

## ORIGINAL ARTICLE

# Chromosome-wide linkage disequilibrium caused by an inversion polymorphism in the white-throated sparrow (*Zonotrichia albicollis*)

LY Huynh<sup>1,2</sup>, DL Maney<sup>3</sup> and JW Thomas<sup>2</sup>

<sup>1</sup>Graduate Program in Population Biology, Ecology and Evolution, Emory University, Atlanta, GA, USA; <sup>2</sup>Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA and <sup>3</sup>Department of Psychology, Emory University, Atlanta, GA, USA

Chromosomal inversions have been of long-standing interest to geneticists because they are capable of suppressing recombination and facilitating the formation of adaptive gene complexes. An exceptional inversion polymorphism (ZAL2<sup>m</sup>) in the white-throated sparrow (*Zonotrichia albicollis*) is linked to variation in plumage, social behavior and mate choice, and is maintained in the population by negative assortative mating. The ZAL2<sup>m</sup> polymorphism is a complex inversion spanning >100 Mb and has been proposed to be a strong suppressor of recombination, as well as a potential model for studying neo-sex chromosome evolution. To quantify and evaluate these features of the ZAL2<sup>m</sup> polymorphism, we generated sequence from 8 ZAL2<sup>m</sup> and 16 ZAL2 chromosomes at 58 loci inside and 4 loci outside the inversion.

Inside the inversion we found that recombination was completely suppressed between ZAL2 and ZAL2<sup>m</sup>, resulting in uniformly high levels of genetic differentiation ( $F_{ST} = 0.94$ ), the formation of two distinct haplotype groups representing the alternate chromosome arrangements and extensive linkage disequilibrium spanning ~104 Mb within the inversion, whereas gene flow was not suppressed outside the inversion. Finally, although ZAL2<sup>m</sup> homozygotes are exceedingly rare in the population, occurring at a frequency of <1%, we detected evidence of historical recombination between ZAL2<sup>m</sup> chromosomes inside the inversion, refuting its potential status as a non-recombining autosome. *Heredity* (2011) **106**, 537–546; doi:10.1038/hdy.2010.85; published online 23 June 2010

**Keywords:** chromosomal inversion; recombination suppression; linkage disequilibrium

## Introduction

Chromosomal inversions are known to occur in a wide variety of organisms, in which they are associated with instances of adaptive evolution, speciation, selfish genes, sex chromosomes, human disease and disease susceptibility (Hartl, 1975; Lahn and Page, 1999; Noor *et al.*, 2001; Hoffmann *et al.*, 2004; Stefansson *et al.*, 2005; Dyer *et al.*, 2007). Although inversions have been studied for nearly a century, only recently have researchers begun to elucidate their effects on patterns of molecular evolution. Their influence on sequence evolution is derived from their unique ability to suppress recombination within the inverted interval in individuals heterozygous for the inversion (Sturtevant, 1921; Sturtevant and Beadle, 1936). Dobzhansky (1937, 1950) proposed that natural selection would favor inversions if they captured a set of positively interacting alleles, which he referred to as a 'supergene' complex (Dobzhansky and Sturtevant, 1938). Alternatively, Kirkpatrick and Barton (2006) showed that an inversion can enhance the fitness of its carrier, in the

absence of positive epistasis, if it simply captures two or more locally adapted alleles. Because inversions suppress recombination, linkage disequilibrium (LD) between the beneficial alleles within the inversion would be preserved even when the alleles are not in close proximity to each other. Thus, the influence of inversions on the rate and pattern of recombination is fundamental to their adaptive significance and evolution.

The suppression of recombination between the alternative chromosomal arrangements along the inverted segment will eventually lead to the formation of two distinct haplotype groups: the standard and the inverted. However, population genetic studies in *Drosophila*, in which inversions have been most intensely studied, have found that genetic differentiation is non-uniform across an inversion and that gene flow usually does occur between the standard and inverted chromosomes inside the inversion (Novitski and Braver, 1954; Hasson and Eanes, 1996; Schaeffer *et al.*, 2003; Schaeffer and Anderson, 2005). In these examples, genetic differentiation is typically highest near the inversion breakpoints. Double recombination events or gene conversion mediated by the formation of an inversion loop between the inverted and standard chromosomes will, over time, lead to significant gene flow within the inversion (Navarro *et al.*, 1997; Kovacevic and Schaeffer, 2000; Andolfatto *et al.*, 2001; Schaeffer *et al.*, 2003; Kennington *et al.*, 2006). The patterns of gene flux associated with an

Correspondence: Dr JW Thomas, Department of Human Genetics, Emory University School of Medicine, 615 Michael Street, Suite 301, Atlanta, GA 30322, USA.

E-mail: jthomas@genetics.emory.edu

Received 15 February 2010; revised 18 May 2010; accepted 28 May 2010; published online 23 June 2010

inversion can be particularly useful in identifying the targets of selection, which are expected to be in LD with each other as well as the inversion (see White *et al.*, 2007). Thus, the extent and pattern of gene flow associated with simple inversion polymorphisms will be influenced by the strength of selection to maintain LD, the size of the inversion, the rate of recombination and the age of the inversion.

Chromosomal polymorphisms involving more than one inversion have been found in natural populations and can be associated with patterns of gene flow that are distinct from the pattern described above for simple inversions. For example, the mouse *t*-complex comprises four non-overlapping inversions on mouse chromosome 17 and has been studied for decades with respect to its effect on recombination and association with meiotic drive (Lyon, 2003). Although gene conversion and some rare recombinants have been reported between wild-type and *t* chromosomes (Erhart *et al.*, 2002; Wallace and Erhart, 2008), strong suppression of recombination extends over the length of the *t*-complex. As a result, genetic differentiation between the *t* and wild-type chromosomes is uniformly high across the entire ~30–40 Mb *t*-complex (Lyon, 2003). Similarly, the X<sup>D</sup> chromosome in *Drosophila recens* is composed of a complex set of inversions and is associated with meiotic drive (Dyer *et al.*, 2007). The X<sup>D</sup> completely suppresses recombination between the X<sup>D</sup> and its non-distorting homolog, X<sup>ST</sup>, resulting in dramatic chromosome-wide LD spanning ~130 cM (Dyer *et al.*, 2007). Unlike the classic model of gene flow for simple inversions in *Drosophila*, chromosomal polymorphisms involving more than one inversion can suppress recombination over the entire length of the rearrangement for prolonged periods of time and lead to genetically distinct haplotypes associated with exceptionally large blocks of LD.

Although chromosomal polymorphisms comprising complex inversions can drastically reduce the frequency of recombination within the rearranged regions in individuals heterozygous for the polymorphism, normal recombination levels are expected within the alternative chromosomal rearrangements in individuals homozygous for the standard or inverted chromosomes. However, if homozygosity for one of the alternative arrangements, for example, the inverted arrangement, causes sterility or early lethality, the region within the inversion will not have the opportunity to recombine during meiosis. As a non-recombining segment of the genome, the inverted chromosome will then become subject to a series of population genetic forces that will result in the accumulation of deleterious mutations and genetic degeneration (Rice, 1994; Charlesworth and Charlesworth, 2000). A dramatic example of the long-term consequence of suppressed recombination is the mammalian Y, which has undergone massive genetic degeneration and lost almost all of its original genes except in the small pseudoautosomal region(s) in which recombination still occurs with the X chromosome (Graves, 2006). Newly arisen neo-Y chromosomes also typically show distinct signatures of genetic degeneration (Filatov *et al.*, 2000; Peichel *et al.*, 2004; Kondo *et al.*, 2006; Marais, 2007; Zhou *et al.*, 2008), as do other rare examples of non-recombining regions of the genome (Slawson *et al.*, 2006). These examples illustrate that under extreme circumstances, complex inversion

polymorphisms can eventually lead to regions of the genome in which recombination is rare or never occurs.

Recently, we described the first modern genetic and genomic characterization of a chromosomal polymorphism in the white-throated sparrow (*Zonotrichia albicollis*) that is extraordinary in respect to its phenotypic effects and genetic properties (Thomas *et al.*, 2008). In particular, the two alternative arrangements of the second chromosome, which will be referred to in this study as ZAL2 and ZAL2<sup>m</sup>, are linked to a plumage polymorphism such that individuals homozygous for ZAL2 are invariably associated with the tan-stripe (TS) morph, whereas individuals of the white-stripe (WS) morph are either heterozygous for the polymorphism (ZAL2/ZAL2<sup>m</sup>) or very rarely ZAL2<sup>m</sup> homozygotes (Thornycroft, 1966, 1975). In addition to the genetic association with plumage, the chromosomal polymorphism is linked to variation in social behavior such that WS individuals are, on average, more aggressive and less parental than their same-sex TS counterparts (Tuttle, 2003). WS and TS individuals occur at similar frequencies in both sexes and show an exceptionally strong negative assortative mating pattern in which >96% of all breeding pairs comprise individuals of opposite morphs (Falls and Kopachena, 1994). As a consequence of this breeding pattern, and perhaps because of reduced viability (Thornycroft, 1975), ZAL2<sup>m</sup>/ZAL2<sup>m</sup> birds are rare in the population with only a single ZAL2<sup>m</sup> homozygote having been detected in studies that combined karyotyped >600 birds (Thornycroft, 1975; Romanov *et al.*, 2009). Another consequence of this mating pattern is that ZAL2<sup>m</sup> is in a near constant state of heterozygosity maintained at the population level by balancing selection. At the molecular level, ZAL2<sup>m</sup> differs from ZAL2 by at least two nested inversions that, together, are predicted to span ~100 Mb and encompass ~1000 genes (Thomas *et al.*, 2008). In addition, limited population-based re-sequencing indicated that recombination was suppressed between ZAL2 and ZAL2<sup>m</sup> within the inverted interval and that as a result of this suppression of recombination between the chromosome types, as well as a paucity of ZAL2<sup>m</sup> homozygotes, ZAL2<sup>m</sup> could be a non-recombining segment of the genome and a model for the early stages of sex chromosome evolution (Thomas *et al.*, 2008). However, these conclusions were based on only a few loci ( $n=10$ ), and thus it remained unclear whether suppression of recombination between the chromosome types extended across the entire inverted interval, and whether or not ZAL2<sup>m</sup> truly contained a non-recombining segment of the genome. We report the results of our study in which we have sampled sequenced diversity at 62 loci distributed at ~1.7 Mb intervals across ZAL2 and ZAL2<sup>m</sup> to definitely establish the patterns of genetic differentiation and recombination associated with this chromosomal polymorphism.

## Materials and methods

### Source of DNA

White-throated sparrows were collected on the campus of Emory University in Atlanta, Georgia, during November and December of 2005–2007. A small blood sample was taken from a wing vein for DNA extraction and the morph of each bird was determined according to

the criteria described in Watt (1986) and Piper and Wiley (1989) and confirmed by PCR using the method described by Michopoulos *et al.* (2007). DNA from a single dark-eyed junco (*Junco hyemalis*) was collected from a locally captured bird. All procedures involving animals were approved by the Institutional Animal Care and Use Committee of the Emory University.

#### DNA sequencing

PCR primers were designed based on publicly available zebra finch genomic sequence (The Genome Center at Washington University, <http://genome.wustl.edu>) in regions conserved with chicken as described by Thomas *et al.* (2008). Primer sequences, orthologous position in zebra finch and specific annealing temperatures are listed in Supplementary Table 1. Each 25  $\mu$ l PCR contained final concentrations of 1  $\times$  PCR buffer, 1.5 mM MgCl<sub>2</sub>, 20 pmol of each primer, 0.2 mM of each deoxynucleotide triphosphate, 1.5 units of Taq or Platinum Taq DNA polymerase (Invitrogen, Carlsbad, CA, USA) and approximately 12.5–25 ng of genomic DNA. PCR cycling started with an initial denaturation at 94 °C for 5 min, followed by 35 cycles of 94 °C for 30 s, 55 °C or 58 °C for 30 s, 72 °C for 1 min and ended with a final extension at 72 °C for 7 min. PCR products were treated with shrimp alkaline phosphatase and exonuclease I (USB) before direct sequencing using the PCR primers or internal primers. All sequences were deposited in GenBank (accession numbers GU449138–GU449818 and GU450220–GU450272).

#### Single-nucleotide polymorphism (SNP) discovery and sequence annotation

Polymorphisms were automatically called using SNP detector (Zhang *et al.*, 2005) and all variants were manually confirmed before further analyses. Insertion and deletion polymorphisms were scored but not included in the subsequent analyses. The sequence at each locus was annotated for gene features based on annotation of the orthologous chicken genomic segments (ICGSC, 2004), as well as the presence of evolutionarily conserved regions as predicted by the PhastCons track on the genome browser of the University of California, Santa Cruz (Siepel *et al.*, 2005). Individual loci were ordered and spaced relative to one another based on a previously established white-throated sparrow–chicken comparative map (Thomas *et al.*, 2008) and the assembled zebra finch genome (taeGut1).

#### DNA sequence analysis

Haplotypes were reconstructed from raw genotype data using PHASE v.2.1.1 (Stephens *et al.*, 2001) under the default parameters. Each locus was phased individually and SNPs with <0.55 phasing confidence (primarily singletons) were assigned randomly to alternative haplotypes that were used to generate population genetic statistics for each locus in Supplementary Table 2. We also phased the concatenated data set to generate a complete haplotype, which was used to build the haplotype network and estimate LD and recombination. Although incorrectly phased haplotypes do not affect measures of diversity, they can subtly influence estimates of LD and recombination. In TS birds, all haplotypes were identified as ZAL2 chromosomes. In WS birds,

within the inversion we used fixed differences to distinguish between the ZAL2 and ZAL2<sup>m</sup> chromosomes.

Population genetic statistics were calculated in DnaSP v.4.50.3 unless otherwise noted (Rozas *et al.*, 2003). We defined neutral sites as non-coding and synonymous sites located outside of segments orthologous to regions detected as being evolutionarily conserved in the chicken genome by PhastCons (Siepel *et al.*, 2005). Summary statistics for the intervals inside and outside the inversion were calculated using the phased haplotype generated from the concatenated sequences except  $\pi$  and  $\theta$ , which were calculated as a weighted average of the individual loci. The neighbor-joining haplotype network was generated in SplitsTree4 (Huson and Bryant, 2006). Pairwise comparisons of non-random associations were performed in Haploview (Barrett *et al.*, 2005) and the  $r^2$  color scheme was used to illustrate LD between pairs of sites. Estimates and standard deviations for the population recombination rate parameter  $\rho = 4N_e r$  were assessed by Monte Carlo coalescent simulations in the *interval* algorithm in LDhat (Auton and McVean, 2007) conditioned on  $\theta_W$  from DnaSP sampling every 2000 of 10<sup>6</sup> iterations after an initial burn-in of 100 000. Finally, we tested for gene conversion using the algorithm developed by Betrán *et al.* (1997) implemented in DnaSP (Rozas *et al.*, 2003).

## Results

#### Data set

Previously, we reported an initial survey of the genetic differentiation between ZAL2 and ZAL2<sup>m</sup> based on sequencing a small number ( $n=10$ ) of loci (Thomas *et al.*, 2008). To carry out a more detailed study of the genetic differentiation, nucleotide diversity and recombination associated with this chromosomal polymorphism, we expanded our study to include a total of 62 loci that were spaced on average every 1.7 Mb along the ZAL2/2<sup>m</sup> chromosome in 4 TS and 8 WS birds, corresponding to a sample size of 16 ZAL2 and 8 ZAL2<sup>m</sup> chromosomes. Of the sequenced loci, 58 mapped within the inversion and totaled ~35 kb, whereas 4 loci mapped outside the inversion and totaled ~1.7 kb (Table 1). On the basis of annotation of the loci and our criteria for defining neutral sequence (see Materials and methods for details) the data set included ~14 and ~0.8 kb of neutral sites inside and outside the inversion, respectively (Table 1). Overall, we identified 297 SNPs, of which 279 mapped within the inversion and 18 mapped outside the inversion (Table 1). After excluding three tri-allelic SNPs, the final data set available for analyses comprised 277 SNPs within and 17 SNPs outside the inversion.

#### Genetic differentiation between ZAL2 and ZAL2<sup>m</sup>

In our earlier study we found high levels of genetic differentiation between ZAL2 and ZAL2<sup>m</sup> within the inversion and evidence of gene flow outside the inversion (Thomas *et al.*, 2008). Consistent with these findings we found that the majority (185/277) of SNPs within the inversion were fixed differences between the chromosomal arrangements (that is, one allele was invariably linked to ZAL2 and the alternative allele to ZAL2<sup>m</sup>), and only one SNP was a shared polymorphism

**Table 1** Diversity, divergence and summary statistics for the standard ZAL2 and inverted ZAL2<sup>m</sup> chromosomes in the white-throated sparrow

Region	Loci	Length (bp)	Segregating sites, S	$\pi$ (All)	$\pi$ (ZAL2)	$\pi$ (ZAL2 <sup>m</sup> )	$F_{ST}$ <sup>a</sup>	$d_{xy}$ <sup>b</sup>	$Tajima's$ $D^c$	$Tajima's$ $D$ (ZAL2)	$Tajima's$ $D$ (ZAL2 <sup>m</sup> )
Inside inversion	58	34 823 (13 789.3)	277 (167)	0.00293 (0.00449)	0.00044 (0.00075)	0.00030 (0.00042)	0.94	0.00583 (0.00871)	0.65	-0.74	-0.84
Outside inversion	4	1655 (777)	17 (14)	0.00191 (0.00296)	0.00160 (0.00311)	0.00238 (0.00455)	0.21	0.00190 (0.00296)	-0.82	-0.90	-0.22

Numbers in parentheses are calculations for neutral sites, noncoding, synonymous sites outside most conserved elements.

<sup>a</sup> $F_{ST}$  is calculated by comparing all ZAL2 chromosomes with all ZAL2<sup>m</sup> chromosomes.

<sup>b</sup>Divergence ( $d_{xy}$ ) between the ZAL2 and ZAL2<sup>m</sup> chromosomes is calculated as the average number of nucleotide substitutions per site.

<sup>c</sup>Tajima's D values were calculated using four tan-stripe (TS) and four white-stripe (WS) birds, thus representing the natural population frequency of the ZAL2 and ZAL2<sup>m</sup> chromosomes.

(that is, both alleles were present in both chromosomal arrangements). In contrast, outside the inversion we identified 17 SNPs, none of which were classified as fixed differences between the chromosomal arrangements and two were shared based on the complete haplotype phase estimate and not the individual locus phasing.

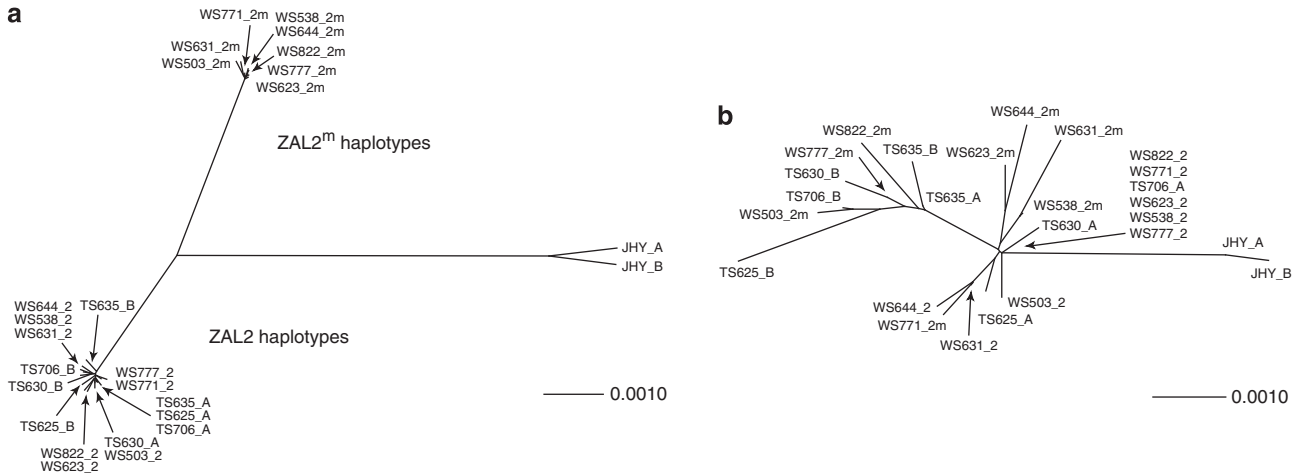
To formally quantify the level of genetic differentiation between the chromosomal arrangements inside and outside the inversion, we calculated the population genetic statistics  $F_{ST}$  and  $d_{xy}$  treating ZAL2 and ZAL2<sup>m</sup> as two separate populations (Table 1). As expected, based on the prevalence of fixed differences between ZAL2 and ZAL2<sup>m</sup>,  $F_{ST}$  inside the inversion was near the theoretical maximum of 1 ( $F_{ST} = 0.94$ ), consistent with a high degree of genetic differentiation and extremely limited gene flow between the two chromosomal arrangements within the inversion. Indeed, the average pairwise divergence in neutral sequence between ZAL2 and ZAL2<sup>m</sup> chromosome inside the inversion was elevated relative to outside the inversion ( $d_{xy} = 0.0087$  versus 0.0029). Outside the inversion, the  $F_{ST}$  value between the arrangements was 0.21. However, because we had no method to reliably assign ZAL2 and ZAL2<sup>m</sup> identity outside the inversion, we tested for significant population structure between TS and WS groups. Inside the inversion, population differentiation between WS and TS was significant ( $F_{ST} = 0.41$ ,  $P = 0.009$ ), whereas outside the inversion, there was no significant structure ( $F_{ST} = 0.06$ ,  $P = 0.09$ ; Excoffier *et al.*, 2005). Thus, these results show that genetic differentiation between ZAL2 and ZAL2<sup>m</sup> is uniformly high across the entire ~104 Mb inverted interval and low within the region outside the inversion, consistent with gene flow between the chromosomal arrangements being restricted to the small segment of the chromosome outside the inversion.

#### Patterns of haplotype and nucleotide diversity within and between ZAL2 and ZAL2<sup>m</sup>

The striking genetic differentiation between the chromosomal rearrangements inside compared with outside the inversion is consistent with the presence of two highly divergent haplotype groups each associated with the inverted segments of the ZAL2 and ZAL2<sup>m</sup> chromosomes. To visualize the relationships among the haplotypes, we constructed separate haplotype networks based on the phased concatenated haplotype inside and outside the inversion (Figure 1). As expected, the ZAL2 and ZAL2<sup>m</sup> haplotypes clustered into two distinct groups that were clearly differentiated from one another, whereas outside the inversion no clustering based on chromosomal arrangement was observed (Figure 1). Moreover, the short branches of the ZAL2 and ZAL2<sup>m</sup> groups illustrate the relatively low diversity observed inside the inversion within both chromosomal rearrangements compared with outside the inversion (Table 1 and Figure 1a).

The haplotype networks clearly indicate distinct differences in the genetic diversities of ZAL2<sup>m</sup>, ZAL2 and the region outside the inversion. To assess this difference we considered that natural populations of white-throated sparrows comprise ~50% TS (ZAL2/ZAL2) individuals and ~50% WS (ZAL2/ZAL2<sup>m</sup>) individuals (Lowther, 1961). Hence, at the population level, the inverted ZAL2<sup>m</sup> region, the homologous standard ZAL2





**Figure 1** ZAL2 and ZAL2<sup>m</sup> haplotype networks. Haplotype networks were constructed using the concatenated phased haplotypes from all loci within (a) and outside (b) the inversion. For the white-striped (WS) birds the ZAL2 and ZAL2<sup>m</sup> haplotypes are labeled as 2 or 2m. For the tan-striped (TS) birds and Junco (JHY), the haplotypes for each individual were arbitrarily labeled A and B.

region and the region outside the inversion exist in a 1:3:4 ratio. Under the neutral model, in which we assume the absence of selection and comparable rates of recombination, nucleotide diversity is directly proportional to the effective population size ( $\theta = 4N_e\mu$ ). In this context, we predicted that the ratio of neutral nucleotide diversity of ZAL2<sup>m</sup>, ZAL2 and the region outside the inversion would also be 1:3:4. The observed ratio of neutral nucleotide diversity between ZAL2<sup>m</sup> and ZAL2 was  $\sim 1:2$ , similar to our expected ratio. However, the ratio of diversity of ZAL2<sup>m</sup>, ZAL2 and the region outside the inversion was 1:2:10 (Table 1), which was lower than predicted by the model. This result could be caused by lower than expected diversity for ZAL2<sup>m</sup> and ZAL2, higher than expected diversity outside the inversion, or both.

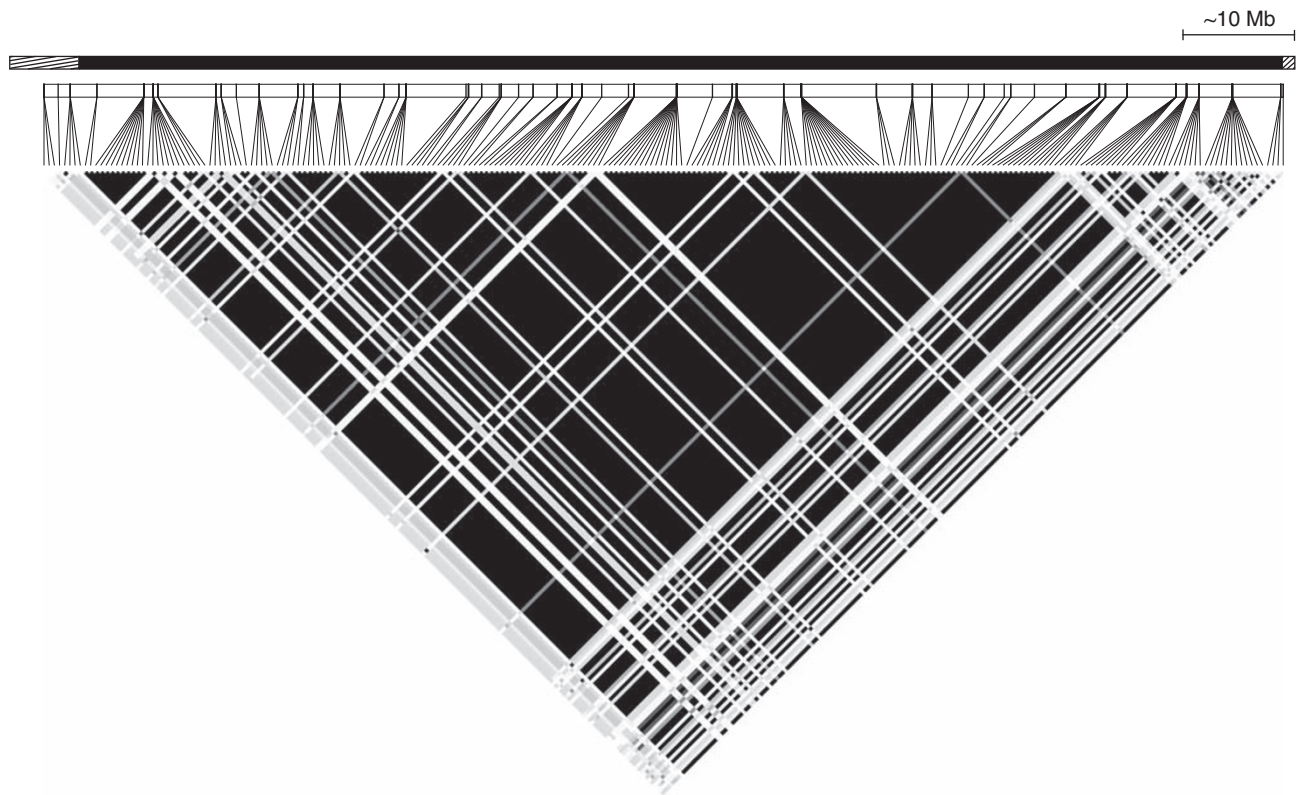
#### Linkage disequilibrium

The suppression of recombination between the chromosomal arrangements across the length of the inverted interval is expected to result in a strong signature of LD extending over  $\sim 100$  Mb inversion. To quantify the extent of LD associated with this chromosomal polymorphism, we measured LD between pairs of informative sites along the entire length of the chromosome (Figure 2). A total of 240 SNPs, 232 inside the inversion and 8 outside the inversion, were informative and, as expected, under a scenario of long-term suppressed recombination between the arrangements, a large fraction of the pairwise comparisons inside the inversion showed significant nonrandom association, whereas outside the inversion we did not detect significant patterns of LD (Figure 2). To summarize the level of LD inside and outside the inversion, we calculated ZnS between informative sites (Kelly, 1997), which was high inside the inversion (ZnS = 0.69) and much lower outside (ZnS = 0.10) the inversion. Thus, these results establish that the suppression of recombination between ZAL2 and ZAL2<sup>m</sup> has resulted in widespread LD that, based on our most proximal and distal markers, spans a minimum of  $\sim 104$  Mb inside the inversion.

#### Rates of recombination within the ZAL2 and ZAL2<sup>m</sup> haplotype groups

Our data strongly support a scenario in which recombination in WS (ZAL2/ZAL2<sup>m</sup>) birds is restricted to the interval outside the inversion. Because ZAL2<sup>m</sup>/ZAL2<sup>m</sup> individuals are extremely rare ( $< 1\%$  of the population; Thorneycroft, 1975; Falls and Kopachena, 1994; Romanov *et al.*, 2009), we had previously predicted that the inverted segment of the ZAL2<sup>m</sup> chromosome might be a non-recombining autosome (Thomas *et al.*, 2008). To test this hypothesis we applied the four-gamete test (Hudson and Kaplan, 1985) to analyse evidence of recombination within ZAL2<sup>m</sup> haplotypes, as well as ZAL2 haplotypes and the haplotypes outside the inversion on both chromosomes. As expected, we detected recombination within the ZAL2 haplotype ( $\geq 12$  events), and within haplotypes outside the inversion ( $\geq 1$  event). In contrast to our prediction, however, the four-gamete test indicated that recombination had also occurred within the ZAL2<sup>m</sup> haplotypes ( $\geq 2$  events).

Given the evidence for recombination within both ZAL2 and ZAL2<sup>m</sup> haplotypes, we next attempted to quantify the difference in population rate of recombination,  $\rho$ , between ZAL2 and ZAL2<sup>m</sup>. In particular, because of the large difference in the population frequency of ZAL2<sup>m</sup>/ZAL2<sup>m</sup> birds ( $< 1\%$ ) to ZAL2/ZAL2 birds ( $\sim 50\%$ ) (Thorneycroft, 1975; Falls and Kopachena, 1994; Romanov *et al.*, 2009), we expected a lower rate of recombination inside the inversion within the ZAL2<sup>m</sup> haplotypes versus within the ZAL2 haplotypes. Estimates of the population recombination rate among all ZAL2<sup>m</sup> chromosomes was similar to that of the ZAL2 chromosomes ( $\rho \pm \text{s.d.} = 8.48 \times 10^{-9} \pm 8.59 \times 10^{-9}$  and  $4.40 \times 10^{-8} \pm 5.05 \times 10^{-8}$  per base, respectively), although it should be noted that diversity levels within the inversion on ZAL2 and ZAL2<sup>m</sup> were low and our small sample size did not support accurate measurements of  $\rho$  (that is, note standard errors). Finally, we also predicted that outside the inversion, the rate of recombination would be greater than within the inversion because the presumed obligatory recombination event required for proper chromosome segregation in  $\sim 50\%$  of the population



**Figure 2** Linkage disequilibrium between informative sites across the ZAL2/ZAL2<sup>m</sup> chromosomes. The concatenated phased haplotypes were used to generate an LD plot across the ZAL2/ZAL2<sup>m</sup> chromosomes. Each of the 239 informative SNPs are plotted to scale with respect to their predicted position along the ZAL2 chromosome. Black squares indicate perfect LD between pairs of SNPs ( $r^2 = 1$ ), gray squares indicate pairs of SNPs with  $r^2$  between 0 and 1 and white squares show no LD. The striped segments of the line above the LD plot indicate regions outside the inversion and the solid (black) segment of the line indicates the region inside the inversion.

that are ZAL2/ZAL2<sup>m</sup> heterozygotes (WS) (Lowther, 1961; Thorneycroft, 1975) must occur in this small interval. Consistent with our prediction,  $\rho$  for this segment was estimated to be higher than observed inside the inversion ( $1.55 \times 10^{-7} \pm 2.18 \times 10^{-7}$  per base); again, note the large confidence intervals.

#### Relative rate of evolution on the ZAL2 versus ZAL2<sup>m</sup> chromosomes

The pattern of extensive recombination suppression between ZAL2 and ZAL2<sup>m</sup> within the inverted region suggests that Hill–Robertson interference among the linked loci could work to reduce the efficacy of selection and subsequently lead to the genetic degeneration of the ZAL2<sup>m</sup> chromosome (Hill and Robertson, 1966; Rice, 1994; Gordo and Charlesworth, 2001). Hence, we were interested in whether ZAL2<sup>m</sup> showed any signs of degeneration. Specifically, we tested for differences in the rates of evolution in the ZAL2 versus ZAL2<sup>m</sup> lineages, applying Tajima’s relative rate method using the  $\chi^2$ -test (Tajima, 1993) to fixed differences that mapped inside the inversion and could be unambiguously polarized using sequence from a closely related outgroup (*J. hyemalis*, Table 2). Although a greater number of mutations were inferred to have occurred in the ZAL2<sup>m</sup> versus ZAL2 lineage in all sequence classes examined, including nonsynonymous and other presumably functional sites, the observed differences were not statistically significant (Table 2). Thus, although the ZAL2<sup>m</sup> lineage may be

**Table 2** Mutations occurring along the ZAL2 and ZAL2<sup>m</sup> lineages

	All	Nonsynon	Synon	Conserved	Neutral
ZAL2	78	1	6	31	47
ZAL2 <sup>m</sup>	98	5	8	36	62
<i>P</i> -value <sup>a</sup>	0.13	0.102	0.593	0.54	0.15

This test includes the subset of fixed differences in which ancestral state could be unambiguously inferred using junco sequence. Conserved category includes all single-nucleotide polymorphisms (SNPs) within most conserved elements and neutral category includes all SNPs outside most conserved elements.

<sup>a</sup>*P*-values were calculated using Tajima’s relative rate test (Tajima, 1993).

associated with a slightly higher overall mutation rate than the ZAL2 lineage, we did not observe any overt signs of genetic degeneration expected on a chromosome with reduced levels of recombination.

## Discussion

### Gene flow, genetic differentiation and population structure between ZAL2 and ZAL2<sup>m</sup>

Unlike most simple inversion polymorphisms in which, over time, gene flow occurs between the standard and inverted arrangements except near the inversion breakpoints, complex inversion polymorphisms can be effective suppressors of recombination (Lyon, 2003; Munté

*et al.*, 2005; Dyer *et al.*, 2007). For example, the  $O_{3+4}$  inversion polymorphism in *Drosophila subobscura* differs from the standard  $O_{ST}$  arrangement by two overlapping inversions and suppresses gene flow between the two chromosome types, producing uniformly high genetic differentiation across the  $\sim 4$ Mb region (Munté *et al.*, 2005). Although evidence for gene conversion between the alternative arrangements is prevalent in this system, the limited gene flow in the absence of double recombination events is clearly shown by the high proportion of fixed differences (35%) and the level of divergence ( $d_{xy} = 1.5\%$ ) between the alternate arrangements (Munté *et al.*, 2005). Similarly, in the case of an even larger complex inversion polymorphism, the  $X_D$  chromosome in *D. recens*, over half of the polymorphisms identified on the  $X_D$  chromosome were fixed differences between the distorting and wild-type chromosomes (Dyer *et al.*, 2007). We also observed a lack of gene flow between ZAL2 and ZAL2<sup>m</sup> as indicated by a high frequency of fixed differences (67%) and divergence between the chromosomes approaching 1%. Moreover, the frequency of shared polymorphisms within the inversion was very low (<1%), and unlike the  $O_{3+4}$  and  $O_{ST}$  system, we failed to detect any gene conversion tracts between the alternative arrangements or between ZAL2<sup>m</sup> chromosomes within the inversion interval. In these examples from *Drosophila* and the ZAL2<sup>m</sup> polymorphism, there is no evidence for genetic exchange toward the center of the inversion polymorphisms, as predicted by models developed by Navarro *et al.* (1997) for single inversions, suggesting that multiple overlapping inversions are likely to completely suppress double cross-overs and can prevent gene flow across the entire inversion interval. Moreover, the very high  $F_{ST}$  value (0.94) between ZAL2 and ZAL2<sup>m</sup> inside the inversion is also quite uncommon. Indeed, similar values of  $F_{ST}$  in excess of >0.8 have been associated with special circumstances, such as homologous loci in non-recombining segments of sex chromosomes (Ironsides and Filatov, 2005) or between cryptic species (Terry *et al.*, 2000). Thus, the degree of genetic differentiation observed between ZAL2 and ZAL2<sup>m</sup> chromosomes is very high even for complex inversions, and is comparable to values observed in the most extreme cases of suppressed gene flow.

#### Dating and origin of ZAL2<sup>m</sup>

Using the level of neutral divergence between ZAL2 and ZAL2<sup>m</sup> chromosomes, we estimated the age of the current ZAL2<sup>m</sup> haplotypes (that is, time since recombination stopped) assuming a genomic mutation rate equivalent to that described for zebra finch,  $2.95 \times 10^{-9}$  substitutions per site per year (Balakrishnan and Edwards, 2009). Our estimate of  $\sim 2.95 \pm 0.3$  MY is similar to our previous estimate based on phylogenetic comparisons of noncoding and synonymous sequence between white-throated sparrows and the *J. hyemalis* outgroup (Thomas *et al.*, 2008). The time of ZAL2<sup>m</sup> origin predates the estimated time of divergence of the white-throated sparrow from other birds in the genus and is at apparent odds with the hypothesis that the inversion occurred in the white-throated sparrow lineage (Thornycroft, 1975). Although the polymorphism may be ancient and was simply lost in the other *Zonotrichia* sparrow lineages, an alternative explanation for its restricted presence in the white-throated sparrow is

through introgression as the result of hybridization. Hybridization in birds is well established (Grant and Grant, 1992; Mallet, 2005), and fertile offspring are common in cases in which the genetic divergence between the species is on the order observed between ZAL2 and ZAL2<sup>m</sup>, that is, <2% (Price and Bouvier, 2002). Although there is precedent for hybridization between the white-throated sparrow and the Junco (Dickerman, 1961), our haplotype networks do not support a recent introgression of either chromosome arrangement from that species. However, relatively recent introgression resulting from the hybridization with another species might explain the high level of genetic divergence between ZAL2 and ZAL2<sup>m</sup>, as well as the lack of recombination within the inversion. Studies of hybrid genomes have shown that inversions can produce patterns of heterogeneous recombination and high genetic divergence (Rieseberg *et al.*, 1999; Feder *et al.*, 2003; Panithanarak *et al.*, 2004; Stump *et al.*, 2005; Noor *et al.*, 2007). In addition, because collinear regions are expected to homogenize more quickly than non-collinear regions (Rieseberg *et al.*, 1999), time since hybridization is another factor to take into consideration. However, note that the accurate molecular dating of the inversion may be confounded by the unique genetic attributes of the polymorphism, and a direct comparison to the speciation events within this genus were not carried out. Thus, we acknowledge that the above scenarios are not the only possible histories for the inversion. Future studies focused on genome-wide patterns of nucleotide diversity and haplotype structures should help clarify whether or not the ZAL2/ZAL2<sup>m</sup> polymorphism could have resulted from a recent hybridization event.

#### Extreme LD associated with the ZAL2/ZAL2<sup>m</sup> system

Theoretically, the adaptive significance of inversions lies in their capability to suppress recombination, thereby maintaining LD between multiple alleles favored by natural selection (Dobzhansky, 1937; Kirkpatrick and Barton, 2006), and the pattern of LD is dependent on the rate of recombination as well as the strength of selection to maintain linkage between the alleles. In the case of ZAL2/ZAL2<sup>m</sup> system, we detected two dominant haplotypes representing each arrangement that spanned  $\sim 100$  Mb inside the inversion, linking alleles from  $\sim 1000$  genes into a super-gene complex. The extended LD we observed in this system is more extreme than that observed in the mouse *t*-complex, which spans  $\sim 30$ – $40$  Mb (Lyon, 2003), and comparable to that observed in the distorting ( $X_D$ ) in *D. recens* (Dyer *et al.*, 2007). In that case, LD was found to essentially extend across the entire X chromosome ( $\sim 130$  cM; Dyer *et al.*, 2007). As far as we are aware, the pattern of LD associated with ZAL2/ZAL2<sup>m</sup> is the most extreme example of long-range LD yet to be reported in vertebrates.

#### Comparison of the ZAL2/ZAL2<sup>m</sup> chromosome pair to sex chromosomes

Previously, we proposed that the ZAL2/ZAL2<sup>m</sup> chromosomes were mimicking the early stages of sex chromosome evolution and that ZAL2<sup>m</sup> shared a number of features with the Y (W) chromosome (Thomas *et al.*, 2008). In particular, these features included suppression of recombination between ZAL2 and ZAL2<sup>m</sup> over most



of their length due to inversions, a negative assortative mating system in which >96% of breeding pairs consist of heterogametic (ZAL2/ZAL2<sup>m</sup>) × homogametic (ZAL2/ZAL2) individuals and the maintenance of ZAL2<sup>m</sup> in a near constant state of heterozygosity (Thomas *et al.*, 2008). Although our results from this study do not require us to reconsider these shared features, the detection of recombination within ZAL2<sup>m</sup> haplotypes refutes the additional prediction that the region inside the inversion on ZAL2<sup>m</sup> may represent a non-recombining autosome. Thus, although ZAL2<sup>m</sup>/ZAL2<sup>m</sup> individuals in the population are rare, the detection of historical recombination events associated with this haplotype suggest that at least some of these individuals are fertile. Indeed, a male ZAL2<sup>m</sup> homozygote has been described that was able to reproduce (Falls and Kopachena, 1994).

Given the presence of recombination, however rare it may be on ZAL2<sup>m</sup>, it was therefore not surprising that we did not detect a significant difference in the accumulation of potentially deleterious mutations on ZAL2<sup>m</sup> compared with ZAL2. The complete cessation of recombination, as with neo-Y chromosomes, such as those in *Drosophila miranda* (Bachtrog, 2004), stickleback and medaka fish (Peichel *et al.*, 2004; Kondo *et al.*, 2006), the black muntjac (Zhou *et al.*, 2008) and *Silene latifolia* (Filatov and Charlesworth, 2002; Marais, 2007), leads to reduced efficacy of selection and accumulation of mutations (Charlesworth and Charlesworth, 2000). However, even very low levels of recombination can effectively prevent the process of genetic degeneration observed in non-recombining chromosomes (Haddrill *et al.*, 2007). In this context, and in light of the results from the loci sampled in this study, we can conclude that ZAL2<sup>m</sup> is unlikely to be mimicking the initial phases of genetic deterioration observed on neo-Y chromosomes.

It should be noted that our inference of recombination within ZAL2<sup>m</sup> chromosomes relies on the accurate reconstruction of haplotypes from genotype information. Because the number of predicted recombination events within ZAL2<sup>m</sup> chromosomes is small and the frequency of the minor alleles of these sites is low, incorrect phase estimation could lead to false positives. In addition, the four-gamete test relies on the assumption of infinite sites, which excludes the possibility of independent recurrent mutations. Considering that the minor allele frequency of the sites involved is generally low, independent mutations within the ZAL2 and ZAL2<sup>m</sup> lineages could produce patterns of variation that resemble those resulting from recombination. However, given that we did not detect the genetic degeneration that is characteristic of non-recombining sequences, we believe that these are true signatures of recombination generated by rare ZAL2<sup>m</sup> homozygotes.

Although we no longer consider ZAL2<sup>m</sup> a non-recombining chromosome, we did note potentially low levels of diversity associated with the ZAL2/ZAL2<sup>m</sup> system that are typical of sex chromosomes. For example, the mammalian and plant Y and the avian W chromosomes, as well as other regions of low recombination, can show 20- to 30-fold reductions in diversity even after correcting for differences in  $N_e$  (Begun and Aquadro, 1992; Filatov *et al.*, 2000; Jensen *et al.*, 2002; Berlin and Ellegren, 2004; Hellborg and Ellegren, 2004; Betancourt *et al.*, 2009). Similarly, the W chromosome has been

associated with lower than expected levels of nucleotide diversity in other birds (Montell *et al.*, 2001; Berlin and Ellegren, 2004). In the case of ZAL2/ZAL2<sup>m</sup> polymorphism we observed that diversity within the inversion on ZAL2 and ZAL2<sup>m</sup> was 5 and 10 times lower, respectively, than outside the inversion. It has been argued that intense sexual selection could be responsible for reductions in diversity on sex chromosomes by further reducing  $N_e$  of the Y chromosome (Caballero, 1995; Nagylaki, 1995; Charlesworth, 1996); however, such a mechanism cannot explain the reduced diversity observed on ZAL2<sup>m</sup>. Hence, natural selection is the only factor that can explain the reduced variability of sex chromosomes and regions of low recombination alike (Hellborg and Ellegren, 2004). Further efforts to quantify patterns of recombination and diversity in the white-throated sparrow genome will provide the necessary context to understand how rates of recombination have influenced the regions inside and outside the inversion.

#### Implications for identifying specific genes underlying the phenotypes associated with the ZAL2<sup>m</sup> polymorphism

One of the compelling reasons to study ZAL2<sup>m</sup> is the opportunity to identify the genetic basis of the phenotypic variation associated with the inversion. Previous studies have localized candidate regions within inversions by examining patterns of recombination between the wild-type and inverted arrangements and identifying region in LD as targets of selection. For example, in a study of the 2La inversion in *Anopheles gambiae*, White *et al.* (2007) examined patterns of divergence and LD to localize candidate regions involved in aridity tolerance. In the case of ZAL2/ZAL2<sup>m</sup> polymorphism, LD across the entire inversion will preclude further localization of candidate regions or genes by standard recombination-based mapping. In addition, if we consider that divergence between ZAL2 and ZAL2<sup>m</sup> within the inversion is on the order of 1%, conservatively estimate that 50% of the differences are fixed or near-fixed differences between the arrangements, and that the inversion is ~100 Mb, then we can predict that there are at least 500 000 single-nucleotide differences that distinguish ZAL2 from ZAL2<sup>m</sup>. Thus, it is possible that the TS and WS birds differ from each other in many traits.

#### Conclusions

Our population genetic analysis reveals that the ZAL2<sup>m</sup> arrangement suppresses recombination in the heterokaryotype, resulting in reduced gene flow, high levels of genetic differentiation, extensive population structure and LD between the alternate arrangements. Although we no longer consider ZAL2<sup>m</sup> a non-recombining chromosome, we believe that it will be valuable as a model for understanding how selection acts to reduce diversity in genomic regions with low recombination rates. Future studies of recombination and diversity at unlinked autosomal and sex-linked loci will likely shed light on the evolution of the ZAL2/ZAL2<sup>m</sup> system.

#### Conflict of interest

The authors declare no conflict of interest.



## Acknowledgements

We thank RT Sullivan, GK Tharp and J Griffin for their computational support, and M Cáceres, D Cutler, K Dyer, T Schlenke and Y Tao for advice on population genetic analyses and comments on the paper. LYH was supported by a Ford Foundation Pre-doctoral Fellowship and a grant from the Center of Behavioral Neuroscience awarded to DLM. DLM and JWT were supported by a grant from the National Institutes of Health (1R21MH082046).

## References

- Andolfatto P, DePaulis F, Navarro A (2001). Inversion polymorphisms and nucleotide variability in *Drosophila*. *Genet Res Camb* **77**: 1–8.
- Auton A, McVean G (2007). Recombination rate estimation in the presence of hotspots. *Genome Res* **17**: 1219–1227.
- Bachtrog D (2004). Evidence that positive selection drives Y-chromosome degeneration in *Drosophila miranda*. *Nat Genet* **36**: 518–522.
- Balakrishnan CN, Edwards SV (2009). Nucleotide variation, linkage disequilibrium and founder-facilitated speciation in wild-populations of the zebra finch (*Taeniopygia guttata*). *Genetics* **181**: 645–660.
- Barrett JC, Fry B, Maller J, Daly MJ (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**: 263–265.
- Begun DJ, Aquadro CF (1992). Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* **356**: 519–520.
- Berlin S, Ellegren H (2004). Chicken W: a genetically uniform chromosome in a highly variable genome. *Proc Natl Acad Sci USA* **101**: 15967–15969.
- Betancourt AJ, Welch JJ, Charlesworth B (2009). Reduced effectiveness of selection caused by a lack of recombination. *Curr Biol* **19**: 655–660.
- Betrán E, Rozas J, Navarro A, Barbadilla A (1997). The estimation of the number and length distribution of gene conversion tracts from population DNA sequence data. *Genetics* **146**: 89–99.
- Caballero A (1995). On the effective size of populations with separate sexes, with particular reference to sex-linked genes. *Genetics* **139**: 1007–1011.
- Charlesworth B (1996). Background selection and patterns of genetic diversity in *Drosophila melanogaster*. *Genet Res* **68**: 131–149.
- Charlesworth B, Charlesworth D (2000). The degeneration of Y chromosomes. *Phil Trans R Soc B* **355**: 1563–1572.
- Dickerman RW (1961). Hybrids among the fringillid genera *Junco-Zonotrichia* and *Melospiza*. *Auk* **78**: 627–632.
- Dobzhansky T (1937). *Genetics and the Origin of Species*. Columbia University Press: New York.
- Dobzhansky T (1950). Genetics of natural populations. XIX. Origin of heterosis through natural selection in populations of *Drosophila pseudoobscura*. *Genetics* **35**: 288–302.
- Dobzhansky T, Sturtevant AH (1938). Inversions in the chromosomes of *Drosophila pseudoobscura*. *Genetics* **23**: 28–64.
- Dyer K, Charlesworth B, Jaenike J (2007). Chromosome-wide linkage disequilibrium as a consequence of meiotic drive. *Proc Natl Acad Sci USA* **104**: 1587–1592.
- Erhart MA, Lekgothoane S, Grenier J, Nadeau JH (2002). Pattern of segmental recombination in the distal inversion of the mouse t-haplotypes. *Mamm Genome* **13**: 438–444.
- Excoffier L, Laval G, Schneider S (2005). Arlequin ver. 3.0: an integrated software package for population genetic data analysis. *Evol Bioinform Online* **1**: 47–50.
- Falls JB, Kopachena JG (1994). White-throated sparrow (*Zonotrichia albicollis*) In: Poole A, Gill F (eds). *The Birds of North America*. The Academy of Natural Sciences: Philadelphia.
- Feder JL, Roethele JB, Filchak KE, Niedbalski J, Romero-Severson J (2003). Evidence for inversions related to sympatric host race formation in the apple maggot fly, *Rhagoletis pomonella* (Diptera: Tephritidae). *Genetics* **163**: 939–953.
- Filatov DA, Charlesworth D (2002). Substitution rates in the X- and Y-linked genes of the plants, *Silene latifolia* and *S. dioica*. *Mol Biol Evol* **19**: 898–907.
- Filatov DA, Moneger F, Negrutiu I, Charlesworth D (2000). Low variability in a Y-linked plant gene and its implications for Y-chromosome evolution. *Nature* **404**: 388–390.
- Gordo I, Charlesworth B (2001). Genetic linkage and molecular evolution. *Curr Biol* **11**: R684–R686.
- Grant PR, Grant BR (1992). Hybridization of bird species. *Science* **256**: 193–197.
- Graves JAM (2006). Sex chromosome specialization and degeneration in mammals. *Cell* **124**: 901–914.
- Hadrill PR, Halligan DL, Tomaras D, Charlesworth B (2007). Reduced efficacy of selection in regions of the *Drosophila* genome that lack crossing over. *Genome Biol* **8**: R18.
- Hartl DL (1975). Genetic dissection of segregation distortion II. Mechanism of suppression of distortion by certain inversions. *Genetics* **80**: 539–547.
- Hasson E, Eanes WF (1996). Contrasting histories of three gene regions associated with *In(3L)Payne* of *Drosophila melanogaster*. *Genetics* **144**: 1565–1575.
- Hellborg L, Ellegren H (2004). Low levels of nucleotide diversity in mammalian Y chromosomes. *Mol Biol Evol* **21**: 158–163.
- Hill WG, Robertson A (1966). The effect of linkage on limits to artificial selection. *Genet Res* **8**: 269–294.
- Hoffmann AA, Sgro CM, Weeks AR (2004). Chromosomal inversion polymorphisms and adaptation. *TREE* **19**: 482–488.
- Hudson RR, Kaplan NL (1985). Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**: 147–164.
- Huson DH, Bryant D (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**: 254–267.
- ICGSC (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**: 695–716.
- Ironside JE, Filatov DA (2005). Extreme population structure and high interspecific divergence of the *Silene* Y chromosome. *Genetics* **171**: 705–713.
- Jensen MA, Charlesworth B, Kreitman M (2002). Patterns of genetic variation at a chromosome 4 locus of *Drosophila melanogaster*. *Genetics* **160**: 493–507.
- Kelly JK (1997). A test of neutrality based on interlocus associations. *Genetics* **146**: 1197–1206.
- Kennington JW, Partridge L, Hoffmann AA (2006). Patterns of diversity and linkage disequilibrium within the cosmopolitan inversion *In(3R)Payne* in *Drosophila melanogaster* are indicative of coadaptation. *Genetics* **172**: 1655–1663.
- Kirkpatrick M, Barton N (2006). Chromosome inversions, local adaptation and speciation. *Genetics* **173**: 419–434.
- Kondo M, Hornung U, Nanda I, Imai S, Sasaki T, Shimizu A *et al.* (2006). Genomic organization of the sex-determining and adjacent regions of the sex chromosomes of medaka. *Genome Res* **16**: 815–826.
- Kovacevic M, Schaeffer SW (2000). Molecular population genetics of X-linked genes in *Drosophila pseudoobscura*. *Genetics* **156**: 155–172.
- Lahn BT, Page DC (1999). Four evolutionary strata on the human X chromosome. *Science* **286**: 964–967.
- Lowther JK (1961). Polymorphism in the white-throated sparrow, *Zonotrichia albicollis* Gmelin. *Can J Zool* **39**: 281–292.
- Lyon MF (2003). Transmission ratio distortion in mice. *Annu Rev Genet* **37**: 393–408.

- Mallet J (2005). Hybridization as an invasion of the genome. *TREE* **20**: 229–237.
- Marais GAB (2007). Sex chromosomes and mitochondrial DNA polymorphism in birds: the Hill-Robertson effects extend from nucleus to mitochondria. *Heredity* **99**: 357–358.
- Michopoulos V, Maney DL, Morehouse CB, Thomas JW (2007). A genotyping assay to determine plumage morph in the white-throated sparrow, *Zonotrichia albicollis*. *Auk* **124**: 1330–1335.
- Montell H, Fridolfsson A-K, Ellegren H (2001). Contrasting levels of nucleotide diversity on the avian Z and W sex chromosomes. *Mol Biol Evol* **18**: 2010–2016.
- Munté A, Rozas J, Aguadé M, Segarra C (2005). Chromosomal inversion polymorphism leads to extensive genetic structure: a multilocus survey in *Drosophila pseudoobscura*. *Genetics* **39**: 1573–1581.
- Nagyaki T (1995). The inbreeding effective population number in dioecious populations. *Genetics* **148**: 1245–1255.
- Navarro A, Betrán E, Barbadilla A, Ruiz A (1997). Recombination and gene flux caused by gene conversion and crossing over in inversion heterozygotes. *Genetics* **146**: 695–709.
- Noor MAF, Garfield DA, Schaeffer SW, Machado CA (2007). Divergence between the *Drosophila pseudoobscura* and *D. persimilis* genome sequences in relation to chromosomal inversions. *Genetics* **177**: 1417–1428.
- Noor MAF, Grams KL, Bertucci LA, Reiland J (2001). Chromosomal inversions and the reproductive isolation of species. *Proc Natl Acad Sci USA* **98**: 12084–12088.
- Novitski E, Braver G (1954). An analysis of crossing over within a heterozygous inversion in *Drosophila melanogaster*. *Genetics* **39**: 197–209.
- Panithanarak T, Hauffe HC, Dallas JF, Glover A, Ward RG, Searle JB (2004). Linkage-dependent gene flow in a house mouse chromosomal hybrid zone. *Evolution* **58**: 37–47.
- Peichel CL, Ross JA, Matson CK, Dickson M, Grimwood J, Schmutz J *et al.* (2004). The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome. *Curr Biol* **14**: 1416–1424.
- Piper WH, Wiley RH (1989). Distinguishing morphs of the white-throated sparrow in basic plumage. *J Field Ornithol* **60**: 73–83.
- Price TD, Bouvier MM (2002). The evolution of F1 postzygotic incompatibilities in birds. *Evolution* **56**: 2083–2089.
- Rice WR (1994). Degeneration of a nonrecombining chromosome. *Science* **263**: 230–232.
- Rieseberg LH, Whitton J, Gardner K (1999). Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* **152**: 713–727.
- Romanov MN, Tuttle EM, Houck ML, Modi WS, Chemnick LG, Korody ML *et al.* (2009). The value of avian genomics to the conservation of wildlife. *BMC Genomics* **10**: S10.
- Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496–2497.
- Schaeffer SW, Anderson WW (2005). Mechanisms of genetic exchange within the chromosomal inversions of *Drosophila pseudoobscura*. *Genetics* **171**: 1729–1739.
- Schaeffer SW, Goetting-Minesky MP, Kovacevic M, Peoples JR, Graybill JL, Miller JM *et al.* (2003). Evolutionary genomics of inversions in *Drosophila pseudoobscura*: evidence for epistasis. *Proc Natl Acad Sci USA* **100**: 8319–8324.
- Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K *et al.* (2005). Evolutionarily conserved elements in vertebrate, insect, worm and yeast genomes. *Genome Res* **15**: 1034–1050.
- Slawson EE, Shaffer CD, Malone CD, Leung W, Kellmann E, Shevchek RB *et al.* (2006). Comparison of dot chromosome sequences from *D.melanogaster* and *D. virilis* reveals an enrichment of DNA transposon sequences in heterochromatic regions. *Genome Biol* **7**: R15.
- Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, Barnard J *et al.* (2005). A common inversion under selection in Europeans. *Nat Genet* **37**: 129–137.
- Stephens M, Smith N, Donnelly P (2001). A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* **68**: 978–989.
- Stump AD, Fitzpatrick MC, Lobo NF, Traoré S, Sagnon N, Costantini C *et al.* (2005). Centromere-proximal differentiation and speciation in *Anopheles gambiae*. *Proc Natl Acad Sci USA* **102**: 15930–15935.
- Sturtevant AH (1921). A case of rearrangement of genes in *Drosophila*. *Proc Natl Acad Sci USA* **7**: 235–237.
- Sturtevant AH, Beadle GW (1936). The relations of inversions in the X chromosome of *Drosophila melanogaster* to crossing over and disjunction. *Genetics* **21**: 554–604.
- Tajima F (1993). Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* **135**: 599–607.
- Terry A, Bucciarelli G, Bernardi G (2000). Restricted gene flow and incipient speciation in disjunct Pacific Ocean and Sea of Cortez populations of a reef fish species, *Girella nigricans*. *Evolution* **54**: 652–659.
- Thomas JW, Cáceres M, Lowman JJ, Morehouse CB, Short ME, Baldwin EL *et al.* (2008). The chromosomal polymorphism linked to variation in social behavior in the white-throated sparrow (*Zonotrichia albicollis*) is a complex rearrangement and a suppressor of recombination. *Genetics* **179**: 1455–1468.
- Thornycroft HB (1966). Chromosomal polymorphism in the white-throated sparrow, *Zonotrichia albicollis* (Gmelin). *Science* **154**: 1571–1572.
- Thornycroft HB (1975). A cytogenetic study of the white-throated sparrow, *Zonotrichia albicollis* Gmelin. *Evolution* **29**: 611–621.
- Tuttle EM (2003). Alternative reproductive strategies in the white-throated sparrow: behavioral and genetic evidence. *Behav Ecol* **14**: 425–432.
- Wallace LT, Erhart MA (2008). Recombination within mouse t haplotypes has replaced significant segments of t-specific DNA. *Mamm Genome* **19**: 263–271.
- Watt DJ (1986). Plumage brightness index for white-throated sparrows. *J Field Ornithol* **57**: 105–115.
- White BJ, Hahn MW, Pombi M, Cassone BJ, Lobo NF, Simard F *et al.* (2007). Localization of candidate regions maintaining a common polymorphic inversion (2La) in *Anopheles gambiae*. *PLoS Genet* **3**: e217.
- Zhang J, Wheeler DA, Yakub I, Wei S, Sood R, Rowe W *et al.* (2005). SNPdetector: a software tool for sensitive and accurate SNP detection. *PLoS Comput Biol* **5**: e53.
- Zhou Q, Wang J, Huang L, Nie W, Wang J, Liu Y *et al.* (2008). Neo-sex chromosomes in the black muntjac recapitulate incipient evolution of mammalian sex chromosomes. *Genome Biol* **9**: R98.

Supplementary Information accompanies the paper on Heredity website (<http://www.nature.com/hdy>)