MOLECULAR ECOLOGY

A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution

Journal:	*Molecular Ecology*
Manuscript ID:	Draft
Manuscript Type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Kawakami, Takeshi; Uppsala University, Department of Evolutionary Biology Smeds, Linnea; Uppsala University, Department of Evolutionary Biology Backstrom, Niclas; Uppsala University, Department of Evolutionary Biology Husby, Arild; Uppsala University, Department of Animal Ecology Qvarnstrom, Anna; Uppsala University, Department of Animal Ecology Mugal, Carina; Uppsala University, Department of Evolutionary Biology Olason, Pall; Uppsala University, Department of Evolutionary Biology Ellegren, Hans; Uppsala University, Department of Evolutionary Biology;
Keywords:	Birds, Genomics/Proteomics, Molecular Evolution, Speciation

A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution

Takeshi Kawakami[1a], Linnéa Smeds[1a], Niclas Backström[1], Arild Husby[2], Anna Qvarnström[2], Carina F. Mugal[1], Pall Olason[1,3] and Hans Ellegren[1*]

[1]Department of Evolutionary Biology, Evolutionary Biology Centre (EBC), Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden

[2]Department of Animal Ecology, Evolutionary Biology Centre (EBC), Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden

[3]Wallenberg Advanced Bioinformatics Infrastructure (WABI), Science for Life Lab, Uppsala University, Husargatan 3, SE-751 23 Uppsala, Sweden

[a]These authors contributed equally

*Correspondence: Hans Ellegren: hans.ellegren@ebc.uu.se

Running head: Genome assembly and genetic map of flycatcher

Keywords: collared flycatcher, linkage map, recombination rate, SNP array

Abstract

Detailed linkage and recombination rate maps are necessary to use the full potential of genome sequencing and population genomic analyses. We used a custom collared flycatcher 50K SNP array to develop a high-density linkage map with 37,262 markers assigned to 34 linkage groups in 33 autosomes and the Z chromosome. The best-order map contained 4,215 markers, with a total distance of 3,132 cM and a mean genetic distance between markers of 0.12 cM. Facilitated by the array being designed to include markers from most scaffolds, we obtained a second-generation assembly of the flycatcher genome that approaches full chromosome sequences (N50 super-scaffold size 20.2 Mb and with 1.042 Gb (out of 1.116 Gb) anchored to and mostly ordered and oriented along chromosomes). We found that flycatcher and zebra finch chromosomes are entirely syntenic but that inversions at mean rates of 1.5-2.0 event (6.6-7.5 Mb) per My have changed the organization within chromosomes, rates high enough for inversions to potentially have been involved with many speciation events during avian evolution. The mean recombination rate was 3.1 cM/Mb and correlated closely with chromosome size, from 2 cM/Mb for chromosomes >100 Mb to >10 cM/Mb for chromosomes <10 Mb. This size-dependence seemed entirely due to an obligate recombination event per chromosome; if 50 cM was subtracted from the genetic lengths of chromosomes, the rate per physical unit DNA was constant across chromosomes. Flycatcher recombination rate showed similar variation along chromosomes as chicken but lacked the large interior recombination desserts characteristic of zebra finch chromosomes.

Introduction

At a time when draft genome sequencing and assembly are practicable for most study organisms (Ellegren 2014), other types of critical genetic information may represent limiting steps in population and evolutionary genetic studies. One such factor is detailed linkage maps and the associated inference of how the rate of recombination varies across the genome (Dumont & Payseur 2008). Linkage maps enable anchoring and ordering of scaffolds along chromosomes (Heliconius Genome Consortium 2012; Huang *et al.* 2013). This is necessary for making full use of the unprecedented power provided by next-generation sequencing technology, which, in the absence of physical mapping approaches (like BAC end sequencing and fingerprinting), does not provide chromosome sequences. Moreover, recombination is a critical parameter in governing the degree and nature of intraspecific diversity as well as interspecific divergence. For example, the rate of recombination is expected to correlate positively with local levels of nucleotide diversity (McGaugh *et al.* 2012; Cutter & Payseur 2013; Campos *et al.* 2014) and with the rate of adaptive evolution (Presgraves 2005; Campos *et al.* 2014), and there is an increasing awareness that recombination moulds the evolution of base composition via GC-biased gene conversion (Duret & Arndt 2008; Webster & Hurst 2012). Also, recombination may be a critical factor in shaping the genomic landscape of species differentiation (Butlin 2005).

Large pedigrees are needed for the development of linkage maps and obtaining such samples can be challenging for many non-model species. Species that are difficult to breed in captivity and/or to monitor and sample in natural settings, or which have long generation times and/or small litter sizes, are examples of organisms that may be problematic in this context. Unfortunately, this applies to many natural populations of species of relevance in ecological or evolutionary research. However, in natural populations of birds, acquiring pedigree

material is greatly facilitated in species that readily accept breeding in artificial nest boxes and display high site fidelity. This is the case for our study species, the collared flycatcher (*Ficedula albicollis*), and it has also made it to be one of the most well-studied avian models for questions such as life history evolution, quantitative genetics and speciation (Ellegren *et al.* 1996; Gustafsson *et al.* 1995; Veen *et al.* 2001; Qvarnstrom *et al.* 2006; Saether *et al.* 2007; Qvarnström *et al.* 2010; Sætre & Sæther 2010; Ellegren *et al.* 2012).

Available evidence, notably from chicken *G. gallus* (ICGSC 2004; Groenen *et al.* 2009) and zebra finch *Taeniopygia guttata* (Stapley *et al.* 2008; Backström *et al.* 2010a), indicates an unusual heterogeneity in the rate of recombination within avian genomes. One determinant of this variation comes from the fact that bird chromosomes differ considerably in size. Moreover, data from zebra finch and to some extent also chicken show a very strong bias for recombination in larger chromosomes to be concentrated to end regions (Groenen *et al.* 2009; Backström *et al.* 2010a). It is not yet known what the underlying mechanism or evolutionary force driving such pattern might be, or whether it is in fact a general feature of bird chromosomes. Another characteristic of avian genomes is an unusual stability of the karyotype (Griffin *et al.* 2007; Ellegren 2010). The majority of species have about 40 pairs of chromosomes and inter-chromosomal rearrangements are rare (Ellegren 2013). For example, only one fusion and one fission event separate the chicken and zebra finch karyotypes despite the fact that these species represent two of the most divergent lineages of contemporary birds (Warren *et al.* 2010). However, there are indications that intra-chromosomal rearrangements occur more frequently (Skinner & Griffin 2012), although the rate and more precise pattern of this remains to be revealed.

Here we present the development of a high-density genetic linkage map of the collared flycatcher based on genotyping with a 50K (50,000) SNP array in a multi-generation pedigree of >600 birds from a natural population. This effort was motivated from several perspectives. First, having recently generated a draft flycatcher genome assembly (Ellegren et al. 2012), we were keen to confidently be able to place, order and orient scaffolds along chromosomes and thereby arrive at an assembly with essentially continuous chromosome sequences. The strategy for achieving this was based on designing an array with SNPs from the majority of all scaffolds, with the aim to place these onto a linkage map. Second, with the access to an updated genome assembly together with detailed information on recombination fractions between markers, we wanted to investigate the recombination landscape in an avian genome at high resolution. Third, with the access to a short read-based genome assembly with unusual continuity, we sought to reveal the character of avian chromosomal evolution by making a high-resolution comparison of flycatcher genome organisation with the only two avian genomes physically assembled and sequenced with Sanger technology, i.e. chicken (ICGSC 2004) and zebra finch (Warren et al. 2010).

Material and methods

Specimens

Blood samples were collected from collared flycatcher (n = 655) families breeding on the Baltic Sea island Öland (56°44′N 16°40′E) from 2002 to 2011. The pedigree consisted of four generations; 204 individuals in the parental generation and 451 F_1-F_3 progenies (Supplementary Figure 1). DNA was extracted from blood samples using a standard proteinase K digestion/phenol-chloroform purification protocol (Sambrook *et al.* 1989).

Genotyping with a 50K SNP array

A 50K SNP array for collared flycatcher has recently been developed by selecting markers from >10 million SNPs identified in genomic re-sequencing of 10 unrelated collared flycatchers (from our study population) and 10 pied flycatchers *Ficedula hypoleuca* (Kawakami *et al.* 2014). The bulk of markers were chosen based on a number of criteria set to maximize the usefulness in collared flycatchers, including polymorphism level in the sequencing sample, even distribution across the genome as judged by comparative map information vis-à-vis the zebra finch linkage map and, if possible, inclusion of at least two SNPs from all scaffolds >25 kb in a preliminary genome assembly version. Five thousand markers on the array were selected to represent potentially fixed differences between the two sister species and were thus generally less informative for intraspecific analyses.

Genotyping was done with an Illumina iScan instrument. Markers that failed to pass the quality filtering for genotype calling were removed from subsequent analysis. Deviation from Hardy-Weinberg Equilibrium (HWE) was tested for in the parental generation using PLINK version 1.07 (Purcell *et al.* 2007). After filtering out SNPs deviating from HWE, Mendelian inheritance was inspected for the remaining markers using GenotypeChecker (Paterson & Law 2011). In total, 38,900 markers were polymorphic in the pedigree, of which 37,443 segregated with a minor allele frequency (MAF) >0.05. Among these there were 845 putative Z-linked markers. The low proportion of loci with rare alleles illustrates the value of selecting markers based on prior information of polymorphism levels, in this case from whole-genome re-sequencing, in the same population.

The inheritance analysis revealed 89 individuals with at least one marker that did not follow Mendelian patterns. Since extra-pair paternity (EPP) is known to occur frequently in the

collared flycatcher (Sheldon & Ellegren 1999), individuals with a high proportion of markers deviating from expected Mendelian segregation likely result from EPP. We therefore removed 46 individuals in which >100 markers showed inconsistent inheritance. The remaining 43 individuals (of the 89 individuals with >1 error) had 1-15 markers with Mendelian inconsistency and were retained, however, the inconsistent markers (181 in total) were removed from the subsequent analysis in all individuals. In the end, we used genotype data from 609 individuals and 37,262 markers for linkage analysis. The average number of informative meioses in the pedigree across all markers was 187.

Linkage analysis

A genetic linkage map of collared flycatcher was constructed using an improved version of CRI-MAP 2.503 (Green *et al.* 1990) developed by Ian Evans and Jill Maddox and implementing the CRI-GEN package provided by Xuelu Liu and Michael Grosz (Monsanto, St. Louis, MO, USA). A detailed account for the different steps in the construction of the map is described in the Supplementary Text. These included calculating pair-wise LOD scores using TWOPOINT and the formation of linkage groups using AUTOGROUP. BUILD was used for making best-order linkage maps.

An updated genome assembly based on high-density genetic linkage data

Markers incorporated in the genetic map were mapped to FicAlb_1.4 with BWA (Li & Durbin 2010). Discrepancies in the form of scaffolds including markers from more than one linkage group were indicative of scaffold chimerism in the assembly and were corrected as described in Supplementary Text. The ends of all new scaffolds were scanned for mate pair reads with their mate on a different scaffold end, representing a means for using mate-pair information that the assembler had failed to automatically integrate in the scaffolding process. This was

done separately for each mate pair library described in Ellegren *et al.* (2012), which had insert sizes of 2.4, 4.1, 5.1, 18 and 21 kb, respectively, where insert size conservatively was let to define the length of what was considered as the scaffold end. For each end, links were sorted and counted and the paired scaffold with most hits was considered for possible adjacency. To infer a physical connection between two scaffolds we then applied a reciprocal criterion requesting that the number of links to the potential neighbour had to be higher than the number of links to the second and third best hit together. We refer to scaffolds connected in this way as super-scaffolds. The new assembly was named FicAlb1.5 and is deposited in GenBank under the accession number AGTO02000000.

Avian karyotypes are notoriously difficult to resolve due to the very large number of minute microchromosomes; most birds have $2n \approx 80$ with the size of about half of the chromosomes <10 Mb. Only chicken karyotype well characterized (Masabanda *et al.* 2004). This, coupled with the observation of a very high degree of synteny conservation among birds (Ellegren 2013), has led to a convention in avian genome sequencing efforts of numbering chromosomes according to homologous chicken chromosomes, even if this does not exactly match decreasing physical size in the focal species (Warren *et al.* 2010). Treatment of fusions or fissions can be illustrated by the nomenclature adopted in the zebra finch genome sequencing project (Warren *et al.* 2010), the second avian genome to be sequenced. For example, chicken chromosome 4 corresponds to two chromosomes in zebra finch, the result of a fusion in the galliform lineage. In zebra finch these chromosomes are referred to as 4 (the larger) and 4A. We have followed this practice as a useful nomenclature for flycatcher chromosomes and this was not least motivated by the observation that flycatcher and zebra finch chromosomes were completely syntenic, without strong evidence for interchromosomal rearrangements.

197

198 **Analyses of chromosomal rearrangements**

199 Ordered and oriented flycatcher scaffolds were concatenated into chromosome sequences
200 with an arbitrary gap size of 5 kb. Repeat masked flycatcher, zebra finch (TaeGut3.2.4) and
201 chicken (WASHUC2) assemblies were aligned with *progressiveMauve* (Darling *et al.* 2010)
202 with default settings, one chromosome at the time. Anchors including all three species were
203 extracted from the backbone file and given as input to GRIMM (Tesler 2002), to be grouped
204 in syntenic blocks with the minimum block size set to 50 kb, unless otherwise stated. MGR
205 (Bourque & Pevzner 2002) was then used for inferring rearrangement events between species,
206 which was essentially only in the form of inversions (see Supplementary Text).

207

208 Flycatcher chromosome sequences were also aligned to zebra finch only with LASTZ.
209 Anchors that overlapped in either of the genomes were filtered, saving the longest one only if
210 the alignment score was more than 1.5 times higher than for the anchors it overlapped with.
211 Regions with several ambiguous overlapping anchors with similar length and alignment score
212 were removed completely. The filtered unique anchors were grouped into syntenic blocks
213 with GRIMM as described above. Unaligned regions between syntenic blocks were
214 considered as breakpoints. To narrow down these regions further we used CASSIS (Baudet *et*
215 *al.* 2010), which attempts to find the precise breakpoint location by a local realignment
216 strategy. In this way most breakpoints decreased in size, however, for a few that CASSIS
217 failed to narrow down we kept the original breakpoint positions. IntersectBed from BEDTools
218 (Quinlan & Hall 2010) was used for extracting overlaps with known repeats and genes. For
219 all rearrangement analyses we only included flycatcher scaffolds that were confidently both
220 oriented and ordered based on direct evidence from linkage or mate-pair data.

221

222 **Recombination rate analysis**

223 We estimated recombination rates in 200 kb windows across the flycatcher genome using the
224 updated assembly version as reference. This was done by calculating recombination fractions
225 between all adjacent markers in the best-order linkage map and assigning window-specific
226 estimates based on the weighted average recombination rate for all marker pairs present
227 within, or flanking, a window. We calculated each window's distance to nearest chromosome
228 end, as well as its gene-density (proportion of exonic sequence), GC-content, repeat content
229 separated into the two classes 'interspersed repeats' and 'microsatellites' (RepeatMasker;
230 Smit, Hubley, and Green; http://repeatmasker.org), and the presence of previously identified
231 (CCNCCNTNNCCNC and CCTCCCT; Myers *et al.* 2010) and *de novo* discovered (see
232 below) sequence motifs associated with high recombination regions. We subsequently omitted
233 all windows spanning a scaffold gap in the genome assembly. This resulted in a set of 4,749
234 windows for which estimates of both recombination rate and the listed genomic parameters
235 were available.
236
237 The variables were transformed to reduce skewness in their distributions; recombination rate
238 was log-transformed to base 10 after adding a constant of 1 to preserve zero rate values,
239 chromosome size was log-transformed to base 10, distance to chromosome end was
240 standardized by chromosome size giving values ranging between 0 and 1, and microsatellite
241 density, repeat density, motif density, gene density and GC content were square-root-
242 transformed. For each parameter we calculated the raw correlation with the recombination
243 rate using the Pearson correlation statistic. We subsequently fitted a multiple linear regression
244 (MLR) model using recombination rate as the response variable to investigate if the variation
245 could be explained by variation in the candidate explanatory variables. As an initial step we
246 investigated the relationship among the candidate explanatory variables by cluster analysis

based on the pair-wise correlations. This revealed that all of the considered genomic features were highly interrelated with each other (Supplementary Figure 2). In particular, microsatellite and motif density both correlated strongly with chromosome size (r = -0.45 and -0.43, respectively). Since small chromosomes only showed limited variation in some of explanatory variables, we focused the analysis on chromosomes larger than 20 Mb.

Correlations between explanatory variables can create biases in regression-like analysis and inference about causal relationships based on MLR analysis thus needs to be made carefully. Beside standard MLR and pair-wise correlation analysis, we thus performed a principal component regression (PCR) analysis using recombination rate as the response variable and the six genomic features as candidate explanatory variables. All regression analyses were performed after Z-transformation of the explanatory variables, which means standardization of the mean value to 0 and of the standard deviation to 1. We also ran MLR and PCR using GC-content as response variable as a proxy for the long-term recombination rate.

In a specific test of the relationship between recombination rate and distance to chromosome end, we performed a locally-weighted polynomial regression (lowess regression) with a smoothing parameter value of 0.5. Based on the lowess regression we classified 'end regions' as regions <5.5 Mb from the chromosome end, and 'centric regions' as regions >5.5 Mb from the chromosome ends (Supplementary Figure 3). The mean recombination rate was then compared between regions.

To investigate sequence context effects on the recombination rate we divided the data into 'hot' and 'cold' recombination regions. The hot regions consisted of the 2.5% (n=112) marker intervals with highest recombination rate and the cold regions consisted of the 2.5% (n=112)

marker intervals with lowest recombination rate (effectively, the latter translates to the subset of marker pairs with the longest physical distance between markers without any evidence for recombination in the pedigree). We used these categories and searched for 6-10 bp sequence motif enrichment in hot regions using the homer2 denovo option in Homer 4.2 (Heinz *et al.* 2010). As suggested by Heinz *et al.* (2010) we applied a stringent significance threshold for enrichment of (10^{-20}), and removed complete redundancies.

Results

A high-density linkage map of the collared flycatcher genome

Linkage analysis first mapped 731 markers to unique positions on 31 linkage groups in a pre-framework map with the stringent threshold of LOD >5. The iterative addition of markers by pair-wise linkage scoring between pre-framework markers and the remaining 36,531 markers subsequently assigned a total of 33,627 markers to 34 different linkage groups, including the three new linkage groups Fal34-Fal36. We then ordered markers within linkage groups and the resulting framework map (marker order supported by LOD >3.0) was composed of 2,456 ordered markers with a total genetic distance of 3,256 cM and a mean genetic distance between adjacent markers of 1.37 cM (± 1.68 SD) (Table 1). We included additional SNPs in this map by step-wise lowering the LOD threshold down to LOD >0.1 (see Supplementary Text), providing a best-order map containing 4,302 markers and spanning 3,256 cM in autosomes and 161 cM in the Z chromosome (Table 1, Figure 1, Supplementary Figure 4). The mean genetic distance between adjacent markers in the best-order map was 0.69 cM (± 1.10 SD). Finally, there were 33,627 markers at this stage that were assigned to one of the linkage groups but not placed on the best-order map. Of these, 31,867 unmapped markers were located in scaffolds containing best-order markers and, therefore, their physical

297 locations could be inferred. When these markers were forced to be included in the map based
298 on their physical position in the respective scaffolds, 76% of markers (24,231) had zero
299 genetic distance with already mapped best-order markers. The total genetic distance of the
300 forced order map was inflated with 13% (3,690 cM), likely at least in part due to small errors
301 in marker order (Table 1). The mean genetic distance between adjacent markers in this forced
302 map was 0.12 cM (± 0.73 SD).
303
304 **A second-generation assembly of the flycatcher genome**
305 The draft assembly of the collared flycatcher genome (version FicAlb_1.4; Ellegren et al.
306 2012) lacks unambiguous information on the order and orientation of scaffolds along most
307 chromosomes. For example, since the draft assembly was based on a coarse linkage map,
308 scaffold ordering had in many cases to be based on indirect information from the assumption
309 of conserved synteny relative to the zebra finch genome. In addition, 55% of the scaffolds
310 remained unanchored to linkage groups/chromosomes. With the aid of the new linkage map
311 we were able to anchor, order and orient scaffolds corresponding to 95.7% (1.013 Gb) out of
312 the final 1.058 Gb assigned to chromosomes (Table 2). We then constructed super-scaffolds
313 by scanning scaffold ends for mate-pair links to all other scaffolds, assigned as well as
314 unassigned, resulting in the incorporation of 43 previously unassigned and mostly small
315 scaffolds (mean size of 68.2 kb, a total of 2.9 Mb) into the assembly. For another 40 scaffolds
316 (mean size 660 kb, 26.4 Mb in total) that had only been indirectly placed in the assembly
317 based on information on the location of homologous sequence in zebra finch, we could
318 confirm ordering and confirm or establish orientation. Finally, and importantly, links were
319 established between adjacent scaffolds for 210 out of the 394 gaps in the assembly (285 out of
320 a total of 437 gaps after the inclusion of the 43 previously unassigned short scaffolds). Since
321 the mate-pair libraries from which these links were established had insert sizes of 2-20 kb,

this indicates the maximum size of the gaps. In no case did we find evidence for links between scaffolds that were not placed immediately adjacent to each other, strongly validating the overall accuracy of the assembly.

The new assembly (FicAlb1.5) has an N50 super-scaffold size of 20.2 Mb (17.4 Mb if including singleton scaffolds) and covers 33 autosomes and the Z chromosome (Table 2, Supplementary Figure 5). The great majority of chromosomes are nearly fully covered by 1-5 super-scaffolds, i.e. not far from continuous chromosome sequences; four chromosomes do in fact correspond to a single super-scaffold and one to a single scaffold (Table 2). The assembly includes sequence data for four microchromosomes that are not represented by defined chromosomes in the chicken assembly (which has sequence data from chromosomes 1-28 and 32). One of these, including 2.1 Mb of flycatcher sequence and covering a genetic distance of 53 cM, has sequence homology to chicken linkage group LGE22. The other three (Fal34 with 16 cM, Fal35 with 37 cM and Fal36 with 10 cM) show no sequence homology to assembled sequence from the chicken genome. The linkage map did not have the same high degree of resolution for the Z chromosome as for autosomes due to the fact that only male meioses were informative. This led to a higher proportion of scaffolds that were not ordered and/or oriented and since such scaffolds were not included in our final assembly, this likely explains why the flycatcher Z chromosome assembly was shorter (59.7 Mb) than that of chicken (74.6 Mb) and zebra finch (72.9 Mb).

In the end, after addition of new scaffolds and scaffold orientation by linkage data and mate-pair linking, the ordered and oriented sequences constitute 98.6% (1.042 Gb) of the scaffolds assigned to chromosomes. Of the total assembly also including unassigned scaffolds (1.116 Gb), 93.4% of the sequence was anchored, ordered and oriented along chromosomes. This

represents a considerable improvement compared to the previous assembly (596 Mb or 56.5% anchored; Table 3).

Highly conserved DNA content of avian chromosomes

The high degree of genome coverage coupled with the unusual continuity in scaffolded sequence along each chromosome give unprecedented power and resolution to study the rate and pattern of chromosomal rearrangements during avian evolution. We made whole-genome alignments of flycatcher, chicken and zebra finch, inferred syntenic blocks >50 kb in size and identified chromosomal rearrangements. Despite the two lineages split ≈40 million years (My) ago (Nabholz *et al.* 2011), flycatcher and zebra finch chromosomes are entirely syntenic without clear-cut evidence of interchromosomal rearrangements (see Supplementary Figure 6 for a comment on the tentative chromosome 1B in zebra finch), witnessing on the rather extreme karyotypic stability of birds. The flycatcher assembly confirms the only two clear cases of interchromosomal rearrangement distinguishing the chicken and zebra finch karyotypes. First, flycatcher and zebra finch have two chromosomes, chromosomes 1 (≈120 Mb) and 1A (≈75 Mb), which correspond to the single chromosome 1 of chicken (201 Mb), a result of a fission in the passeriform lineage. Second, chicken chromosome 4 (94 Mb) corresponds in both flycatcher and zebra finch to two chromosomes, chromosomes 4 (≈70 Mb) and 4A (≈21 Mb), resulting from a fusion in the galliform lineage.

The assembly sizes of individual chromosomes were remarkably similar among chicken, zebra finch and flycatcher (Table 2), and did not differ by more than 2.5 Mb in size for 26 out 29 autosomes. As a consequence, the total amount of sequence assigned to chromosomes was nearly identical in the three bird species (1.02-1.04 Gb), again testifying on an overall

371 evolutionary stasis of avian chromosomes. We note that, exactly like for chicken and zebra
372 finch, flycatcher chromosome 16 was difficult to sequence and assemble.
373
374 **Frequent intrachromosomal rearrangements during avian evolution**
375 The evolutionary stability in the size and content of avian chromosomes stands in sharp
376 contrast to frequent changes in the genomic organisation within chromosomes
377 (intrachromosomal rearrangements). We found a total of 343 inversions, which can explain
378 the current organization of chromosomal segments in chicken, zebra finch and flycatcher
379 (Table 4). As expected, most of these (203) can be traced back to the long lineage connecting
380 chicken and the common ancestor of flycatcher and zebra finch in an unrooted tree. The
381 number of rearrangements in the flycatcher and zebra finch lineages was 61 and 79,
382 respectively, and can be readily seen in circular visualization of sequence homologies
383 between the two species (Figure 2). Based on these numbers, we estimate the rate of inversion
384 at 1.5 (flycatcher lineage), 2.0 (zebra finch) and 1.7 (chicken-passeriform ancestor) events per
385 My. This corresponds to rates of 0.0014-0.0019 per My per Mb. We note that all three genome
386 assemblies used for making this inference are based on genetic linkage data.
387
388 The size of inversions was biased towards the lower end of detectable events (Supplementary
389 Figure 7), with median size of 3.34 Mb (chicken), 2.62 Mb (zebra finch), and 0.78 Mb
390 (flycatcher). With the propensity for inversions to be short, it was clear that many events
391 would have been missed with lower assembly continuity and at higher thresholds for
392 minimum size of syntenic blocks. This was confirmed when we increased block size to 100
393 kb, 250 kb or 1 Mb to make inference about the number of rearrangements (Supplementary
394 Table 1); the total number of rearrangements decreased from 343 at the resolution of 50 kb to
395 87 at the resolution of 1 Mb. Accordingly, the estimated rates of inversion decreased from 1.5-

2.0 to 0.2-0.5 per My. This highlights the importance of the level of resolution for characterization of chromosome rearrangements.

Another way of quantifying the inversion rate is to also take the amount of inverted sequence into account. The total length of all inversions was 476.1 Mb (49.1% of the aligned sequence), 299.2 Mb (30.2%), and 265.4 Mb (26.3%) in the chicken-passeriform ancestor, zebra finch and flycatcher lineage, respectively. This gives inversion rates of 4.0, 7.5, and 6.6 Mb inverted DNA per My, respectively. Note that these numbers are based on the amount of unique sequence involved in rearrangements; the sequence of nested inversions was only considered once. Also note that the rate estimates cannot be expected to increase linearly with time since, with a constant rate of rearrangement, the amount of sequence not yet inverted will decrease over time. This may explain the lower rate estimate for the long chicken-passeriform ancestor branch.

We next examined chromosomal breakpoints and sought to elucidate their characteristics. Here we used pair-wise alignments between flycatcher and zebra finch to get higher resolution (due to the shorter evolutionary distance). One hundred sixty-five breakpoint regions were identified, with a median size of 2.4 kb for measurable regions (see below; Supplementary Figure 8). Of these, 28 regions were re-used twice (17%). There was a very strong association between the location of scaffold junctions in the flycatcher assembly and chromosomal breakpoints. If the 165 breakpoints would have been randomly distributed in the genome, we should have expected to find <1 to coincide with the location of scaffold junctions. However, we observed 42 scaffold junctions inside breakpoints, clearly showing that some regions of the genome are both resistant to sequence assembly and prone to chromosomal mutation.

The distribution of breakpoints across the genome was non-random with clusters of multiple inversion events interspersed with large chromosomal regions of structural stasis (Figure 3). There was a propensity for breakpoints to be located toward the ends of chromosomes, with a significant deviation from a uniform distribution along chromosomes (Supplementary Figure 9; Goodness-of-fit test, chi-square = 22.46, d.f. = 9, $p < 0.05$). Moreover, there was a negative correlation between chromosome size and the rate of inversion per Mb (Wilcoxon's test, z=6.06, $p < 0.001$). Furthermore, several genomic parameters differed significantly between breakpoint regions and the rest of genome, including recombination rate (mean 5.83 vs. 3.25 cM/Mb, z = 5.74, $p = 4.8 \times 10^{-9}$), GC content (mean 0.513 vs. 0.416, z = 11.28, $p = 1.6 \times 10^{-29}$), and repeat density (mean 0.221 vs. 0.096, z = 4.79, $p = 8.4 \times 10^{-7}$).

Recombination rate variation

With a high-density linkage map and a genome assembly with a high degree of sequence continuity along chromosomes it is possible to obtain detailed recombination rate estimates across the flycatcher genome. We divided the genome into 200 kb windows and observed a mean sex-averaged recombination rate of 3.1±4.1 cM/Mb across windows. The genomic landscape of recombination was highly heterogeneous, with two major, large-scale trends of recombination rate variation. First, the mean recombination rate per chromosome was considerably higher for small chromosomes than for large chromosomes (Table 5, Figure 4). The rate was in excess of 10 cM/Mb for chromosomes <10 Mb; for the new linkage group Fal35, with only 230 kb of assembled sequence (and a genetic distance of 36.8 cM), recombination reached an extreme estimated rate of 160 cM/Mb. For the three chromosomes >100 Mb, the rate was uniformly ≈2.0 cM/Mb while for chromosome size classes in the range

445 of 10-100 Mb, recombination rate was intermediate and increased with decreasing
446 chromosome size.
447
448 It is interesting to note that the effect of chromosome size on rate of recombination gradually
449 diminished with increasing chromosome size. In fact, if subtracting 50 cM from the length of
450 each linkage group (reflecting one obligate inter-chromatid crossing-over per chromosome,
451 see Discussion) before calculating the chromosome-average recombination rate as map length
452 divided by physical size, the rate of recombination seemed largely independent of
453 chromosome size (Table 5) and increased with 1.5-2.0 cM for every Mb of increased physical
454 size. In addition, it could also be noted from Figure 4 that the mean recombination rate of the
455 Z chromosome (2.7 cM/Mb over 60 Mb) as measured in male meiosis was very similar to that
456 of similarly sized autosomes (chromosomes 1A, 4 and 5, sized 65-75 Mb, have a mean
457 recombination rate of 2.7 cM/Mb).
458
459 Second, there was a significant increase in recombination rate towards chromosome ends, a
460 pattern consistent irrespective of chromosome size (Figure 5). For instance, the average
461 recombination rate in the ends, defined as the distal 5.5 Mb of each chromosome end (see
462 Material and Methods and Supplementary Figure 3 for motivation) of chromosomes 1-6 was
463 5.7 cM/Mb while the internal regions of these chromosomes had a mean rate of 2.3 cM/Mb.
464 Similarly, the average recombination rate at chromosome ends and interior regions of smaller
465 chromosomes 7-28 were 7.0 and 2.4 cM/Mb, respectively.
466
467 The total map length was on average 10% longer in males than females (3,300 cM and 2,997
468 cM in the best-order autosomal map, respectively; Wilcoxon's test for matched pair of
469 windows, V = 104, p = 0.002). There was limited regional variation in sex-specific

recombination rates (Supplementary Figure 10); however, chromosome 17 and 27 made exceptions by showing marked differences between the sexes in 2-3 Mb regions (Figure 6; Table 1). To test if these differences were repeatable we divided the pedigree into six subsets of individuals (n=100 each) and estimated sex-specific recombination rates in each subset. In chromosome 27, all six subsets showed larger total genetic distance in males (total genetic distance = 34.2 ~ 76.4 cM in females and 81.0 ~ 129.0 cM in males), and in chromosome 17, five out of six subsets showed larger total genetic distance in males (total genetic distance = 59.1 ~ 85.5 cM in females and 76.4 ~ 89.9 cM in males). This suggests that there is a true signal of sex differences in recombination rate in these chromosomes.

In order to search for sequence motifs potentially associated with high recombination rates we partitioned the rate between all marker pairs into two extreme classes, representing the regions with the 2.5% highest ('hot regions') and 2.5% lowest ('cold regions') rates. The previously described (Myers *et al.* 2008; Winckler *et al.* 2005) sequence motifs CCNCCNTNNCCNC and CCTCCCT associated with high recombination were both present at higher density (2.1 and 1.6 times, respectively) in the hot regions than in the cold regions, although this was not statistically significant. We also searched for enrichment of previously unidentified sequence motifs in hot regions and, after correcting for redundancy and multiple testing, we found evidence for enrichment of six different sequence motifs 6-9 bp long: GATGAGATG, AATCAATC, GAAGGAGA, CCATATC, GGATCC and TCGAGG; Supplementary Table 2).

Several genomic parameters have previously been shown to be associated with recombination rate variation in other organisms (Coop & Przeworski 2007; Webster & Hurst 2012; Cutter & Payseur 2013). Focusing on chromosomes >20 Mb, we found significant pair-wise

correlations between recombination rate and chromosome size, distance to chromosome end (as shown for all chromosomes, described above), microsatellite density, sequence motif density and gene density (Table 6). To disentangle the relative effect of each of these parameters we performed multiple linear regression (MLR) analysis and principal component regression (PCR) analysis using recombination rate as response variable. This showed that microsatellite density, motif density and distance to chromosome end explained most of the variation in recombination rate, while the impact of chromosome size, gene density and interspersed repeat density was of minor importance (Table 6). The relative limited effect of chromosome size was probably related to the fact that we only analysed chromosomes >20Mb (see Discussion). The PCR further allowed us to disentangle two independent effects (PC I and PC II), which contributed separately to the variation in recombination rate (Table 6, Figure 7). Distance to chromosome end clustered together with microsatellite and motif density in PC I. Chromosome size built the main contribution to PC II, which points towards an independent effect of chromosome size on the recombination rate.

Recombination rate conservation

A broad-scale overview of the recombination landscape in flycatcher compared to zebra finch and chicken is given in Figure 8, which depicts the relationship between physical position and cumulative genetic map length for each chromosome. Clearly, the flycatcher landscape is more similar to that in chicken than to that in zebra finch. Although all three species show an increased recombination rate towards chromosome ends, this trend is much more pronounced in the zebra finch than in the other two species. This difference is reinforced by the very low rate of recombination in the interior regions of zebra finch chromosomes.

519 As recombination impacts the patterns of local base composition in avian genomes via GC-
520 biased gene conversion (Mugal *et al.* 2013), GC content might be a good indicator of long-
521 term global recombination rate variation. A correlation between GC content and current
522 recombination rate might thus be indicative of long-term conservation in recombination rate
523 variation across the genome. This was indeed observed ($r = 0.47$, $p < 10^{-15}$; Supplementary
524 Figure 11). In light of this we repeated the calculations of pair-wise correlation, MLR and
525 PCR using GC content as the response variable. MLR showed that motif density and
526 microsatellite density explained most of the variation in GC content, followed by distance to
527 chromosome end and chromosome size (Table 6). PCR showed that variation in GC content
528 was explained by two major principal components, PC I was composed primarily of motif
529 density, microsatellite density, and distance to chromosome end while PC II was composed of
530 chromosome size, repeat density and gene density (Table 6). This trend was thus consistent
531 with the variation in pedigree-based recombination rate estimates. In fact, when
532 recombination rate and GC content were used as a combined response variable, more than a
533 half of the variation was explained by these variables.
534

535 ## Discussion

536

537 We have capitalized on the power of contemporary DNA sequencing technology to develop a
538 high-resolution genetic map of the collared flycatcher genome. This allowed the construction
539 of an improved genome assembly and downstream analyses of recombination rate variation
540 and chromosomal evolution at high resolution. High-throughput sequencing was critical in the
541 process of map construction for at least two reasons. First, markers for the map were well
542 distributed across the genome. This owes to the fact that we had a draft assembly of the
543 flycatcher genome, constructed using high-throughput sequencing but without physical

mapping tools such as BAC or fosmid clones, from which suitably distributed markers could be selected. Second, the availability of polymorphism data from whole-genome re-sequencing of population samples meant that we could select highly variable markers. We also capitalized on new technology for the development of a genetic map with unusually high marker density in a non-model organism by performing array-based SNP genotyping using a custom 50K SNP array, purposely developed for this endeavour (Kawakami *et al.* 2014).

By integrating high-density linkage map data with scaffold sequences from the draft genome assembly we obtained a significantly improved assembly of the collared flycatcher genome. The assembly has 98.5% of the anchored sequence ordered and oriented along chromosomes and a super-scaffold N50 size of 20.2 Mb. It covers 33 autosomes and the Z chromosome, which compares well with the two Sanger-sequenced avian genomes (chicken: 28 autosomes with >0.1Mb of assembled sequence (ICGSC 2004); zebra finch: 31 autosomes with >0.1Mb of assembled sequence (Warren *et al.* 2010)).

The karyotype of collared flycatcher has not been characterized. For 25 other bird species of the order Passeriformes, chromosome number is in the range of 2n = 72-84, with 19 species showing 2n = 78-80 (which is also the most common number across the whole class of Aves; Gregory 2011). This could suggest that there are at most 5-6 minute chromosomes for which we still have not anchored scaffolds to linkage groups. Compared to a random process of marker selection for genotyping, an informed strategy of using SNPs from the vast majority of all scaffolds was of obvious benefit for linkage-based scaffold ordering and orientation. Together, this illustrates that it is feasible to obtain a *de novo* assembly of a vertebrate genome with nearly continuous chromosome sequences, without additional genomic resources or

568 molecular tools. The latter represents the default situation for essentially all non-model
569 organisms.
570
571 **A 'core' avian genome**
572 It is known that birds show less variation in genome size than other amniote lineages (Griffin
573 *et al.* 2007; Ellegren 2010). However, it is remarkable that with flycatcher now added to the
574 avian genomes so far sequenced at high sequence continuity all have assemblies of ≈1.10 Gb,
575 with 1.02-1.04 Gb assigned to chromosomes (ICGSC 2004; Dalloul *et al.* 2010; Warren *et al.*
576 2010; Huang *et al.* 2013; Shapiro *et al.* 2013). Moreover, the amount of sequence assigned to
577 syntenic chromosomes showed very limited variation among species (Table 2). This indicates
578 that the overall DNA content of birds is highly conserved across divergent lineages, although
579 there may be occasional genome size expansions from increased transposon activity in certain
580 lineages (Dyke & Kaiser 2011). Cytometric estimates of total DNA content of birds vary
581 more, between 1-2 pg and with the majority in the range 1.2-1.5 pg (1 pg ≈ 0.98 Gb; Gregory
582 2011). However, these estimates have been obtained by several different methods and are
583 sensitive to calibration, experimental error and gender. For chicken, recent estimates tend to
584 converge at 1.20-1.25 pg (Mendonca *et al.* 2010). It thus remains to be seen from other
585 species how much genome size actually vary across birds; it may very well be that the
586 variation is even more limited than previously indicated by cytometry.
587
588 **The rate of chromosomal evolution in birds**
589 It is clear that the avian karyotype has remained largely stable during the evolution of modern
590 birds (Griffin *et al.* 2007; Ellegren 2010), which is in sharp contrast to frequent
591 interchromosomal rearrangements occurring during, for example, mammalian evolution
592 (Murphy *et al.* 2005). However, it is less clear whether the rate of intrachromosomal

rearrangements also varies among vertebrate lineages and if avian chromosomes are slowly evolving also in this respect. Quantitative analyses have largely been lacking and comparisons among taxa are sensitive to methodology and resolution. Our data demonstrate that the rate of inversion in the sampled avian lineages (1.5-2.0 inversion per My) is similar to many mammalian lineages analysed with the same algorithms and resolution (Supplementary Table 3). In fact, if one takes into account that the DNA content of avian genomes is generally <50% of that of mammalian genomes, the rate of inversions per Mb is higher in the sampled avian lineages than in many mammals, like primates (Zhao & Bourque 2009). Thus, a stable avian karyotype does not translate into an overall stability of the organisation within bird chromosomes.

Variation in the rate of inversion from 1.5 events per My (flycatcher lineage) to 2.0 events per My (zebra finch) gives some indication that there is rate variation among avian lineages for intrachromosomal rearrangements, just as there is substitution rate variation; for example, for the two avian orders in focus here, the substitution rate in Passeriformes is higher than in Galliformes (Nam *et al.* 2010; Nabholz *et al.* 2011). There are rare examples of avian species/families with unusually small (2n=40-50) or large (2n=130-140) number of chromosomes (Gregory 2011) and it will be interesting to see if the dynamic karyotype evolution (with fusions and fissions) in these lineages are associated with a high rate of intrachromosomal rearrangements.

Birds have less repetitive DNA than other amniotes, with a repeat content of the avian genomes so far sequenced of ≈10% (ICGSC 2004; Dalloul *et al.* 2010; Warren *et al.* 2010; Ellegren *et al.* 2012). It has been tempting to associate the karyotypic stability of birds with the low repeat content under the scenario that fewer (transposable) repeats provides less

618 opportunity for nonallelic homologous recombination (Burt *et al.* 1999). However, why then
619 would the rate of inversion be at least as high in birds as in repeat-rich mammalian genomes?
620 One explanation could be that the role of repeats in mediating chromosomal mutations differs
621 between inversions and interchromosomal rearrangements, such as translocations or
622 fusions/fissions. However, there is strong evidence for the involvement of transposable
623 elements in generating inversions, consistent with our observation of increased repeat density
624 in avian intrachromosomal breakpoints (Kidd *et al.* 2008; Lee *et al.* 2008; Zhao & Bourque
625 2009). Moreover, repetitive sequences such as gene duplicates, gene clusters or other forms of
626 segmental duplications (Armengol *et al.* 2003; Bailey *et al.* 2004; Zhao & Bourque 2009) are
627 frequently found at sites of breakpoints, including in birds (Dalloul *et al.* 2010; Völker *et al.*
628 2010). This would suggests that repeat density is in fact unrelated to karyotypic stability and
629 that the conserved chromosome structure so characteristic for birds owes to other factors.
630 Perhaps the mechanisms of chromosome replication, recombination or segregation at avian
631 meiosis are less prone to interchromosomal rearrangements in the first place. Alternatively,
632 the negative fitness effects of such mutations could be more severe than in other vertebrate
633 lineages, meaning that they are to a larger extent removed by selection in birds.
634
635 It is interesting to note that despite a stable karyotype, there has been a dynamic process of
636 sequences changing their relative position within chromosomes during avian evolution. For
637 example, since the split of flycatcher and zebra finch lineages 40 My ago, 25-30% of all
638 sequence has been repositioned by inversions. This provides an unusual opportunity to
639 compare molecular evolutionary parameters between sequences that have remained in the
640 same chromosomal position for a long time and sequences that have become integrated into
641 another context of the genomic landscape, yet remaining on the same chromosome. This

includes aspects such as the evolution of base composition, substitution rates and recombination rates.

Characteristics of chromosomal breakpoints

Previous work has suggested that chromosomal breakpoints are re-used during evolution, representing hot-spot regions for chromosome instability (Pevzner & Tesler 2003; Larkin *et al.* 2009; Skinner & Griffin 2012). However, the case for evolutionary re-use of breakpoints is an issue of discussion (Sankoff & Trinh 2005; Peng *et al.* 2006; Alekseyev & Pevzner 2007) and may in the end be a matter of resolution (Becker & Lenhard 2007; Larkin *et al.* 2009; Attie *et al.* 2011). Still, our results demonstrate a concentration of breakpoints to certain regions of the avian genome with 17% of breakpoint regions being re-used. Observations of independently occurring rearrangements at approximately the same chromosomal position in different avian lineages have generally been made with much lower resolution than applied herein (Griffin *et al.* 2007; Kemkemer *et al.* 2009; Dalloul *et al.* 2010; Völker *et al.* 2010; Skinner & Griffin 2012), and cannot unambiguously distinguish between re-use of sites or regions. Our data point at the former since the observed breakpoints were generally small (median size 2.4 kb).

Molecular evolutionary analyses often reveal that several genomic parameters are interrelated. Similarly, we found several parameters to correlate with the location of chromosomal breakpoints in the flycatcher-zebra finch comparison, including recombination rate, distance to chromosome end, chromosome size, repeat density and GC content. Although it is difficult to dissect the causal relationships between these correlations, we note that recombination events as well as chromosome rearrangements are initiated by the formation of double-strand breaks (DBS; Baudat *et al.* 2013). An association between recombination rate and

667 chromosomal breakpoints has been independently demonstrated in a comparison of the
668 chicken and turkey genomes (Völker *et al.* 2010).
669
670 **The role of chromosome rearrangements during avian evolution**
671 Chromosomal speciation models posit that rearrangements distinguishing diverging
672 populations will promote speciation via underdominance (due to fitness reduction of
673 unbalanced gametes in heterozygotes; White 1973) or by reducing interspecific recombination
674 in the rearranged regions hindering gene flow and facilitating the build up of genetic
675 incompatibilities (Noor *et al.* 2001; Navarro & Barton 2003; Coyne & Orr 2004; Kirkpatrick
676 & Barton 2006). Empirical evidence that suppressed recombination in regions of inversions
677 are associated with speciation is accumulating in both animals and plants (Hoffmann &
678 Rieseberg 2008; Nachman & Payseur 2012). However, there are so far only few, if any, well-
679 documented examples of inversions contributing to speciation in birds. On the other hand,
680 inversion polymorphisms associated with distinct phenotypic differences have been detected,
681 like a nearly 100 Mb inversion in one of the macrochromosomes of the white-throated
682 sparrow (*Zonotrichia albicollis*) associated with a suite of traits including behavioural
683 phenotypes (Thorneycroft 1966; Thomas *et al.* 2008).
684
685 Diversification rates are likely to differ over time and estimating speciation rates is
686 notoriously difficult, even for the most recent divergences. Speciation durations in both
687 mammals and birds may entail at least two million years on average (Avise *et al.* 1998; Coyne
688 & Orr 2004) and the loss of hybrid fertility in birds may be of the order of millions of years
689 (Price & Bouvier 2002; Fitzpatrick 2004). We thus note that the observed rate of
690 intrachromosomal rearrangements (1.5-2.0 per My) has been sufficiently high for inversions
691 to potentially play a significant role in the build up of reproductive incompatibility in birds.

692 We encourage further research on the genetics of speciation in birds that specifically seek to
693 address this question.
694
695 **Recombination rates in the flycatcher genome**
696 The considerable variation in chromosome size in avian genomes is associated with
697 systematic variation in chromosome-specific recombination rates: recombination rate shows a
698 clear increase with decreasing chromosome size (ICGSC 2004; Stapley *et al.* 2008;
699 Backström *et al.* 2010a). There is evidence from several organisms of one obligate crossing-
700 over per chromosome, often thought to be necessary for proper segregation of chromosomes
701 at meiosis (Fledel-Alon *et al.* 2009; Wang *et al.* 2012). The observation that the intercept of a
702 linear correlation between flycatcher chromosome size and genetic length was at ≈50 cM
703 (Figure 4a) shows that a genetic distance of 50 cM applies regardless of the size of
704 chromosomes and leads to very high rates of recombination per physical unit of DNA. This is
705 entirely consistent with our observations: both MLR and PCR showed that chromosome size
706 had a strong impact on the rate of recombination. When 50 cM was subtracted from the
707 genetic length of each chromosome (reflecting the genetic length accrued by one crossing-
708 over), we found that recombination rate was nearly constant across chromosomes and thus
709 independent of chromosome size. This would suggest that the number of additional
710 recombination events per chromosome solely reflects variation in chromosome size and need
711 not be related to inherent differences among chromosomes in the rate of recombination per
712 physical unit of DNA.
713
714 Many species across different groups of organisms are heterochiasmic; that is, they show
715 genome-wide differences in the sex-specific rates of recombination (Burt *et al.* 1991;
716 Lenormand 2003; Lenormand & Dutheil 2005). The observation of on average 10% higher

717 recombination in flycatcher males than in females is in line with the idea that suppressed sex
718 chromosome recombination in the heterogametic sex somehow 'spill over' on autosomes, to
719 reduce the genome-wide rate of recombination in that sex (Burt *et al.* 1991). However, there
720 are exceptions to this and there are also several alternative explanations to why sex
721 differences in recombination evolve (Otto & Lenormand 2002; Lenormand 2003; Hansson *et*
722 *al.* 2005). One interesting possibility is that epistatic interactions between loci can favour the
723 spread of sexually antagonistic alleles when recombination differs between males and females
724 (Mank 2009; Connallon & Clark 2010; Wyman & Wyman 2013). Under this scenario one
725 could potentially expect localized regions with pronounced sex-differences in the rate of
726 recombination, such as on flycatcher chromosomes 17 and 27 in collared flycatcher and as
727 chromosomes 9 and 19 in humans (Kong *et al.* 2010), as candidate regions under sexually
728 antagonistic selection.
729
730 The recombination rate of the Z chromosome (2.7 cM/Mb) was essentially identical to that of
731 similarly sized autosomes. Birds have female heterogamety (males ZZ, females ZW) so the Z
732 chromosome does not recombine in females, with exception of the pseudoautosomal region,
733 and the estimated rate comes from male meiosis only. The effective recombination rate of the
734 Z chromosome is thus 2/3 x 2.7 = 1.8 cM/Mb (not ½ x 2.7 since two of the three potentially
735 transmitted Z chromosome per breeding pair will recombine, Lohmueller *et al.* 2010).
736 Moreover, this is independent of any difference in the effective population size of males and
737 females, and of the female-to-male breeding ratio. Sex chromosomes are often considered as
738 hot-spots for speciation, i.e. the large-X effect (Coyne & Orr 2004) or Coyne's rule (Turelli &
739 Moyle 2007). One of several explanations for this is that the rate of recombination of the X/Z
740 chromosome is lower than of autosomes due to reduced recombination in the heterogametic
741 sex, thereby facilitating the maintenance of combinations of diverged gene variants (see

742 Qvarnstrom & Bailey 2008). However, our quantitative analysis shows that the effective
743 recombination rate of the Z chromosome is not much different from the sex-averaged rate of
744 the three largest chromosomes (2.0 cM/Mb), which encompass more than 35% of the
745 flycatcher genome. If generally applicable, this would suggest that the large-X effect mainly
746 attributes to other factors, such as dominance (Coyne & Orr 2004).
747
748 The unusual heterogeneity in the rate of recombination in avian genomes, in particular the
749 high rate of recombination in microchromosomes, will impact on several aspects of molecular
750 ecological and molecular evolutionary analyses. For example, higher marker densities will be
751 required for detection of linkage in regions with high recombination rate in QTL mapping and
752 genome-wide association studies. However, when this is done, causative loci are likely to be
753 in closer physical vicinity in those high recombination rate regions than in low recombination
754 rate regions. Another aspect is that a heterogeneous recombination landscape can provide
755 increased power in detecting correlations between the rate of recombination and genomic
756 parameters potentially associated with recombination. For example, recombination rate is
757 expected to correlate with both nucleotide diversity and the rate of protein evolution, in the
758 latter case related to the efficacy of selection (Webster & Hurst 2012). Much focus is
759 currently put on the question if Hill-Robertson interference - the counteracting effect on
760 genetic variation at linked sites by selection - is mainly caused by selective sweeps for
761 advantageous alleles or background selection against slightly deleterious mutations (Campos
762 *et al.* 2014). Avian genomic data may be useful in resolving this issue, by comparing
763 sequence evolution in regions with markedly different recombination rates.
764
765 **Conservation of rates and patterns of recombination**

766 Comparative studies provide evidence for a phylogenetic signal in recombination rate
767 variation among species (i.e., conservation of recombination rates; Dumont & Payseur 2008;
768 Dumont & Payseur 2011; Smukowski & Noor 2011; Segura *et al.* 2013). In line with this, we
769 previously found that the rate of recombination in orthologous regions of the chicken and
770 zebra finch genomes was correlated (Backström *et al.* 2010a). The strong correlation between
771 GC content and flycatcher recombination rate seen in the present study is consistent with
772 long-term conservation of the recombination landscape in birds, with GC-biased gene
773 conversion driving GC content in regions of high recombination. However, there is also
774 evidence from other studies that the total amount of recombination can vary among related
775 species, or even subspecies (Dumont *et al.* 2011). For example, the length of the human
776 genetic map is more than two times longer than that of mouse and rat (Jacob *et al.* 1995;
777 Dietrich *et al.* 1996; Cox *et al.* 2009), although genome size is only 10% larger in humans
778 than in rodents. With one obligate crossing-over per chromosome or chromosome arm (Pardo-
779 Manuel de Villena & Sapienza 2001), variation in number of chromosomes or number of
780 chromosome arms can explain at least part of the variation in total amount of recombination
781 among species, as is the case in the comparison of primates and rodents. We found that the
782 total amount of recombination in flycatcher was 200% of that of zebra finch despite both
783 species belonging to the same order of birds and their karyotypes probably being very similar.
784 In contrast, the amount of recombination in flycatcher was similar to that in the more distantly
785 related chicken (Groenen *et al.* 2009); chicken and flycatcher lineages diverged about 80 My
786 ago (Nabholz *et al.* 2011). Our data thus point both at long-term conservation in the amount
787 of avian recombination and that there can be relatively short-term changes.
788
789 Domestication may select for increased recombination by favouring the generation of new
790 haplotypes and new gene combinations in the face of drastically changed selection pressures

(Burt & Bell 1987; Ross-Ibarra 2004). It has been hypothesized that this could explain the higher total amount of recombination in chicken (and turkey, which appears similar to chicken with respect to recombination, Aslam *et al.* 2010) than in zebra finch (Backström *et al.* 2010a; van Oers *et al.* 2014). However, with similar recombination rates in chicken and flycatcher, our data do not support this hypothesis. If anything, the fact that zebra finch linkage map data comes from birds held in captivity for many generations does not support increased recombination as a response to artificial selection. It has also been suggested that passerine birds would have lower recombination rates than galliforms (van Oers *et al.* 2014), a view tentatively supported by low-density linkage map data from some species (Åkesson *et al.* 2007; Hansson *et al.* 2009; Jaari *et al.* 2009; van Oers *et al.* 2014). This might be true although it was not supported by our data as both flycatcher and zebra finch belong to the order Passeriformes. In general, we caution against taking interpretations from low-density linkage maps of species without an assembled genome sequence too far. With increased recombination rate towards chromosome ends, which might not necessarily be covered in linkage maps based on random markers, and by an additional 50 cM added to the total map length for every inclusion of another microchromosome, low-density linkage maps may grossly underestimate the total amount of recombination.

Although homologous chromosomes of flycatcher and chicken are differently organised due to inversions, the broad-scale recombination landscape in these two distantly related birds was similar (Figure 8). This stands in sharp contrast to the recombination landscape in zebra finch. Large zebra finch chromosomes are characterized by the presence of extensive recombination desserts spanning the most of the interior parts of these chromosomes, not seen in flycatcher (Stapley *et al.* 2008; Backström *et al.* 2010a). As much as 80% of the total amount of recombination is concentrated on the 20% distal parts of several large chromosomes.

816 Although there is an increase in recombination rate towards the ends of chromosomes in
817 flycatcher and chicken, this effect is far from as dramatic as in zebra finch (Supplementary
818 Table 4). As far as we aware of, it is not known what factors may affect differences in the
819 distribution of crossing-over events along vertebrate chromosomes.
820

821 ## Perspectives and conclusions

822

823 Developments in the use of genetic approaches for addressing ecological and evolutionary
824 questions in *Ficedula* flycatchers well illustrate the overall developments in the field of
825 molecular ecology and provides a timeline for its progress. For the *Ficedula* system, this
826 began about 25 years ago with the analysis of allozymes and restriction fragment length
827 polymorphisms of pied flycatcher mtDNA (Gelter *et al.* 1989; Tegelstrom & Gelter 1990),
828 and was soon followed by the introduction of microsatellite (Ellegren 1991, 1992) and DNA
829 fingerprinting markers (Gelter & Tegelstrom 1992; Ratti *et al.* 1995). Questions at this time
830 were mainly related to behavioural ecology, like the fitness return of extra-pair paternity, but
831 also focused on speciation and hybridization. It also included tests of sex allocation theory,
832 using PCR-based approaches for molecular sexing (Ellegren *et al.* 1996; Sheldon & Ellegren
833 1996). DNA sequencing of flycatcher mtDNA came into use around year 2000 (Saetre *et al.*
834 2001) and provided phylogenetic perspectives and increased resolution for the detection of
835 hybridization between flycatcher species. This was subsequently augmented with the use of
836 nuclear single nucleotide polymorphisms, SNPs (Primmer *et al.* 2002) and opened a venue for
837 studying gene flow, introgression and population structure in further detail (Saetre *et al.* 2003;
838 Borge *et al.* 2005; Lehtonen *et al.* 2009). In the mid of the last decade, genetic mapping
839 efforts of flycatchers began and introduced a genomic perspective, although the genome itself
840 was considered only far away in the horizon (Backström *et al.* 2006; Backström *et al.* 2008;

841 Backström *et al.* 2010b). However, the tremendous power offered by next-generation
842 sequencing technology meant that a draft sequence of the collared flycatcher genome could be
843 presented in 2012 (Ellegren *et al.* 2012), providing novel insights into genomic divergence
844 during lineage splitting. Moreover, this provided a platform for genome-wide studies of gene
845 expression (Uebbing *et al.* 2013) and flycatcher population history (Nadachowska-Brzyska *et*
846 *al.* 2013). Furthermore, recent proteomic analysis suggests that functional genomic studies are
847 on their way (Leskinen *et al.* 2012). All in all, this makes *Ficedula* flycatcher a prime model
848 organism in molecular ecology.
849
850 This study provides a genome assembly with nearly continuous chromosome sequences and a
851 detailed genetic map of the flycatcher genome. Together, this information allowed us to
852 conclude that the highly conserved nature of the avian karyotype stands in sharp contrast to
853 the observation of frequent intrachromosomal rearrangements during avian evolution. The
854 rate of these rearrangements is high enough for inversions to potentially have been involved
855 with many events of speciation. We envision that this idea could be tested by mapping
856 inversion events onto a phylogeny of closely related species of birds. Moreover, we found
857 significant variation in the rate of recombination across the genome and concluded that the
858 large effect of chromosome size could mainly be explained by obligate recombination events
859 per chromosome. Surprisingly the overall level and patterns of flycatcher recombination was
860 more similar to chicken than to the more closely related zebra finch.
861
862 We recently showed that the genomic landscape of divergence between pied and collared
863 flycatchers is characterized by the presence on numerous 'differentiation islands', with
864 markedly higher F_{ST} than in the genomic background (Ellegren *et al.* 2012). This observation
865 was based on data from whole-genome re-sequencing of a relatively limited number of

individuals of each species. A direction that now should be taken is to sequence multiple sympatric and allopatric populations of both species and with these data integrate information on recombination rate variation across the genome. This would allow addressing if differentiation islands in multiple population comparisons coincide with recombination cold-spot regions. Such association would be compatible with a scenario of incidental islands where selection at linked sites locally reduces the effective population size and thereby enhance the rate of lineage sorting. Essentially, this means distinguishing a scenario of genomic islands of speciation from a scenario of genomic islands *and* speciation, quoting (Turner & Hahn 2010).

Acknowledgements

Financial support was obtained from the European Research Council, a Knut and Alice Wallenberg Scholar Grant and the Swedish Research Council. Genotyping was performed by the SNP&SEQ Technology Platform, Science for Life Laboratory at Uppsala University, a national infrastructure supported by the Swedish Research Council (VR-RFI) and the Knut and Alice Wallenberg Foundation.

Author Contributions

TK constructed the linkage map, with input from NB. LS constructed the updated assembly, with input from TK, and performed the analyses of genome evolution. TK, NB, and CFM performed recombination rate analysis. AH provided the samples. AQ organized the long-term flycatcher study on Öland. PO contributed to marker development. TK, LS, NB and HE wrote the manuscript with input from the other authors. HE contributed to data analysis, and

conceived of and led the study.

References

Alekseyev MA, Pevzner PA (2007) Are there rearrangement hotspots in the human genome? *PLoS Computational Biology* **3**, e209.

Armengol Ls, Pujana MA, Cheung J, Scherer SW, Estivill X (2003) Enrichment of segmental duplications in regions of breaks of synteny between the human and mouse genomes suggest their involvement in evolutionary rearrangements. *Human Molecular Genetics* **12**, 2201-2208.

Aslam M, Bastiaansen J, Crooijmans R*, et al.* (2010) A SNP based linkage map of the turkey genome reveals multiple intrachromosomal rearrangements between the turkey and chicken genomes. *BMC Genomics* **11**, 647.

Attie O, Darling A, Yancopoulos S (2011) The rise and fall of breakpoint reuse depending on genome resolution. *BMC Bioinformatics* **12**, S1.

Avise JC, Walker D, Johns GC (1998) Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **265**, 1707-1712.

Backström N, Brandstrom M, Gustafsson L*, et al.* (2006) Genetic mapping in a natural population of collared flycatchers (*Ficedula albicollis*): Conserved synteny but gene order rearrangements on the avian Z chromosome. *Genetics* **174**, 377-386.

Backström N, Forstmeier W, Schielzeth H*, et al.* (2010a) The recombination landscape of the zebra finch *Taeniopygia guttata* genome. *Genome Research* **20**, 485-495.

Backström N, Karaiskou N, Leder EH*, et al.* (2008) A gene-based genetic linkage map of the collared flycatcher (*Ficedula albicollis*) reveals extensive synteny and gene-order conservation during 100 million years of avian evolution. *Genetics* **179**, 1479-1495.

Backström N, Palkopoulou E, Qvarnstrom A, Ellegren H (2010b) No evidence for Z-chromosome rearrangements between the pied flycatcher and the collared flycatcher as judged by gene-based comparative genetic maps. *Molecular Ecology* **19**, 3394-3405.

Bailey J, Baertsch R, Kent W, Haussler D, Eichler E (2004) Hotspots of mammalian chromosomal evolution. *Genome Biology* **5**, R23.

Baudat F, Imai Y, de Massy B (2013) Meiotic recombination in mammals: localization and regulation. *Nature Reviews Genetics* **14**, 794-806.

Baudet C, Lemaitre C, Dias Z*, et al.* (2010) Cassis: detection of genomic rearrangement breakpoints. *Bioinformatics* **26**, 1897-1898.

Becker T, Lenhard B (2007) The random versus fragile breakage models of chromosome evolution: a matter of resolution. *Molecular Genetics and Genomics* **278**, 487-491.

Borge T, Lindroos K, Nadvornik P, Syvänen AC, Saetre GP (2005) Amount of introgression in flycatcher hybrid zones reflects regional differences in pre and post-zygotic barriers to gene exchange. *Journal of Evolutionary Biology* **18**, 1416-1424.

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

Burt A, Bell G (1987) Mammalian chiasma frequencies as a test of two theories of recombination. *Nature* **326**, 803-805.

Burt A, Bell G, Harvey PH (1991) Sex-differences in recombination. *Journal of Evolutionary Biology* **4**, 259-277.
Burt DW, Bruley C, Dunn IC*, et al.* (1999) The dynamics of chromosome evolution in birds and mammals. *Nature* **402**, 411-413.
Butlin RK (2005) Recombination and speciation. *Molecular Ecology* **14**, 2621-2635.
Campos JL, Halligan DL, Haddrill PR, Charlesworth B (2014) The relation between recombination rate and patterns of molecular evolution and variation in *Drosophila melanogaster*. *Molecular Biology and Evolution*.
Connallon T, Clark AG (2010) Sex linkage, sex-specific selection, and the role of recombination in the evolution of sexually dimorphic gene expression. *Evolution* **64**, 3417-3442.
Consortium HG (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **487**, 94-98.
Coop G, Przeworski M (2007) An evolutionary view of human recombination. *Nature Reviews Genetics* **8**, 23-34.
Cox A, Ackert-Bicknell CL, Dumont BL*, et al.* (2009) A new standard genetic map for the laboratory mouse. *Genetics* **182**, 1335-1344.
Coyne JA, Orr HA (2004) *Speciation* Sinauer Associates, Inc., Sunderland, MA.
Cutter AD, Payseur BA (2013) Genomic signatures of selection at linked sites: unifying the disparity among species. *Nature Reviews Genetics* **14**, 262-274.
Dalloul RA, Long JA, Zimin AV*, et al.* (2010) Multi-platform next-generation sequencing of the domestic turkey *Meleagris gallopavo*: genome assembly and analysis. *PLoS Biology* **8**, e1000475.
Darling AE, Mau B, Perna NT (2010) progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* **5**, e11147.
Dietrich WF, Miller J, Steen R*, et al.* (1996) A comprehensive genetic map of the mouse genome. *Nature* **380**, 149-152.
Dumont BL, Payseur BA (2008) Evolution of the genomic rate of recombination in mammals. *Evolution* **62**, 276-294.
Dumont BL, Payseur BA (2011) Evolution of the genomic recombination rate in murid rodents. *Genetics* **187**, 643-657.
Dumont BL, White MA, Steffy B, Wiltshire T, Payseur BA (2011) Extensive recombination rate variation in the house mouse species complex inferred from genetic linkage maps. *Genome Research* **21**, 114-125.
Duret L, Arndt PF (2008) The impact of recombination on nucleotide substitutions in the human genome. *PLoS Genetics* **4**, e1000071.
Dyke G, Kaiser G (2011) Major Events in Avian Genome Evolution. In: *Living Dinosaurs: The Evolutionary History of Modern Birds* (eds. Dyke G, Kaiser G, Organ C, Edwards S). John Wiley & Sons Ltd, Chichester, UK.
Ellegren H (1991) DNA typing of museum birds. *Nature* **354**, 113.
Ellegren H (1992) Polymerase chain reaction (PCR) analysis of microsatellites - a new approach to studies of genetic relationships in birds. *Auk* **109**, 886-895.
Ellegren H (2010) Evolutionary stasis: the stable chromosomes of birds. *Trends in Ecology & Evolution* **25**, 283-291.
Ellegren H (2013) The evolutionary genomics of birds. *Annual Review of Ecology, Evolution, and Systematics* **44**, 239-259.
Ellegren H (2014) Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution* **29**, 51-63.

Ellegren H, Gustafsson L, Sheldon BC (1996) Sex ratio adjustment in relation to paternal attractiveness in a wild bird population. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 11723-11728.

Ellegren H, Sméds L, Burri R, *et al.* (2012) The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* **491**, 756-760.

Fitzpatrick BM (2004) Rates of evolution of hybrid inviability in birds and mammals. *Evolution* **58**, 1865-1870.

Fledel-Alon A, Wilson DJ, Broman K, *et al.* (2009) Broad-scale recombination patterns underlying proper disjunction in humans. *PLoS Genetics* **5**, e1000658.

Gelter HP, Tegelstrom H (1992) High-frequency of extra-pair paternity in Swedish pied flycatchers revealed by allozyme electrophoresis and DNA fingerpringting. *Behavioral Ecology and Sociobiology* **31**, 1-7.

Gelter HP, Tegelstrom H, Stahl G (1989) Alloyme similarity between the pied and collared flycatchers (Aves, *Ficedula-hypoleuca* and *Ficedula-albicollis*). *Hereditas* **111**, 65-72.

Green P, Falls K, Crook S (1990) Documentation for CRIMAP, version 2.4. . Washington Univ. School of Medicine, St. Louis, MO.

Gregory TR (2011) *Animal Genome Size Database.* http://www.genomesize.com

Griffin DK, Robertson LBW, Tempest HG, Skinner BM (2007) The evolution of the avian genome as revealed by comparative molecular cytogenetics. *Cytogenetic and Genome Research* **117**, 64-77.

Groenen MAM, Wahlberg P, Foglio M, *et al.* (2009) A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. *Genome Research* **19**, 510-519.

Gustafsson L, Qvarnström A, Sheldon BC (1995) Trade-offs between life-history traits and a secondary sexual character in male collared flycatchers. *Nature* **375**, 311-313.

Hansson B, √Ökesson M, Slate J, Pemberton JM (2005) Linkage mapping reveals sex-dimorphic map distances in a passerine bird. *Proceedings of the Royal Society B: Biological Sciences* **272**, 2289-2298.

Hansson B, Ljungqvist M, Dawson DA, *et al.* (2009) Avian genome evolution: insights from a linkage map of the blue tit (*Cyanistes caeruleus*). *Heredity* **104**, 67-78.

Heinz S, Benner C, Spann N, *et al.* (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell* **38**, 576-589.

Hoffmann AA, Rieseberg LH (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**, 21-42.

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

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

Jaari S, Li M-H, Merila J (2009) A first-generation microsatellite-based genetic linkage map of the Siberian jay (*Perisoreus infaustus*): insights into avian genome evolution. *BMC Genomics* **10**, 1.

Jacob HJ, Brown DM, Bunker RK, *et al.* (1995) A genetic linkage map of the laboratory rat, *Rattus norvegicus*. *Nature Genetics* **9**, 63-69.

Kawakami T, Backström N, Burri R, *et al.* (2014) Estimation of linkage disequilibrium and interspecific gene flow in *Ficedula* flycatchers by a newly developed 50K SNP array. *Molecular Ecology Resources* **in revision**.

Kemkemer C, Kohn M, Cooper D, *et al.* (2009) Gene synteny comparisons between different vertebrates provide new insights into breakage and fusion events during mammalian karyotype evolution. *BMC Evolutionary Biology* **9**, 84.
Kidd JM, Cooper GM, Donahue WF, *et al.* (2008) Mapping and sequencing of structural variation from eight human genomes. *Nature* **453**, 56-64.
Kirkpatrick M, Barton N (2006) Chromosome inversions, local adaptation and speciation. *Genetics* **173**, 419-434.
Kong A, Thorleifsson G, Gudbjartsson DF, *et al.* (2010) Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* **467**, 1099-1103.
Larkin DM, Pape G, Donthu R, *et al.* (2009) Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. *Genome Research* **19**, 770-777.
Lee J, Han K, Meyer TJ, Kim H-S, Batzer MA (2008) Chromosomal inversions between human and chimpanzee lineages caused by retrotransposons. *PLoS ONE* **3**, e4047.
Lehtonen PK, Laaksonen T, Artemyev AV, *et al.* (2009) Geographic patterns of genetic differentiation and plumage colour variation are different in the pied flycatcher (*Ficedula hypoleuca*). *Molecular Ecology* **18**, 4463-4476.
Lenormand T (2003) The evolution of sex dimorphism in recombination. *Genetics* **163**, 811-822.
Lenormand T, Dutheil J (2005) Recombination difference between sexes: A role for haploid selection. *PLoS Biol* **3**, e63.
Leskinen PK, Laaksonen T, Ruuskanen S, Primmer CR, Leder EH (2012) The proteomics of feather development in pied flycatchers (*Ficedula hypoleuca*) with different plumage coloration. *Molecular Ecology* **21**, 5762-5777.
Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595.
Lohmueller KE, Degenhardt JD, Keinan A (2010) Sex-averaged recombination and mutation rates on the X chromosome: A comment on Labuda et al. *The American Journal of Human Genetics* **86**, 978-980.
Mank JE (2009) The evolution of heterochiasmy: the role of sexual selection and sperm competition in determining sex-specific recombination rates in eutherian mammals. *Genetical Research* **91**, 355-363.
Masabanda JS, Burt DW, O'Brien PCM, *et al.* (2004) Molecular cytogenetic definition of the chicken genome: The first complete avian karyotype. *Genetics* **166**, 1367-1373.
McGaugh SE, Heil CSS, Manzano-Winkler B, *et al.* (2012) Recombination modulates how selection affects linked sites in *Drosophila*. *PLoS Biol* **10**, e1001422.
Mendonca MAC, Carvalho CR, Clarindo WR (2010) DNA content differences between male and female chicken (*Gallus gallus domesticus*) nuclei and Z and W chromosomes resolved by image cytometry. *Journal of Histochemistry & Cytochemistry* **58**, 229-235.
Mugal CF, Arndt PF, Ellegren H (2013) Twisted signatures of GC-biased gene conversion embedded in an evolutionary stable karyotype. *Molecular Biology and Evolution* **30**, 1700-1712.
Murphy WJ, Larkin DM, van der Wind AE, *et al.* (2005) Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science* **309**, 613-617.
Myers S, Bowden R, Tumian A, *et al.* (2010) Drive against hotspot motifs in primates implicates the *PRDM9* gene in meiotic recombination. *Science* **327**, 876-879.

Myers S, Freeman C, Auton A, Donnelly P, McVean G (2008) A common sequence motif associated with recombination hot spots and genome instability in humans. *Nature Genetics* **40**, 1124-1129.
Nabholz B, Künstner A, Wang R, Jarvis E, Ellegren H (2011) Dynamic evolution of base composition: causes and consequences in avian phylogenomics. *Molecular Biology and Evolution* **28**, 2197-2210.
Nachman MW, Payseur BA (2012) Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal Society B-Biological Sciences* **367**, 409-421.
Nadachowska-Brzyska K, Burri R, Olason PI, *et al.* (2013) Demographic divergence history of pied flycatcher and collared flycatcher inferred from whole-genome re-sequencing data. *PLoS Genetics* **9**.
Nam K, Mugal C, Nabholz B, *et al.* (2010) Molecular evolution of genes in avian genomes. *Genome Biology* **11**, R68.
Navarro A, Barton NH (2003) Chromosomal speciation and molecular divergence--accelerated evolution in rearranged chromosomes. *Science* **300**, 321-324.
Noor MAF, Grams KL, Bertucci LA, Reiland J (2001) Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 12084-12088.
Otto SP, Lenormand T (2002) Resolving the paradox of sex and recombination. *Nature Reviews Genetics* **3**, 252-261.
Pardo-Manuel de Villena F, Sapienza C (2001) Recombination is proportional to the number of chromosome arms in mammals. *Mammalian Genome* **12**, 318-322.
Paterson T, Law A (2011) Genotypechecker: an interactive tool for checking the inheritance consistency of genotyped pedigrees. *Animal Genetics* **42**, 560-562.
Peng Q, Pevzner PA, Tesler G (2006) The fragile breakage versus random breakage models of chromosome evolution. *PLoS Computational Biology* **2**, e14.
Pevzner P, Tesler G (2003) Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 7672-7677.
Presgraves DC (2005) Recombination enhances protein adaptation in *Drosophila melanogaster*. *Current Biology* **15**, 1651-1656.
Price TD, Bouvier MM (2002) The evolution of F1 postzygotic incompatibilities in birds. *Evolution* **56**, 2083-2089.
Primmer CR, Borge T, Lindell J, Saetre GP (2002) Single-nucleotide polymorphism characterization in species with limited available sequence information: high nucleotide diversity revealed in the avian genome. *Molecular Ecology* **11**, 603-612.
Purcell S, Neale B, Todd-Brown K, *et al.* (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559-575.
Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842.
Qvarnstrom A, Bailey RI (2008) Speciation through evolution of sex-linked genes. *Heredity* **102**, 4-15.
Qvarnstrom A, Brommer JE, Gustafsson L (2006) Testing the genetics underlying the co-evolution of mate choice and ornament in the wild. *Nature* **441**, 84-86.
Qvarnström A, Rice AM, Ellegren H (2010) Speciation in *Ficedula* flycatchers. *Philosophical Transactions of the Royal Society B-Biological Sciences* **365**, 1841-1852.

Ratti O, Hovi M, Lundberg A, Tegelstrom H, Alatalo RV (1995) Extra-pair paternity and male characteristics in the pied flycatcher. *Behavioral Ecology and Sociobiology* **37**, 419-425.
Ross-Ibarra J (2004) The evolution of recombination under domestication: A test of two hypotheses. *The American Naturalist* **163**, 105-112.
Saether SA, Saetre G-P, Borge T, *et al.* (2007) Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. *Science* **318**, 95-97.
Saetre GP, Borge T, Lindell J, *et al.* (2001) Speciation, introgressive hybridization and nonlinear rate of molecular evolution in flycatchers. *Molecular Ecology* **10**, 737-749.
Saetre GP, Borge T, Lindroos K, *et al.* (2003) Sex chromosome evolution and speciation in Ficedula flycatchers. *Proceedings of the Royal Society B-Biological Sciences* **270**, 53-59.
Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbour Laboratory Press, New York.
Sankoff D, Trinh P (2005) Chromosomal breakpoint reuse in genome sequence rearrangement. *Journal of Computational Biology* **12**, 812-821.
Segura J, Ferretti L, Ramos-Onsins S, *et al.* (2013) Evolution of recombination in eutherian mammals: insights into mechanisms that affect recombination rates and crossover interference. *Proceedings of the Royal Society B-Biological Sciences* **280**.
Shapiro MD, Kronenberg Z, Li C, *et al.* (2013) Genomic diversity and evolution of the head crest in the rock pigeon. *Science* **339**, 1063-1067.
Sheldon B, Ellegren H (1999) Sexual selection resulting from extrapair paternity in collared flycatchers. *Animal Behaviour* **57**, 285-298.
Sheldon BC, Ellegren H (1996) Offspring sex and paternity in the collared flycatcher. *Proceedings of the Royal Society B-Biological Sciences* **263**, 1017-1021.
Skinner BM, Griffin DK (2012) Intrachromosomal rearrangements in avian genome evolution: evidence for regions prone to breakpoints. *Heredity* **108**, 37-41.
Smukowski CS, Noor MAF (2011) Recombination rate variation in closely related species. *Heredity* **107**, 496-508.
Stapley J, Birhead TR, Burke T, Slate J (2008) A linkage map of the zebra finch *Taeniopygia guttata* provides new insights into avian genome evolution. *Genetics* **179**, 651-667.
Sætre GP, Sæther SA (2010) Ecology and genetics of speciation in *Ficedula* flycatchers. *Molecular Ecology* **19**, 1091-1106.
Tegelstrom H, Gelter HP (1990) Haldane rule and sex biased gene flow between 2 hybridizing flycatcher species (*Ficedula-albicollis and F-hypoleuca*, Aves, Muscicapidae). *Evolution* **44**, 2012-2021.
Tesler G (2002) GRIMM: genome rearrangements web server. *Bioinformatics* **18**, 492-493.
Thomas JW, Cáceres M, Lowman JJ, *et al.* (2008) The chromosomal polymorphism linked to variation in social behavior in the white-throated sparrow (*Zonotrichia albicollis*) is a complex rearrangement and suppressor of recombination. *Genetics* **179**, 1455-1468.
Thorneycroft HB (1966) Chromosomal polymorphism in white-throated sparrow *Zonotrichia albicollis* (Gmelin). *Science* **154**, 1571-1572.
Turelli M, Moyle LC (2007) Asymmetric postmating isolation: Darwin's corollary to Haldane's rule. *Genetics* **176**, 1059-1088.
Turner TL, Hahn MW (2010) Genomic islands of speciation or genomic islands and speciation? *Molecular Ecology* **19**, 848-850.
Uebbing S, Kunstner A, Makinen H, Ellegren H (2013) Transcriptome sequencing reveals the character of incomplete dosage compensation across multiple tissues in flycatchers. *Genome Biology and Evolution* **5**, 1555-1566.

van Oers K, Santure AW, De Cauwer I*, et al.* (2014) Replicated high-density genetic maps of two great tit populations reveal fine-scale genomic departures from sex-equal recombination rates. *Heredity* **112**, 307-316.
Veen T, Borge T, Griffith SC*, et al.* (2001) Hybridization and adaptive mate choice in flycatchers. *Nature* **411**, 45-50.
Völker M, Backström N, Skinner BM*, et al.* (2010) Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution. *Genome Research* **20**, 503-511.
Wang J, Fan HC, Behr B, Quake SR (2012) Genome-wide single-cell analysis of recombination activity and *de novo* mutation rates in human sperm. *Cell* **150**, 402-412.
Warren WC, Clayton DF, Ellegren H*, et al.* (2010) The genome of a songbird. *Nature* **464**, 757-762.
Webster MT, Hurst LD (2012) Direct and indirect consequences of meiotic recombination: implications for genome evolution. *Trends in Genetics* **28**, 101-109.
White MJD (1973) *Animal Cytology and Evolution* Cambridge University Press.
Winckler W, Myers SR, Richter DJ*, et al.* (2005) Comparison of fine-scale recombination rates in humans and chimpanzees. *Science* **308**, 107-111.
Wyman MJ, Wyman MC (2013) specific recombination rates and allele frequencies affect the invasion of sexually antagonistic variation on autosomes. *Journal of Evolutionary Biology* **26**, 2428-2437.
Zhao H, Bourque G (2009) Recovering genome rearrangements in the mammalian phylogeny. *Genome Research* **19**, 934-942.
Åkesson M, Hansson B, Hasselquist D, Bensch S (2007) Linkage mapping of AFLP markers in a wild population of great reed warblers: importance of heterozygosity and number of genotyped individuals. *Molecular Ecology* **16**, 2189-2202.

Figure legends

Figure 1. A genetic linkage map of the collared flycatcher genome. The horizontal bars on each chromosome or linkage group represent mapped SNP markers based on best-order map. The scale bar to the left shows the lengths of linkage groups as measured in Kosambi cM. A more detailed map with marker names is presented as Supplementary Figure 4.

Figure 2. Comparative circular visualization of the organization of homologous chromosomes in collared flycatcher and zebra finch. Collared flycatcher is shown to the left, zebra finch to the right. Scale is indicated on the zebra finch side of plots, in Mb.

Figure 3. Genomic distribution of breakpoint regions. Resolution is 50 kb synteny blocks in pairwise whole-genome alignments of flycatcher and zebra finch.

Figure 4. The relationship a) between the genetic distance (cM) and the chromosome size (Mb) and b) between recombination rate (cM/Mb) and the chromosome size (Mb) in the best-order map of the collared flycatcher genome. Red is the Z chromosome.

Figure 5. Scatter plot showing the recombination rate as a function of distance to chromosome end for a) chromosomes >20 Mb and b) chromsomes < 20 Mb. Values > 45 cM/Mb are omitted from the graph to increase detail.

Figure 6. Sex-specific relationships between the genetic (cM) and the physical (Mb) distance in the best-order map of collared flycatcher chromosomes 17 and 27. Red and blue circles indicate the female- and male-specific genetic map, respectively.

Figure 7. Barplot showing the fractions explained by each parameter in the Principal Component Regression analysis with a) recombination and b) GC-content as response variables. In each analysis the proportional contribution of each explanatory variable was very similar but the total amount of the variance explained was higher when using the GC-content as a response variable (notice the difference in scaling of the y-axis).

Figure 8. The relationship between the genetic (cM) and the physical (Mb) distance in the best-order map of the collared flycatcher genome (blue circle, this study), zebra finch (orange square, Backström *et al*. 2010) and chicken (black cross, Groenen *et al*. 2009). For comparison purposes, chromosomes 1 and 4 in chicken were split into two chromosomes, corresponding respectively to chromosomes 1 and 1A, and chromosomes 4 and 4A, based on the zebra finch genome. Data were not available for chromosomes 2, 21, 22, 24, 26, 27, and 28 in zebra finch.

Table 1. Number of mapped SNP markers and total genetic distance of each chromosomes of the collared flycatcher.

	Number of markers			Framework map (cM)			Best-order map (cM)			Forced map (cM)		
Chrom.	Frame-work map	Best-order map	Forced map	Average	Female	Male	Average	Female	Male	Average	Female	Male
1	147	341	2,247	252.8	238.5	266.9	246.2	231.1	261.6	261.9	248.4	276.5
1A	166	317	1,976	204.9	188.8	220.4	206.1	189.3	224.2	227.5	202.8	256.5
2	172	301	1,608	325.9	314.4	335.1	316.1	307.8	323.4	364.7	343.4	391.7
3	123	180	1,518	232.5	224.7	240.8	225.9	214.4	237.0	243.4	236.1	252.0
4	58	102	363	168.6	166.7	170.4	167.2	164.3	169.9	151.2	146.4	155.1
4A	100	158	1,498	81.7	74.0	89.5	80.3	73.2	88.2	117.4	109.6	127.1
5	95	199	1,389	168.4	157.3	181.3	170.4	156.6	185.1	199.5	175.3	226.2
6	87	138	1,022	125.4	119.1	131.9	121.3	115.9	126.5	121.7	115.8	128.1
7	101	203	1,209	122.0	118.4	125.7	122.5	119.1	128.9	214.4	141.3	299.5
8	91	116	752	95.7	94.4	98.7	96.0	93.1	99.3	126.2	122.4	131.3
9	104	184	1,207	96.8	90.0	103.8	96.6	90.0	103.5	116.2	110.0	123.4
10	97	160	1,529	93.2	93.7	93.3	94.0	91.2	97.5	106.7	106.9	106.8
11	95	160	1,593	84.0	72.6	95.2	81.0	68.9	93.6	101.8	106.2	99.9
12	79	138	1,072	83.5	70.1	97.0	84.7	71.2	99.6	90.8	82.6	100.4
13	39	65	451	85.5	86.3	84.5	87.8	87.0	86.1	86.6	90.9	81.8
14	95	132	1,227	88.1	83.3	92.1	87.6	82.5	92.3	95.3	90.1	101.2
15	88	146	1,149	60.6	55.9	65.6	59.3	54.1	65.2	64.2	56.7	72.7
17	108	146	1,169	74.8	58.3	93.1	73.8	58.1	90.4	77.5	60.3	96.7
18	70	128	1,115	79.2	76.3	81.5	79.7	78.9	79.6	88.8	82.9	95.6

19	66	126	1,163	55.8	59.9	51.3	58.0	59.5	56.6	57.8	59.3	56.5
20	75	129	1,257	52.5	55.5	48.6	53.7	56.1	51.1	54.4	57.4	51.0
21	39	64	695	46.6	46.0	46.8	48.3	48.6	47.9	48.8	48.7	49.0
22	15	32	44	51.0	60.2	42.3	53.2	56.0	50.5	51.0	53.8	48.6
23	45	96	929	47.1	51.7	41.8	49.1	53.2	44.7	50.2	52.6	47.6
24	71	98	1,149	50.6	49.2	52.0	50.5	51.3	50.2	54.7	51.3	58.6
25	12	22	60	46.2	43.7	51.0	47.9	45.6	53.0	50.0	48.5	54.0
26	48	81	950	46.7	50.7	41.6	46.3	48.1	43.2	88.5	98.0	79.6
27	42	73	534	74.9	69.4	83.6	73.5	68.0	82.2	85.1	80.6	97.7
28	24	39	188	48.9	49.9	49.2	48.2	49.3	49.5	66.2	72.3	60.4
LGE22	16	32	38	52.3	52.1	51.6	53.3	53.0	53.6	49.8	43.5	55.8
Fal34	2	94	10	12.1	31.0	6.9	16.3	20.7	11.9	15.4	19.1	11.5
Fal35	5	8	8	36.8	35.5	39.3	37.2	34.7	39.6	36.6	34.7	39.0
Fal36	2	5	5	2.9	0	5.6	9.7	6.5	13.8	9.7	6.5	13.8
Total[a]	2,377	4,213	31,124	3,148	3,038	3,278	3,142	2,997	3,300	3,574	3,354	3,846
Z	79	89	743	107.7[b]	-	161.6	107.5[b]	-	161.2	115.9[b]	-	173.9
Total[c]	2,456	4,302	31,867	3,256	3,038	3,440	3,249	2,997	3,461	3,690	3,354	4,020

a Autosomes.

b Sex-average genetic distance for chromosome Z calculated as male genetic distance * 2/3.

c Autosomes plus Z chromosome.

Table 2. Number of ordered and oriented scaffolds assigned to each collared flycatcher chromosome in the FicAlb1.5 assembly version. Also shown are assembly size of homologous chromosomes of flycatcher, zebra finch and chicken.

	Collared flycatcher					Zebra finch (Mb)	Chicken (Mb)
Chromosome	Scaffolds	Super-scaffolds	Singleton scaffolds[a]	Not oriented scaffolds	Size (Mb)		
1	43	6	3	1	119.8	119.6[b]	201.0
1A	38	8	6	2	74.8	73.7	-
2	35	8	3	1	157.4	156.4	154.9
3	28	6	5	1	115.7	112.6	113.7
4A	5	2	1	1	21.2	69.8	94.2
4	26	4	4	1	70.3	20.7	-
5	22	5	5	0	64.6	62.4	62.2
6	11	2	1	0	37.2	36.3	37.4
7	14	3	0	0	39.3	39.8	38.4
8	12	3	0	0	32.0	28.0	30.7
9	7	3	0	0	26.8	27.2	25.6
10	11	2	0		21.3	20.8	22.6
11	5	2	1	0	21.7	21.4	21.9
12	11	2	1	0	21.9	21.6	20.5
13	6	2	2	0	18.6	17.0	18.9
14	2	1	0	0	17.4	16.4	15.8
15	1	0	1	0	14.9	14.4	13.0
16	0	0	0	0	-	<0.01	0.43
17	4	2	0	0	12.4	11.6	11.2
18	13	1	1	1	13.1	11.2	10.9
19	6	2	0	0	11.9	11.6	9.9
20	8	2	1	0	15.6	15.7	14.0
21	5	1	2	1	8.1	6.0	7.0
22	8	3	1	3	5.7	3.4	3.9

23	5	1	0	0	7.9	6.2	6.0
24	4	1	0	0	8.0	8.0	6.4
25	19	3	2	1	2.7	1.3	2.0
26	6	1	2	1	7.6	4.9	5.1
27	19	4	2	7	5.5	4.6	4.8
28	12	3	0	3	6.1	5.0	4.5
LGE22[c]	10	3	3	4	2.1	0.9	0.9
Fal34	4	1	1	4	0.11[d]	-	-
Fal35	0	0	0	7	0[e]	-	-
Fal36	1	0	1	3	0.18[f]		
Z	32	8	5	7	59.7	74.6	72.9
Total					1042	1023	1031

[a] scaffolds which could not be joined to other scaffolds in the super-scaffolding process.

[b] 1.1 Mb from the tentative chromosome 1B in zebra finch has been added to 1A of this species.

[c] In the most recent chicken genome assembly (Galgal4), the full name of this linkage group is LGE22C19W28_E50C23.

[d] Fal34 has an assembly size of 0.29 Mb when including un-oriented scaffolds.

[e] Fal35 has an assembly size of 0.32 Mb when including un-oriented scaffolds.

[f] Fal36 has an assembly size of 0.46 Mb when including un-oriented scaffolds.

Table 3. Summary assembly statistics for the second-generation assembly version of the collared flycatcher genome (FicAlb1.5; present study) and the previous FicAlb_1.4 version (Ellegren *et al.* 2012). All data from scaffolds >200 bp are included, which explains the large number of unassigned scaffolds and the large total number of scaffolds. Data for scaffolds >100 kb are shown in parentheses; note that excluding scaffolds <100 kb has little influence on total assembly size. 'Inferred' means scaffolds indirectly assigned to chromosomes based on conserved synteny with zebra finch.

	FicAlb_1.4		FicAlb1.5	
	# Scaffolds	Size (Mb)	# Scaffolds	Size (Mb)
Ordered and oriented	67	596	441	1042
Ordered	67	224	46	15
Inferred	164	182	2	0
Unassigned	21,467 (109)	114 (73)	21,354 (73)	59 (23)
Total	21,765 (404)	1,116 (1076)	21,843 (451)	1,116 (1075)

Table 4. Number of intrachromosomal rearrangements (inversions) per chromosome in three avian lineages detected with a resolution of 50 kb synteny blocks.

Chromosome	Chicken	Zebra finch	Flycatcher
1+1A	29	11	18
2	9	10	6
3	18	6	4
4+4A	33	5	2
5	11	3	3
6	8	1	0
7	4	5	1
8	6	1	1
9	5	1	0
10	2	0	1
11	8	4	0
12	6	1	0
13	3	0	0
14	6	1	1
15	4	1	0
17	0	1	0
18	4	2	0
19	2	1	0
20	5	1	2
21	4	2	0
22	3	0	0
23	5	2	0
24	5	2	0
25	0	0	4
26	4	2	3
27	2	2	3

28	7	2	1
Z	10	12	6
Total	203	79	61

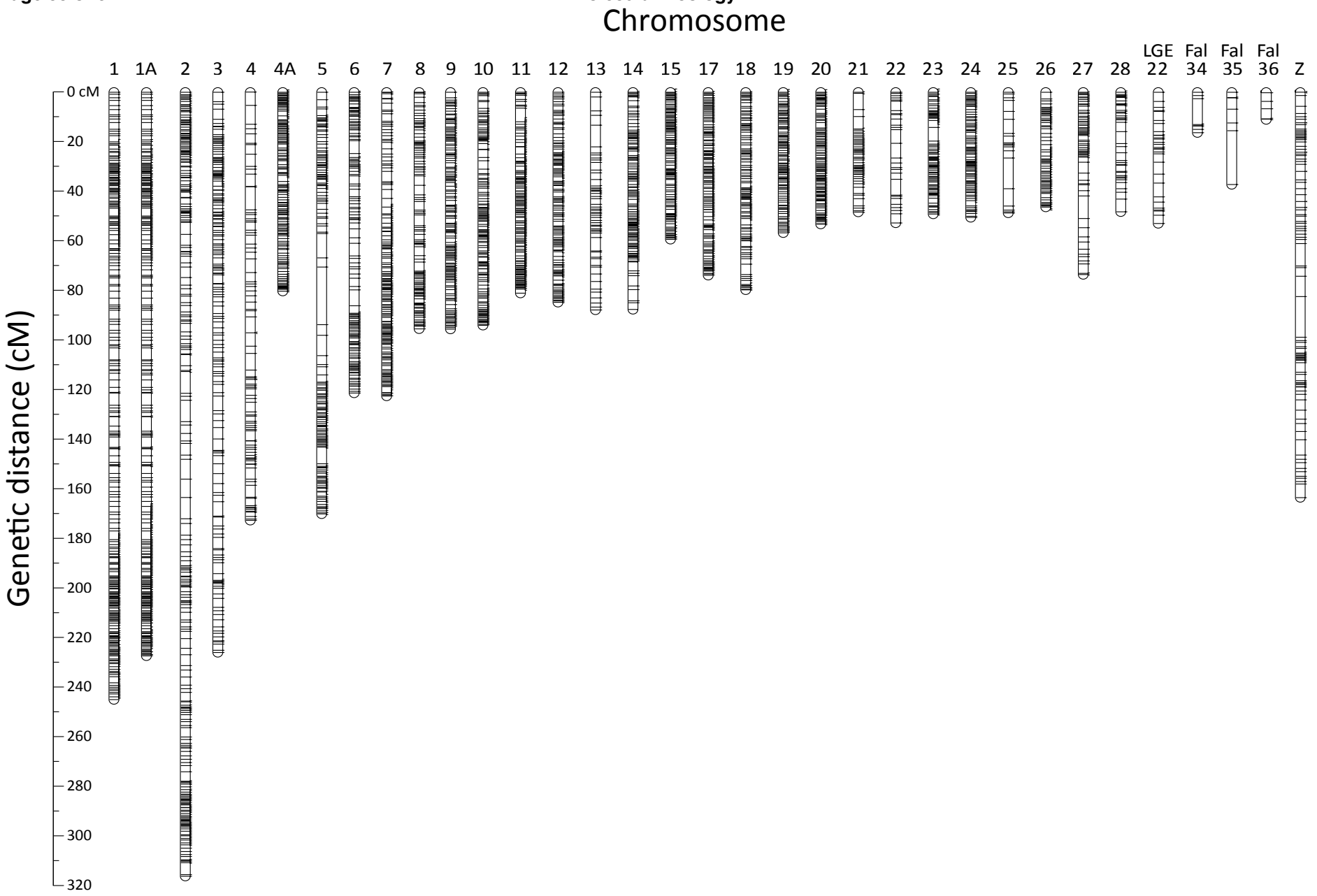
Table 5. Mean (and standard deviation, S.D.) sex-average recombination rate per chromosome for different autosomal size categories. Also shown are recombination rates after 50 cM (corresponding to one obligate recombination event per chromosome) has been subtracted from the genetic length of each chromosome.

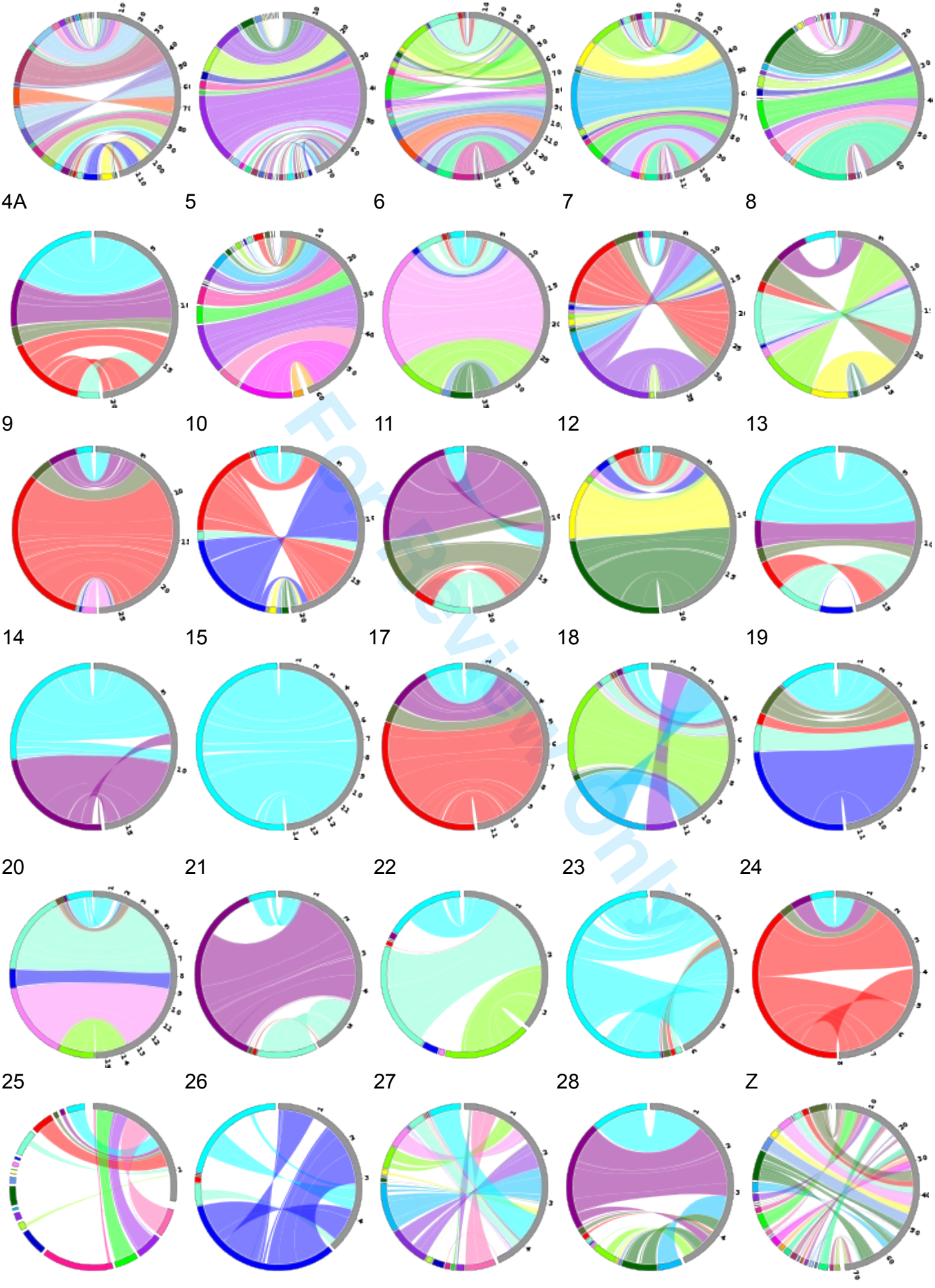
Chromosomal size category	n	Recombination rate (cM/Mb)		Recombination rate (cM/Mb) after subtracting 50 cM	
		Mean	S.D.	Mean	S.D.
>100 Mb	3	2.0	0.0	1.6	0.1
50-100 Mb	3	2.7	0.3	2.0	0.3
25-50 Mb	4	3.2	0.2	1.7	0.2
10-25 Mb	11	4.5	0.9	1.5	0.7
< 10 Mb	9	11.1	6.9	-[a]	-[a]

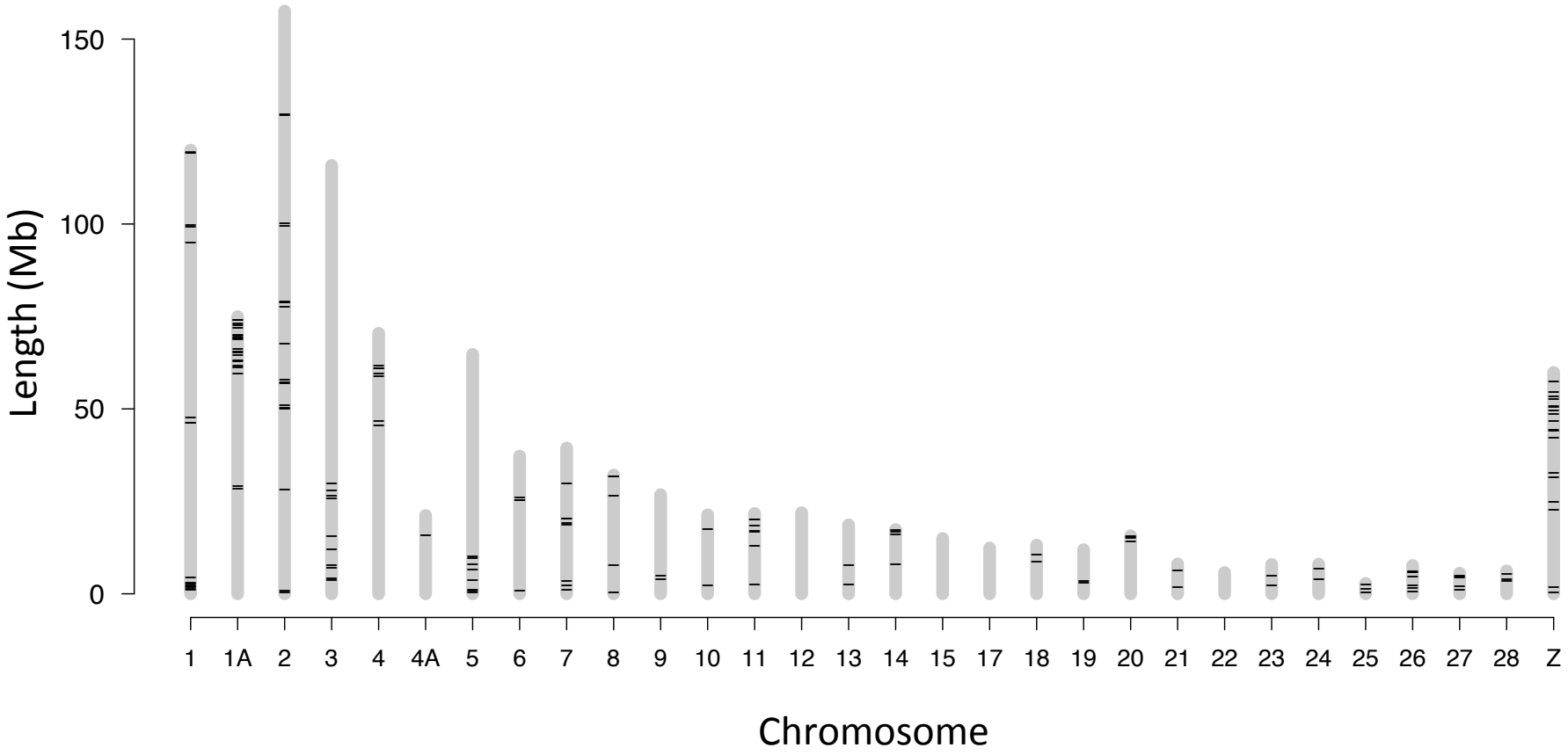
[a] Most of these chromosomes have a genetic distance of less than 50 cM.

Table 6. Estimates (Est) and statistical significance (*p*-value) of multi-linear regression (MLR) analysis for six candidate explanatory variables of variation in recombination rate and GC-content, respectively. Included in the table is also the raw Pearson's pair-wise correlation coefficient (r^2) between each explanatory variable and recombination rate, and the amount of variation explained by each explanatory variable according to the principal component regression (PCR). Data are from chromosomes >20 Mb.

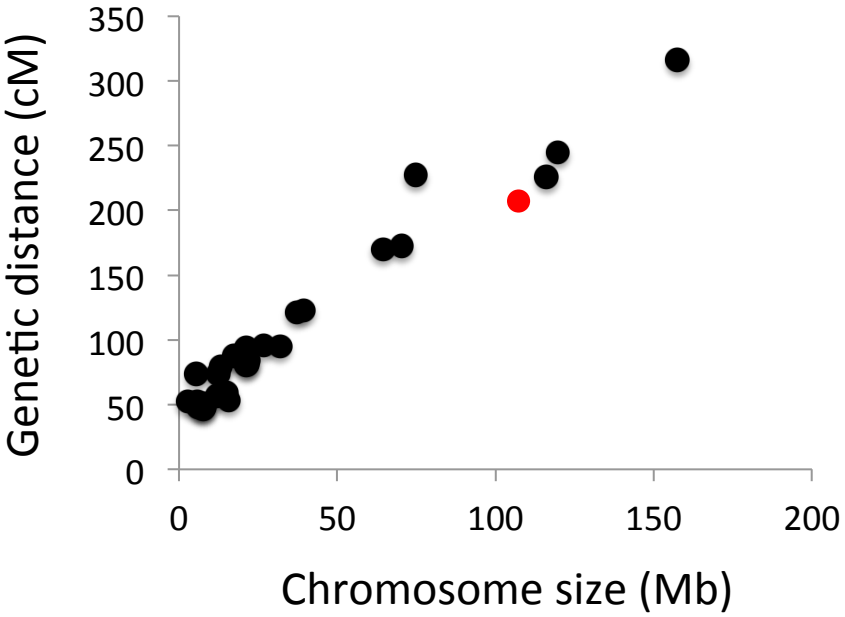
Parameter	Recombination				GC-content			
	Est	*p*-value	r^2	PCR (%)	Est	*p*-value	r^2	PCR (%)
Distance to end	-0.069	$< 10^{-15}$	-0.37	5.37	-0.048	$< 10^{-15}$	-0.42	12.21
Chromosome size	-0.036	$< 10^{-15}$	-0.21	1.89	-0.037	$< 10^{-15}$	-0.37	9.34
Microsatellites	0.002	0.67	0.23	5.75	0.107	$< 10^{-15}$	0.58	13.21
Repeat density	-0.001	0.81	0.01	2.60	-0.059	$< 10^{-15}$	-0.11	10.24
Motif density	0.061	$< 10^{-15}$	0.38	5.78	0.133	$< 10^{-15}$	0.65	15.40
Gene density	0.022	$9.0 \cdot 10^{-9}$	0.09	0.78	0.073	$< 10^{-15}$	0.29	5.52
Total				22.2				65.5


Chromosome
1
1A
2
3
4
4A
5
6
7
8
9
10
11
12
13
14
15
17
18
19
20
21
22
23
24
25
26
27
28
LGE 22
Fal 34
Fal 35
Fal 36
Z
Genetic distance (cM)
0 cM
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
320

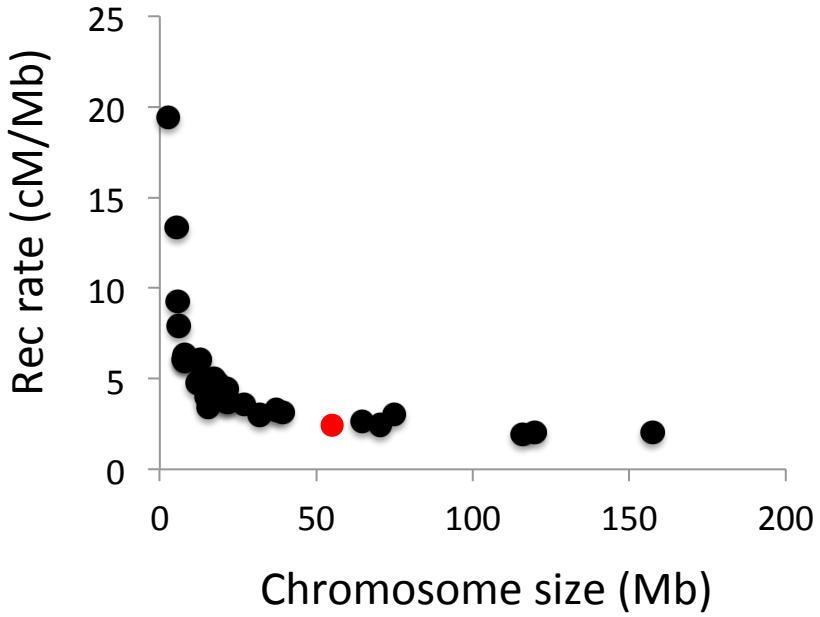


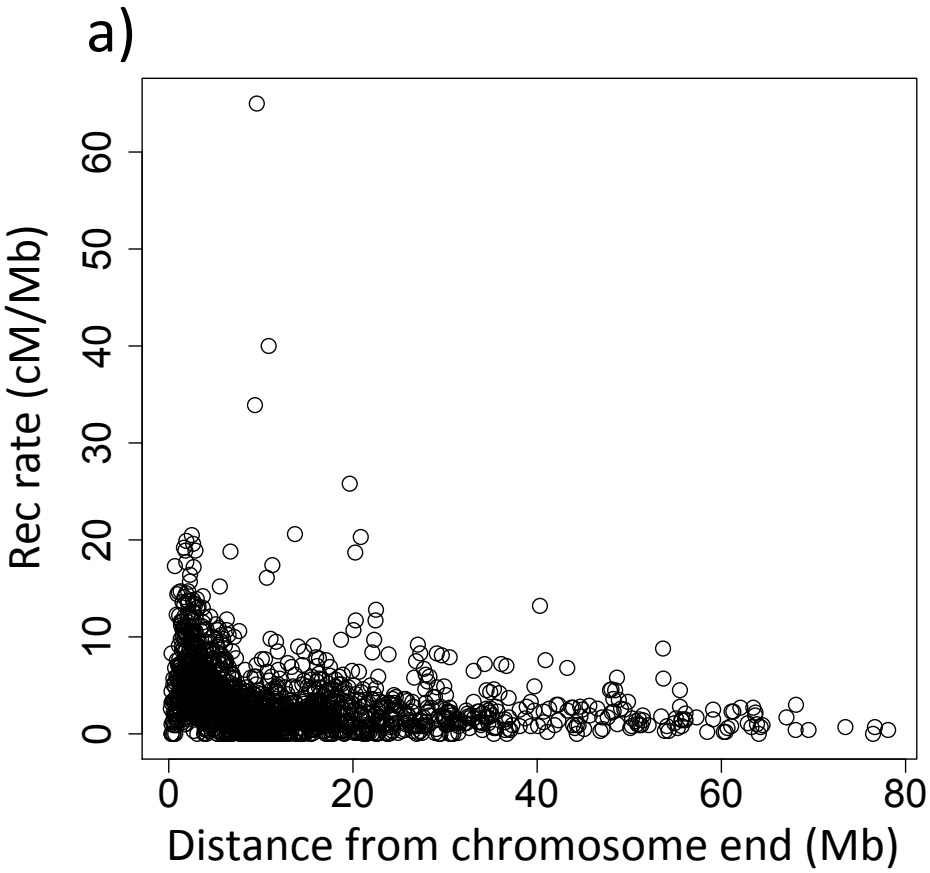

150
100
50
0
Length (Mb)
1
1A
2
3
4
4A
5
6
7
8
9
10
11
12
13
14
15
17
18
19
20
21
22
23
24
25
26
27
28
Z
Chromosome

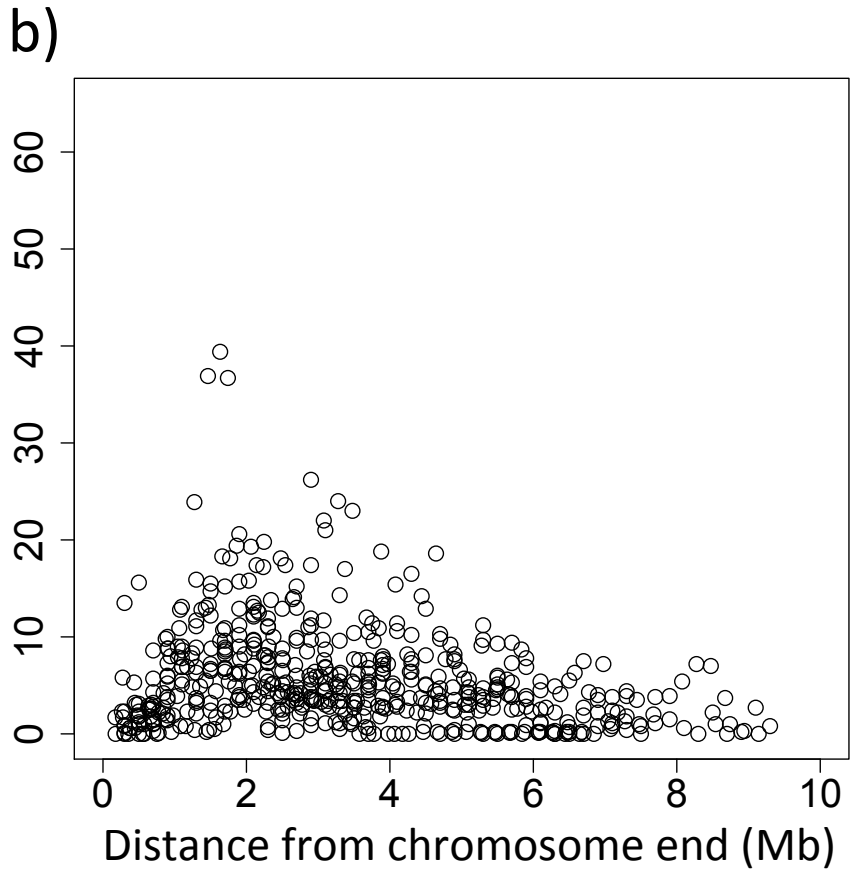
a)

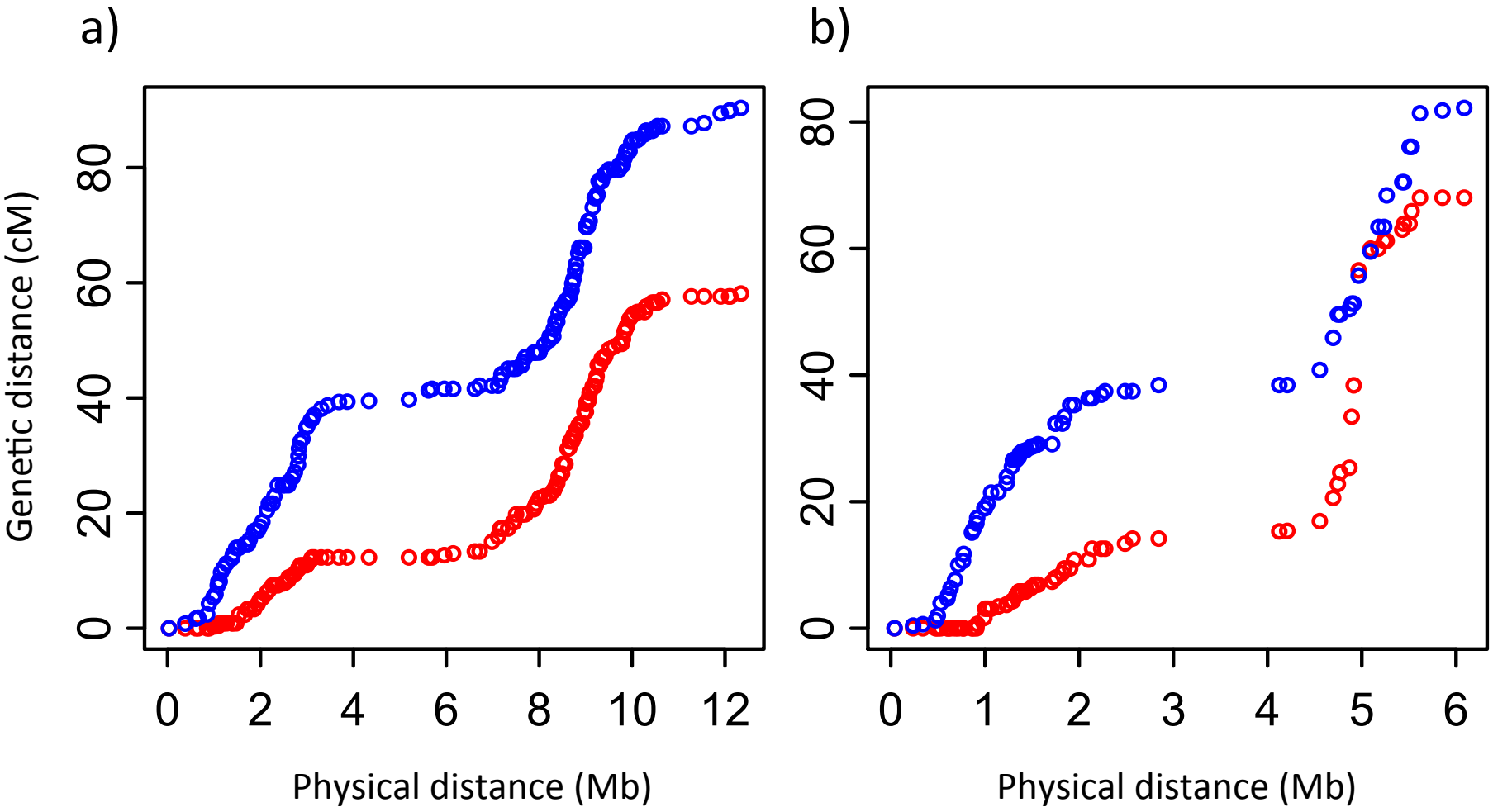

Genetic distance (cM)
350
300
250
200
150
100
50
0
0
50
100
150
200
Chromosome size (Mb)

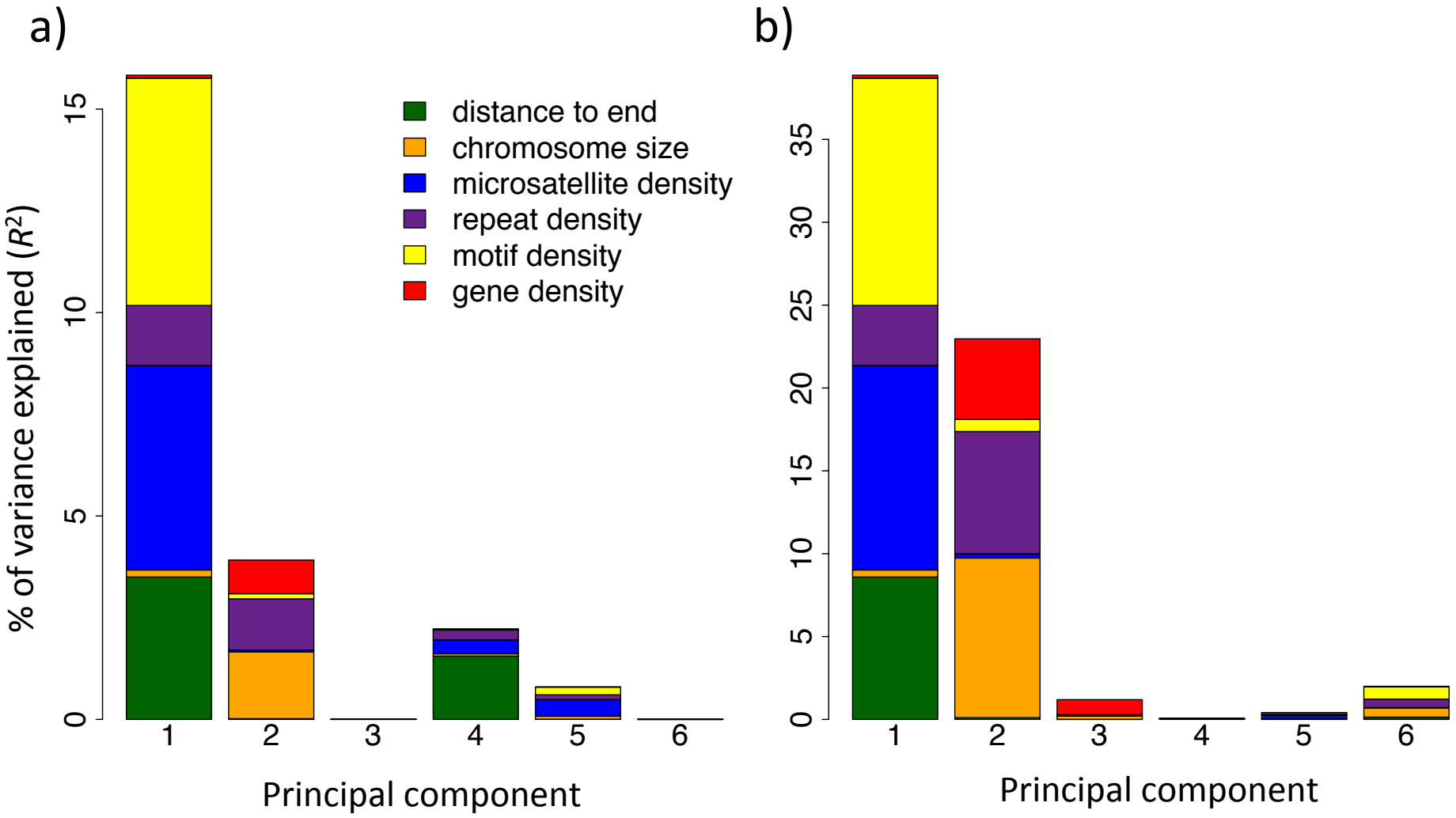
b)

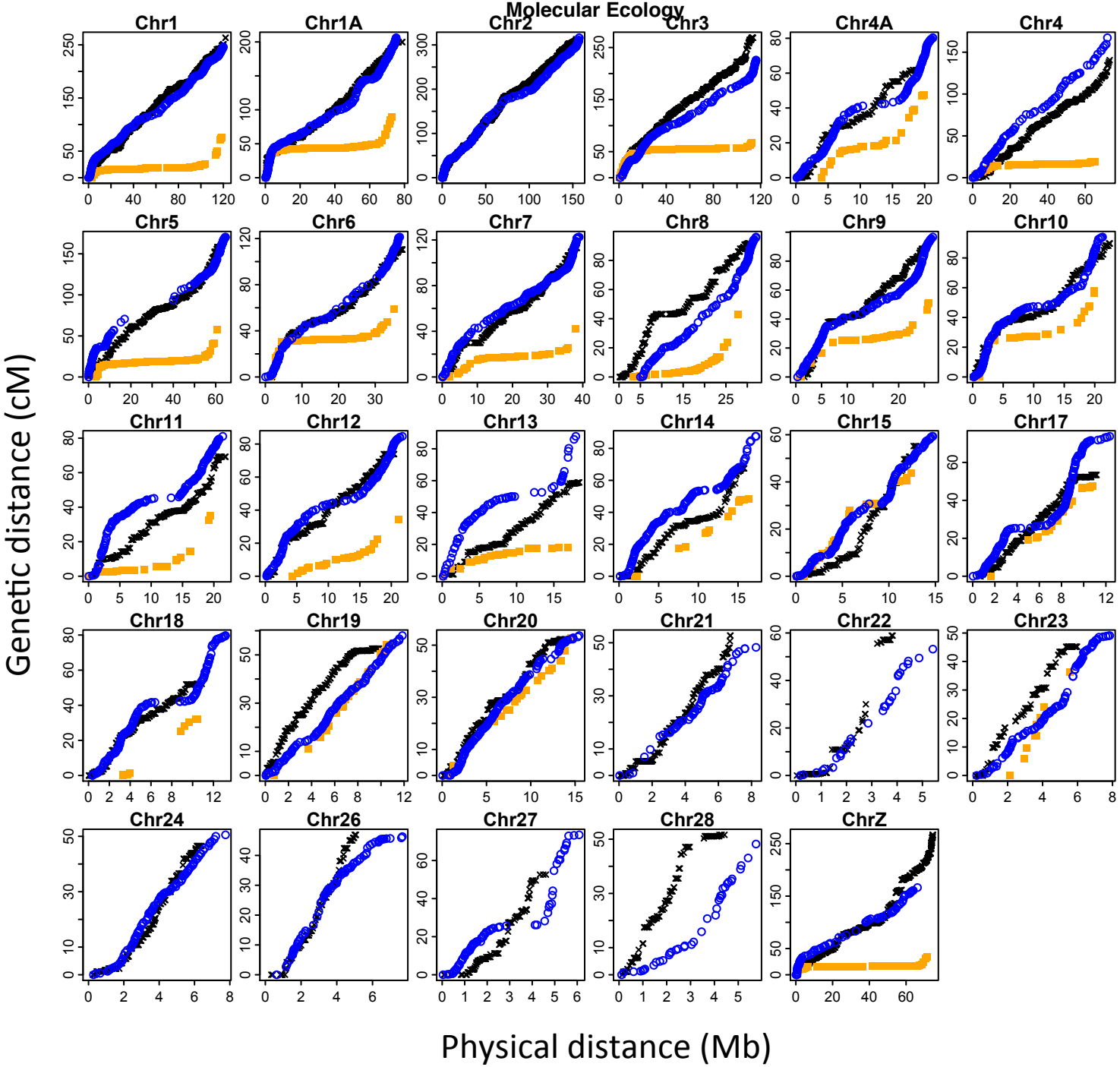

Rec rate (cM/Mb)
25
20
15
10
5
0
0
50
100
150
200
Chromosome size (Mb)


a)
Rec rate (cM/Mb)
0
10
20
30
40
50
60
0
20
40
60
80
Distance from chromosome end (Mb)


b)
0
10
20
30
40
50
60
0
2
4
6
8
10
Distance from chromosome end (Mb)


a)
b)
Genetic distance (cM)
0
20
40
60
80
0
2
4
6
8
10
12
Physical distance (Mb)
0
1
2
3
4
5
6
Physical distance (Mb)


a)
b)
distance to end
chromosome size
microsatellite density
repeat density
motif density
gene density
% of variance explained (R^2)
0
5
10
15
20
25
30
35
1
2
3
4
5
6
Principal component
Principal component


Chr1
Chr1A
Chr2
Chr3
Chr4A
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
Chr14
Chr15
Chr17
Chr18
Chr19
Chr20
Chr21
Chr22
Chr23
Chr24
Chr26
Chr27
Chr28
ChrZ
Genetic distance (cM)
Physical distance (Mb)