	CAG	CCG	Q	P	basic	nonpolar	LOC105760877
63,391,465	TTC	ATC	F	I	nonpolar	nonpolar	LOC105760877
63,483,480	CGG	CAG	R	Q	basic	basic	RCL1
63,717,621	GTT	ATT	V	I	nonpolar	nonpolar	RIC1
63,722,614	ATG	TTG	M	L	nonpolar	nonpolar	RIC1
63,724,558	ATG	ACG	M	T	nonpolar	polar	RIC1
63,724,559	ATA	ATG	I	M	nonpolar	nonpolar	RIC1
63,733,039	ATA	CTA	I	L	nonpolar	nonpolar	RIC1
63,733,141	TCT	GCT	S	A	polar	nonpolar	RIC1

Position	Codons (Minor)	Codons (Major)	Protein (Minor)	Protein (Major)	Polarity (Minor)	Polarity (Major)	Gene Name
63,735,404	ATC	ATG	I	M	nonpolar	nonpolar	RIC1
63,735,936	TGT	TCT	C	S	polar	polar	RIC1
63,745,971	AGG	AAG	R	K	basic	basic	ERMP1
63,811,763	CGA	CAA	R	Q	basic	basic	KIAA2026
64,219,047	TTC	CTC	F	L	nonpolar	nonpolar	KDM4C
67,115,679	ACA	GCA	T	A	polar	nonpolar	NFIB
69,065,102	TTC	TTA	F	L	nonpolar	nonpolar	ADAMTSL1
69,084,955	ACG	ATG	T	M	polar	nonpolar	ADAMTSL1
69,299,983	GGA	TGA	G	*	nonpolar	stop codon	ACER2

Table S3.3:*List of synonymous fixed differences between the A and B inversion haplotypes.*

Position	Codons (Minor)	Codons (Major)	Gene Name	Position	Codons (Minor)	Codons (Major)	Gene Name
7202417	GAT	GAC	SYK	29193354	AGG	CGG	SVEP1
7644540	GTG	GTA	GADD45G	29283131	TCT	TCC	SVEP1
7765261	GCT	GCC	SEMA4D	29830895	ATT	ATC	GRAMD2B
8076255	GAT	GAC	SPIN1	30034013	GCA	GCT	FEM1C
9654798	GCG	GCA	LRRC2	30157355	AGT	AGC	CCDC112
10065734	CGA	CGG	ERCC6L2	31932657	CGA	CGG	TEK
10086004	CTA	TTA	ERCC6L2	32023548	CCA	CCG	LOC115491242
10289896	CTG	TTG	PTCH1	32049474	TTA	TTG	LOC115491125
10595594	AGA	AGG	AOPEP	34316808	ACG	ACA	NCBP1
10659988	TAT	TAC	LOC100219012	34333697	AGA	AGG	NCBP1
10715701	ATA	ATC	DAPK1	34355500	TAT	TAC	TMOD1
10715782	ACA	ACG	DAPK1	34391033	TTG	CTG	TDRD7
10733355	AAT	AAC	DAPK1	34453050	TTC	TTT	DGKQ
11297719	AGT	AGC	CENPH	34599934	TCC	TCT	DDX58
12048526	AAA	AAG	SMC5	34600538	GCC	GCT	DDX58
12152629	CAT	CAC	TRPM3	34968663	AAA	AAG	BDP1
12591008	TAT	TAC	CEMIP2	35021148	TTA	CTA	LOC100217858
12610868	GGC	GGG	CEMIP2	35398906	TCG	TCA	CZH18orf25
13544530	CCA	CCG	RORB	40065100	ACA	ACG	CELF4
13656889	GGT	GGC	TRPM6	40537896	TCT	TCC	NOL6
13660851	AGT	AGC	TRPM6	44327807	AAG	AAA	RANBP3L
13734231	AGG	AGA	NMRK1	44342459	CCT	CCC	RANBP3L
14070540	ATT	ATC	ARSK	44768383	TTT	TTC	CPLANE1
14090554	GAA	GAG	TTC37	45199467	AGC	AGT	EGFLAM
14113399	GCA	GCG	TTC37	45270129	TGC	TGT	LIFR
14263810	AAA	AAG	MCTP1	45447097	CAG	CAA	RICTOR
14992172	GCC	GCT	NR2F1	46155240	AAA	AAG	RPL37
15962465	GGA	GGC	ADGRV1	46491242	GCC	GCG	LOC121468186
17148663	TCC	TCG	TMEM161B	46813086	GAC	GAT	SELENOP
18076376	CAA	CAG	ALDOB	49009112	CCG	CCA	ITGA2
18941954	AAT	AAC	GIN1	49570126	GAT	GAC	HSPB3
19018582	AAA	AAG	MACIR	49861214	CCA	CCG	CDC20B
20482949	AGA	CGA	EFNA5	49887739	AGG	AGA	MCIDAS
20755332	GGG	GGA	FBXL17	49899841	GCG	GCA	DHX29
21246751	CTA	CTT	FER	49948506	AAT	AAC	MTREX
22343007	ACC	ACG	EPB41L4A	50106496	CTG	CTA	DDX4
22590622	GCA	GCT	SRP19	50247941	AGC	AGT	ANKRD55
22669633	ACA	ACG	MCC	50548238	TCA	TCG	MAP3K1
22917990	GAA	GAG	APBA1	50554613	GAA	GAG	MAP3K1
24352022	CTA	TTA	CSNK1G3	50554616	ACG	ACA	MAP3K1
28550447	TAC	TAT	KIAA1958	50571546	TTG	CTG	SETD9

Position	Codons (Minor)	Codons (Major)	Gene Name
50675787	GAA	GAG	GPBP1
51708950	CAA	CAG	PDE4D
51942635	CAT	CAC	ELOVL7
52001699	GCT	GCC	LOC115491269
52021164	TAT	TAC	ERCC8
52074882	CCA	CCG	NDUFAF2
52166510	GAG	GAA	ZSWIM6
52251767	CAT	CAC	ZSWIM6
53626334	AAG	AAA	PPWD1
53970334	AAT	AAC	LOC115491052
54336965	CGT	CGC	MAST4
54797317	CTA	CTG	GOLM1
54887658	GAA	GAG	AGTPBP1
55115973	TTA	CTA	NTRK2
55373901	AGC	AGT	SLC28A3
55668614	CCA	CCG	FRMD3
55710609	CCT	CCC	FRMD3
56228444	CAT	CAC	LOC100221996
56230466	GCC	GCA	LOC100221996
56232370	TCG	TCA	LOC100221996
57711939	CTA	CTG	VPS13A
57717319	AAT	AAC	VPS13A
57838929	GAG	GAA	PRUNE2
57964582	GAG	GAA	GCNT1
58189527	GGC	GGA	LOC116807240
58243678	ACC	ACG	SERINC5
58256067	CAG	CAA	SERINC5
58281189	CCC	CCT	THBS4
58414240	GTA	GTC	CMYA5
58435917	GCA	GCC	CMYA5
58529253	TAT	TAC	HOMER1
58546658	GAC	GAT	HOMER1
58613502	CCC	CCG	JMY
58734184	CCT	CCC	DMGDH
58747197	GAG	GAA	DMGDH
58945362	AGC	AGT	LHFPL2
58959450	TCC	TCT	LHFPL2
59655718	CCC	CCA	LOC100224500
59692900	GTG	GTA	F2RL1
59726866	TTC	TTT	F2R
59771621	CCG	CCA	IQGAP2
59780976	GCG	GCT	F2RL2
60182231	GGT	GGC	ANKDD1B
60247706	CAG	CAA	CERT1

Position	Codons (Minor)	Codons (Major)	Gene Name
60296762	AAC	AAT	CERT1
60325361	GCA	GCC	HMGCR
60345975	GTA	GTG	ANKRD31
60346077	CTA	CTC	ANKRD31
60367230	TAC	TAT	ANKRD31
60377923	GAG	GAA	ANKRD31
60488068	GCA	GCG	FAM169A
60489717	CAA	CAG	FAM169A
60499398	AGG	AGA	FAM169A
60506351	GCC	GCT	FAM169A
60519004	CGC	CGG	NSA2
60519430	AGG	AGA	GFM2
60530485	GCC	GCT	GFM2
60531232	ACA	ACG	GFM2
60554928	GTA	GTG	HEXB
60909887	CTT	CTA	ARHGEF28
60914271	GTT	GTC	ARHGEF28
60919721	CCC	CCA	ARHGEF28
61008100	CCT	CCC	UTP15
61014393	GTA	GTG	UTP15
61014531	GGA	GGT	UTP15
61024800	CAG	CAA	ANKRA2
61220780	GTA	GTG	TMEM171
61245059	GTC	GTA	FCHO2
61281539	AAA	AAG	FCHO2
61506974	TTG	CTG	ZNF366
61551772	GAG	GAA	PTCD2
61559312	TAT	TAC	MRPS27
61589924	ATC	ATA	MRPS27
61617942	CTG	CTT	MAP1B
61621479	GAG	GAA	MAP1B
61878849	GGT	GGA	PGM5
61918735	TTC	TTT	CBWD1
62175646	CCA	CCG	DMRT1
62175823	CTA	CTG	DMRT1
62243169	GCG	GCT	DMRT3
62636841	AAC	AAT	SMARCA2
62894886	CTG	CTT	PUM3
62938651	GAC	GAT	LOC105760879
63156209	GAA	GAG	GLIS3
63388074	CTG	CTA	LOC105760877
63391400	AGG	AGA	LOC105760877
63391509	GCG	GCA	LOC105760877
63461624	GGT	GGC	RCL1

Position	Codons (Minor)	Codons (Major)	Gene Name
63572632	CTG	TTG	JAK2
63732035	CGA	CGG	RIC1
63733194	GCG	GCA	RIC1
63745970	AAA	AAG	ERMP1
63756450	CAA	CAG	ERMP1
64019801	TGC	TGT	GLDC

Position	Codons (Minor)	Codons (Major)	Gene Name
67433079	AAG	AAA	FREM1
67441651	CTG	CTC	FREM1
67718013	CTT	CTC	CCDC171
69026555	CCC	CCT	ADAMTSL1
69922423	GCA	GCG	FXN

Figure S3.1:

Nucleotide diversity (π) plots showing the average number of nucleotide differences per site across 100Kb windows with a 10Kb step. **a.** only using SNPs within A/AA karyotype birds (number of individuals = 6, number of Z chromosomes = 8) **b.** only using SNPs within B/BB karyotype birds (number of individuals = 7, number of Z chromosomes = 8) **c.** using SNPs within a subset of birds of a mixture of karyotypes (number of individuals = 6, number of Z chromosomes = 8, karyotypes = A, AA, AB, B, B, C).

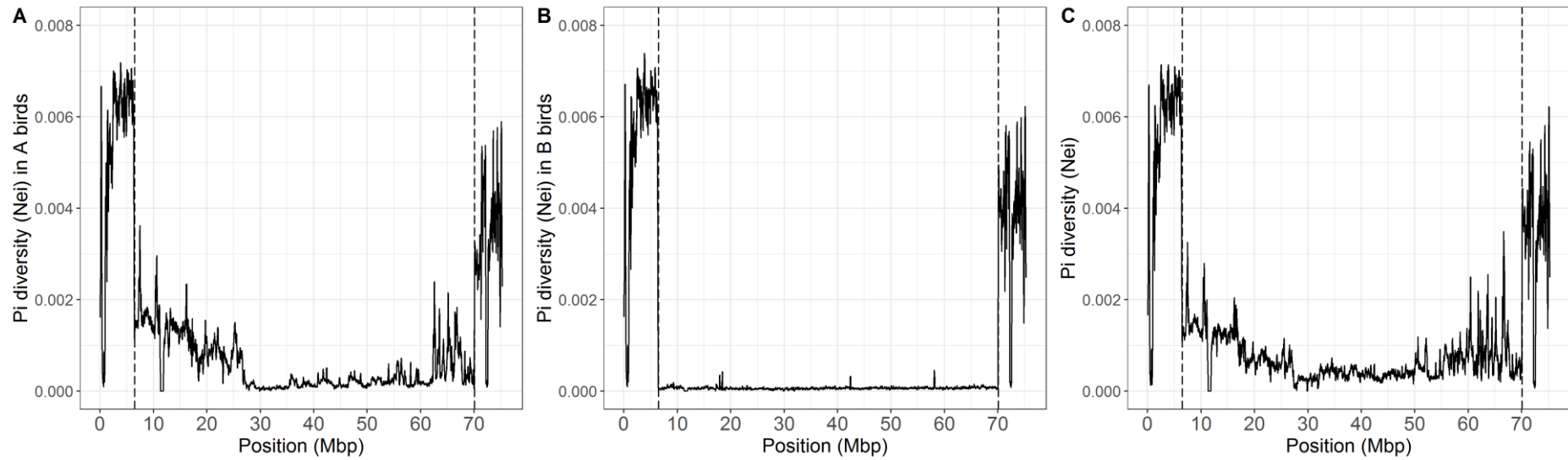


Figure S3.2:

Fixation index (F_{ST}) between zebra finches of inversion karyotype A or AA (number of individuals = 6, number of Z chromosomes = 8) and of inversion karyotype B or BB (number of individuals = 7, number of Z chromosomes = 8). SNPs were found within all 13 birds when aligning reads against the zebra finch reference genome (bTG1.4). F_{ST} is averaged for all SNP positions in 100Kb sliding windows of sequence across the bTG1.4 Z chromosome, with each window overlapping by 10Kb. Dashed lines indicate predicted breakpoints of the Z chromosome inversion polymorphism.

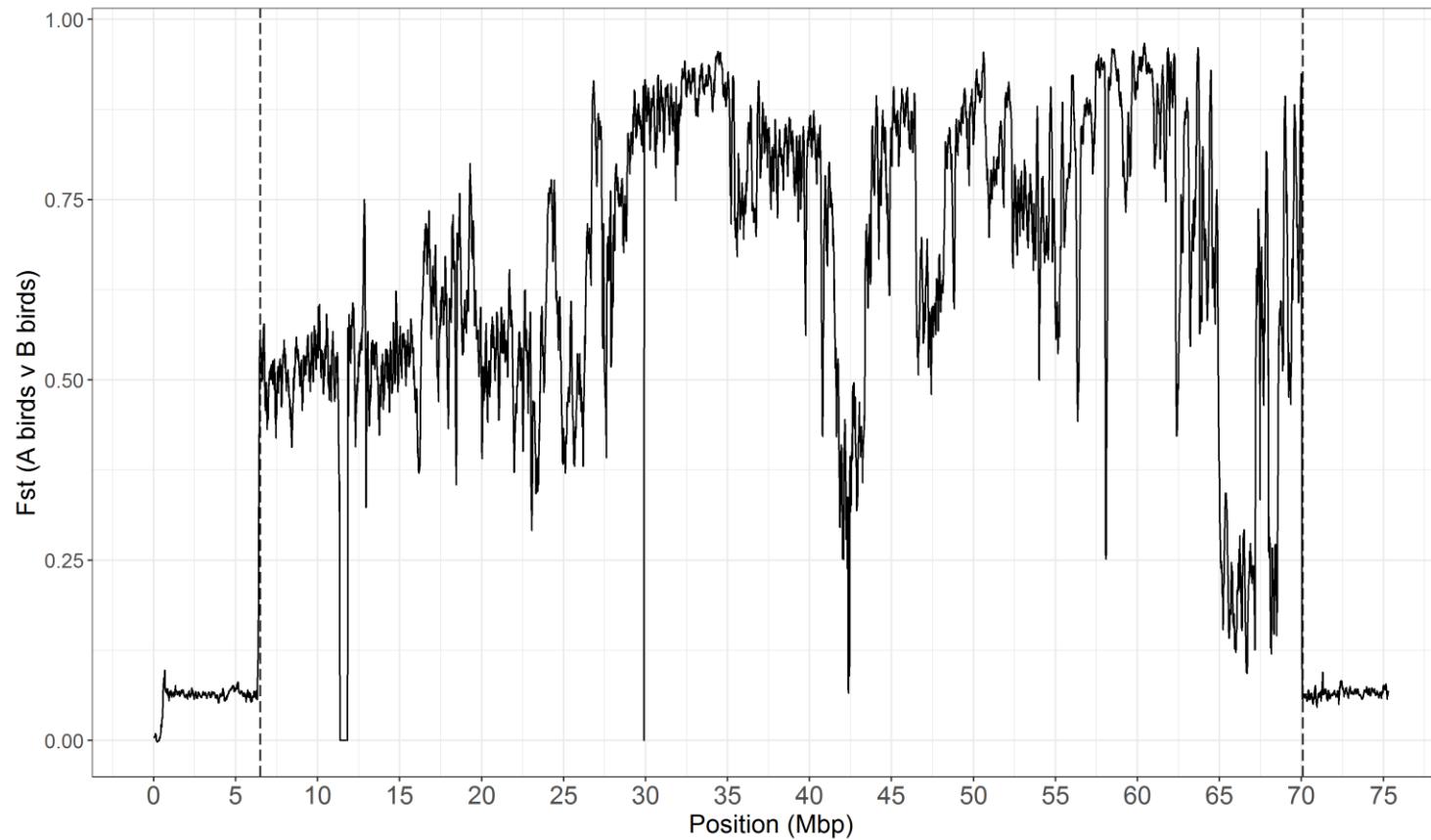


Figure S3.3:

Neutrality statistics calculated using SNPs called using sequence data from 6 karyotype A/AA finches (number of Z chromosomes = 8), 7 B/BB finches (number of Z chromosomes = 8) and 6 finches of mixed karyotypes (number of Z chromosomes = 8, karyotypes = A, AA, AB, B, B, C). Tajima's D calculated using **a.** A/AA finches **b.** B/BB finches **c.** finches of mixed karyotypes. Fay & Wu's H calculated using **d.** A/AA finches **e.** B/BB finches **f.** finches of mixed karyotypes. Zeng's E calculated using **g.** A/AA finches **h.** B/BB finches **i.** finches of mixed karyotypes. Reads were aligned against the reference genome (bTG1.4) to give physical positions. Sequence from a female long-tailed finch was used as an outgroup sequence for calculating H and E . Dashed lines indicate the predicted approximate breakpoints of the Z chromosome inversion region. All statistics were calculated from 100kbp sliding windows across the genome with a 10kbp step between windows.

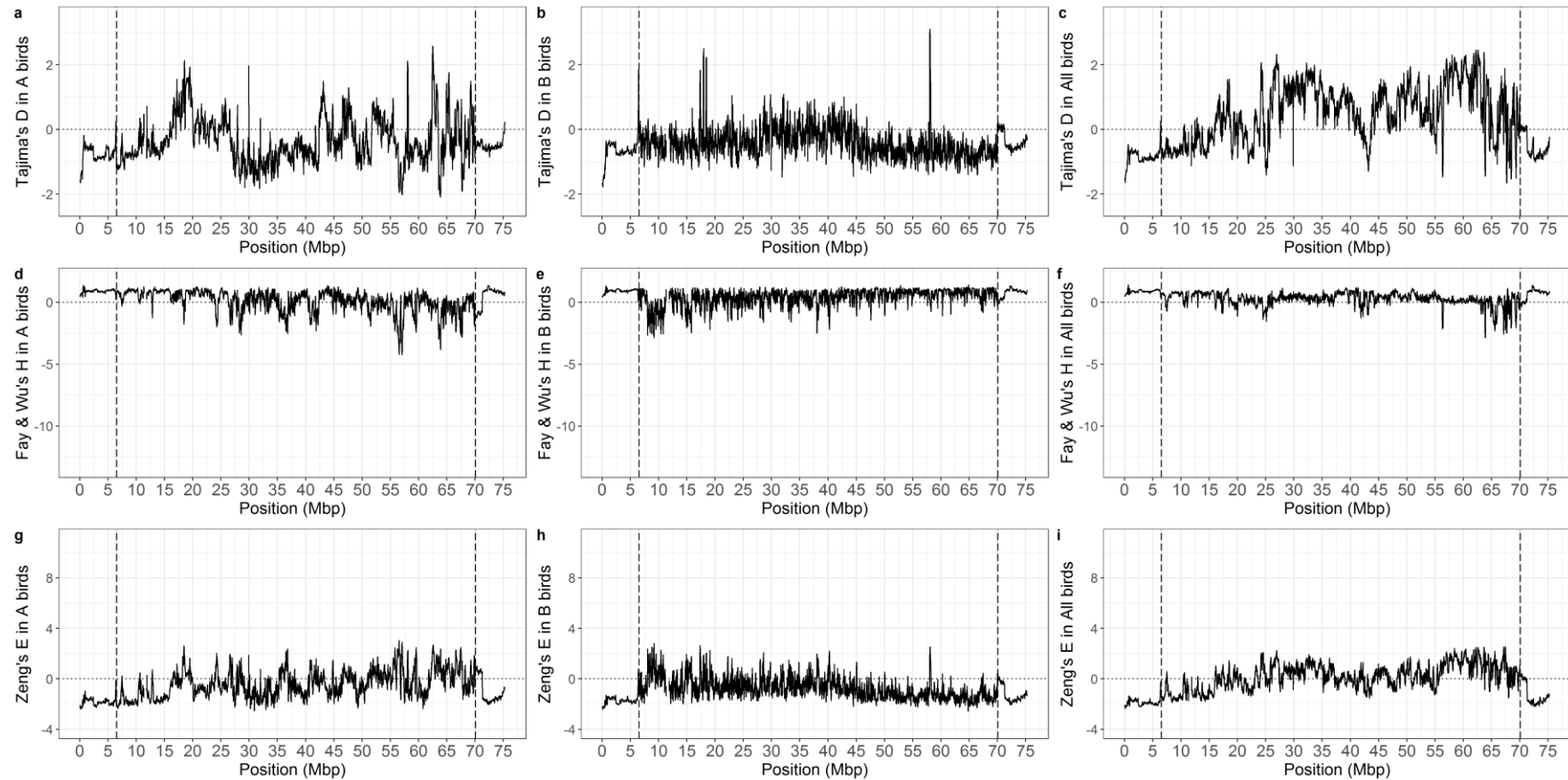


Figure S4.1:

A timetree inferred by applying the RelTime method in MEGA11 to the ML phylogenetic tree produced using all SNPs in Region I (shown in Figure 4.6a). The timetree was computed using the divergence time between zebra finches and long-tailed finches (6.4mya) as a calibration constraint. Bars around each node represent 95% confidence intervals. Sections of the tree showing divergence between A type individuals and between B type individuals are collapsed.

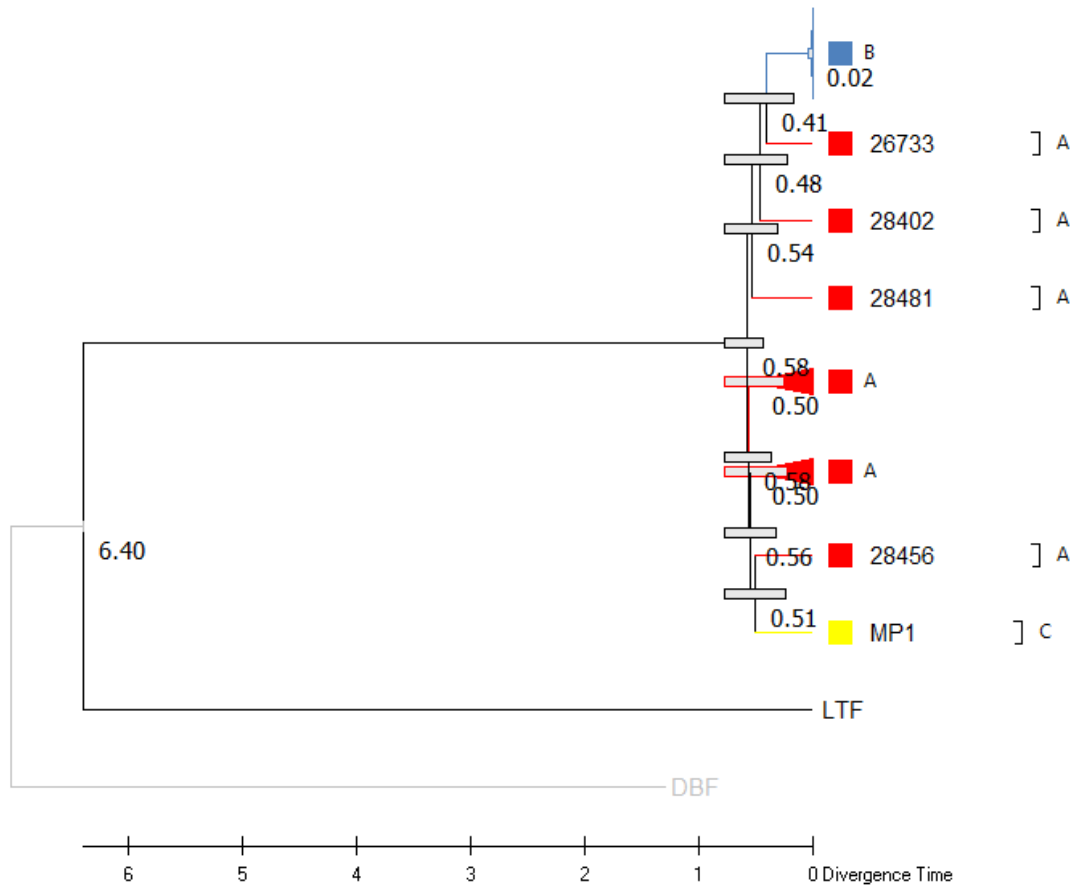


Figure S4.2:

A timetree inferred by applying the RelTime method in MEGA11 to the ML phylogenetic tree produced using all SNPs in Region II (shown in Figure 4.6b). The timetree was computed using the divergence time between zebra finches and long-tailed finches (6.4mya) as a calibration constraint. Bars around each node represent 95% confidence intervals. Sections of the tree showing divergence between A type individuals and between B type individuals are collapsed.

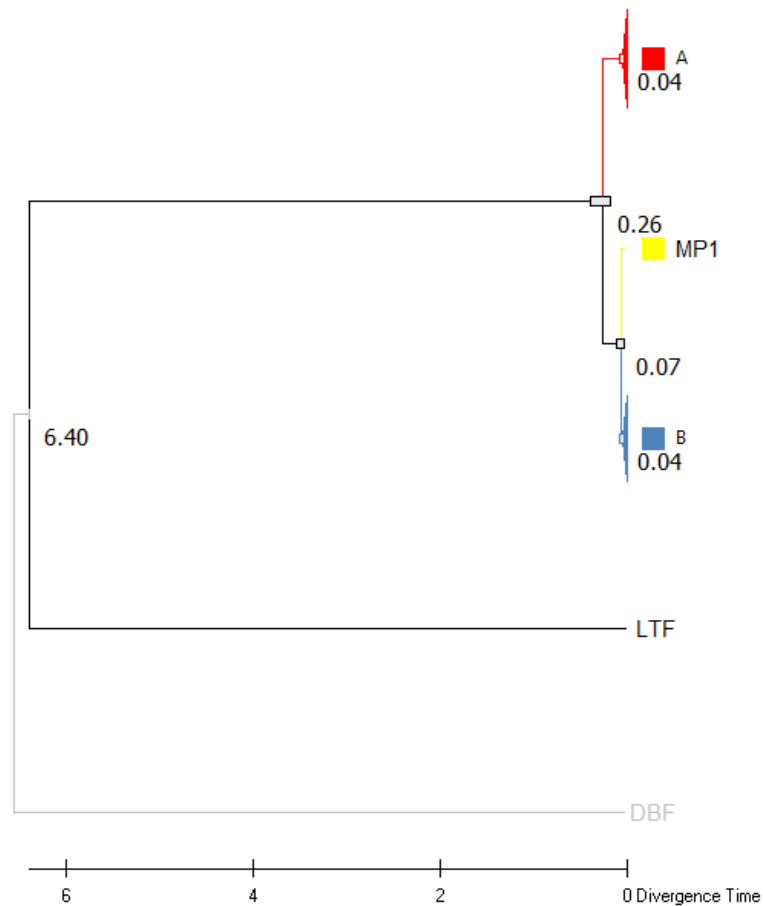


Figure S4.3:

A timetree inferred by applying the RelTime method in MEGA11 to the ML phylogenetic tree produced using all SNPs in Region III (shown in Figure 4.6c). The timetree was computed using the divergence time between zebra finches and long-tailed finches (6.4mya) as a calibration constraint. Bars around each node represent 95% confidence intervals. Sections of the tree showing divergence between A type individuals and between B type individuals are collapsed.

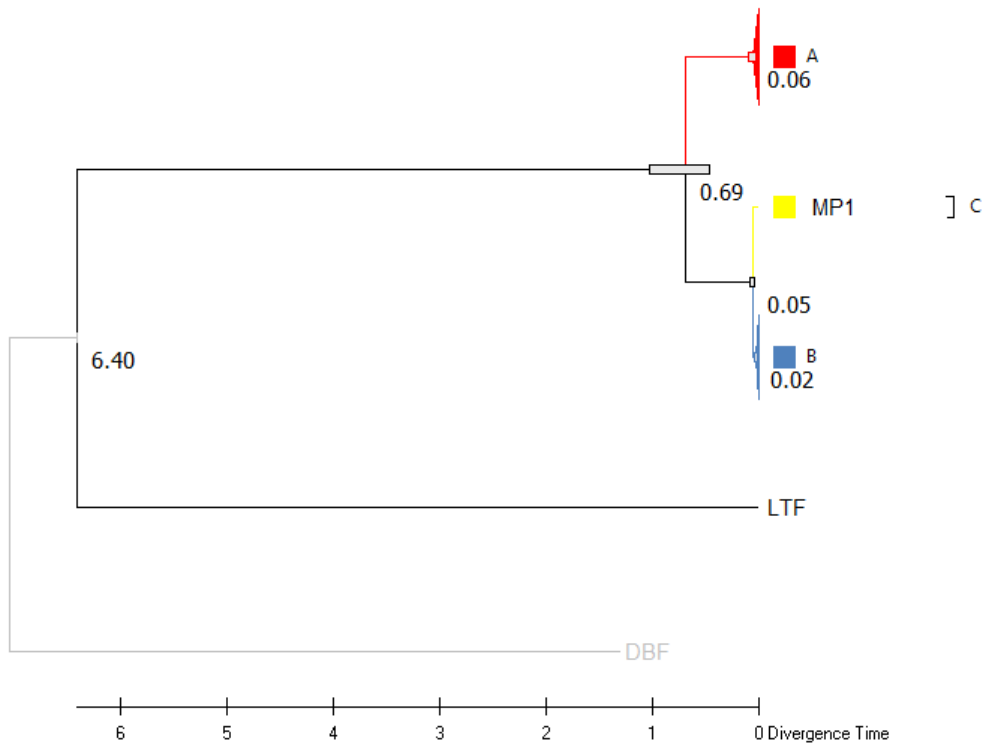


Figure S4.4:

A timetree inferred by applying the RelTime method in MEGA11 to the ML phylogenetic tree produced using all SNPs in Region IV (shown in Figure 4.6d). The timetree was computed using the divergence time between zebra finches and long-tailed finches (6.4mya) as a calibration constraint. Bars around each node represent 95% confidence intervals. Sections of the tree showing divergence between A type individuals and between B type individuals are collapsed.

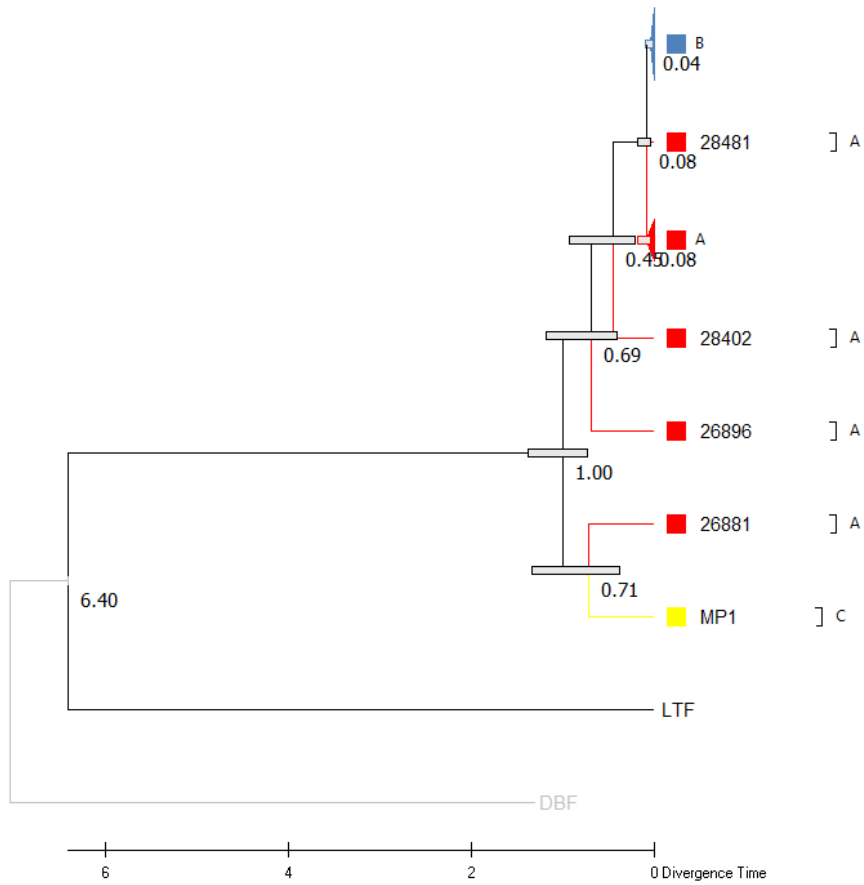


Figure S4.5:

A timetree inferred by applying the RelTime method in MEGA11 to the ML phylogenetic tree produced using all SNPs in Region V (shown in Figure 4.6e). The timetree was computed using the divergence time between zebra finches and long-tailed finches (6.4mya) as a calibration constraint. Bars around each node represent 95% confidence intervals. Sections of the tree showing divergence between A type individuals and between B type individuals are collapsed.

