Rapid regulatory evolution of a nonrecombining autosome linked to divergent behavioral phenotypes

Dan Suna, Iksoo Huhb, Wendy M. Zinzow-Kramera, Donna L. Maneyb,1 and Soojin V. Yia

aSchool of Biological Sciences, Institute for Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, GA 30332; and bDepartment of Psychology, Emory University, Atlanta, GA 30322

Edited by Hopi E. Hoekstra, Harvard University, Cambridge, Massachusetts, and approved February 2, 2018 (received for review October 10, 2017)

In the white-throated sparrow (Zonotrichia albicollis), the second chromosome bears a striking resemblance to sex chromosomes. First, within each breeding pair of birds, one bird is homozgyous for the standard arrangement of the chromosome (ZAL2/ZAL2) and its mate is heterozygous for a different version (ZAL2/ZAL2m). Second, recombination is profoundly suppressed between the two versions, leading to genetic differentiation between them. Third, the ZAL2m version is linked with phenotypic traits, such as bright plumage, high aggression, and low parental behavior, which are usually associated with males. These similarities to sex chromosomes suggest that the evolutionary mechanisms that shape sex chromosomes, in particular, genetic degeneration of the heterogametic version due to the suppression of recombination, are likely important in this system as well. Here, we investigated patterns of protein sequence evolution and gene expression evolution between the ZAL2 and ZAL2m chromosomes by whole-genome sequencing and transcriptome analyses. Patterns of protein evolution exhibited only weak signals of genetic degeneration, and few genes harbored signatures of positive selection. We found substantial evidence of transcriptome evolution, such as significant expression divergence between ZAL2 and ZAL2m alleles and signatures of dosage compensation for highly expressed genes. These results suggest that, early in the evolution of heteromorphic chromosomes, gene expression divergence and dosage compensation can prevail before large-scale genetic degeneration. Our results show further that suppression of recombination between heteromorphic chromosomes can lead to the evolution of alternative (sex-like) behavioral phenotypes before substantial genetic degeneration.

The white-throated sparrow (Zonotrichia albicollis) occurs in two alternative morphs, white-striped and tan-striped (hereafter referred to as white and tan) (Fig. 1A), that differ in both plumage and social behavior (1–5). The morphs are completely linked to a large rearrangement on the second chromosome, called ZAL2m (6). White birds are heterozygous for ZAL2m, whereas tan birds are homozygous for a different version of the chromosome, ZAL2, which shares little genetic exchange with ZAL2m. Intriguingly, nearly all matings in the white-throated sparrow occur between birds of different morphs (7, 8). Because of this disassortative mating, ZAL2m is rarely homozygous and is, like the W and Y chromosomes in diverse species (9), excluded from homologous recombination. A group of closely linked loci that control a polymorphic phenotype, such as those found on sex chromosomes, is sometimes referred to as a “supergene” (10–16). The ZAL2/ZAL2m system represents an intriguing example of a non-sex-linked chromosome functioning as a supergene responsible for phenotypic divergence.

Evolutionary analyses have established that the ZAL2m rearrangement arose at most 2–3 million years ago (7, 17, 18). Indeed, putatively neutral sites in the two versions have differentiated by only 1% over that time period (17, 18). Thus, this system affords a rare look at the early evolution of heteromorphic chromosomes. Previously, Thomas et al. (18) utilized both population genetic and cytogenetic mapping approaches to investigate this system and revealed an ∼100-Mb rearrangement consisting of at least two inversions. Subsequent studies have found strong linkage disequilibrium as well as reduced genetic diversity within the rearranged ZAL2m region compared with ZAL2 counterparts for dozens of loci (19, 20), which is consistent with suppressed recombination within the rearranged region. Interestingly, despite the evidence of reduced recombination, Davis et al.’s (20) study of 62 genes did not find any pseudogenization or truncation of proteins in the rearranged region of ZAL2m. Recently, Tuttle et al. (7) characterized the divergence between ZAL2 and ZAL2m chromosomes by low-coverage pooled sequencing of heterozygous individuals and reported slightly reduced levels of genetic diversity and an excess of nonsynonymous polymorphisms on the ZAL2m. However, reliably distinguishing ZAL2 and ZAL2m alleles in heterozygous individuals was challenging, which limited the interpretation of the results (7).

In this study, we overcame the challenges associated with distinguishing ZAL2m and ZAL2 allelic differences in heterozygous individuals (7, 18) and offer a comprehensive genome-wide comparison of the two chromosomes. We took advantage of a fortuitous finding of a rare individual homozygous for the ZAL2m chromosome (21). This “superwhite” female, a product of an extremely rare mating between white morphs, exhibited brighter plumage and higher aggression compared with both white and tan birds (21), consistent with the role of the ZAL2m chromosome in alternative phenotypes.

Significance

The evolution of nonrecombining chromosomes such as sex chromosomes involves degeneration leading to loss of genetic information. We do not know, however, what happens during the incipient stages of such chromosomes, before appreciable degeneration. We studied this process in white-throated sparrows, a species that occurs in two alternative behavioral phenotypes determined by a nonrecombining autosomal rearrangement. We report that this rearrangement shows few signs of large-scale genetic degeneration. Instead, substantial changes have evolved at the level of gene expression, some of them consistent with adaptive evolution. Our work with this chromosome reveals that rapid changes in gene expression and dosage compensation, not necessarily large-scale genetic degeneration, characterize the early evolution of het- eromorphic chromosomes and the associated divergent phenotypes.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

Data deposition: The sequence reported in this paper has been deposited in the Sequence Read Archive database (accession no. SRP191732).

1To whom correspondence may be addressed. Email: dmaney@emory.edu or soojinyyi@gatech.edu.

This article contains information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1717721115/-/DCSupplemental.

Published online February 26, 2018.
in this species. We sequenced the genome of this rare bird to compare the ZAL2 sequence with the ZAL2 m sequence. Moreover, we extended our genome comparison with study expression divergence by integrating deep-sequencing data from the superwhite female with brain transcriptome data from behaviorally characterized tan and white birds (22). We then proceeded to look for evidence of genetic degeneration, which would be predicted for nonrecombining chromosomes (23–26), as well as evidence of divergence at the level of gene expression. Interestingly, comparison of protein-coding sequences between ZAL2 and ZAL2 m suggests that a large-scale degeneration is yet to occur in ZAL2 m. On the other hand, gene expression divergence between ZAL2 and ZAL2 m is extensive.

Results
Genomic Differentiation of ZAL2 and ZAL2 m. We generated ~35X coverage sequence data from a rare superwhite female homozygous for ZAL2 m as described by Horton et al. (21) [Sequence Read Archive (SRA) accession no. SRR4191732]. We mapped these data to the existing Z. albicollis reference assembly from a tan individual (ZAL2/ ZAL2) (7) and identified scaffolds that are likely to reside on the ZAL2 m chromosomes on the basis of previous fluorescent in situ hybridization (FISH) studies (7, 18, 20) and interchromosomal homology in birds (27–30) (SI Appendix). Divergence measures between ZAL2 and ZAL2 m alleles were computed using the zebra finch (Taeniopygia guttata) (29) as an outgroup. We were able to distinguish ZAL2 scaffolds that reside within versus outside of the rearrangement based upon the number of nucleotide substitutions per site (dXY) between ZAL2 and ZAL2 m, population differentiation (FST) between the tan and white morphs, and the presence of putatively fixed differences between ZAL2 and ZAL2 m in the 21 white-throated sparrows used in this study (Materials and Methods and SI Appendix).

We identified 35 (105.96 Mb) ZAL2 scaffolds within the rearrangement and five ZAL2 scaffolds (5.38 Mb) outside the rearrangement (SI Appendix, Table S1). The average dXY between ZAL2 and ZAL2 m inside the rearrangement was 1.202 ± 0.006% (Fig. 1 B and C). The variability of regions outside the rearrangement was statistically indistinguishable from that of the genomic background (Fig. 1 C), reflecting the level of intraspecies polymorphism. We did not observe clear clustering of nucleotide divergence, or so-called “evolutionary strata” (31), within the rearranged region. Population differentiation between the morphs was reflected in an average FST of 0.381 ± 0.006 within the rearranged region (Fig. 1 D). Regions outside the rearrangement, as well as in the rest of the genome, showed little differentiation, as expected (Fig. 1 D). The rearranged regions in ZAL2 and ZAL2 m also differed with respect to insertion-deletion (indel) frequencies; regions inside the rearrangement exhibited significantly more indels compared with regions outside or in the rest of the genome (Fig. 1E). These results confirm the high degree of genetic differentiation between the rearranged ZAL2 m region and ZAL2

Fig. 1. Genetic divergence between the ZAL2 and ZAL2 m chromosomes in the white-throated sparrow. (A) Two plumage morphs in the white-throated sparrow, white and tan, are completely linked to the presence or absence of a rearrangement in the second pair of chromosomes (karyotypes adapted from ref. 6, with permission from the Society for the Study of Evolution). (B) Nucleotide divergence measures per 500-kb windows between ZAL2 and ZAL2 m are shown for scaffolds longer than 1 Mbp. A red arrow indicates the inferred major breakpoint (18). Scaffolds residing outside the rearrangement (n = 3, orange background) show little divergence between the two chromosomes. Divergence is elevated in regions within the rearrangement. Scaffolds with known, precise locations (n = 13) are shown on a light green background. Additional scaffolds (n = 7) that reside within the rearrangement are shown on a dark green background. (C–E) Pairwise nucleotide divergences (dXY), degrees of population differentiation (FST), and indel frequencies, all measured in 10-kb nonoverlapping windows, were significantly higher in scaffolds within the rearrangement than in those outside it (Mann-Whitney U test, ***p < 0.001; NS, not significant). Including all scaffolds that are putatively on the second chromosome in the calculation yielded highly similar results (SI Appendix, Fig. S1).

Weak Signatures of Degeneration and Positive Selection in Protein-Sequencing Coding Sequences. We examined protein-coding sequences for signatures of genetic degeneration, such as nonsense mutations and frame-shift mutations. Of 1,007 protein-coding loci linked to the rearranged ZAL2 m, only 28 contained premature stop codons or had lost start or stop codons (SI Appendix, Table S2). We found no frame-shift mutations or deletions of genes from the rearranged portion of ZAL2 m (SI Appendix).

Several genome-scale features of the ZAL2 m-linked loci were consistent with a subtle level of genetic degeneration. Rates of nonsynonymous substitution (dN), radical amino acid substitution (dK), and promoter (defined as 1 kb upstream from transcription start site) substitution (dP), the ratio of radical-to-conservative amino acid substitution rates (dN/dC), as well as the ratio of nonsynonymous-to-synonymous substitution rates (dN/dS), were all elevated on ZAL2 m compared with ZAL2 (Fig. 2 A). These results could be explained by the accumulation of slightly deleterious mutations on the ZAL2 m chromosome. On the other hand, rates of synonymous substitution (dS) and whole-genome divergence (dXY) did not show chromosomal bias (Fig. 2 A). The numbers of putatively fixed substitutions across nonsynonymous sites, synonymous sites, and 5′ and 3′ untranslated transcribed regions (UTRs) varied in concordance with the degree of purifying selection (Fig. 2 B). Synonymous sites showed the highest level of divergence. The level of divergence was progressively lower in 3′ UTRs, 5′ UTRs, and nonsynonymous sites. The number of fixed substitutions thus correlated inversely with the expected degree of purifying selection (e.g., ref. 32). Rates of conservative amino acid substitution (dC) were also similar between the two chromosomes, consistent with strong purifying selection (Fig. 2 A). Using a branch-site model in PAML (33) with sequences from 13 other Passeriformes and a simulation-based approach (34) (Materials and Methods), we examined signatures of positive selection in the coding sequences. We found signs of positive selection for three genes on ZAL2 and two genes on ZAL2 m by using a cutoff of false discovery rate (FDR)-adjusted Q < 0.2 (SI Appendix, Table S3). Because incomplete lineage sorting may result in discordance between gene trees and the species tree and therefore interfere with identification of positively selected genes, we constructed maximum-likelihood gene trees for the five genes and ran PAML. The
signatures of positive selection remained for all five genes after Bonferroni correction (Bonferroni-corrected P < 0.05). We also tested whether genes in specific functional categories were evolving under different selective constraints (35, 36) (Materials and Methods). Interestingly, some categories of genes showed evidence of significantly different selection pressure between the two chromosomes (SI Appendix; Table S4), for example, rRNA processing [Gene Ontology (GO): 0006364] on ZAL2 and positive regulation of neuron projection development (GO: 0010970) and the Wnt-signaling pathway (GO: 0016055) on ZAL2m.

**Differential Allelic Expression Between ZAL2 and ZAL2m.** Genetic differentiation of regulatory regions could cause divergence of gene expression even in the absence of coding sequence divergence (37). We examined gene expression divergence between ZAL2 and ZAL2m alleles using RNA-seq data of 9 tan and 10 white individuals from two brain regions implicated in social behavior in birds: the hypothalamus and nucleus taeniae (called medial amygdala in some publications) (22). These data came from breeding birds that had been confirmed to exhibit the typical morph differences in behavior. We first tested for allele-specific expression of ZAL2 and ZAL2m alleles in white birds using DESeq2 (38) (Materials and Methods and SI Appendix). We found 335 and 375 genes (41.41 and 46.53% of all ZAL2/ZAL2m-linked genes examined) with significant allele-specific expression in the hypothalamus and nucleus taeniae, respectively (FDR-adjusted Q < 0.05; Table 1). Among all of the genes with allele-specific expression, 236 were common between the two brain regions, a significant enrichment of shared allele-specific expression (hypergeometric test, P < 0.001). This supports the idea that allele-specific expression of ZAL2 and ZAL2m alleles is driven by divergence of regulatory sequences. As expected, there was a bias toward reduced ZAL2m expression (one-tailed, paired Mann–Whitney U test, hypothalamus: P = 0.066 and nucleus taeniae: P = 0.049).

**Potential Dosage Compensation of Highly Expressed Genes.** We then asked whether allele-specific expression in white birds has led to differential expression between the tan and white morphs (henceforth referred to as “morph-biased gene expression”). For example, if morph-biased expression of a given gene is caused entirely by allele-specific expression, reduced expression of the ZAL2m allele compared with the ZAL2 allele (ZAL2 > ZAL2m) could lead to lower total expression in the white birds (ZAL2/ZAL2m) relative to tan birds (ZAL2/ZAL2) (Fig. 3A and B). Selection could further favor increased expression of ZAL2 compared with the degenerated ZAL2m allele, leading to increased expression in tan birds (23, 39). Conversely, genes for which ZAL2m expression is elevated compared with ZAL2 (ZAL2m > ZAL2) could exhibit higher expression in white birds compared with tan birds (SI Appendix, Fig. S2).

Interestingly, a large number of genes that exhibit significant allele-specific expression were similarly expressed between the morphs (Table 1). We tested the hypothesis that a dosage-rebalancing mechanism may be in play for these genes. In other words, disruption of ZAL2m allelic expression could be compensated by changes in ZAL2 expression in white birds to bring total levels of expression close to those seen in tan birds (Fig. 3C). In this scenario, ZAL2 allelic expression in white birds should be higher than that in tan birds (white-ZAL2/tan-ZAL2 > 1; Fig. 3C), a result that we observed in our data (Mann–Whitney U test, P < 0.001 for both brain regions; Fig. 3D and E). Likewise, an up-regulation of ZAL2m allelic expression could be compensated by decreased ZAL2 expression in white birds (SI Appendix, Fig. S2).

Recent studies have demonstrated that dosage compensation may be limited to a subset of genes (40–43), in which highly expressed genes are particularly enriched (44–47). To determine whether dosage rebalancing is observed preferentially for highly expressed genes, we compared expression levels of candidate dosage-compensated genes (tan ≈ white and ZAL2 > ZAL2m or tan ≈ white and ZAL2m > ZAL2) with those of the background (i.e., genes that do not exhibit allele-specific or morph-biased expression patterns). For ZAL2 > ZAL2m genes, candidate dosage-compensated genes displayed significantly higher expression levels relative to the background (SI Appendix, Fig. S3). For ZAL2m > ZAL2 genes, this effect was weak and limited to the nucleus taeniae (SI Appendix, Fig. S4).

**Table 1. Patterns of morph-biased expression for genes showing allele-specific expression**

<table>
<thead>
<tr>
<th>Brain region</th>
<th>tan &gt; white</th>
<th>white &gt; tan</th>
<th>Other (tan ≈ white)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothalamus</td>
<td>ZAL2 &gt; ZAL2m</td>
<td>63 (87)</td>
<td>1 (2)</td>
</tr>
<tr>
<td></td>
<td>ZAL2m &gt; ZAL2</td>
<td>0 (0)</td>
<td>109 (146)</td>
</tr>
<tr>
<td>Nucleus taeniae</td>
<td>ZAL2 &gt; ZAL2m</td>
<td>54 (86)</td>
<td>1 (1)</td>
</tr>
<tr>
<td></td>
<td>ZAL2m &gt; ZAL2</td>
<td>0 (0)</td>
<td>99 (123)</td>
</tr>
</tbody>
</table>

Among genes that show allele-specific expression between ZAL2 and ZAL2m in white birds, some also show morph-biased expression. Above, morph-biased genes are separated into tan > white and white > tan. Genes that do not exhibit morph-biased expression are shown in the “Other” category. Differences were assessed by DESeq2, and the numbers of genes with FDR-corrected Q < 0.05 (or Q > 0.05 for the “Other” category) are shown in the table, with P < 0.05 (or P > 0.05 for the “Other” category) in parentheses.
Allele-Specific and Morph-Biased Genes Play Central Roles in Gene Coexpression Networks. We previously identified weighted coexpression networks from our RNA-seq data (22). Genes in two large modules, namely the “black” module in the hypothalamus (containing 511 genes, 226 genes on ZAL2/ZAL2m and the “green-yellow” module in nucleus taeniae (containing 157 genes, 115 genes on ZAL2/ZAL2m), were significantly correlated with territorial singing, a behavior that differs between the morphs (22). In other words, expression of these genes was morph-biased. We hypothesized that genes that are allele-specific and morph-biased, compared with genes that are only morph-biased, play more central roles in the gene coexpression networks. Consistent with our hypothesis, genes with allele-specific expression exhibited significantly higher intramodule connectivities compared with other genes in the modules (Fig. 4).

Discussion
In sex chromosomal supernumerary, the heterogametic chromosome (e.g., the Y or W chromosome) is prevented from recombining because it is transmitted nearly always via heterogametocytes. Lack of recombination reduces the effective population size of the heterogametic chromosome, leading to its degeneration (24, 26). Here, we have investigated the early evolution of a chromosomal system with striking similarity to sex chromosomes. In white-throated sparrows, the ZAL2m is almost always heterozygous due to strong disassortative mating (6, 7, 21, 49, 49). Consequently, it has been hypothesized that the ZAL2m chromosome may undergo genetic degeneration (7, 18). Previous analyses have been inconclusive in this regard. Karyotyping studies did not identify any signs of heterochromatinization in the ZAL2m chromosome (6, 18). Population genetic analyses have revealed slightly reduced levels of genetic diversity within ZAL2m-linked loci (7, 17, 19). A recent genome-wide study further reported a weak but significant excess of nonsynonymous polymorphisms on ZAL2m (7). However, other signs of degeneration, such as pseudogenization and accumulation of repetitive sequences, have not been observed (20). Our genome-wide comparison of protein-coding sequences between ZAL2 and ZAL2m shows that the degree of degeneration of ZAL2m is weak at most, consistent with a young supergene system. Despite the pronounced phenotypic differentiation associated with ZAL2m, we did not observe a massive accumulation of pseudogenes or evidence of a substantial increase in repetitive sequences on this chromosome (SI Appendix).

The low level of degeneration is probably due to the infrequent yet present recombination between ZAL2m chromosomes in rare homozygotes. Unlike other non-sex-linked supernumeraries in fire ants (Solenopsis invicta) and ruffs (Philomachus pugnax) (10–12, 50), the ZAL2m is not homozygous-lethal. Healthy adult ZAL2m/ZAL2m homozygotes of both sexes have been observed in nature (6, 21, 49), indicating that ZAL2m is functional. The rarity of ZAL2m/ZAL2m homozygotes can be explained by the disassortative mating system. According to Tuttle et al. (7), 0.8% of total matings occur between white birds; thus, the expected frequency of ZAL2m/ZAL2m homozygotes is ~0.2% of total offspring. Based upon the five reports of ZAL2m/ZAL2m homozygotes (6, 7, 21, 48, 49), the frequencies of such individuals range between 0 and 1%, and the frequency based on all data is 0.2% (6 of 3,057 birds genotyped to date), exactly the same as the expected frequency. The more pressing question, then, is why and how the disassortative mating has evolved in this species. It is possible that, because the two morphs complement each other with respect to territorial aggression and parenting, same-morph pairs are not as successful (1).

While we did not find evidence for substantial degeneration in protein-coding regions, there were robust signals of regulatory evolution in substantial numbers of genes. More than 40% of ZAL2/ZAL2m-linked genes exhibited allele-specific expression in white birds. One explanation for this finding could be that ZAL2m allelic expression is disrupted due to degeneration of regulatory sequences. Indeed, we observed an overall reduction of ZAL2m.

Fig. 3. Potential dosage compensation in the ZAL2m system. (A–C) A potential scenario for dosage compensation. (A) Initially, expression dosage (shown as black waves) is similar between ZAL2 and ZAL2m and between tan and white. (B) If the expression of ZAL2m allelic is reduced and there is no dosage compensation, heterozygous (white) individuals should show reduced expression. (C) Dosage between the morphs may be rebalanced via overexpression of the ZAL2 allele in white birds. Consequently, expression of the ZAL2 allele should be greater in white than in tan birds. (D and E) Levels of compensation (white-ZAL2/tan-ZAL2) are significantly elevated for tan = white (dosage-compensated) genes compared with nondosage-compensated genes (tan > white). The definitions of tan = white and tan > white are based on P values from DESeq2 (Table 1). Using FDR-corrected Q values yielded similar results (SI Appendix, Fig. 53). Mann–Whitney U test, ***P < 0.001.

Fig. 4. Allele-specific and morph-biased genes occupy more central positions in gene coexpression networks. Intramodular connectivity (kIN) for genes exhibiting allele-specific and morph-biased expression compared with genes in the rest of the module (Rest) (22) for (A) the hypothalamus and (B) the nucleus taeniae. Mann–Whitney U test, ∗P < 0.05; ∗∗P < 0.01; ∗∗∗P < 0.001; NS: P > 0.05.
allelic expression compared with ZAL2, although this trend was weak and not significant in the hypothalamus (Table 1). Importantly, a large proportion of the genes showing allele-specific expression are shared between the two brain regions that we sampled, suggesting that the allele-specific expression could be driven by genetic divergence of regulatory regions (37, 51, 52). Alternatively, the differences in allelic expression could be driven by relatively small changes in the protein sequences of important transcription factors (53).

Despite substantial allele-specific expression of some genes, many genes exhibited similar expression levels between the morphs. This result is reminiscent of dosage compensation in sex chromosomes (9, 54–56). Dosage compensation may evolve as a response to reduced allelic dose (23, 39), particularly for highly expressed genes (44–47). Dosage compensation can also be achieved without selective up- or down-regulation of specific genes through passive feedback or other buffering in regulatory networks (57–59). This compensation may potentially occur via epigenetic mechanisms such as DNA methylation (60). Interestingly, we observed that within the ZAL2\textsuperscript{m} down-regulated genes (ZAL2 > ZAL2\textsuperscript{m}), those that are dosage-compensated tend to be more highly expressed than the background genes. In addition, the proportion of genes that appear to be dosage-compensated is significantly higher for ZAL2\textsuperscript{m} down-regulated genes compared with ZAL2\textsuperscript{m} up-regulated genes (Table 1; proportional test, \(P < 0.001\) in both brain regions). These observations are consistent with the idea that degeneration of the ZAL2\textsuperscript{m} chromosome drives dosage compensation (23, 39). The fact that we observe gene expression changes resembling dosage compensation is particularly notable given that avian sex chromosomes generally exhibit weak levels of dosage compensation (61, 62).

Supergenes that are formed via chromosomal rearrangements are emerging as key determinants for alternative reproductive phenotypes in diverse taxa (e.g., fire ant, ruff, and some plants) (10–12, 15, 50). The pivotal step in the evolution of a supergene is the suppression of recombination (13, 14, 63). The best-understood consequence of suppression of recombination is genetic degeneration of the nonrecombining region (24, 64). Previous studies of white-throated sparrows, such as the recent work by Tuttle et al. (7), have shown evidence consistent with degeneration. Those signals were weak, however, which was viewed at odds with the pervasive phenotypic differentiation between the ZAL2\textsuperscript{m} and ZAL2\textsuperscript{m} chromosomes, which are emerging as key determinants for alternative reproductive phenotypes in diverse taxa (e.g., fire ant, ruff, and some plants) (10–12, 15, 50).

Gene Expression. We sequenced a “superwhite” bird homozygous for the rearrangement (ZAL2/ZAL2\textsuperscript{m}) (21). High-molecular-weight genomic DNA was extracted from liver and sequenced using HiSeq2500 at the Roy Carver Genome Center of the University of Illinois. Approximately 240 million reads of 150 bp were generated, which are available from the SRA database (accession no. SRX4191732).

Identification of Scaffolds on the Second Chromosome. We mapped scaffolds from the genome of a tan bird in National Center for Biotechnology Information (NCBI) onto that of the zebra finch using LASTZ 1.03.73 (68) (more details in SI Appendix). We identified ZAL2 scaffolds using homology to the corresponding zebra finch chromosome (commonly referred to as TGJU due to its homology to chicken chromosome 3 [29]), previous FISH studies (7, 18, 20), as well as homology with two other passerine birds with chromosome-level assemblies, the collared flycatcher (Ficedula albicollis) and the great tit (Parus major) (28, 30).

Genetic Divergence Between ZAL2 and ZAL2\textsuperscript{m} Alleles. To measure the divergence between the ZAL2 and ZAL2\textsuperscript{m} chromosomes, we called variants (SNPs and indels) by following best practices for the Genome Analysis Toolkit (GATK) for variant calling in genome-sequencing data (69–71). To calculate \(F_{st}\), we called variants with available transcriptome data from the hypothalamus and nucleus taeniae of 9 tan and 11 white individuals (22) following the GATK best practices for variant calling in RNA-seq data (69–71). A variant was considered as putatively fixed in the sampled individuals if (i) the variant was biallelic; (ii) individual was homozygous for the reference allele (AA); (iii) individual was heterozygous (Aa); and (iv) the superwhite (ZAL2/ZAL2\textsuperscript{m}) individual was homozygous for the alternative allele (aa). Scaffolds residing inside versus outside the rearrangement were identified on the basis of distinctive patterns of genetic divergence (more details in SI Appendix).

Analyses of Protein-Coding Sequences. For each ZAL2-linked gene, we extracted the longest transcript and constructed the ZAL2\textsuperscript{m} version with the putatively fixed differences. Additionally, we utilized all available genome annotations in NCBI for 13 other avian species in the order of Passeriformes (the same order as the white-throated sparrow) as outgroup species (SI Appendix, Fig. S6). We obtained codon alignments for a total of 800 genes. We calculated \(d_{s}\) and \(d_{\nu}\) for the ZAL2 and ZAL2\textsuperscript{m} branches with a free-ratio model using codeml from the PAML 4.8 package (33). The Ron-New package (72), which adopts the amino acid classification system that considers charge and polarity, was used to estimate \(d_{s}\) and \(d_{\nu}\). To test for positive selection, a branch-site model was run with codeml in PAML, which was then compared with the null model following the simulation approach used by Nielsen et al. (34). Positively selected functional categories were identified by first assigning genes according to the PANTHER Classification System version 10 (73). For each category with more than 10 genes, the difference between the cumulative distributions of \(P\) for genes in the category, versus that of genes not in that category, was examined using a one-tailed Mann–Whitney U test (35).

Gene Expression. We used RNA-Seq data from 9 tan and 10 white males (22) to detect morph-biased and allele-specific gene expression. To account for potential mapping bias toward the reference (ZAL2/ZAL2) genome caused by mismatches between ZAL2 and ZAL2\textsuperscript{m}, we N-masked (74) putatively fixed differences in the reference. We additionally checked potential heterozygous bias by aligning whole-genome sequences of three white birds described by Tuttle et al. (7) to this N-masked genome. ZAL2 and ZAL2\textsuperscript{m} alleles exhibited roughly equal coverage per site, as expected if there is little mapping bias (SI Appendix, Fig. S7). Read counts per gene at the morph or allele level were calculated by htseq-count 0.9.1 (75). Morph-biased and allele-specific expression analyses were performed using DESeq2 1.12.3 (38) (more details in SI Appendix).

ACKNOWLEDGMENTS. We thank the Roy Carver Genome Center of the University of Illinois for help with sequencing; Dr. Kathleen Grogan, members of the S.V.Y. laboratory, Brian Charlesworth, and two anonymous reviewers for comments on the manuscript; and Xiulan Pan and Chen Feng for help with illustration. This work was supported by National Institutes of Health Grants R01 MH082833 (to D.L.M.) and 1R21 NIMH102677 (to D.L.M. and S.V.Y.) and by Georgia Tech (S.V.Y.).

Materials and Methods

Detailed descriptions of materials and methods are provided in SI Appendix.


