

RESEARCH ARTICLES

NONHUMAN GENETICS

Stable recombination hotspots in birds

Sonal Singhal,^{1,2*†} Ellen M. Leffler,^{3,4*} Keerthi Sannareddy,³ Isaac Turner,⁴ Oliver Venn,⁴ Daniel M. Hooper,⁵ Alva I. Strand,¹ Qiye Li,⁶ Brian Raney,⁷ Christopher N. Balakrishnan,⁸ Simon C. Griffith,⁹ Gil McVean,⁴ Molly Przeworski^{1,2,†}

The DNA-binding protein PRDM9 has a critical role in specifying meiotic recombination hotspots in mice and apes, but it appears to be absent from other vertebrate species, including birds. To study the evolution and determinants of recombination in species lacking the gene that encodes PRDM9, we inferred fine-scale genetic maps from population resequencing data for two bird species: the zebra finch, *Taeniopygia guttata*, and the long-tailed finch, *Poephila acuticauda*. We found that both species have recombination hotspots, which are enriched near functional genomic elements. Unlike in mice and apes, most hotspots are shared between the two species, and their conservation seems to extend over tens of millions of years. These observations suggest that in the absence of PRDM9, recombination targets functional features that both enable access to the genome and constrain its evolution.

Meiotic recombination is a ubiquitous and fundamental genetic process that shapes variation in populations, yet our understanding of its underlying mechanisms is based on a handful of model organisms, scattered throughout the tree of life. One pattern shared among most sexually reproducing species is that meiotic recombination tends to occur in short segments of hundreds to thousands of base pairs, termed “recombination hotspots” (1). In apes and mice, the location of hotspots is largely determined by PRDM9, a zinc-finger protein that binds to specific motifs in the genome during meiotic prophase and generates histone H3 lysine 4 trimethylation (H3K4me3) marks, eventually leading to double-strand breaks (DSBs) and both crossover and noncrossover resolutions (2–5). In mammals, the zinc-finger domain of the gene *PRDM9* evolves quickly, with evidence of positive selection on residues in contact with DNA (2, 6); as a result, there is rapid turnover of hotspot locations across populations, subspecies, and species (7–10).

Although PRDM9 plays a pivotal role in controlling recombination localization in mice and apes, many species lacking *PRDM9* nonetheless have hotspots (6). An artificial example is provided by *Prdm9* knockout mice. Despite being sterile, they make similar numbers of DSBs as wild-type mice make, and their recombination

hotspots appear to default to residual H3K4me3 mark locations, notably at promoters (10). A natural but puzzling example is provided by canids, which carry premature stop codons in *PRDM9* yet are able to recombine and remain fertile (11, 12). As with *Prdm9* knockout mice, in dogs and in other species without *PRDM9*—such as the yeast *Saccharomyces cerevisiae* and the plant *Arabidopsis thaliana*—hotspots tend to occur at promoters or other regions with promoter-like features (11, 13, 14). In yet other taxa without *PRDM9*, including *Drosophila* species (15), honeybees (16), and *Caenorhabditis elegans* (17), short intense recombination hotspots appear to be absent altogether.

To further explore how the absence of *PRDM9* shapes the fine-scale recombination landscape and influences its evolution, we turned to birds, because an analysis of the chicken genome suggested that it may not have *PRDM9* (6). We first confirmed the absence of *PRDM9* across reptiles by querying the genomes of 48 species of birds, three species of crocodylians, two species of turtles, and one species of lizard for *PRDM9* (18), finding that only the turtle genomes contain putative orthologs with all three *PRDM9* domains (fig. S1). We also found no expression of any *PRDM9*-like transcripts in RNA sequencing data from testis tissue of the zebra finch (*Taeniopygia guttata*) (18). Given the likely absence of *PRDM9* in birds, we asked: Is recombination nonetheless concentrated in hotspots in these species? If so, how quickly do the hotspots evolve? Where does recombination tend to occur in the genome? To address these questions, we generated whole-genome resequencing data for wild populations

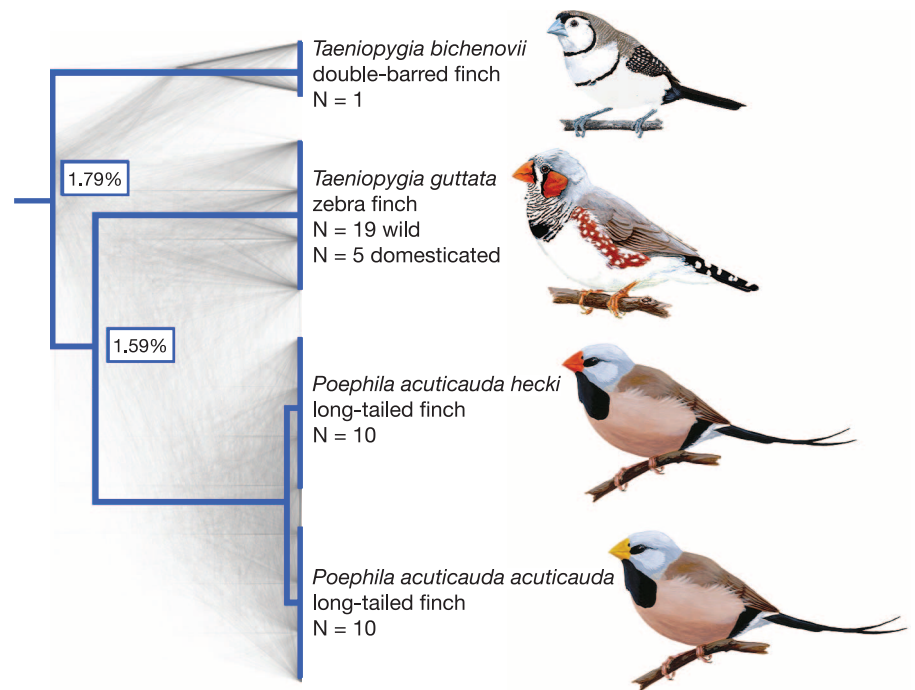


Fig. 1. Species tree for the finch species in this study. Species sampled were double-barred finch, zebra finch, and the two long-tailed finch subspecies. The tree was rooted with the medium ground finch and collared flycatcher (full phylogeny is shown in Fig. 4). Shown in gray are 1000 gene trees, which were used to infer the species tree (18). The pairwise divergence between species is indicated at nodes, as measured by the genome-wide average across autosomes.

¹Department of Biological Sciences, Columbia University, New York, NY 10027, USA. ²Department of Systems Biology, Columbia University, New York, NY 10032, USA. ³Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA. ⁴Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. ⁵Committee on Evolutionary Biology, University of Chicago, Chicago, IL 60637, USA. ⁶China National Genebank, BGI-Shenzhen, Shenzhen 518083, China. ⁷Center for Biomolecular Science and Engineering, University of California–Santa Cruz, Santa Cruz, CA 95064, USA. ⁸Department of Biology, East Carolina University, Greenville, NC 27858, USA. ⁹Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia.

*These authors contributed equally to this work. †Corresponding author. E-mail: sonal.singhal@gmail.com (S.S.); molly.przeworski@gmail.com (M.P.)

of two bird species and inferred fine-scale genetic maps from patterns of linkage disequilibrium.

Inferring fine-scale recombination maps

We sampled three species of finch in the family Estrildidae: zebra finch (*Taeniopygia guttata*; $n = 19$ wild unrelated birds and $n = 5$ birds from a domesticated nuclear family); long-tailed finch (*Poephila acuticauda*; $n = 20$, including 10 of each of two similar subspecies with average autosomal $F_{ST} = 0.039$); and, for use as an outgroup, double-barred finch (*Taeniopygia bichenovii*; $n = 1$) [Fig. 1 and table S1 (18)]. Despite extensive incomplete lineage sorting between the species, they do not appear to have diverged with gene flow (fig. S2). Moreover, nucleotide divergence among the three finch species is similar to that among humans, chimpanzees, and gorillas, providing a well-matched comparison to apes (Fig. 1) (8, 9).

We mapped reads from all individuals to the zebra finch reference genome [1 Gb assembled across 34 chromosomes (19)] and generated de novo single-nucleotide polymorphism (SNP) calls for all three species. After filtering for quality, we identified 44.6 million SNPs in the zebra finch, 26.2 million SNPs in the long-tailed finch, and 3.0 million SNPs in the double-barred finch (table S2). These SNP numbers correspond to autosomal nucleotide diversities of $\pi = 0.82\%$ and $\theta_w = 1.37\%$ in the zebra finch and $\pi = 0.55\%$ and $\theta_w = 0.73\%$ in the long-tailed finch, about 10 times higher than estimates in apes (20). Assuming a mutation rate per base pair per generation of 7×10^{-10} (18), these diversity levels suggest a long-term effective population size (N_e) of 4.8×10^6 and 2.5×10^6 for the zebra finch and long-tailed finch, respectively. Thus, these two species have much larger N_e than most other species for which there exist fine-scale recombination maps, with N_e being more reflective of biodiversity at large (fig. S3).

Next, we inferred haplotypes for the zebra finch and long-tailed finch, using a linkage-disequilibrium approach that incorporated phase-informative reads and family phasing. From the haplotypes, we estimated fine-scale recombination maps using the program LDhelmet, which works well for species with higher nucleotide diversity (15). The resulting maps estimated median recombination rates in the zebra finch and long-tailed finch genomes as $\hat{\rho} = 26.2/\text{kb}$ and $14.0/\text{kb}$, respectively, which corresponds to a median rate of 0.14 centimorgans (cM)/Mb in both species (18). Simulations indicated that we had limited power to identify hotspots in regions with high recombination rates (fig. S4), so we restricted our analyses to the 18 largest chromosomes in the reference genome (930 Mb; 91% of the assembled genome). For these 18 chromosomes, our results accord well with recombination maps inferred from a more limited pedigree-based study of zebra finch (21), with a correlation of 0.90 for rates estimated at the 5-Mb scale (fig. S5), providing confidence in our rate inferences.

Hotspots and their evolution

To identify hotspots in the genome, we operationally defined them as regions that are at least

2 kb in length; have at least five times the background recombination rate as estimated across the 80 kb of sequence surrounding the region; and are statistically supported as hotspots by a likelihood ratio test (18). This approach yielded 3949 hotspots in the zebra finch genome and 4933 hotspots in the long-tailed finch genome (Fig. 2 and figs. S6 and S7), with one hotspot detected on average every 215 and 179 kb in the two species, respectively. Both the lower density of hotspots in the zebra finch relative to the long-tailed finch and the lower density of hotspots in the finches relative to humans are consistent with simulations that indicate decreased power to detect hotspots when the background population recombination rate is higher [figs. S4 and S8 (18)]. The hotspots were detected after aggressively filtering our SNP data sets and show no evidence of having higher phasing error rates than the rest of the genome (fig. S9 and tables S3 and S4).

Considering hotspots to be shared if their midpoints occur within 3 kb of each other, 73% of

zebra finch hotspots (2874 of 3949 hotspots) were detected as shared between the two species (fig. S10) when only 4.4% were expected to overlap by chance (figs. S10 and S11); similar results were obtained under different criteria for hotspot sharing (table S5). The true fraction of shared hotspots between the zebra finch and long-tailed finch is probably higher than observed, because we do not have complete statistical power (fig. S4) and because simulations suggest that we are unlikely to detect spurious cases of hotspot sharing (18). On the other hand, the observed levels of sharing are somewhat lower than expected, compared with a model in which all hotspots are identical in the two species (fig. S12).

This conservation of hotspots contrasts sharply with comparative analyses in apes and mice, where, even across populations with modest levels of genetic differentiation, there is no hotspot sharing (8–10). In fact, if we apply the same criterion for hotspot sharing to humans and chimpanzees, only 10.5% of chimpanzee hotspots overlap

Fig. 2. Recombination rates across hotspots in zebra finch and long-tailed finch genomes.

Average relative recombination rate ($\hat{\rho}$ /bp divided by the background $\hat{\rho}$ of 20 kb on either side of the hotspot) across (A) hotspots detected only in the zebra finch genome ($n = 1075$), (B) those detected only in the long-tailed finch genome ($n = 2059$), and (C) those inferred as shared in the two species ($n = 2874$). Shared hotspots are those whose midpoints occur within 3 kb of each other. Recombination rates in the zebra finch are shown in blue, and those in the long-tailed finch are shown in red. The orientation of hotspots is with respect to the genomic sequence.

Fig. 3. Equilibrium GC content (GC*) and broad-scale recombination rates in zebra finch and long-tailed finch genomes.

(A and B) Relationship between GC* (18) and $\hat{\rho}$ /bp for the zebra finch (A) and long-tailed finch (B) across all autosomal chromosomes. Both GC* and $\hat{\rho}$ were calculated across 50-kb windows with local regression curves shown for a span of 0.2. (C and D) GC* and PAR for the zebra finch (C) and long-tailed finch (D). The histogram shows GC* for chromosome Z across 500-kb windows; GC* for the 450-kb PAR is shown by the vertical line.

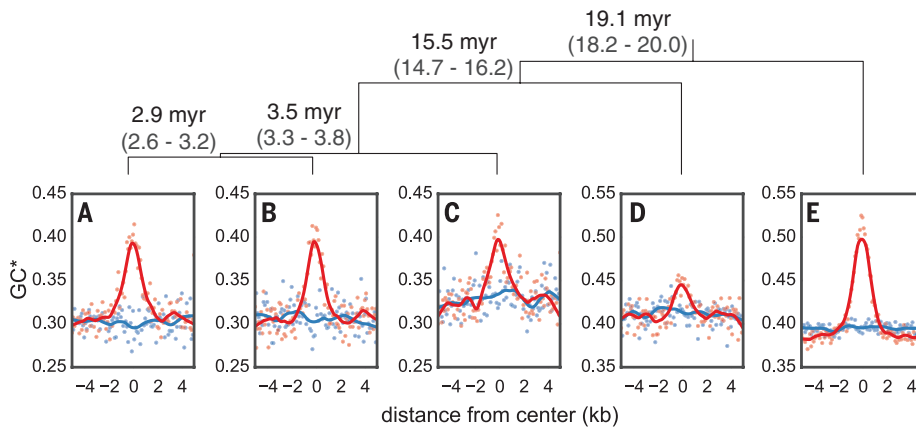


Fig. 4. Expected GC* around hotspots and matched coldspots for five bird species. Points (hotspots in red and coldspots in blue) represent GC* estimated from the lineage-specific substitutions aggregated in 100-bp bins from the center of all hotspots in (A) zebra finch and (B) long-tailed finch. GC* for (C) the double-barred finch, (D) the medium ground finch, and (E) the collared flycatcher was calculated around hotspots identified as shared between the zebra finch and long-tailed finch. Local regression curves are shown for a span of 0.2. The orientation of hotspots is with respect to the genomic sequence. The species tree (18) above the panels is shown with estimated divergence times in millions of years (myr) and its 95% highest posterior density in gray.

with human hotspots when a 7.2% overlap is expected by chance (fig. S11).

To provide further support for the validity of the inferred hotspots, we tested whether they show evidence for GC-biased gene conversion (gBGC), measured as higher expected equilibrium levels of GC content (GC*) (18). Because evidence for gBGC in birds is somewhat indirect (22), we first looked for support for gBGC at broad genomic scales, finding a positive relationship between recombination rate and GC* (Fig. 3, A and B). Narrowing our focus to the regions surrounding hotspots, we observed that hotspots exhibit peaked GC* relative to both flanking sequences and “coldspots” (regions without peaks in recombination) matched for the same overall GC and CpG content (Fig. 4, A and B). A similar phenomenon is evident in intraspecies variation data: At hotspots but not at matched coldspots, derived alleles segregate at a higher frequency at AT-to-GC polymorphisms than at GC-to-AT polymorphisms (fig. S13). Thus, two independent signatures of recombination—namely, patterns of linkage disequilibrium and of base composition—converge in demonstrating that finches have recombination hotspots and that these are conserved over much longer time scales than in apes and mice (8–10).

After observing the pattern of gBGC at hotspots in the zebra finch and long-tailed finch genomes, we tested how far the conservation of hotspot locations extends across the avian phylogeny by additionally considering the genomes of the double-barred finch [an estimated ~3.5 million years diverged from the zebra finch (18)], medium ground finch *Geospiza fortis* [~15.5 million years diverged from the zebra finch (23)], and collared flycatcher *Ficedula albicollis* [~19.1 million years diverged from the zebra finch (24)]. Because we only had a single diploid genome from these species, we tested for hotspot conservation indirectly by determining whether these species

had peaks in GC* at the hotspot locations that we had inferred to be shared between the zebra finch and long-tailed finch. We found localized GC* peaks at hotspots in all three species (Fig. 4, C to E), suggesting that the conservation of hotspots extends across tens of millions of years of evolution. These findings mirror those obtained from four species of *Saccharomyces* yeast, which show nearly complete conservation of hotspot locations and intensities across species that are 15 million years diverged (25). Almost all hotspots in *Saccharomyces* yeast occur at promoters, which are evolutionarily stable, suggesting that how hotspot locations are specified influences how they evolve (12, 26).

The localization of hotspots in the genome

Hotspots in the zebra finch and long-tailed finch genomes are enriched near transcription start sites (TSSs), transcription stop sites (TESs), and CpG islands (CGIs), with close to half of all hotspots occurring within 3 kb of one of these features (~17% occur within 3 kb of both an annotated TSS and a CGI, 3% within 3 kb of both a TES and a CGI, and ~26% within 3 kb of a CGI only; fig. S14). In particular, the hotspots near CGIs are more likely to be shared between species and exhibit stronger evidence for gBGC, compared with hotspots distant from CGIs (fig. S15), providing further support for the importance of these elements in the targeting of recombination. Consistent with the findings about hotspots, recombination rates are nearly two times higher near annotated TSSs and TESs (Fig. 5, A and B). This pattern appears to be driven mainly by their colocalization with CGIs (Fig. 5, A and B, and fig. S16): Rates near CGIs are more than three times higher, with only a small further increase if they are near a TSS or a TES (Fig. 5, C and D, and fig. S17).

A positive association between proximity to the TSS and recombination rate has been pre-

viously reported in a number of species without *PRDM9*, including *S. cerevisiae*, the monkey flower *Mimulus guttatus*, dogs, and *A. thaliana* (11, 13, 14, 27), and an association between TES and recombination rate has been shown in *A. thaliana* (14). In turn, the link between CGIs and recombination rates has been found both in species without *PRDM9*, including dogs (11), and, albeit more weakly, in species with *PRDM9*, including humans and chimpanzees (9). Moreover, the relationship between distance to CGIs and recombination rate remains significant after controlling for expression levels in zebra finch testes (Spearman's $r = -0.1$; $P = 4.32 \times 10^{-27}$; fig. S18). This increase in recombination rates near TSSs, TESs, and CGIs supports a model in which, particularly in the absence of *PRDM9*-binding specificity, recombination is concentrated at functional elements that are accessible to the recombination machinery. TSSs, TESs, and CGIs all coincide with destabilization of nearby nucleosome occupancy (28, 29), and both TSSs and CGIs serve as sites of transcription initiation (30). One implication is that the structure of linkage disequilibrium may differ systematically between species with and without *PRDM9*, with tighter coupling between regulatory and exonic variants in species with *PRDM9*.

Under a model in which the recombination machinery tends to target accessible genomic elements, we would not necessarily expect to see enrichment of specific binding motifs associated with hotspot activity. Accordingly, when we tested for motifs enriched in hotspots relative to coldspots, the top motifs in both species were strings of adenines that are also enriched in *A. thaliana* and yeast hotspots and that may be nucleosome-depleted or facilitate nucleosome removal (fig. S19) (13, 31). We also found a number of additional motifs that are GC-rich and perhaps indicative of CGIs.

At even finer resolution, recombination rates are higher in exonic than in intronic regions, as is observed in to *A. thaliana* (14), dogs (11), and *M. guttatus* (27), and higher toward the ends of the gene than in the middle (Fig. 5, E and F). One possibility for these patterns is that DSBs preferentially initiate in exons near the TSS and TES, and their resolution occurs in intervening exons and introns. The specific mechanism by which DSBs would preferentially initiate in exons is unknown, but the pattern is consistent with an important role for chromatin marks that distinguish exons from introns (28).

Contrasting tempos of broad- and fine-scale recombination rate evolution

Median recombination rates across and within chromosomes vary over nearly six orders of magnitude (figs. S8 and S20), creating a heterogeneous landscape of broad-scale recombination rates across the genome, with regions of elevated recombination near telomeres and large intervening deserts [as inferred from zebra finch pedigree data (21)]. Most of the recombination events in the zebra finch and long-tailed finch occur in a narrow portion of the genome, with 82 and 70%

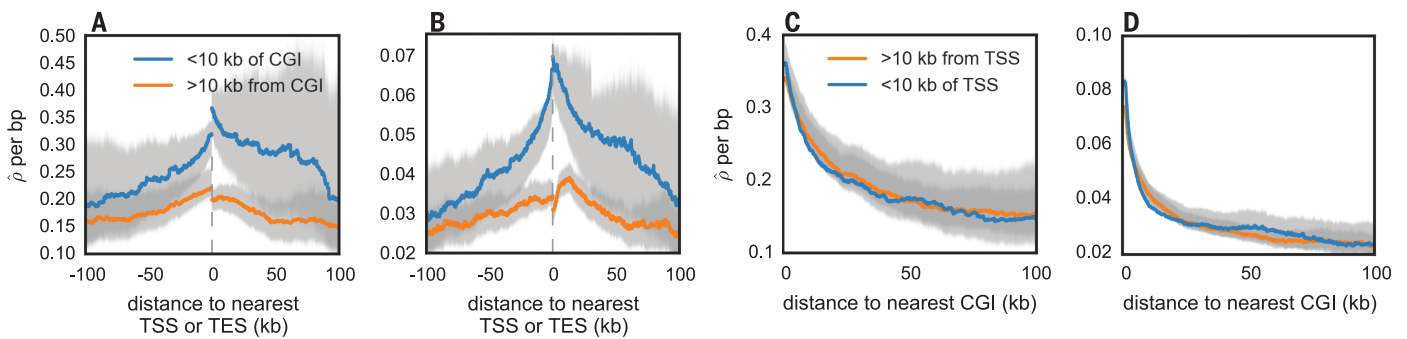


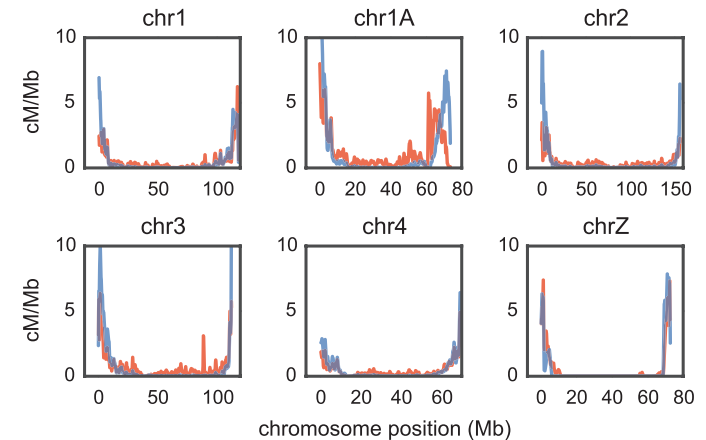
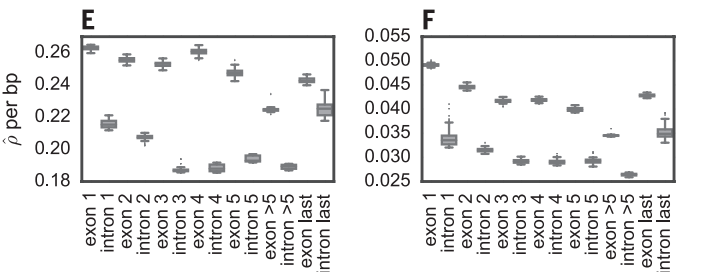
Fig. 5. Recombination rates across genomic features in zebra finch and long-tailed finch genomes. (A and B) Estimated recombination rates ($\hat{\rho}$ /bp) around annotated TSSs and TESs in zebra finch (A) and long-tailed finch (B) genomes, conditional on whether the sites are within 10 kb of a CGI. The gray dotted line represents the location of the gene, and the distances are shown accounting for the 5' → 3' orientation of genes. (C and D) $\hat{\rho}$ shown as a function of distance to the nearest CGI in zebra finch (C) and long-tailed finch (D) genomes, conditional on whether the CGI is within 10 kb of an annotated TSS. Figure S17 shows the pattern of CGIs relative to TESs. For (A) to (D), uncertainty in rate estimates (shown in gray) was estimated by drawing 100 bootstrap samples and recalculating means. (E and F) $\hat{\rho}$ within exons and introns for genes that have ≥ 5 exons ($n = 7131$) in zebra finch (E) and long-tailed finch (F) genomes. Figure S28 shows simulation results that suggest that the inference of higher background $\hat{\rho}$ in exons does not reflect differences in diversity levels between exons and introns.

of events localized to 20% of the genome in the zebra finch and long-tailed finch, respectively (fig. S21). In particular, recombination rates for the Z sex chromosome are two orders of magnitude lower than those for the most similarly sized autosome, chromosome 1A, even after accounting for the lack of recombination in females (fig. S8) (27). Although cytological data indicate that both zebra finches and long-tailed finches harbor a pericentric inversion polymorphism over much of chromosome Z (32, 33), such an inversion is unlikely to explain this extreme a difference (18).

Between the zebra finch and long-tailed finch, broad-scale rates are highly similar, with genome-wide correlations of 0.82 and 0.86 at the 10-kb and 1-Mb scales, respectively (Fig. 6 and fig. S20). Despite this broad-scale concordance, we infer that some genomic regions between the two species have very different rates of recombination (fig. S22), and we found tentative support for some of these changes in the derived allele frequency spectra (fig. S23). Moreover, at a greater evolutionary distance, broad-scale patterns differ markedly; the collared flycatcher (~19 million years diverged) has a relatively homogeneous recombination landscape compared with the zebra finch and long-tailed finch (24). This evolution of broad-scale rates is particularly notable because, in many species, shifts in broad-scale recombination patterns can be explained almost entirely by chromosomal rearrangements, shifts in karyotypes, and changes in chromosome lengths (9, 34, 35). However, there is no obvious pattern by which chromosomal rearrangements drive differences in recombination rates between the zebra finch and long-tailed finch (fig. S22), and, despite harboring a number of small inversions between them, the collared flycatcher and zebra

Fig. 6. Comparative recombination rates in zebra finch and long-tailed finch genomes. Zebra finch rates are shown in red; long-tailed finch rates are shown in blue.

Estimated rates [cM/Mb; obtained from $\hat{\rho}$ /bp (18)] are shown as rolling means calculated across 100-kb windows. We show here the five largest autosomal chromosomes and chromosome Z (fig. S20 shows all chromosomes). Rate estimates for chromosome Z should be taken with caution for both biological and technical reasons [more information is given in (18)].



finch have similar karyotypes and syntenic genomes (24). The evidence that broad-scale recombination patterns have changed across the same phylogenetic breadth for which we see hotspot conservation suggests two nonexclusive possibilities that merit further investigation: The heats or locations of some hotspots have evolved, or rates have changed in regions that fall outside of our operational definition of hotspots.

The impact of recombination on the genome

Given the marked variation in recombination rates across the genome, we consider the consequences for genome evolution. Increased recombination rates drive increasing GC content in the genome, presumably via gBGC, and we see this phenomenon both at the genome-wide scale (Fig. 3, A and

B) and at the scale of hotspots (Fig. 4). An extreme example is provided by the pseudoautosomal region (PAR), which we identified on an unassembled scaffold from chromosome Z, using estimates of coverage in males and females. We confirmed the PAR by inferring homology to PARs identified in the medium ground finch and collared flycatcher (fig. S24). The PAR is short, estimated to be just 450 kb, and subject to an obligate crossover in every female meiosis (36); as such, it has very high recombination rates. The consequence is visible in the high GC* for the PAR, which exceeds estimates of GC* across most of the rest of chromosome Z in both species (Fig. 3, C and D).

Further, as has been reported for many other organisms, including chickens (37–39), our results suggest that recombination is positively correlated with levels of nucleotide diversity, particularly on the Z chromosome (figs. S25 to S27). This

observation is consistent with widespread effects of linked selection in these species (40).

Conclusion

Finches lack *PRDM9*, yet they nonetheless harbor hotspots, with recombination concentrated at functional elements (TEs, TSSs, and CGIs) that likely denote greater accessibility to the cellular recombination machinery. In sharp contrast to apes and mice, the hotspot locations are conserved among species several millions of years diverged and probably over tens of millions of years. These results suggest that the genetic architecture of recombination influences the rate at which hotspots evolve. Whereas the binding specificity of *PRDM9* drives rapid turnover, the reliance on accessible functional genomic features leads to stasis. This hypothesis dovetails with recent results in yeast, in which recombination is concentrated at promoters and hotspots are stable in intensity and location over tens of millions of years (25). To further investigate how deeply this stasis extends and to explore the taxonomic generality of these findings, the approaches illustrated here can be applied to other sequenced bird species (41) and beyond. In doing so, we will begin to better understand why species differ so drastically in their specification of hotspots and, in particular, why a subset relies on *PRDM9*.

REFERENCES AND NOTES

1. B. de Massy, *Annu. Rev. Genet.* **47**, 563–599 (2013).
2. S. Myers *et al.*, *Science* **327**, 876–879 (2010).
3. I. L. Berg *et al.*, *Nat. Genet.* **42**, 859–863 (2010).
4. F. Baudat *et al.*, *Science* **327**, 836–840 (2010).
5. A. L. Williams *et al.*, *eLife* **4**, e04637 (2015).
6. P. L. Oliver *et al.*, *PLOS Genet.* **5**, e1000753 (2009).
7. A. G. Hinch *et al.*, *Nature* **476**, 170–175 (2011).
8. L. S. Stevison *et al.*, <http://biorxiv.org/search/013755> (2015).
9. A. Auton *et al.*, *Science* **336**, 193–198 (2012).
10. K. Brick, F. Smagulova, P. Khil, R. D. Camerini-Otero, G. V. Petukhova, *Nature* **485**, 642–645 (2012).
11. A. Auton *et al.*, *PLOS Genet.* **9**, e1003984 (2013).
12. E. Axelsson *et al.*, *Genome Res.* **22**, 51–63 (2012).
13. J. Pan *et al.*, *Cell* **144**, 719–731 (2011).
14. K. Choi *et al.*, *Nat. Genet.* **45**, 1327–1336 (2013).
15. A. H. Chan, P. A. Jenkins, Y. S. Song, *PLOS Genet.* **8**, e1003090 (2012).
16. A. Wallberg, S. Glémin, M. T. Webster, *PLOS Genet.* **11**, e1005189 (2015).
17. T. Kaur, M. V. Rockman, *Genetics* **196**, 137–148 (2014).
18. Materials and methods are available as supplementary materials on Science Online.
19. W. C. Warren *et al.*, *Nature* **464**, 757–762 (2010).
20. E. M. Leffler *et al.*, *PLOS Biol.* **10**, e1001388 (2012).
21. N. Backström *et al.*, *Genome Res.* **20**, 485–495 (2010).
22. C. K. Weber, B. Boussau, J. Romiguier, E. D. Jarvis, H. Ellegren, *Genome Biol.* **15**, 549 (2014).
23. G. Zhang, P. Parker, B. Li, J. Wang, *GigaScience* 10.5524/100040 (2012).
24. T. Kawakami *et al.*, *Mol. Ecol.* **23**, 4035–4058 (2014).
25. I. Lam, S. Keeney, *Science* **350**, 932–937 (2015).
26. A. Nicolas, L. Treco, N. P. Schultes, J. W. Szostak, *Nature* **338**, 35–39 (1989).
27. U. Hellsten *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 19478–19482 (2013).
28. P. A. Jones, *Nat. Rev. Genet.* **13**, 484–492 (2012).
29. N. Kaplan *et al.*, *Nature* **458**, 362–366 (2009).
30. A. M. Deaton, A. Bird, *Genes Dev.* **25**, 1010–1022 (2011).
31. E. Wijner *et al.*, *eLife* **2**, e01426 (2013).
32. L. Christidis, *Genetica* **71**, 81–97 (1986).
33. Y. Itoh, K. Kamp, C. N. Balakrishnan, A. P. Arnold, *Chromosoma* **120**, 255–264 (2011).
34. B. L. Dumont, B. A. Payseur, *Evolution* **62**, 276–294 (2008).
35. M. I. Jensen-Seaman *et al.*, *Genome Res.* **14**, 528–538 (2004).
36. S. P. Otto *et al.*, *Trends Genet.* **27**, 358–367 (2011).
37. M. W. Nachman, *Trends Genet.* **17**, 481–485 (2001).
38. D. J. Begun, C. F. Aquadro, *Nature* **356**, 519–520 (1992).

39. C. F. Mugal, P. F. Arndt, H. Ellegren, *Mol. Biol. Evol.* **30**, 1700–1712 (2013).
40. B. Charlesworth, M. T. Morgan, D. Charlesworth, *Genetics* **134**, 1289–1303 (1993).
41. G. Zhang *et al.*, *Science* **346**, 1311–1320 (2014).

ACKNOWLEDGMENTS

This project was started when M.P. was a Howard Hughes Medical Institute Early Career Scientist and was funded, in part, by Wellcome Trust grants 086786/Z/08/Z to O.V. and 090532/Z/09/Z to the Wellcome Trust Centre for Human Genetics. We thank B. de Massy, C. Grey, S. Myers, T. Price, M. Schumer, J. Wall, A. Williams, and J. Willis for helpful discussions and/or comments on the manuscript; K. Argoud and P. Piazza at the Genomics Core at the Wellcome Trust Centre for Human Genetics for assistance with laboratory work; and M. T. Gilbert for sharing the zebra finch gene annotations in advance of publication. We thank I. Lam and S. Keeney for sharing their unpublished manuscript with us, S. Keeney for many helpful discussions, and A. Johnson for the illustrations used in Fig. 1.

Binary Alignment Map files for genomic data and filtered variant call files for the zebra finch, long-tailed finch, and double-banded finch are available at www.ebi.ac.uk/ena/data/view/PRJEB10586. Sequence reads for zebra finch RNA sequencing experiments are available at www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA295077. Masked genome files, the reconstructed ancestral genome, and recombination maps for both species are available at DataDryad (doi: 10.5061/dryad.f24). All scripts and an electronic laboratory notebook for this work are available at <https://github.com/singhal/postdoc> and <https://github.com/singhal/labnotebook>, respectively.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/350/6263/928/suppl/DC1
Materials and Methods
Figs. S1 to S35
Tables S1 to S6
References (42–94)

23 July 2015; accepted 5 October 2015
10.1126/science.aad0843

NONHUMAN GENETICS

Nonparadoxical evolutionary stability of the recombination initiation landscape in yeast

Isabel Lam^{1,2} and Scott Keeney^{1,2,3*}

The nonrandom distribution of meiotic recombination shapes heredity and genetic diversification. Theoretically, hotspots—favored sites of recombination initiation—either evolve rapidly toward extinction or are conserved, especially if they are chromosomal features under selective constraint, such as promoters. We tested these theories by comparing genome-wide recombination initiation maps from widely divergent *Saccharomyces* species. We find that hotspots frequently overlap with promoters in the species tested, and consequently, hotspot positions are well conserved. Remarkably, the relative strength of individual hotspots is also highly conserved, as are larger-scale features of the distribution of recombination initiation. This stability, not predicted by prior models, suggests that the particular shape of the yeast recombination landscape is adaptive and helps in understanding evolutionary dynamics of recombination in other species.

DNA double-strand breaks (DSBs) generated by the SpoII protein initiate meiotic recombination, which alters genetic linkage and promotes pairing and accurate chromosome segregation (1). DSBs are distributed nonrandomly across genomes, occurring often within narrow regions called hotspots (2). Theoretical work exploring evolutionary dynamics of recombination has led to a prevailing hypothesis, the “hotspot paradox,” that predicts rapid hotspot extinction (3–7). This view rests on biased gene conversion, in which the broken chromosome copies genetic information from its uncult homolog and may generate an extra copy of a genetic variant (Fig. 1A). Consequently, hotspot alleles with different DSB activity deviate from a Mendelian segregation ratio, with recombinationally less active alleles overrepresented among the offspring.

This type of meiotic drive is observed in yeast (8) and humans (9) and predicts that mutations that reduce or eliminate hotspot activity will be rapidly fixed in populations, whereas hotspot-activating mutations are rapidly extinguished (3, 5, 10). The paradox is that hotspots exist at all despite this drive against them.

One answer to this paradox comes from *PRDM9*, a mammalian histone methyltransferase with an array of Zn-finger modules that rapidly evolve new DNA binding specificity (11). *PRDM9* targets DSB formation near its binding sites and thus dictates hotspot positions. *PRDM9* recognition motifs, which have no known intrinsic function, are lost quickly from genomes of humans and mice because of meiotic drive from biased gene conversion (12–14), but the appearance of new *PRDM9* alleles with different sequence specificity creates new hotspots and redraws the recombination landscape (11). This hotspot-targeting mechanism confirms the rapid extinction predicted by the hotspot paradox and explains how hotspots can, nonetheless, exist. However, most taxa (including yeast and some mammals) lack such a system, so it has remained unclear how generalizable this solution is.

¹Louis V. Gerstner, Jr., Graduate School of Biomedical Sciences, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA. ²Molecular Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA.

³Howard Hughes Medical Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA.

*Corresponding author. E-mail: s-keeney@ski.mskcc.org (S.K.).

Stable recombination hotspots in birds

Sonal Singhal, Ellen M. Leffler, Keerthi Sannareddy, Isaac Turner, Oliver Venn, Daniel M. Hooper, Alva I. Strand, Qiye Li, Brian Raney, Christopher N. Balakrishnan, Simon C. Griffith, Gil McVean and Molly Przeworski

Science **350** (6263), 928-932.
DOI: 10.1126/science.aad0843

Recombination: The birds and the yeast

Apes and mice have a specific gene, *PRDM9*, that is associated with genomic regions with high rates of recombination, called hotspots. In species with *PRDM9*, hotspots move rapidly within the genome, varying among populations and closely related species (see the Perspective by Lichten). To investigate recombination hotspots in species lacking *PRDM9*, Singhal *et al.* examined bird genomes, which lack a *PRDM9* gene. They looked closely at the genomes of finch species and found that recombination was localized to the promoter regions of genes and highly conserved over millions of years. Similarly, Lam and Keeney examined recombination localization within yeast, which also lacks *PRDM9*. They found a similar more-or-less fixed pattern of hotspots. Thus, recombination in species lacking a *PRDM9* gene shows similar patterns of hotspot localization and evolution.

Science, this issue p. 913, p. 928; see also p. 932

ARTICLE TOOLS

<http://science.sciencemag.org/content/350/6263/928>

SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2015/11/18/350.6263.928.DC1>

RELATED CONTENT

<http://science.sciencemag.org/content/sci/350/6263/913.full>
<http://science.sciencemag.org/content/sci/350/6263/932.full>

REFERENCES

This article cites 88 articles, 27 of which you can access for free
<http://science.sciencemag.org/content/350/6263/928#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)