Evolution, 2024, **XX(XX)**, 1–16
https://doi.org/10.1093/evolut/qpae169
Advance access publication 26 November 2024
Original Article

Rapid evolution of recombination landscapes during the divergence of cichlid ecotypes in Lake Masoko

Marion Talbi[1,2], George F. Turner[3], Milan Malinsky[1,2]

[1]Biology Department, Institute of Ecology and Evolution, University of Bern, Bern, Switzerland
[2]Department of Fish Ecology and Evolution, EAWAG, Kastanienbaum, Switzerland
[3]School of Natural & Environmental Sciences, Bangor University, Bangor, United Kingdom

Corresponding authors: Institute of Ecology and Evolution, Baltzerstrasse 6, University of Bern, 3012 Bern, Switzerland. Email: marion.talbi@gmail.com; Institute of Ecology and Evolution, Baltzerstrasse 6, University of Bern, 3012 Bern, Switzerland. Email: millanek@gmail.com

Abstract

Variation of recombination rate along the genome is of crucial importance to rapid adaptation and organismal diversification. Many unknowns remain regarding how and why recombination landscapes evolve in nature. Here, we reconstruct recombination maps based on linkage disequilibrium and use subsampling and simulations to derive a new measure of recombination landscape evolution: the Population Recombination Divergence Index (PRDI). Using PRDI, we show that fine-scale recombination landscapes differ substantially between two cichlid fish ecotypes of *Astatotilapia calliptera* that diverged only ~2,500 generations ago. Perhaps surprisingly, recombination landscape differences are not driven by divergence in terms of allele frequency (F_{ST}) and nucleotide diversity ($\Delta(\pi)$): although there is some association, we observe positive PRDI in regions where F_{ST} and $\Delta(\pi)$ are zero. We found a stronger association between the evolution of recombination and 47 large haplotype blocks that are polymorphic in Lake Masoko, cover 21% of the genome, and appear to include multiple inversions. Among haplotype blocks, there is a strong and clear association between the degree of recombination divergence and differences between ecotypes in heterozygosity, consistent with recombination suppression in heterozygotes. Overall, our work provides a holistic view of changes in population recombination landscapes during the early stages of speciation with gene flow.

Keywords: recombination, rapid evolution, simulations, cichlids, speciation

Introduction

Meiotic recombination is central to genetics and to evolution in sexually reproducing organisms. It facilitates rapid adaptation by generating new combinations of alleles (McDonald et al., 2016; Nielsen, 2006; Rice & Chippindale, 2001), but in some contexts, it can also hinder adaptation by breaking up locally adapted haplotypes (Ortiz-Barrientos et al., 2016; Schluter & Rieseberg, 2022). Recombination is itself subject to selection (Otto & Barton, 2001), and a substantial body of theory has been developed describing genetic variants that regulate recombination rates, so-called "recombination modifiers," and the conditions under which such modifier variants would be selected for or against (Coop & Przeworski, 2007; Feldman et al., 1996; Nei, 1967). Many genetic variants, both cis- and trans- acting, are known to affect the recombination rates and the positioning of recombination events (Brand et al., 2018; Halldorsson et al., 2019; Rowan et al., 2019; Samuk et al., 2017). For example, recombination suppression is often facilitated by larger structural genetic variants, especially inversions (Jay et al., 2018; Todesco et al., 2020), although insertions, deletions, and sequence translocations have also been implicated (Kent et al., 2017; Rowan et al., 2019; Schluter & Rieseberg, 2022). Even single-nucleotide polymorphisms (SNPs) can modify recombination rates at specific loci, as in the case of the ade6-M26 mutation, which creates a hotspot of 10 to 15× elevated recombination in yeast (Ponticelli et al., 1988; Szankasi et al., 1988).

Recombination is also subject to forces that appear largely decoupled from organismal adaptation or diversification. First, it must fulfill its essential role in meiosis and chromosome segregation (Petronczki et al., 2003). This virtually ubiquitous requirement provides a lower bound of one recombination event per chromosome (Henderson & Bomblies, 2021) and contributes to limiting average recombination rates to a relatively narrow range above this minimum via the mechanism of "crossover interference" (Otto & Payseur, 2019). Second, in some vertebrate species, principally in mammals, recombination is directed towards binding sites of the zinc-finger (ZF) protein PRDM9 (Baker et al., 2017; Baudat et al., 2010; Cavassim et al., 2022; Myers et al., 2010). In these species, the rapid evolution of recombination landscapes is mediated by intra-genomic conflict (Baker et al., 2023; Latrille et al., 2017; Úbeda & Wilkins, 2010), and genetic variants altering recombination landscapes in this way are therefore usually studied through the prism of internal genome dynamics and not within the framework of traditional recombination modifier theory (Genestier et al., 2024).

Outside of mammals, in most other vertebrates, recombination does not appear to be associated with PRDM9 binding sites (Cavassim et al., 2022). Species lacking the PRDM9

Received April 30, 2024; revisions received November 6, 2024; accepted November 25, 2024

Associate Editor: Andrea Betancourt; Handling Editor: Jason Wolf

© The Author(s) 2024. Published by Oxford University Press on behalf of The Society for the Study of Evolution (SSE).
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/evolut/advance-article/doi/10.1093/evolut/qpae169/7908996 by guest on 06 December 2024

mechanism have elevated recombination rates at and around genomic features such as CpG islands and promoters, likely due to the greater chromatin accessibility in these regions (Baker et al., 2017), and there is evidence that species lacking PRDM9 have more conserved recombination landscapes (Lam & Keeney, 2015; Singhal et al., 2015). However, the association with these genomic features is only partial (Singhal et al., 2015), and recombination rates do also evolve in species without PRDM9 (Déserts et al., 2021; Ritz et al., 2017; Samuk et al., 2020). It has even been suggested that in stickleback fish, hotspots may evolve at similar rates to those observed in species with PRDM9 (Shanfelter et al., 2019).

In the presence of gene flow, recombination counteracts the buildup of linkage among genetic loci that contribute to population divergence and, ultimately, speciation (Barton, 2020; Butlin et al., 2021; Felsenstein, 1981). Consistent with the important role in this context, a recombination landscape, together with natural selection, shapes the distribution of genomic regions of divergence along each chromosome (Duranton et al., 2018; Martin et al., 2019; Ravinet et al., 2017; Schumer et al., 2018). The recombination landscape itself is an evolving dynamic parameter (Déserts et al., 2021; Ritz et al., 2017; Samuk et al., 2020), which should be taken into account in genomic studies of speciation (Ortiz-Barrientos & James, 2017; Ortiz-Barrientos et al., 2016). In recent years, there has been a growing appreciation of the role of recombination suppression in organismal diversification (Schluter & Rieseberg, 2022). These efforts often concentrated on specific large inversions or other low-recombining haplotype blocks (Faria et al., 2019a; Jay et al., 2018; Reeve et al., 2023; Todesco et al., 2020). In addition, a recent study based on genome-wide estimates of recombination rates supports the notion that cis-acting recombination modifiers play an important role in promoting adaptive divergence between populations (Venu et al., 2024). Building upon this initial progress, additional comparisons of recombination landscapes in different species and across different levels of divergence will be required to understand where in the genome, how fast, and by what mechanisms recombination rates evolve and, ultimately, the interplay with natural selection and organismal evolution.

The Lake Masoko system presents a well-suited opportunity to study the evolution of recombination rates in the context of organismal diversification. Lake Masoko is a small (~670 m in diameter) crater lake in Southern Tanzania (Figure 1A) and is approximately 50k years old (Barker et al., 2003). Two ecotypes of the cichlid fish species *Astatotilapia calliptera* have evolved within this lake—the shallow-water "littoral" and the deep-water "benthic." They differ from each other in several ecologically important traits and, while almost half of the sites have zero F_{ST}, there is elevated allele frequency divergence at about a hundred of well-demarcated genomic regions—islands of divergence (Malinsky et al., 2015). Several other fish species belonging to the same clade (Percomorpha) lack a functional PRDM9, although *A. calliptera* itself has not been tested (Baker et al., 2017; Cavassim et al., 2022).

Over the last decade or so, many empirical studies used recombination estimates based on linkage disequilibrium (LD)—patterns of nonrandom association of genetic variants in a sample of individuals from a population (Auton et al., 2012; Shanfelter et al., 2019; Singhal et al., 2015; Spence & Song, 2019). LD-based methods estimate a "population recombination rate" (ρ)—a population genetic parameter which is a product of effective population size (N_e) and the historical per-generation recombination rate averaged over the ancestry of the sampled individuals. Differences between LD-based recombination maps can arise from sampling variance, methodological limitations, different ancestries across samples, and factors affecting N_e such as demographic change and natural selection (Coop & Przeworski, 2007; Peñalba & Wolf, 2020; Samuk et al., 2020). Therefore, it is challenging to use LD-based methods to compare recombination landscapes between populations and species to track their evolution.

In this study, we reconstruct genetic maps from patterns of LD in whole genome population genetic data of 70 benthic and 69 littoral individuals to investigate the evolution of recombination landscapes in Lake Masoko. By carefully controlling for confounding factors, we demonstrate that the population recombination landscapes are considerably different despite the recent split time between these ecotypes and we quantify the degree of this population recombination rate evolution. The regions where recombination rates differ significantly are not distributed equally across the genome. We show a link with genetic differentiation, as measured, for example, by F_{ST}, and with larger haplotype blocks, although neither of these fully explain the recombination rate divergence. We found a partial copy of PRDM9 in the *A. calliptera* genome. However, its predicted binding sites do not show any relationship with recombination rates, which is consistent with previous studies in fishes with partial PRDM9 (Baker et al., 2017). Overall, our study quantifies and provides new insights into rapid population recombination rate evolution in the context of sympatric ecotype divergence.

Methods

Variant calling and filtering

Genomic DNA from a total of 336 individuals from Lake Masoko was sequenced on the Illumina HiSeq X Ten platform, obtaining 150bp paired-end reads (NCBI Short Read Archive, BioProject ID: PRJEB27804). All the reads, without filtering or trimming, were aligned to the *Astatotilapia calliptera* reference genome fAstCal1.5 (GenBank ID: GCA_900246225.6) using bwa-mem v.0.7.17 (Li, 2013). The reference sequence is based on an *A. calliptera* sample from the Itupi stream, which is a close outgroup to Lake Masoko. We used the MarkDuplicates tool from the Picard package v.2.26.6 to tag PCR and optical duplicate reads and GATK v.4.2.3 (DePristo et al., 2011) to call variants, using HaplotypeCaller in GVCF mode for each individual separately followed by joint genotyping using GenotypeGVCFs with the --include-non-variant-sites option.

Next, we generated a *callability mask* to identify and filter out the regions of the genome where we were unable to confidently call variants. The mask included: (i) sites with an overall read depth cutoffs based on examining a depth histogram (<3,800, corresponding to the 20.7th percentile or >5,700, corresponding to the 97.9th percentile; Supplementary Figure 1); (ii) sites where more than 10% of individuals had missing genotypes; (iii) sites identified by GATK as low quality (with the LowQual tag) and (iv) sites with poor mappability. Specifically, to obtain the mappability information, we broke down the genome into overlapping k-mers of 150bp (matching the read length), mapped these k-mers back to the genome, and masked all sites where fewer than 90% of k-mers mapped back to their original location perfectly and uniquely. In total,

Downloaded from https://academic.oup.com/evolut/advance-article/doi/10.1093/evolut/qpae169/7908996 by guest on 06 December 2024

the callability mask comprised 311 million bp, or about 35% of the genome. In addition to applying the callability mask, we used several hard filters based on GATK best practices, specifically focusing on mapping quality (MQ < 40), mapping strand bias (FS > 40), variant quality normalized by depth (QD < 2) and excess heterozygosity when compared with Hardy–Weinberg equilibrium (ExcessHet > 40). These additional filters removed fewer than 1% of the remaining SNPs.

Sample selection for recombination analyses

We used the full set of 336 available individuals for variant calling because the inclusion of more samples leads to more accurate genotyping. However, in this study, we were specifically interested in differences between the littoral and benthic ecotypes of Lake Masoko. Therefore, we retained 80 individuals assigned in the field as benthic and 79 assigned as littoral and excluded 201 other individuals who were not assigned to either ecotype because they were juveniles, females (neither category show the ecotype-distinct male breeding colors), or putative hybrids. To check the validity of these field assignments, we first built a neighbor-joining tree based on a genetic distance matrix, i.e., the average number of single-nucleotide differences between each pair of individuals, using the stats command from the evo package v.0.1 r28 and the --diff-matrix option. The pairwise difference matrix was divided by the callable genome size to obtain pairwise distances per base pair, and this was then used as input into the nj() tree-building function implemented in the package *ape* in R (Paradis et al., 2004). Next, for principal component analysis (PCA) we used smartPCA (Patterson et al., 2006) on data filtered for minor allele frequency >=0.05 using plink v1.9 (Purcell et al., 2007) with the --maf 0.05 option and LD pruned using the plugin + prune from bcftools v.1.16 (Danecek et al., 2021) with the -m 0.8 -w 1000 options.

Genome annotation

Because the fAstCal1.5 assembly was not annotated by NCBI, we "lifted over" the annotation from the older fAstCal1.2 assembly (GenBank: GCA_900246225.3). We used the UCSC paradigm (Miller et al., 2007) to generate a pairwise whole genome alignment between fAstCal1.2 and fAstCal1.5 assembly. Afterwards, we used the UCSC liftOver tool to translate the NCBI Annotation Release 100 to the new coordinates.

Inference of demographic history and estimation of the level of gene flow

To estimate split time between the ecotypes and changes in effective population size (N_e) through time, we used smc++ v.1.15.4 (Terhorst et al., 2017), using the sequence of smc++ commands: vcf2smc -> estimate -> split. To translate the time axis into a number of generations, we used the cichlid-specific mutation rate estimate of $\mu = 3.5 \times 10^{-9}$ per bp per-generation with 95% CI (1.6×10^{-9}, 4.6×10^{-9}) (Malinsky et al., 2018). Next, we used fastsimcoal2.7 (Excoffier et al., 2021) to estimate the level of gene flow between the two ecotypes. In fastsimcoal, we entered the split time and changes in N_e as inferred by smc++ as fixed parameters and estimated continuous asymmetrical migration rates after the ecotype split. We ran 30 simulations with different starting parameter values, which revealed two local peaks in the likelihood surface (Supplementary Figure 2). To reduce the confounding effects of selection in these demographic analyses, we used only sites from noncoding regions of the genome, masking all annotated exons, introns, and promoters.

Subsampling and permutations over individuals

We first used the `shuf -n` command to randomly draw a set of 35 individuals from each ecotype for the first subset. The remaining individuals (35 littoral and 34 benthic) then formed the second subset. Therefore, these (a) and (b) subsets (Figure 1D) are independent in the sense that they are composed of nonoverlapping sets of individuals. We then generated a separate VCF file for each subset using the bcftools v.1.16 view command and used these VCFs for recombination map reconstruction. We repeated this random sampling procedure (and the following genetic map reconstruction) to obtain nine permutation replicates over individuals.

Inference of recombination rates

We used the pyrho software (Spence & Song, 2019) to infer recombination rates along the genome based on patterns of LD. We choose pyrho because it accounts for demography, i.e., changes in N_e through time, and because its performance does not depend on haplotype phasing—it performs equally well with phased and unphased data [as described in fig. S8 of (Spence & Song, 2019) and confirmed by our own simulations (data not shown)].

To build likelihood tables for pairs of biallelic sites, we used the make_table command with $\mu = 3.5 \times 10^{-9}$, demographic history for each of the ecotypes as inferred by smc++, and the Moran approximation with the –approx and –moran_pop_size N flags where N equals 1.5× the number of haplotypes in each subset. Recombination inference is highly influenced by the block penalty parameter, which affects the smoothness of the map. To determine the best value to use, we processed a set of simulations with evolutionary parameters corresponding to the one of our cichlid species (e.g., μ, sample size, N_e) and chose the value that was minimizing the quantity of false negatives and false positives (Supplementary Figure 3). The best results were obtained for block penalty of 15, which was then used in all runs of the optimize command with a window size of 50 SNPs to infer the recombination maps. The output of pyrho contained estimates of recombination rate between each pair of SNPs.

Neutral coalescent simulations

We used msprime v.1.0.2 (Baumdicker et al., 2021) to simulate genetic data matching the population and demographic histories (split time, N_e, and gene flow) that we inferred from empirical data as described above. Because recombination landscapes were the same for both simulated populations and natural selection was absent in these simulations, the results from analyzing the simulated data allowed us to better evaluate and interpret the empirical results. We ran 23 simulations—one for each chromosome—using $\mu = 3.5 \times 10^{-9}$ and the empirical recombination maps as input. From each simulation, we sampled 70 individuals from each population, labeled them as "benthic" and "littoral," randomly subsampled the (a) and (b) subsets and further processed the VCF output in the same way as we did for empirical data.

To confirm that the conclusions of this manuscript are robust to the specifics of demographic inference, we ran the simulations with an extended set of split time values (1,000, 2,500, 5,000 and 10,000 generations ago) and with migration rates corresponding to the two local likelihood peaks in fastsimcoal2 inference ("low migration rates": 11.5×10^{-5} for littoral to benthic and 7.01×10^{-5} for benthic to littoral;

Downloaded from https://academic.oup.com/evolut/advance-article/doi/10.1093/evolut/qpae169/7908996 by guest on 06 December 2024

and "high migration rates": 0.0339 for littoral to benthic and 0.0368 for benthic to littoral).

We also used msprime to estimate how the inferred recombination maps reflect the relative contributions of recombination events that happened in the common history of the ecotypes vs. events that happened after their split. To do this, we counted the number of local genealogies, which reflects the recombination events that changed the local genealogy of the sample. We used the end_time option in the sim_ancestry() function of msprime to stop the simulation at the split time and counted the distinct genealogies at that time point in benthic (N_{tb}) and in littoral (N_{tl}) ecotypes—these counts reflect the genealogy-changing events that happened in the ecotypes after their split. Then, we continued the simulation backward in time all the way to the common ancestor of all samples and counted the total number of genealogies (N_t). Finally, we calculated the proportion of genealogy-changing events that happened in the ecotypes after their split (P_{as}) as:

$$P_{as} = \frac{(N_{tb} - 1) \quad + \quad (N_{tl} - 1)}{(N_t - 1)}$$

The scripts used to run these simulations are available from GitHub https://github.com/MarionTalbi/MasokoPaper.

Processing and comparisons of recombination maps

We used the PhysicalWindowAverages command from the evo package v.0.1 r28 to obtain mean recombination rates in identical 2 Kb windows for all datasets, which facilitated easy comparisons between different maps. The correlations, map distances, and other comparisons were then calculated using R scripts, available from GitHub (https://github.com/MarionTalbi/MasokoPaper).

The first set of comparisons among recombination maps was based on correlations, which allowed us to define a measure of divergence which we call the Population Recombination Divergence Index (PRDI). Briefly, based on pairwise Spearman correlations of recombination maps in 2 kb windows, binned per 5 Mb regions along the genome, we first obtained median correlations for the within-ecotype replicate maps (denoted m_{wb} and m_{wl} for within-benthic and within-littoral replicates, respectively), and the median of all map comparisons between ecotypes, denoted m_b. We used medians rather than means to avoid undue impact of very high recombination hotspot values. Between-ecotype correlations are generally lower than correlations for within-ecotype replicates and we measure this difference as:

$$d_{me} = \min\,(m_{wb}, m_{wl}) - m_b$$

The use of minimum over m_{wb} and m_{wl} is the most conservative choice. The d_{me} measure is derived from the empirical data. We also define an analogous measure, d_{ms} based on the simulated data to account for the separation of recent benthic vs. littoral ancestry. Finally, PRDI, our measure of recombination landscape divergence is defined as the difference between these two values:

$$PRDI = d_{me} - d_{ms}$$

The second set or comparisons was based on recombination map distances. Map distances were calculated in nonoverlapping 100 kb windows (i.e., vectors of 50 values for 2 kb each) using the dist() function in R. We use log_{10} transformed absolute (Manhattan) distances because these are straightforward to interpret: a log_{10} distance of 1 signifies an average difference in recombination estimates of one order of magnitude. To find the regions of the genome where the recombination distance between the ecotypes was significantly elevated, which we refer to as $\Delta(r)$ outliers, we used the following procedure. We first calculated the recombination distances in comparisons of within-ecotype replicate maps (denoted $\Delta(r)\_w$). Then, we calculated the analogous measure for map comparisons between ecotypes (denoted $\Delta(r)\_b$). Finally, we calculated the standard deviation of the log_{10} transformed $\Delta(r)\_w$ measure across the permutation replicates (we denote this standard deviation as sd_w) and used the cutoff of three standard deviations to define outliers. Specifically, we refer to any 100 kb interval of the genome as a $\Delta(r)$ outlier if it satisfies the following inequality:

$$-\log_{10}\left[\Delta(r)_b\right] \quad > \quad -\log_{10}\left[\Delta(r)_w\right] + 3 * sd_w$$

Filtered maps

The reliability of LD-based recombination rate inference varies across the genome, depending on several factors, including miscalled variants, errors in the reference genome, and amount of genetic variation (i.e., amount of data available for inference). To understand how our results are affected by these factors, we generated filtered recombination maps where the less reliable regions of the genome were masked. While the main figures of this manuscript report the results for the raw maps, we conducted many of the key analyses also using the filtered maps and present these results as Supplementary Figures.

First, errors in the reference genome can mistakenly place in physical proximity genetic variants that have large genetic distances between them. Therefore, we masked intervals 50 kb upstream and 50 kb downstream of each run of unspecified bases ("N" characters). These runs of Ns mark contig joints, i.e., joints between contiguous sequences in the assembly. There were 514 joints on the 22 chromosomes, leading to masking of 50.46 Mb of sequence. Second, we used the callability mask produced for variant filtering (see above). A lack of data can make recombination inference difficult. We reduced this effect by masking each 100 kb region within which more than 70% were not callable. Third, to exclude regions where the recombination maps showed especially elevated sampling noise, we took advantage of the permutation replicates and masked all 100 kb regions where the difference in inferred recombination rates between the replicate maps was greater than one order of magnitude. Overall, the filtered recombination maps had masked a total of 264 Mb, or approximately 30.6% of the chromosomes, which is about 5% less than what was filtered out by the callability mask during genotype filtering.

Hotspot analyses

A recombination hotspot is a narrow region of unusually elevated recombination rate. When searching for hotspots in our data, we required the local recombination rate estimate in any inter-SNPs interval to be at least five times higher than a background rate, in keeping with common practice (Raynaud et al., 2023; Singhal et al., 2015). For the background recombination rate, we applied three definitions the (i) mean rate in 40 kb around the interval (20 kb before and 20 kb after), (ii) mean rate in 1Mb around the interval, and

Downloaded from https://academic.oup.com/evolut/advance-article/doi/10.1093/evolut/qpae169/7908996 by guest on 06 December 2024

(iii) mean recombination rate for the whole chromosome as in (Halldorsson et al., 2019). The use of these three different backgrounds resulted in three sets of hotspots. In most cases, several neighboring intervals were identified as being a part of a hotspot, and these intervals were merged using the bedtools v2.29.2 merge command if the distance between such intervals was less than 1 kb. The proportion of overlap between hotspots from different maps was calculated using the intersect command from bedtools v2.29.2 with default parameters, meaning that hotspots were considered overlapping if they shared at least 1bp. When considering the mean recombination rates around hotspots, we normalized the highest point in each hotspot to equal 1.0, so that all hotspots were considered equal. Some hotspots were very long and contained implausibly large fractions of recombination, a phenomenon also reported by other studies (Auton et al., 2012; Hoge et al., 2024). We removed hotspots longer than 5kb from the above analyses and from the search for sequence motifs described below.

Measures of genetic differentiation

To assess the degree of genetic differentiation between the benthic and littoral ecotypes for windows along the genome we calculated F_{ST} and the difference in nucleotide diversity (π), which we call $\Delta(\pi)$. Our F_{ST} calculation implements the Hudson estimator, as defined in equation 10 in Bhatia et al. (2013), using "ratio of averages" to combine estimates across multiple variants. To calculate nucleotide diversity for each ecotype, we divide the average number of differences between any two haplotypes by the number of callable sites in each genomic window. These calculations are implemented in the Fst command of the evogenSuite software, with the --accessibleGenomeBED option providing an inverse of the callability mask. We did this (i) for physical windows of 2Kb (-f option) and (ii) for 20 SNPs windows along the genome (-w option).

Distance from CpG islands and transcription start sites

We used two different definitions of CpG islands (CpGi). First, consistent with (Baker et al., 2017), we used the `maskOutFa` and `cpg_lh` command from UCSC utils. This approach stipulates a minimum of 50% GC content for defining a CpGi. This resulted in 17 000 CpG islands representing a total of 8.5 Mb. Because GC content in the genome is lower in percomorpha that in vertebrates due to less biased gene conversion toward GC (Escobar et al., 2011) and constraining the definition of CpG island by GC content may not be the most appropriate here. We thus also used the `cpgplot -minoe 0.6 -minpc 0` command from `EMBOSS` software (Rice et al., 2000), which resulted in a much greater number (228,069) of CpGi. For the Transcription Starting Sites (TSS), we used the genome annotation described before. We then used the `intersect -v` and `closest` command from `bcftools v.2` to obtain the mean recombination rates in 2Kb and 10Kb associated with the distance of the closest TSS nonoverlapping with a CpGi and vice versa. We finally used the `loess` function in R to obtain the relative recombination rate depending on the distance with the closest TSS or CpGi.

Inference of haplotype blocks

To discover large-scale variation shared by loci along the genome, we used the program `lostruct` that visualizes the local effect of population structure (Li & Ralph, 2018). `lostruct` summarize the pattern of relatedness in a local PCA for nonoverlapping windows along the genome and calculate the dissimilarities between each pair of local PCAs. It then uses multidimensional scaling (MDS) to visualize relationships between windows. We ran `lostruct` for each chromosome separately on 100 SNPs windows. We then plot the first and second axis of the MDS against the genome position. We manually identified 47 regions with high MDS values and high difference in the MDS scores (Supplementary Figure 4). To visualize population structure in `lostruct` outliers, we used `smartPCA` (Patterson et al., 2006) on data filtered for minor allele frequency >=0.05 using plink v1.9 (Purcell et al., 2007) with the `--maf 0.05` option but did not filter based on LD.

Characterization of haplotype blocks

To better understand the evolutionary history of each of the 47 haplotype blocks we calculated several population genetic statistics in these regions. First, we calculated the proportion of heterozygous sites for each individual (H_{ind}) and used the `kmeans` function in R with `k = 2` to assign these values into two clusters. Then, we used the binomial distribution to test if the proportion of high-heterozygosity individuals (in the H_{ind} cluster) was different between the ecotypes. The inbreeding coefficient F was calculated for each SNP using our R code. To estimate the time to the most recent common ancestor (TMRCA) we calculated pairwise d_{xy} (i.e., the number of differences in the nucleotide sequence) among all individuals. We took the maximum value of pairwise d_{xy} within each of the 47 local PCA outliers to approximate the TMRCA as $\widehat{TMRCA} = \frac{Max(d_{xy})}{2 * \mu}$ (acknowledging that this can be a slight underestimate of the true TMRCA of individual sequences because of using unphased data).

We then manually assigned ten control regions of a length of 3.9Mb, corresponding to the mean length of the local PCA outlier regions, that we chose to lie in regions of low MDS1 and MDS2 scores. The relative enrichment for each negative value of F in the local PCA outliers in comparison with the control regions was calculated as the relative proportion of SNPs from each category (outliers vs. controls) in each interval of F values using our R code.

The highly differentiated regions (HDRs) were defined analogously to ref. (Malinsky et al., 2015). We took the top 1% of the F_{ST} values in 20 SNP windows (calculated as described above) and merged the windows that were within 10 kb of each other. This resulted in 352 HDRs, which is very similar to the 344 HDRs found in (Malinsky et al., 2015). The permutation test used to assess significance of mean F_{ST} per each local PCA outlier was implemented using custom R script. Briefly, for each local PCA outlier we sampled 1,000 random genomic windows of the same size, obtaining a null distribution of F_{ST} values, and considered the F_{ST} significantly elevated when it fell within the 5% of this distribution.

PRDM9 ortholog research and distance from ZF binding motif

To find the PRDM9 ortholog in *A. calliptera*, we used the `blastp` command (Altschul et al., 1990) using as query the PRDM9 protein sequence from the Atlantic salmon *Salmo salar* (Gene ID 100380788) against the NCBI RefSeq database (O'Leary et al., 2016) for the species *Astatotilapia calliptera*. Using the best matching protein sequence (XP_026034002.1), we applied a ZF motif predictor (Persikov & Singh, 2014;

Downloaded from https://academic.oup.com/evolut/advance-article/doi/10.1093/evolut/qpae169/7908996 by guest on 06 December 2024

Table 1. Genomic regions underlying recombination rate evolution.

Feature	Δ (r) outlier proportion	Δ (r) outlier excess	Permutation p-value	Mean Δ (r) (log10)
High F_{ST} (top 10%)	13.7%	+58.3%	<0.001	0.109
High Δ (r) (top 10%)	12.8%	+47.8%	<0.001	0.153
Local PCA outliers	31.6%	+47.5%	0.031	0.133
PRDM9 binding sites	0.08%	−8.31%	1	0.106

Persikov et al., 2009) to obtain the position weight matrix representing the binding site prediction based on 11 identified ZF domains. We then used the `FIMO` command from the `MEME Suite` (Bailey et al., 2015) on the reference genome of our species to localize the binding sites. We then used the `intersect -v` and `closest` command from `bcftools v.2` to obtain the mean recombination rates in 2Kb with the distance of the closest ZF binding DNA motif.

Contribution of genomic features to recombination divergence

In Table 1, we summarize how recombination outliers ($\Delta(r)$ outliers) coincide with different genomic regions (10 % regions of higher F_{ST}, high $\Delta(\pi)$, `lostruct` outliers, and predicted PRDM9 binding sites). The excess of $\Delta(r)$ outliers in these regions was calculated by dividing the proportion of $\Delta(r)$ outlier sequence length overlapping these regions by the proportion of the genome taken up by the regions. The significance of excess overlap was calculated using 1,000 permutations with the R package `regioneR v. 1.34.0` (Gel et al., 2016).

Results

Study system and demographic history

We obtained whole genome short read sequences of 159 male individuals, each of which was assigned to either the benthic or littoral ecotype based on a field photograph. Median sequencing coverage was 15.8 × (min = 12.4×, max = 22.1×). After variant calling and filtering, we discarded 1.4 million of multiallelic sites and indels and kept 3.86 million SNPs. To check the validity of the field assignment we used the genetic data to run a PCA (Figure 1B) and reconstructed a Neighbor-Joining tree (Supplementary Figure 5). While no individuals were misassigned, we identified 20 individuals that appeared to be genetically intermediate, which may be the result of recent hybridization. Because our goal was to focus on the differences between the ecotypes, we removed the intermediate individuals (gray in Figure 1B). The final VCF with 139 individuals was composed of 3.3 million biallelic SNPs, and this was used for all the following analyses.

To obtain a more accurate understanding of the historical demographic context of the ecotype divergence, we first used the program SMC++ (Terhorst et al., 2017) to infer the recent changes in effective population size. Consistent with previous results (Malinsky et al., 2015), we find a bottleneck that may be related to the lake colonization, followed by recent demographic expansions in both ecotypes (Figure 1C). This approach also allowed us to reestimate the split time between the two ecotypes, which we now put at ∼2,500 generations ago (95% CI: 1,902 to 5,469 generations). While this is considerably older than reported previously (Malinsky et al., 2015), this difference is primarily a result of using the cichlid-specific mutation rate from (Malinsky et al., 2018) in place of the human mutation rate used in the previous study. We also estimated the amount of gene flow between the ecotypes using fastsimcoal2, with best migration rate estimates being 11.5×10^{-5} for littoral to benthic and 7.01×10^{-5} for benthic to littoral, although higher migration rates of 0.0339 for littoral to benthic and 0.0368 for benthic to littoral had almost equivalent likelihoods and appear more realistic given the number of intermediate individuals we found (Figure 1B; Supplementary Figure 5; see *Methods* section).

Population recombination landscapes differ between ecotypes

First, to quantify the effects of sampling variability and methodological limitations, we divided the individuals from each ecotype into two independent subsets and reconstructed a separate recombination map for each subset (Figure 1D). Thus, we obtained a total of two replicate maps for each ecotype. Spearman correlation between the replicate maps from the same ecotype (within-ecotype) was 0.77 for within-littoral and 0.71 for within-benthic comparisons at 2 kb scale (Figure 2A; see *Methods* section). The relatively low correlation coefficients for the within-ecotype replicates reflect a sensitivity of recombination rate inference to sampling variance. Next, we made recombination landscape comparisons between ecotypes (in gray in Figure 2A) and found that the correlation coefficients were considerably lower than within ecotypes (mean Spearman correlation = 0.57). The difference in median correlations (within vs. between ecotypes) which we denote d_{me} was thus 0.71–0.57 = 0.14 (see *Methods* section). The key result, i.e., that between-ecotype correlations are consistently and considerably lower than for within-ecotype replicates holds across genomic scales from 2 kb to 5 Mb (Supplementary Figure 6). It is also highly consistent across replicate maps obtained by splitting the ecotypes into different subsamples of individuals by permutation (Supplementary Figure 6).

To account for the separation of recent benthic vs. littoral ancestry, which leads to nonrandom sampling of recombination histories, we conducted neutral coalescent simulations assuming that recombination landscapes were the same for both ecotypes. First, we used msprime to simulate genetic data matching our best estimates of population and demographic histories (split time, N_e, and gene flow) that we inferred from empirical data as described above. Based on data simulated under these best-estimate parameters, the within-ecotype vs. between-ecotype difference in median recombination map correlations, which we denote d_{ms} was 0.046 (Figure 2B), which was approximately three times lower than in empirical data (empirical d_{me} = 0.140 vs. simulation d_{ms} = 0.046). We consider this difference ($d_{me} - d_{ms}$) to be a meaningful measure

Downloaded from https://academic.oup.com/evolut/advance-article/doi/10.1093/evolut/qpae169/7908996 by guest on 06 December 2024

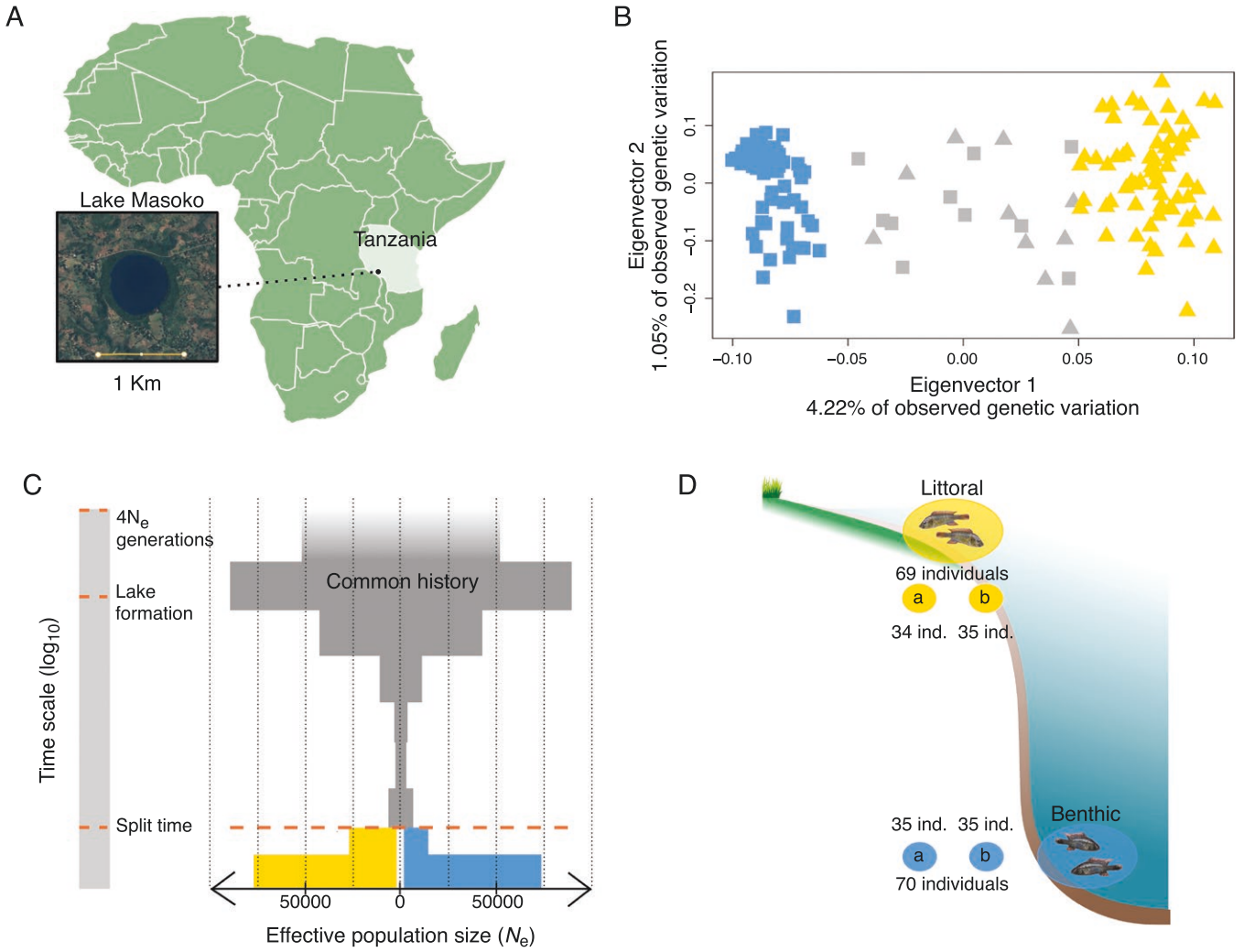

Figure 1. Study system and demographic history. (A) Lake Masoko is a circular small (~670 m diameter) maar-type volcanic crater lake located in the East African rift valley in southern Tanzania. (B) A principal component analysis based on 784,974 SNPs. Some individuals labeled in the field as "benthic" or "littoral" turned out to be genetically admixed. The genetic maps presented and examined in this study are based on individuals with little to no admixture, highlighted in yellow for littoral and blue for benthic. (C) SMC++ inference of demographic history—the changes in effective population sizes (N_e; *x*-axis)—through time (*y*-axis). After the split time, the population sizes for the littoral ecotype are shown in yellow and for the benthic in blue. (D) We divided the individuals from each cichlid ecotype (littoral, benthic) of Lake Masoko into two independent subsets, "subset a" and "subset b." Each subset and the recombination maps inferred using the subsets represent biological replicates.

of recombination landscape evolution and refer to it as the PRDI (see *Methods* section; Figure 2C). The expected value of PRDI is zero if there is no difference in the population recombination landscapes between the ecotypes, and the magnitude of a positive PRDI indicates the degree of change.

Because the PRDI estimate depends on the parameters used in the simulations and because there are uncertainties regarding the correct values for these parameters, we conducted additional simulations across a broad range of split times and using both the higher and lower migration rate estimates from fastsimcoal2. We found that with the higher migration rates, PRDI estimates were consistently high, around 0.14 across the full range of split times, because d_{ms} hovered around zero reflecting the intertwined ancestries of the ecotypes in the presence of high migration (Figure 2C). With the low migration rates, PRDI was lower and depended on the simulated split time, reaching a minimum of 0.07 when we simulated ecotype divergence 10,000 generations ago (Figure 2C). Overall, these results provide evidence that the population recombination landscapes differ between the ecotypes, with the magnitude of divergence estimated by PRDI to be between 0.07 and 0.14.

To further confirm that the observed recombination differences between benthic and littoral were not driven by technical artifacts (e.g., implausibly large false positive hotspots, Auton et al., 2012) in regions of the genome where inference is especially error-prone, we applied a stringent filtering mask (see *Methods* section). This eliminated a substantial proportion of noise from the inferred recombination maps across all genomic scales from 2 kb to 5 Mb (Supplementary Figure 6). We repeated all the analyses above using these filtered maps and found that, despite the strictness of the filtering, the results, and specifically the d_{me} and PRDI estimates were virtually the same as for the raw maps (Supplementary Figures 6 and 7A). We also verified that the d_{ms} estimates and PRDI do not depend on the reference map that is used as input for the coalescent simulations (Supplementary Figure 7B).

In Lake Masoko *A. calliptera*, the ancestry for the sampled individuals extends substantially beyond the ecotype split

Downloaded from https://academic.oup.com/evolut/advance-article/doi/10.1093/evolut/qpae169/7908996 by guest on 06 December 2024

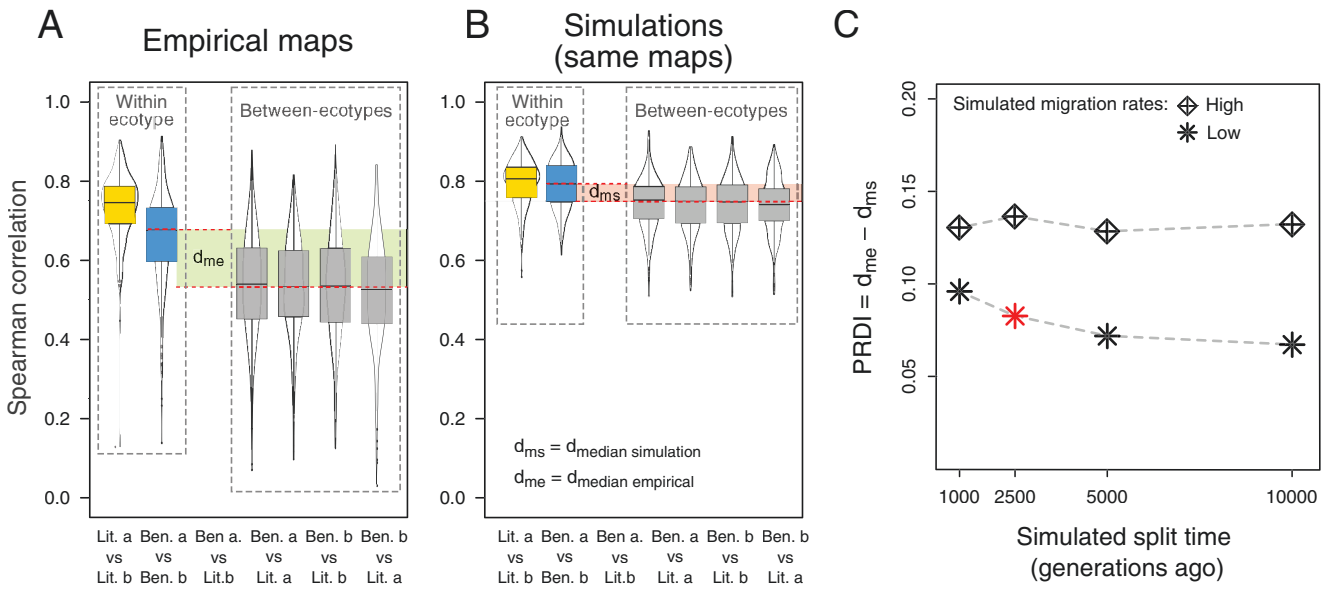

Figure 2. Rapid evolution of the recombination landscapes between the ecotypes. (A) Spearman correlation between recombination maps on 2 kb scale with each datapoint representing a 5 Mb genomic interval. The correlations in between-ecotype comparisons are substantially lower than in the within-ecotype replicate comparisons, with the difference measured by d_{me} (see *Methods* section). Blue and yellow colors refer to the benthic and littoral ecotypes, respectively. (B) Correlations among recombination maps inferred from simulations that used the same (benthic) recombination landscape as input/reference for the simulations. The simulation parameters were based on our best estimates, matching the demographic history of Lake Masoko. The between-ecotype comparisons also show lower correlations than within-ecotype replicates, reflecting the recent separate ancestries of the ecotypes, but this difference (d_{ms}) is lower than d_{me}. (C) A measure of recombination landscape evolution, the Population Recombination Divergence Index (PRDI) as a function of different split times and migration rates used for the simulations. Across the entire parameter space, PRDI varies between 0.07 and 0.14. The datapoint highlighted in red is based on the best-estimate parameters shown in (B).

time (Figure 1C), which means that the inferred recombination maps for each ecotype can be interpreted as a mixture of two time periods: (i) recombination events that happened in the common history of the ecotypes and (ii) events that happened after their split. To estimate the contribution of each of these epochs, we used coalescent simulations and found that, across ten simulations with the best-estimate split time of $\sim$ 2,500 generations ago, an average 34.9% of recombination events that changed the genealogy of the sample occurred after the split (min = 34.4%; max = 35.7%; see *Methods* section). These are the recombination events that make up the differences between the recombination landscapes between the ecotypes.

Recombination landscapes are highly heterogenous, and a large proportion of events tends to occur in so-called "hotspots" (Coop & Przeworski, 2007; Peñalba & Wolf, 2020). We quantified the heterogeneity of recombination along the genome in the *A. calliptera* of Lake Masoko and found that 50% of all events were concentrated in less than 9.6% of the genome (Supplementary Figure 8A; min: 8.9%; max: 10.4%, depending on the ecotype and the subsample). Therefore, the concentration of recombination in hotspots can be considered intermediate—substantially lower than, for example, in humans, but higher than, for example, in the plant Arabidopsis thaliana (Supplementary Figure 8A). Using a definition of hotspots as having at least 5x higher recombination rate than the 500 kb of surrounding sequence, we found on average 2,322 hotspots in each recombination map (between 2,275 and 2,345; Supplementary Table 1). Only 41.5% of hotspots were shared when comparing "a" and "b" replicate maps within the same ecotype, showing that hotspot detection is particularly sensitive to sampling variance. Nevertheless, we again observed the same pattern as for correlations – the comparisons between benthic and littoral maps showed even less hotspot overlap than expected based on simulations (Supplementary Figure 8). Qualitatively similar results were obtained using three different hotspot definitions (Supplementary Figure 8C).

Characterizing recombination rate evolution

The changes in recombination landscapes measured by LD-based maps are based on effective recombination—the recombination events present in the ancestry of the sampled individuals—and are directly influenced by divergent selection between the ecotypes. Therefore, we next investigated the relationship between recombination rates and genetic divergence between the ecotypes in terms of allele frequencies (F_{ST}) and levels of nucleotide diversity ($\Delta(\pi)$) along the genome. As expected, we found that PRDI generally increased with F_{ST} and $\Delta(\pi)$, consistent with an effect of divergent selection on genotypes (Figure 3A). However, we observed positive PRDI not only in regions of particularly high F_{ST} or $\Delta(\pi)$, but also at low levels of genetic divergence (Figure 3A). Therefore, the evolution of recombination landscapes is not driven by divergent selection on genotypes (e.g., selective sweeps), and/or the effect that differences in π can have on the accuracy of recombination inference (Supplementary Figure 3; Raynaud et al., 2023). Moreover, equivalent results were obtained for filtered maps, providing additional confidence regarding the robustness of these conclusions (Supplementary Figure 9A).

To explore how the differences in recombination between the ecotypes are distributed across the genome, we calculated the mean difference in recombination rates between the inferred genetic maps in 100 kb windows (see *Methods* section). The average recombination distance between benthic and littoral maps (denoted $\Delta(r)_b$) was greater than the

Downloaded from https://academic.oup.com/evolut/advance-article/doi/10.1093/evolut/qpae169/7908996 by guest on 06 December 2024

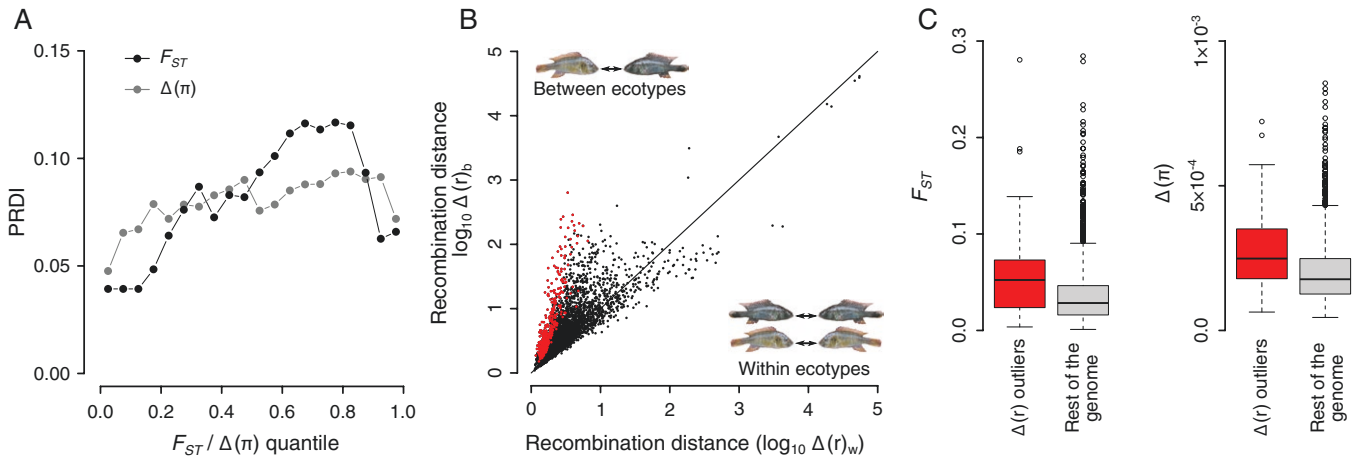

Figure 3. Interplay between recombination and genomic differentiation. (A) The Population Recombination Divergence Index (PRDI)_as a function of benthic-littoral F_{ST} and of $\Delta(\pi)$. (B) A scatterplot of average recombination map distances in 100 kb windows, within biological replicates—$\Delta(r)_w$ and between the ecotypes—$\Delta(r)_b$. Datapoints corresponding to "$\Delta(r)$ outliers" are highlighted in red color. For details, see text. (C) Comparing the distributions of benthic-littoral F_{ST} values and of $\Delta(\pi)$ within and outside of $\Delta(r)$ outliers.

distance for within-ecotype replicates (denoted $\Delta(r)_w$) in 83.1% of the windows (Figure 3B). This metric allowed us to identify genomic regions with rapidly diverging recombination rates. In the following, we use the "net recombination distance" $\Delta(r) = \Delta(r)_b - \Delta(r)_w$ and give particular focus to "$\Delta(r)$ outliers"—regions where the between-ecotype distance is more than three standard deviations higher than within ecotypes. These outliers correspond to 42.7Mb of sequence, which is about 5% of the genome.

We found that $\Delta(r)$ outliers are not uniformly distributed across chromosomes; for example, they cover only 1.2% of chromosome 15 (LS420033.2) but over 11.9% of chromosome 1 (LS420019.2) (Supplementary Figure 10A). Furthermore, the proportion of outliers across chromosomes is positively correlated with average per-chromosome F_{ST}. Although this chromosome-wide link is only moderately strong and not statistically significant (Pearson correlation = 0.18; $p = 0.42$; Supplementary Figure 10B), when looking directly at $\Delta(r)$ outliers, we found that both F_{ST} and $\Delta(\pi)$ were significantly elevated (Mann–Whitney U test: $p = 1.88 \times 10^{-6}$ for F_{ST} and $p = 5.5 \times 10^{-7}$ for $\Delta(\pi)$), clearly confirming that there is an association between allele frequency divergence and the most rapidly evolving population recombination landscapes (Figure 3C).

Given the rapid evolution of recombination rates across the genome, we wanted to verify whether the PRDM9 mechanism may be active in Lake Masoko *A. calliptera*. As in several other percomorph species (Cavassim et al., 2022), we found one incomplete PRDM9 ortholog missing the KRAB and SSXRD domains that appear to be necessary for PRDM9 to direct recombination (Baker et al., 2017). Consistent with this, recombination rates were elevated at and near CpG islands (~1.2× higher; Supplementary Figure 11) and transcription start sites (TSS; ~1.3× higher; Supplementary Figure 11), a pattern that is similar to that reported previously for swordtail fish (Baker et al., 2017). Because the PRDM9 ZF array in *A. calliptera* was intact, we predicted its binding sites across the genome and found no increase in recombination rates at or near the binding sites (Supplementary Figure 11). Overall, these results confirm that PRDM9 does not direct recombination in *A. calliptera* and, therefore, cannot contribute to the rapid evolution of recombination rates in this system.

Large haplotype blocks contribute to evolution of recombination rates

Ecotype divergence and speciation in the face of gene flow are often facilitated by regions of suppressed recombination, which allow a buildup of linkage between multiple loci under divergent selection (Faria et al., 2019b). Nonrecombining haplotype blocks can be revealed as extended regions of the genome with distinct population structures substantially different from the genome-wide average (Ma & Amos, 2018; Mérot, 2020; Todesco et al., 2020). To look for such regions, we used a local PCA approach (Li & Ralph, 2018) and identified a total of 47 outliers (Supplementary Figures 4 and 12), ranging in size between 550 kb and 25.7 Mb (mean 3.9 Mb) and covering a total 21.4% of the genome. Importantly, these regions contain 31.6% of $\Delta(r)$ outliers and thus contribute disproportionately to the observed differences in recombination rates between the ecotypes (Table 1). Consistent with this, we found that the signal of recombination rate evolution measured by PRDI is about 10% stronger in the local PCA outliers than in the rest of the genome (Supplementary Figure 13). At the same time, it should be emphasized that there is also a strong PRDI signal of recombination rate evolution outside of these blocks (Supplementary Figure 13).

Large haplotype blocks can be the result of a lack of recombination between alternative haplotypes segregating in Lake Masoko. We hypothesized that, on average, recombination would be reduced in the ecotype with a higher proportion of individuals who are heterozygous for such nonrecombining haplotypes. Therefore, for each local PCA outlier region, we clustered the individuals based on individual heterozygosity, that is, the proportion of heterozygous sites per individual (H_{ind}). An example of a local PCA outlier region associated with differences in recombination is shown in Figure 4, with strong benthic vs. littoral clustering by H_{ind} illustrated in Figure 4C.

As predicted, across the 47 haplotype blocks, we found a significant positive association between $\Delta(r)$, the net recombination distance between ecotypes and the degree of ecotype clustering by H_{ind} (Figure 5A). Additionally, PRDI was substantially elevated in the haplotype blocks with significant of ecotype clustering by H_{ind} (Supplementary Figure 13). In contrast, the association of $\Delta(r)$ with allele frequency

Downloaded from https://academic.oup.com/evolut/advance-article/doi/10.1093/evolut/qpae169/7908996 by guest on 06 December 2024

divergence (mean F_{ST} per haplotype block) with was less pronounced and not statistically significant (Figure 5B).

Next, we focused on gathering evidence regarding the nature and origin of individual haplotype blocks. Local PCA outliers could be explained, for example, by linked selection, recent admixture from outside of Lake Masoko, or by locally low recombination rates. We found that some of the regions we identified by local PCA show signatures characteristic of polymorphic inversions, including (i) long haplotypes with consistent sharp edges in multiple individuals, (ii) distinct homozygote vs. heterozygote clusters in PCA, and (iii) unusually high values of individual heterozygosity (Supplementary Figures 14 and 15).

To investigate if the heterozygous state is overrepresented within local PCA outliers, which would be consistent with a form of balancing selection, we compared the distribution of inbreeding coefficient (F) per SNP against SNPs from a set of control regions (see *Methods* section). We found a significant enrichment of SNPs with negative F—that is, excess of heterozygotes—in the local PCA outliers, with up to 9-fold enrichment for the SNPs at the lowest values of F (Figure 5C). We expected that the haplotype blocks with the lower value of inbreeding coefficient might have been maintained as polymorphism for a long time through the effect of balancing selection. Therefore, for each PCA outlier, we estimated the time TMRCA. Perhaps surprisingly, we found that the ancestry of most of these regions was of a similar age as the genome-wide average, with only a few that were clearly much older (Figure 5D). There was a weak and not statistically significant trend of older regions having a lower inbreeding coefficient. Therefore, we conclude that the effect of balancing selection on maintaining polymorphic haploblocks appears to be limited in the Lake Masoko ancestry.

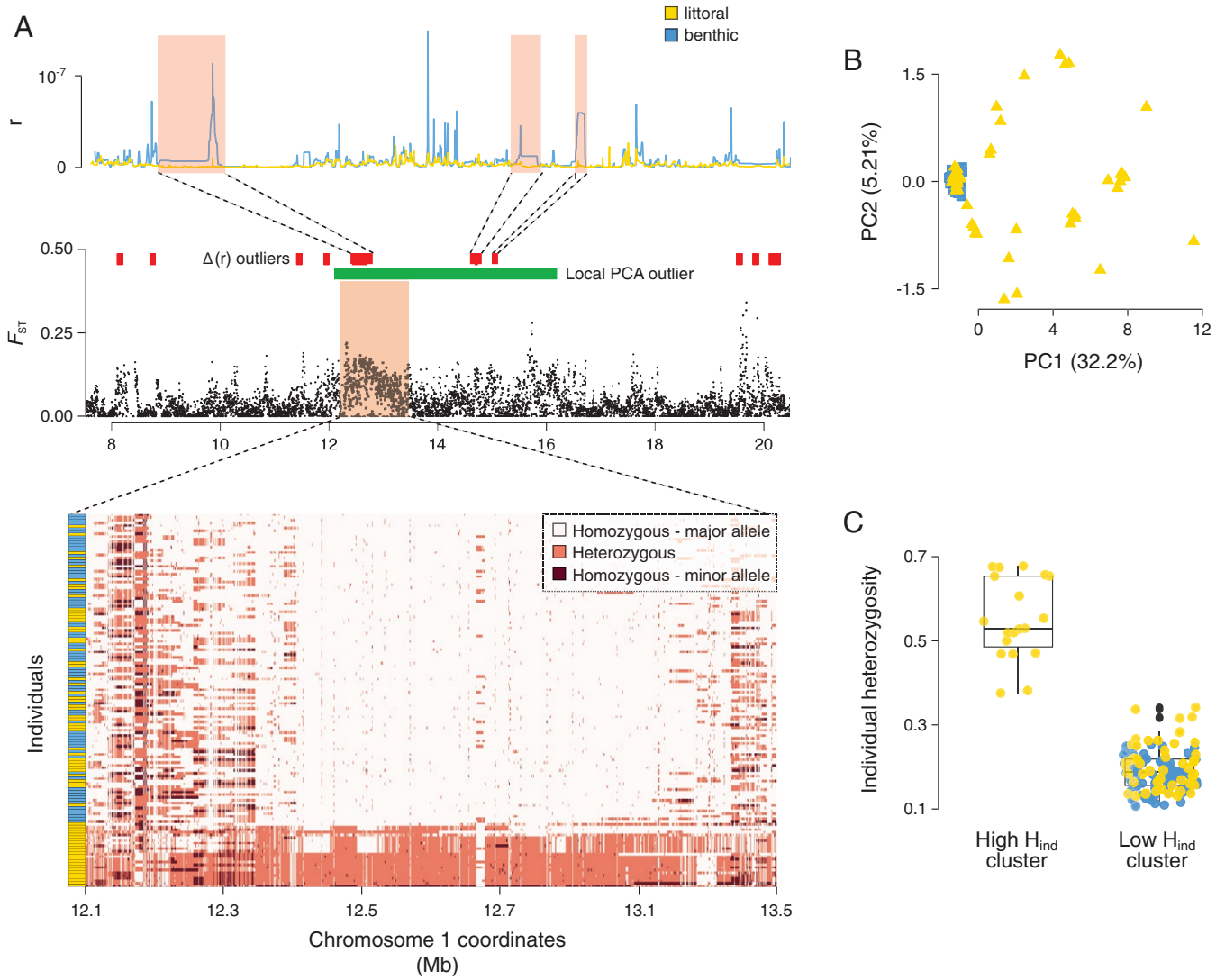

Figure 4. A local principal component analysis (PCA) outlier associated with a difference in recombination. (A) An example region of chromosome 1 (LS420019.2) illustrates the link between a large haplotype block, F_{ST}, and recombination rate evolution. Top: average recombination rates in the two ecotypes. There is lower recombination in the littoral ecotype in this region. Middle: F_{ST} is moderately elevated in the haplotype block region. Bottom: Genotypes at SNPs with minor allele frequency >5%. (B) Local PCA in the haplotype block region from (A). The genetic structure within this region differs considerably from the PCA obtained from whole-genome data (Figure 1B). (C) Clustering based on individual heterozygosity (H_{ind})—i.e., the proportion of variable sites that are heterozygous in an individual—in the haplotype block region from (A). All the individuals in the cluster with high H_{ind} belong to the littoral ecotype.

Downloaded from https://academic.oup.com/evolut/advance-article/doi/10.1093/evolut/qpae169/7908996 by guest on 06 December 2024

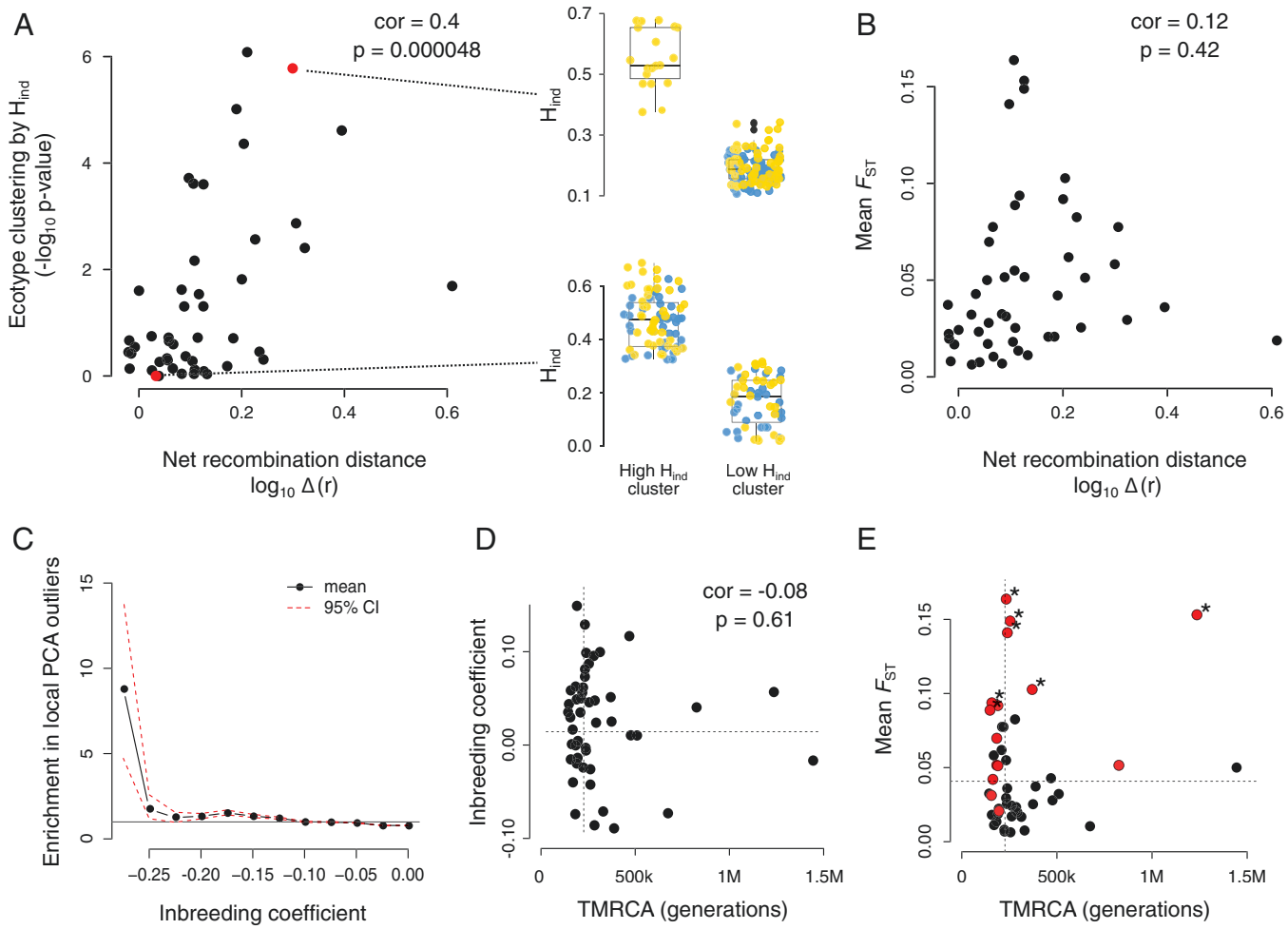

Figure 5. Characterizing haplotype blocks and their link with recombination in Lake Masoko. (A) Relationship, across the 47 local principal component analysis (PCA) outliers, between the degree of ecotype clustering by H_{ind} and net between-ecotype recombination map distances. We highlighted examples of strong (top) and weak (bottom) ecotype clustering by H_{ind}. (B) Relationship between mean F_{ST} and net between-ecotype recombination map distances. (C) For each value of the inbreeding coefficient, we show the relative enrichment of SNPs in the PCA outlier regions in comparison with the control regions. (D) Relationship between inbreeding coefficient and TMRCA. The horizontal dashed line represents the mean inbreeding coefficient calculated for the control regions, while the vertical dashed line represents an estimate of genome-wide average TMRCA. (E) Relationship between TMRCA and mean F_{ST}. The local PCA outliers that contain island(s) of differentiation are colored in red. We used stars to mark the regions with significantly elevated F_{ST} (permutation test p-value < 0.05).

Finally, we investigated in more detail the link between allele frequency divergence and the local PCA outliers in Lake Masoko. We found that only seven out of the 47 regions had a significantly elevated average level of F_{ST}; of these, there were six regions of average age and one old region whose ancestry dates back almost 1.5M generations ago (Figure 5E). Islands of differentiation—also referred to as highly diverged regions (HDRs), defined as in ref. (Malinsky et al., 2015)—also did not appear in many of the local PCA outliers. While four of the local PCA outliers contained more than ten HDRs each, in total, only 15 local PCA outliers contained at least one HDR (Figure 5E; Supplementary Table 2).

Discussion

The landscape of recombination across the genome is not static but evolves through time. In this study, we undertook a holistic investigation of recombination rate evolution between two ecotypes that diverged very recently, in sympatry with gene flow, and adapted rapidly across multiple phenotypic traits to new lake environments (Malinsky et al., 2015). Given ecotype divergence at many—at least a hundred or so—genomic loci (Malinsky et al., 2015), we could expect that recombination, or the lack thereof, would be important in bringing together and keeping together the alleles that are beneficial in each environment (Battlay et al., 2023; Ortiz-Barrientos et al., 2016; Schluter & Rieseberg, 2022; Todesco et al., 2020). Our findings reveal, characterize, and quantify substantial differences in population recombination rates between the ecotypes, complementing previous studies of divergence in ecology, mate choice, and allele frequencies (Malinsky et al., 2015), methylation (Vernaz et al., 2022), gene expression (Carruthers et al., 2022), and in sex determination (Munby et al., 2021).

The use of LD-based maps, integrating over a large number of recombination events throughout the ancestry of the ecotypes, allowed us to analyze recombination landscape evolution at a fine scale. Despite the challenges inherent in such comparisons, we used subsampling (replicate maps from the same ecotype) and extensive simulations to derive

Downloaded from https://academic.oup.com/evolut/advance-article/doi/10.1093/evolut/qpae169/7908996 by guest on 06 December 2024

meaningful measures of evolution in population landscapes. We show how the recombination differences are distributed across the genome, how they are linked to allele frequency divergence (F_{ST}; $\Delta(\pi)$) and to large haplotype blocks identified by local PCA.

As we expected, we found a link between recombination rate divergence and regions of high allele frequency divergence. This link can be expected for a variety of reasons—for example, local recombination rates can impact inference of F_{ST} (Booker et al., 2020), they partly determine the extent of background selection which can affect F_{ST} (Burri et al., 2015; Matthey-Doret & Whitlock, 2019), and they play a major role in in the formation of genomic islands of divergence during speciation with gene flow (Cruickshank & Hahn, 2014; Feder et al., 2012). Moreover, selection on genotypes (e.g., incomplete selective sweeps) is directly reflected in LD-based recombination maps (in contrast to gamete-based or pedigree-based maps) (Coop & Przeworski, 2007; Peñalba & Wolf, 2020). Given these considerations, it is surprising that allele frequency divergence does not appear to be the main driver of the observed recombination rate differences. For example, only 13.7% of significant differences in recombination, i.e., $\Delta(r)$ outliers, are colocated with the top 10% of F_{ST} (Table 1), and PRDI is substantially positive in regions where F_{ST} and $\Delta(\pi)$ are zero.

A stronger link was identified between recombination rate evolution and local PCA outliers, which are characterized by large haplotype blocks and comprise more than a fifth of the genome. Some of these blocks carry clear signatures characteristic of inversions. However, it remains to be seen what mechanisms give rise to the remaining blocks where the haplotype structure and local relationships among individuals are more complex. In many cases we see many clusters in the local PCA, which is consistent with signatures of multiple overlapping inversions segregating locally in the population (Faria et al., 2019a). There is growing evidence that haplotype blocks caused by large inversions are present in many species (Wellenreuther & Bernatchez, 2018), often comprise considerable proportions of the genome, and have clear links to adaptation and diversification in both animals (Blumer et al., 2024; Faria et al., 2019a; Harringmeyer & Hoekstra, 2022; Reeve et al., 2023) and plants (Battlay et al., 2023; Todesco et al., 2020). Therefore, future work should focus on exploring the dynamics of inversions in the Lake Masoko system in greater detail.

We have focused on large haplotype blocks because the local PCA approach facilitates their study from short-read data in a population genomic context. However, the recombination suppression effect of structural variants does not depend on the size of the variant region, and other types of structural variation can also suppress recombination (Kent et al., 2017; Mérot et al., 2020; Rowan et al., 2019). Therefore, it is likely that shorter structural variants are responsible for at least some of the remaining $\Delta(r)$ outliers, which are not accounted for in our current study. Further reduction in the cost of long-read sequencing will, among other benefits, enable more unbiased population-scale analyses of structural variants and their roles in the evolution of recombination landscapes (Coster et al., 2021).

We found substantial recombination landscape evolution where the ecotypes cluster by individual heterozygosity levels (Figure 5A; Supplementary Figure 13). This could be because inversions prevent crossover formation only in the gametes of heterozygous individuals (Faria et al., 2019b). A recent study using pedigree-based recombination maps raises an additional possibility. Venu et al. (2024) found that regions with higher heterozygosity have lower recombination rates due to haplotype incompatibilities between diverging ecotypes. Therefore, the same mechanism could play a role in the divergence of recombination landscapes seen in Lake Masoko.

More general and important open questions concerning the nature of selection on recombination. Which of the observed differences in population recombination rates are a result of changes in the distribution of crossovers during gamete formation? And which changes are an indirect effect of subsequent selection for or against specific genetic variants and recombinant haplotypes? Is there any effect of recombination plasticity (Henderson & Bomblies, 2021) arising from the ecotypes occupying different environments? These questions are not possible to answer conclusively with LD-based estimates alone. Future comparison of our LD-based maps against recombination landscapes obtained by sequencing of gametes and/or individuals related by pedigrees will likely shed further light on this question (Peñalba & Wolf, 2020).

Our finding of rapid recombination rate evolution, while consistent with some previous studies (Déserts et al., 2021; Shanfelter et al., 2019), seems to conflict with the current paradigm of evolutionary stability of recombination landscapes in species lacking the PRDM9 mechanism (Lam & Keeney, 2015; Singhal et al., 2015). However, this is not as surprising as it may appear because it is common that different tempos of molecular evolution are observed between micro- and macro-evolutionary timescales (Rolland et al., 2023). For example, recombination suppression by inversions may be a temporary phenomenon and may disappear once one of the inversion alleles rises to fixation. At the same time, evidence is emerging that even in many species with an intact PRDM9 mechanism, a large fraction of recombination can take place outside of PRDM9 directed hotspots, so the dichotomy of mechanisms may not be as clear as previously thought (Hoge et al., 2024; Joseph et al., 2024).

Adaptation and organismal diversification are increasingly seen as multidimensional and combinatorial, typically with the involvement of multiple polygenic traits and epistasis (Barton, 2022; Marques et al., 2019; Yeaman, 2022), and the relative genetic distances between the loci involved constitute key parameters. Comparative studies are starting to shed light on recombination landscape evolution across populations and species with different demographic histories, genomic architectures, ecological contexts, and divergence times. However, this is still typically done at a rough Megabase-scale resolution (Brazier & Glémin, 2022; Haenel et al., 2018). We adopted the LD-based approach, enabling us to infer fine-scale rates and show that they can evolve rapidly. We envisage that the large and growing amount of population genomic data available will enable the construction and comparisons of many LD-based maps, such as in our current study. Together with advances in gamete typing and pedigree-based methods, this will make recombination rates and their fine-scale evolution into integral parts of future genomic studies of adaptation and speciation.

Supplementary material

Supplementary material is available online at *Evolution*.

Downloaded from https://academic.oup.com/evolut/advance-article/doi/10.1093/evolut/qpae169/7908996 by guest on 06 December 2024

Data availability

Genomic DNA from all 336 individuals used in this study is available from the NCBI Short Read Archive (BioProject ID: PRJEB27804). The VCF file and recombination maps are available on DataDryad (https://doi.org/10.5061/dryad.v15dv425w). The code used to analyze the data is available on GitHub (https://github.com/MarionTalbi/MasokoPaper).

Author contributions

M.M. conceived and designed the study with input from M.T.; M.T. conducted the analyses with guidance from M.M.; G.F.T. led the fieldwork, identified samples, and obtained the genome data. M.T. and M.M. wrote the manuscript, with comments from G.F.T.

Funding

This work was funded by a Swiss National Foundation (SNSF) award to M.M. (grant: 193464), and a Leverhulme Trust award to G.F.T. (grant: RPG2014-214).

Conflict of interest: The authors declare no conflict of interest.

Acknowledgements

We would like to thank Claire Mérot, Catherine Peichel, Molly Przeworski, and Ole Seehausen for helpful discussions and comments on the manuscript and Jerome Kelleher and Georgia Tsambos for helpful tips on using msprime. We thank the Sanger Institute sequencing core for DNA sequencing. We also thank the Tanzania Fisheries Research Institute for their assistance and support and COSTECH for a series of research permits. Finally, we would like to thank the reviewers for their time and thoughtful suggestions, which helped us to considerably improve this manuscript.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Auton, A., Fledel-Alon, A., Pfeifer, S., Venn, O., Ségurel, L., Street, T., Leffler, E. M., Bowden, R., Aneas, I., Broxholme, J., Humburg, P., Iqbal, Z., Lunter, G., Maller, J., Hernandez, R. D., Melton, C., Venkat, A., Nobrega, M. A., Bontrop, R., … McVean, G. (2012). A fine-scale chimpanzee genetic map from population sequencing. *Science (New York, N.Y.)*, *336*(6078), 193–198. https://doi.org/10.1126/science.1216872

Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME suite. *Nucleic Acids Research*, *43*(W1), W39–W49. https://doi.org/10.1093/nar/gkv416

Baker, Z., Przeworski, M., & Sella, G. (2023). Down the Penrose stairs, or how selection for fewer recombination hotspots maintains their existence. *eLife*, *12*, e83769. https://doi.org/10.7554/eLife.83769

Baker, Z., Schumer, M., Haba, Y., Bashkirova, L., & Elife, C. H. (2017). Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates. *eLife*, *6*, e24133. https://doi.org/10.7554/eLife.24133

Barker, P., Williamson, D., Gasse, F., & Gibert, E. (2003). Climatic and volcanic forcing revealed in a 50,000-year diatom record from Lake Massoko, Tanzania. *Quaternary Research*, *60*, 9–9. https://doi.org/10.1016/j.yqres.2003.07.001

Barton, N. H. (2020). On the completion of speciation. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *375*(1806), 20190530. https://doi.org/10.1098/rstb.2019.0530

Barton, N. H. (2022). The "New synthesis". *Proceedings of the National Academy of Sciences of the United States of America*, *119*(30), e2122147119. https://doi.org/10.1073/pnas.2122147119

Battlay, P., Wilson, J., Bieker, V. C., Lee, C., Prapas, D., Petersen, B., Craig, S., Boheemen, L. van, Scalone, R., de Silva, N. P., Sharma, A., Konstantinović, B., Nurkowski, K. A., Rieseberg, L. H., Connallon, T., Martin, M. D., & Hodgins, K. A. (2023). Large haploblocks underlie rapid adaptation in the invasive weed *Ambrosia artemisiifolia*. *Nature Communications*, *14*, 1717. https://doi.org/10.1038/s41467-023-37303-4

Baudat, F., Buard, J., Grey, C., Fledel-Alon, A., Ober, C., Przeworski, M., Coop, G., & Massy, B. de (2010). PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science (New York, N.Y.)*, *327*, 836–840. https://doi.org/10.1126/science.1183439

Baumdicker, F., Bisschop, G., Goldstein, D., Gower, G., Ragsdale, A. P., Tsambos, G., Zhu, S., Eldon, B., Ellerman, E. C., Galloway, J. G., Gladstein, A. L., Gorjanc, G., Guo, B., Jeffery, B., Kretzschumar, W. W., Lohse, K., Matschiner, M., Nelson, D., Pope, N. S., … Kelleher, J. (2021). Efficient ancestry and mutation simulation with msprime 1.0. *Genetics*, *220*, iyab229. https://doi.org/10.1093/genetics/iyab229

Bhatia, G., Patterson, N., Sankararaman, S., & Price, A. L. (2013). Estimating and interpreting FST: The impact of rare variants. *Genome Research*, *23*(9), 1514–1521. https://doi.org/10.1101/gr.154831.113

Blumer, L. M., Burskaia, V., Artiushin, I., Saha, J., Garcia, J. C., Elkin, J., Fischer, B., Zhou, C., Gresham, S., Malinsky, M., Linderoth, T., Sawasawa, W., Bista, I., Hickey, A., Kucka, M., Louzada, S., Zatha, R., Yang, F., Rusuwa, B., … Svardal, H. (2024). Introgression dynamics of sex-linked chromosomal inversions shape the Malawi cichlid adaptive radiation. *bioRxiv* 2024.07.28.605452

Booker, T. R., Yeaman, S., & Whitlock, M. C. (2020). Variation in recombination rate affects detection of outliers in genome scans under neutrality. *Molecular Ecology*, *29*(22), 4274–4279. https://doi.org/10.1111/mec.15501

Brand, C. L., Cattani, M. V., Kingan, S. B., Landeen, E. L., & Presgraves, D. C. (2018). Molecular evolution at a meiosis gene mediates species differences in the rate and patterning of recombination. *Current Biology*, *28*(8), 1289–1295.e4. https://doi.org/10.1016/j.cub.2018.02.056

Brazier, T., & Glémin, S. (2022). Diversity and determinants of recombination landscapes in flowering plants. *PLoS Genetics*, *18*(8), e1010141. https://doi.org/10.1371/journal.pgen.1010141

Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., Suh, A., Dutoit, L., Bures, S., Garamszegi, L. Z., Hogner, S., Moreno, J., Qvarnström, A., Ružić, M., Sæther, S. -A., Saetre, G. -P., Török, J., & Ellegren, H. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. *Genome Research*, *25*, 1656–1665. http://doi.org/10.1101/gr.196485.115

Butlin, R. K., Servedio, M. R., Smadja, C. M., Bank, C., Barton, N. H., Flaxman, S. M., Giraud, T., Hopkins, R., Larson, E. L., Maan, M. E., Meier, J., Merrill, R., Noor, M. A. F., Ortiz-Barrientos, D., & Qvarnström, A. (2021). Homage to Felsenstein 1981, or why are there so few/many species? *Evolution*, *75*(5), 978–988. https://doi.org/10.1111/evo.14235

Carruthers, M., Edgley, D. E., Saxon, A. D., Gabagambi, N. P., Shechonge, A., Miska, E. A., Durbin, R., Bridle, J. R., Turner, G. F., & Genner, M. J. (2022). Ecological speciation promoted by divergent regulation of functional genes within African cichlid fishes. *Molecular Biology and Evolution*, *39*(11), msac251. https://doi.org/10.1093/molbev/msac251

Cavassim, M. I. A., Baker, Z., Hoge, C., Schierup, M. H., Schumer, M., & Przeworski, M. (2022). PRDM9 losses in vertebrates are

Downloaded from https://academic.oup.com/evolut/advance-article/doi/10.1093/evolut/qpae169/7908996 by guest on 06 December 2024

coupled to those of paralogs ZCWPW1 and ZCWPW2. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(9), e2114401119. https://doi.org/10.1073/pnas.2114401119

Coop, G., & Przeworski, M. (2007). An evolutionary view of human recombination. *Nature Reviews. Genetics*, *8*(1), 23–34. https://doi.org/10.1038/nrg1947

Coster, W. D., Weissensteiner, M. H., & Sedlazeck, F. J. (2021). Towards population-scale long-read sequencing. *Nature Reviews Genetics*, *22*, 572–587. https://doi.org/10.1038/s41576-021-00367-3

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*(13), 3133–3157. https://doi.org/10.1111/mec.12796

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008. https://doi.org/10.1093/gigascience/giab008

DePristo, M. A. M., Banks, E. E., Poplin, R. R., Garimella, K. V. K., Maguire, J. R. J., Hartl, C. C., Philippakis, A. A. A., del Angel, G. G., Rivas, M. A. M., Hanna, M. M., McKenna, A. A., Fennell, T. J. T., Kernytsky, A. M. A., Sivachenko, A. Y. A., Cibulskis, K. K., Gabriel, S. B. S., Altshuler, D. D., & Daly, M. J. M. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*, 491–498. https://doi.org/10.1038/ng.806

Déserts, A. D. des, Bouchet, S., Sourdille, P., & Servin, B. (2021). Evolution of recombination landscapes in diverging populations of bread wheat. *Genome Biology and Evolution*, *13*, evab152. https://doi.org/10.1093/gbe/evab152

Duranton, M., Allal, F., Fraïsse, C., Bierne, N., Bonhomme, F., & Gagnaire, P. -A. (2018). The origin and remolding of genomic islands of differentiation in the European sea bass. *Nature Communications*, *9*(1), 2518. https://doi.org/10.1038/s41467-018-04963-6

Escobar, J. S., Glémin, S., & Galtier, N. (2011). GC-biased gene conversion impacts ribosomal DNA evolution in vertebrates, angiosperms, and other eukaryotes. *Molecular Biology and Evolution*, *28*(9), 2561–2575. https://doi.org/10.1093/molbev/msr079

Excoffier, L., Marchi, N., Marques, D. A., Matthey-Doret, R., Gouy, A., & Sousa, V. C. (2021). fastsimcoal2: Demographic inference under complex evolutionary scenarios. *Bioinformatics*, *37*(24), 4882–4885. https://doi.org/10.1093/bioinformatics/btab468

Faria, R., Chaube, P., Morales, H. E., Larsson, T., Lemmon, A. R., Lemmon, E. M., Rafajlović, M., Panova, M., Ravinet, M., Johannesson, K., Westram, A. M., & Butlin, R. K. (2019a). Multiple chromosomal rearrangements in a hybrid zone between *Littorina saxatilis* ecotypes. *Molecular Ecology*, *28*(6), 1375–1393. https://doi.org/10.1111/mec.14972

Faria, R., Johannesson, K., Butlin, R. K., & Westram, A. M. (2019b). Evolving inversions. *Trends in Ecology and Evolution*, *34*(3), 239–248. https://doi.org/10.1016/j.tree.2018.12.005

Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics: TIG*, *28*(7), 342–350. https://doi.org/10.1016/j.tig.2012.03.009

Feldman, M. W., Otto, S. P., & Christiansen, F. B. (1996). Population genetic perspectives on the evolution of recombination. *Annual Review of Genetics*, *30*, 261–295. https://doi.org/10.1146/annurev.genet.30.1.261

Felsenstein, J. (1981). Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution; International Journal of Organic Evolution*, *35*(1), 124–138. https://doi.org/10.1111/j.1558-5646.1981.tb04864.x

Gel, B., Díez-Villanueva, A., Serra, E., Buschbeck, M., Peinado, M. A., & Malinverni, R. (2016). regioneR: An R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics*, *32*(2), 289–291. https://doi.org/10.1093/bioinformatics/btv562

Genestier, A., Duret, L., & Lartillot, N. (2024). Bridging the gap between the evolutionary dynamics and the molecular mechanisms of meiosis: A model based exploration of the PRDM9 intra-genomic Red Queen. *PLOS Genetics*, *20*, e1011274.

Haenel, Q., Laurentino, T. G., Roesti, M., & Berner, D. (2018). Meta-analysis of chromosome-scale crossover rate variation in eukaryotes and its significance to evolutionary genomics. *Molecular Ecology*, *27*(11), 2477–2497. https://doi.org/10.1111/mec.14699

Halldorsson, B. V., Palsson, G., Stefansson, O. A., Jonsson, H., Hardarson, M. T., Eggertsson, H. P., Gunnarsson, B., Oddsson, A., Halldorsson, G. H., Zink, F., Gudjonsson, S. A., Frigge, M. L., Thorleifsson, G., Sigurdsson, A., Stacey, S. N., Sulem, P., Masson, G., Helgason, A., Gudbjartsson, D. F., … Stefansson, K. (2019). Characterizing mutagenic effects of recombination through a sequence-level genetic map. *Science (New York, N.Y.)*, *363*(6425), eaau1043. https://doi.org/10.1126/science.aau1043

Harringmeyer, O. S., & Hoekstra, H. E. (2022). Chromosomal inversion polymorphisms shape the genomic landscape of deer mice. *Nature Ecology and Evolution*, *6*(12), 1965–1979. https://doi.org/10.1038/s41559-022-01890-0

Henderson, I. R., & Bomblies, K. (2021). Evolution and plasticity of genome-wide meiotic recombination rates. *Annual Review of Genetics*, *55*(1), 23–43. https://doi.org/10.1146/annurev-genet-021721-033821

Hoge, C., Manuel, M. de, Mahgoub, M., Okami, N., Fuller, Z., Banerjee, S., Baker, Z., McNulty, M., Andolfatto, P., Macfarlan, T. S., Schumer, M., Tzika, A. C., & Przeworski, M. (2024). Patterns of recombination in snakes reveal a tug of war between PRDM9 and promoter-like features. *Science*, *383*, eadj7026. https://doi.org/10.1126/science.adj7026

Jay, P., Whibley, A., Frézal, L., de Cara, M. R., Nowell, R. W., Mallet, J., Dasmahapatra, K. K., & Joron, M. (2018). Supergene evolution triggered by the introgression of a chromosomal inversion. *Current Biology*, *28*, 1839–1845.e3.

Joseph, J., Prentout, D., Laverré, A., Tricou, T., & Duret, L. (2024). High prevalence of Prdm9-independent recombination hotspots in placental mammals. *Proceedings of the National Academy of Sciences*, *121*, e2401973121. https://doi.org/10.1073/pnas.2401973121

Kent, T. V., Uzunović, J., & Wright, S. I. (2017). Coevolution between transposable elements and recombination. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *372*(1736), 20160458. https://doi.org/10.1098/rstb.2016.0458

Lam, I., & Keeney, S. (2015). Nonparadoxical evolutionary stability of the recombination initiation landscape in yeast. *Science (New York, N.Y.)*, *350*(6263), 932–937. https://doi.org/10.1126/science.aad0814

Latrille, T., Duret, L., & Lartillot, N. (2017). The Red Queen model of recombination hot-spot evolution: A theoretical investigation. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *372*(1736), 20160463. https://doi.org/10.1098/rstb.2016.0463

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv.org q-bio.GN*.

Li, H., & Ralph, P. (2018). Local PCA shows how the effect of population structure differs along the genome. *Genetics*, *211*(1), 289–304. https://doi.org/10.1534/genetics.118.301747

Ma, J., & Amos, C. I. (2018). Investigation of inversion polymorphisms in the human genome using principal components analysis. *PLoS One*, *7*(7), e40224. https://doi.org/10.1371/journal.pone.0040224

Malinsky, M., Challis, R. J., Tyers, A. M., Schiffels, S., Terai, Y., Ngatunga, B. P., Miska, E. A., Durbin, R., Genner, M. J., & Turner, G. F. (2015). Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science*, *350*(6267), 1493–1498. https://doi.org/10.1126/science.aac9927

Malinsky, M., Svardal, H., Tyers, A. M., Miska, E. A., Genner, M. J., Turner, G. F., & Durbin, R. (2018). Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology & Evolution*, *2*(12), 1940–1955. https://doi.org/10.1038/s41559-018-0717-x

Downloaded from https://academic.oup.com/evolut/advance-article/doi/10.1093/evolut/qpae169/7908996 by guest on 06 December 2024

Marques, D. A., Meier, J. I., & Seehausen, O. (2019). A combinatorial view on speciation and adaptive radiation. *Trends in Ecology and Evolution*, *34*(6), 531–544. https://doi.org/10.1016/j.tree.2019.02.008

Martin, S. H., Davey, J. W., Salazar, C., & Jiggins, C. D. (2019). Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLoS Biology*, *17*(2), e2006288. https://doi.org/10.1371/journal.pbio.2006288

Matthey-Doret, R., & Whitlock, M. C. (2019). Background selection and FST: Consequences for detecting local adaptation. *Molecular Ecology*, *28*, mec.15197-3914. https://doi.org/10.1111/mec.15197

McDonald, M. J., Rice, D. P., & Desai, M. M. (2016). Sex speeds adaptation by altering the dynamics of molecular evolution. *Nature*, *531*(7593), 233–236. https://doi.org/10.1038/nature17143

Mérot, C. (2020). Making the most of population genomic data to understand the importance of chromosomal inversions for adaptation and speciation. *Molecular Ecology*, *29*(14), 2513–2516. https://doi.org/10.1111/mec.15500

Mérot, C., Oomen, R. A., Tigano, A., & Wellenreuther, M. (2020). A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends in Ecology and Evolution*, *35*(7), 561–572. https://doi.org/10.1016/j.tree.2020.03.002

Miller, W., Rosenbloom, K., Hardison, R. C., Hou, M., Taylor, J., Raney, B., Burhans, R., King, D. C., Baertsch, R., Blankenberg, D., Pond, S. L. K., Nekrutenko, A., Giardine, B., Harris, R. S., Tyekucheva, S., Diekhans, M., Pringle, T. H., Murphy, W. J., Lesk, A., … Kent, W. J. (2007). 28-way vertebrate alignment and conservation track in the UCSC Genome Browser. *Genome Research*, *17*, 1797–1808. http://www.genome.org/cgi/doi/10.1101/gr.6761107

Munby, H., Linderoth, T., Fischer, B., Du, M., Vernaz, G., Tyers, A. M., Ngatunga, B. P., Shechonge, A., Denise, H., McCarthy, S. A., Bista, I., Miska, E. A., Santos, M. E., Genner, M. J., Turner, G. F., & Durbin, R. (2021). Differential use of multiple genetic sex determination systems in divergent ecomorphs of an African crater lake cichlid. *bioRxiv* 2021.08.05.455235

Myers, S., Bowden, R., Tumian, A., Bontrop, R. E., Freeman, C., MacFie, T. S., McVean, G., & Donnelly, P. (2010). Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science (New York, N.Y.)*, *327*(5967), 876–879. https://doi.org/10.1126/science.1182363

Nei, M. (1967). Modification of linkage intensity by natural selection. *Genetics*, *57*(3), 625–641. https://doi.org/10.1093/genetics/57.3.625

Nielsen, R. (2006). Why sex? *Science*, *311*(5763), 960–961. https://doi.org/10.1126/science.1124663

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., … Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, *44*(D1), D733–D745. https://doi.org/10.1093/nar/gkv1189

Ortiz-Barrientos, D., Engelstädter, J., & Rieseberg, L. H. (2016). Recombination rate evolution and the origin of species. *Trends in Ecology & Evolution*, *31*(3), 226–236. https://doi.org/10.1016/j.tree.2015.12.016

Ortiz-Barrientos, D., & James, M. E. (2017). Evolution of recombination rates and the genomic landscape of speciation. *Journal of Evolutionary Biology*, *30*(8), 1519–1521. https://doi.org/10.1111/jeb.13116

Otto, S. P., & Barton, N. H. (2001). Selection for recombination in small populations. *Evolution*, *55*(10), 1921–1931. https://doi.org/10.1111/j.0014-3820.2001.tb01310.x

Otto, S. P., & Payseur, B. A. (2019). Crossover interference: Shedding light on the evolution of recombination. *Annual Review of Genetics*, *53*(1), 19–44. https://doi.org/10.1146/annurev-genet-040119-093957

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics (Oxford, England)*, *20*(2), 289–290. https://doi.org/10.1093/bioinformatics/btg412

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, *2*(12), e190–e190. https://doi.org/10.1371/journal.pgen.0020190

Peñalba, J. V., & Wolf, J. B. W. (2020). From molecules to populations: Appreciating and estimating recombination rate variation. *Nature Reviews Genetics*, *21*(8), 476–492. https://doi.org/10.1038/s41576-020-0240-1

Persikov, A. V., Osada, R., & Singh, M. (2009). Predicting DNA recognition by Cys2His2 zinc finger proteins. *Bioinformatics*, *25*(1), 22–29. https://doi.org/10.1093/bioinformatics/btn580

Persikov, A. V., & Singh, M. (2014). De novo prediction of DNA-binding specificities for Cys2His2 zinc finger proteins. *Nucleic Acids Research*, *42*(1), 97–108. https://doi.org/10.1093/nar/gkt890

Petronczki, M., Siomos, M. F., & Nasmyth, K. (2003). Un ménage à Quatre the molecular biology of chromosome segregation in meiosis. *Cell*, *112*(4), 423–440. https://doi.org/10.1016/s0092-8674(03)00083-7

Ponticelli, A. S., Sena, E. P., & Smith, G. R. (1988). Genetic and physical analysis of the M26 recombination hotspot of *Schizosaccharomyces pombe*. *Genetics*, *119*(3), 491–497. https://doi.org/10.1093/genetics/119.3.491

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*, 559–575. https://doi.org/10.1086/519795

Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M. A. F., Mehlig, B., & Westram, A. M. (2017). Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, *30*(8), 1450–1477. https://doi.org/10.1111/jeb.13047

Raynaud, M., Gagnaire, P. -A., & Galtier, N. (2023). Performance and limitations of linkage-disequilibrium-based methods for inferring the genomic landscape of recombination and detecting hotspots: A simulation study. *Peer Community Journal*, *3*, e27. https://doi.org/10.24072/pcjournal.254

Reeve, J., Butlin, R. K., Koch, E. L., Stankowski, S., & Faria, R. (2023). Chromosomal inversion polymorphisms are widespread across the species ranges of rough periwinkles (*Littorina saxatilis* and *L. arcana*). *Molecular Ecology*. https://doi.org/10.1111/mec.17160

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: The European molecular biology open software suite. *Trends in Genetics*, *16*(6), 276–277. https://doi.org/10.1016/s0168-9525(00)02024-2

Rice, W. R., & Chippindale, A. K. (2001). Sexual recombination and the power of natural selection. *Science*, *294*(5542), 555–559. https://doi.org/10.1126/science.1061380

Ritz, K. R., Noor, M. A. F., & Singh, N. D. (2017). Variation in recombination rate: Adaptive or not? *Trends in Genetics*, *33*(5), 364–374. https://doi.org/10.1016/j.tig.2017.03.003

Rolland, J., Henao-Diaz, L. F., Doebeli, M., Germain, R., Harmon, L. J., Knowles, L. L., Liow, L. H., Mank, J. E., Machac, A., Otto, S. P., Pennell, M., Salamin, N., Silvestro, D., Sugawara, M., Uyeda, J., Wagner, C. E., & Schluter, D. (2023). Conceptual and empirical bridges between micro- and macroevolution. *Nature Ecology and Evolution*, *7*(8), 1181–1193. https://doi.org/10.1038/s41559-023-02116-7

Rowan, B. A., Heavens, D., Feuerborn, T. R., Tock, A. J., Henderson, I. R., & Weigel, D. (2019). An ultra high-density *Arabidopsis thaliana* crossover map that refines the influences of structural variation and epigenetic features. *Genetics*, *213*(3), 771–787. https://doi.org/10.1534/genetics.119.302406

Samuk, K., Manzano-Winkler, B., Ritz, K. R., & Noor, M. A. F. (2020). Natural selection shapes variation in genome-wide recombination rate in *Drosophila pseudoobscura*. *Current Biology*, *30*(8), 1517–1528.e6. https://doi.org/10.1016/j.cub.2020.03.053

Samuk, K., Owens, G. L., Delmore, K. E., Miller, S. E., Rennison, D. J., & Schluter, D. (2017). Gene flow and selection interact to promote

Downloaded from https://academic.oup.com/evolut/advance-article/doi/10.1093/evolut/qpae169/7908996 by guest on 06 December 2024

adaptive divergence in regions of low recombination. *Molecular Ecology*, *26*(17), 4378–4390. https://doi.org/10.1111/mec.14226

Schluter, D., & Rieseberg, L. H. (2022). Three problems in the genetics of speciation by selection. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(30), e2122153119. https://doi.org/10.1073/pnas.2122153119

Schumer, M., Xu, C., Powell, D. L., Durvasula, A., Skov, L., Holland, C., Blazier, J. C., Sankararaman, S., Andolfatto, P., Rosenthal, G. G., & Przeworski, M. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science (New York, N.Y.)*, *360*(6389), 656–660. https://doi.org/10.1126/science.aar3684

Shanfelter, A. F., Archambeault, S. L., & White, M. A. (2019). Divergent fine-scale recombination landscapes between a freshwater and marine population of Threespine stickleback fish. *Genome Biology and Evolution*, *11*(6), 1573–1585. https://doi.org/10.1093/gbe/evz090

Singhal, S., Leffler, E. M., Sannareddy, K., Turner, I., Venn, O., Hooper, D. M., Strand, A. I., Li, Q., Raney, B., Balakrishnan, C. N., Griffith, S. C., McVean, G., & Przeworski, M. (2015). Stable recombination hotspots in birds. *Science*, *350*(6263), 928–932. https://doi.org/10.1126/science.aad0843

Spence, J. P., & Song, Y. S. (2019). Inference and analysis of population-specific fine-scale recombination maps across 26 diverse human populations. *Science Advances*, *5*(10), eaaw9206. https://doi.org/10.1126/sciadv.aaw9206

Szankasi, P., Heyer, W. -D., Schuchert, P., & Kohli, J. (1988). DNA sequence analysis of the ade6 gene of *Schizosaccharomyces pombe* wild-type and mutant alleles including the recombination hot spot allele ade6-M26. *Journal of Molecular Biology*, *204*(4), 917–925. https://doi.org/10.1016/0022-2836(88)90051-4

Terhorst, J., Kamm, J. A., & Song, Y. S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature Genetics*, *49*(2), 303–309. https://doi.org/10.1038/ng.3748

Todesco, M., Owens, G. L., Bercovich, N., Légaré, J. -S., Soudi, S., Burge, D. O., Huang, K., Ostevik, K. L., Drummond, E. B. M., Imerovski, I., Lande, K., Pascual-Robles, M. A., Nanavati, M., Jahani, M., Cheung, W., Staton, S. E., Muños, S., Nielsen, R., Donovan, L. A., … Rieseberg, L. H. (2020). Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature*, *584*(7822), 602–607. https://doi.org/10.1038/s41586-020-2467-6

Úbeda, F., & Wilkins, J. F. (2010). The Red Queen theory of recombination hotspots. *Journal of Evolutionary Biology*, *24*(3), 541–553. https://doi.org/10.1111/j.1420-9101.2010.02187.x

Venu, V., Harjunmaa, E., Dreau, A., Brady, S., Absher, D., Kingsley, D. M., & Jones, F. C. (2024). Fine-scale contemporary recombination variation and its fitness consequences in adaptively diverging stickleback fish. *Nature Ecology and Evolution*, *8*, 1337–1352. https://doi.org/10.1038/s41559-024-02434-4

Vernaz, G., Hudson, A. G., Santos, M. E., Fischer, B., Carruthers, M., Shechonge, A. H., Gabagambi, N. P., Tyers, A. M., Ngatunga, B. P., Malinsky, M., Durbin, R., Turner, G. F., Genner, M. J., & Miska, E. A. (2022). Epigenetic divergence during early stages of speciation in an African crater lake cichlid fish. *Nature Ecology and Evolution*, *6*, 1940–1951. https://doi.org/10.1038/s41559-022-01894-w

Wellenreuther, M., & Bernatchez, L. (2018). Eco-evolutionary genomics of chromosomal inversions. *Trends in Ecology and Evolution*, *33*(6), 427–440. https://doi.org/10.1016/j.tree.2018.04.002

Yeaman, S. (2022). Evolution of polygenic traits under global vs local adaptation. *Genetics*, *220*(1), iyab134. https://doi.org/10.1093/genetics/iyab134

Downloaded from https://academic.oup.com/evolut/advance-article/doi/10.1093/evolut/qpae169/7908996 by guest on 06 December 2024