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

Dan Sun, Georgia Institute of Technology
Iksoo Huh, Georgia Institute of Technology
Wendy M. Zinzow-Kramer, Emory University
Donna Maney, Emory University
Soojin V. Yi, Georgia Institute of Technology

Journal Title: Proceedings of the National Academy of Sciences
Volume: Volume 115, Number 11
Publisher: National Academy of Sciences | 2018-01-01, Pages 2794-2799
Type of Work: Article | Final Publisher PDF
Publisher DOI: 10.1073/pnas.1717721115
Permanent URL: https://pid.emory.edu/ark:/25593/s93c7

Final published version: http://dx.doi.org/10.1073/pnas.1717721115

Copyright information:
© 2018 National Academy of Sciences. All Rights Reserved.
This is an Open Access work distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Accessed April 28, 2019 11:14 AM EDT
Rapid regulatory evolution of a nonrecombining autosome linked to divergent behavioral phenotypes

Dan Sun*, Iksoo Huh*, Wendy M. Zinzow-Kramerb, Donna L. Maneyb,1, and Soojin V. YPa,1

*School 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 homogamous 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 hetero- genetic 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 specific 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.

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 heteromorphic 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. SRX6191732).

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

This article contains supporting 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 ZAL2m 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 ZAL2m suggests that a large-scale degeneration is yet to occur in ZAL2m. On the other hand, gene expression divergence between ZAL2 and ZAL2m is extensive.

Results

Genomic Differentiation of ZAL2 and ZAL2m. We generated ~35x coverage sequence data from a rare superwhite female homozygous for ZAL2m 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 ZAL2m 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 ZAL2m 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 ZAL2m, population differentiation (Fst) between the tan and white morphs, and the presence of putatively fixed differences between ZAL2 and ZAL2m 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 ZAL2m inside the rearrangement was 1.202 ± 0.0065% (Fig. 1 B and C). The variability of regions outside the rearrangement was statistically indistinguishable from that of the genomic background (Fig. 1C), 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. 1D). Regions outside the rearrangement, as well as in the rest of the genome, showed little differentiation, as expected (Fig. 1D). The rearranged regions in ZAL2 and ZAL2m 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 ZAL2m region and ZAL2.

Weak Signatures of Degeneration and Positive Selection in Protein-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 ZAL2m, 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 ZAL2m (SI Appendix).

Several genome-scale features of the ZAL2m-linked loci were consistent with a subtle level of genetic degeneration. Rates of nonsynonymous substitution (ds), radical amino acid substitution (dka), and promoter (defined as 1 kb upstream from transcription start site) substitution (dps), the ratio of radical-to-conservative amino acid substitution rates (dka/dc), as well as the ratio of nonsynonymous-to-synonymous substitution rates (dka/ds), were all elevated on ZAL2m compared with ZAL2 (Fig. 2A). These results could be explained by the accumulation of slightly deleterious mutations on the ZAL2m chromosome. On the other hand, rates of synonymous substitution (ds) and whole-genome divergence (dxy) did not show chromosomal bias (Fig. 2B). 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. 2B). 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 (dca) were also similar between the two chromosomes, consistent with strong purifying selection (Fig. 2A).

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 ZAL2m 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 reran 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, RNA processing [Gene Ontology (GO): 0006364] on ZAL2 and positive regulation of neuron projection development (GO: 0010976) 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>63 (87)</td>
<td>1 (2)</td>
<td>111 (118)</td>
</tr>
<tr>
<td>ZAL2 &gt; ZAL2m</td>
<td>0 (0)</td>
<td>109 (146)</td>
<td>51 (40)</td>
</tr>
<tr>
<td>ZAL2m &gt; ZAL2</td>
<td>54 (86)</td>
<td>1 (1)</td>
<td>153 (136)</td>
</tr>
<tr>
<td>Nucleus taeniae</td>
<td>0 (0)</td>
<td>99 (123)</td>
<td>68 (65)</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 supernumeraries, the heterogametic chromosome (e.g., the Y or W chromosome) is prevented from recombining because it is transmitted nearly always via heterozygotes. 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, 48, 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 superfine 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 expression 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.
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 dosage (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 down-regulated genes (ZAL2 > ZAL2w), 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 down-regulated genes compared with ZAL2w up-regulated genes (Table 1; proportional test, P < 0.001, across both brain regions). These observations are consistent with the idea that degeneration of the ZAL2 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 and ZAL2w chromosomes.

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/ZAL2w) genome caused by mismatches between ZAL2 and ZAL2w, we N-masked (74) putatively fixed differences in the reference. We additionally checked potential leftover bias by aligning whole-genome sequences of three white birds described by Tuttle et al. (7) to this N-masked genome. ZAL2 and ZAL2w 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.).