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

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

Affiliations:
[1] Institute of Ecology and Evolution, University of Bern, Switzerland
[2] EAWAG, Kastanienbaum, Switzerland
[3] School of Natural & Environmental Sciences, Bangor University, UK

Corresponding authors:
Milan Malinsky: millanek@gmail.com
Marion Talbi: marion.talbi@gmail.com

Abstract:

Meiotic recombination is fundamental to evolution of sexually reproducing organisms and differences in recombination rates are important during 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 show that fine-scale recombination landscapes differ substantially between two cichlid fish ecotypes of *Astatotilapia calliptera* that diverged only ~2,500 generations ago. The observed results are not driven by PDRM9, whose binding sites do not show any relationship to recombination rates in this species. We show that regions where recombination histories differ between ecotypes have non-random distribution across chromosomes. They are associated with, but only partially explained, by regions of high divergence between ecotypes in allele frequency (F_{ST}) and / or nucleotide diversity. We also found 47 large haplotype blocks that are polymorphic in Lake Masoko, cover 21% of the genome, appear to include inversions, and contribute disproportionately to the evolution of recombination. Only a small number of them have elevated F_{ST}. While some haplotype blocks are old and likely maintained by balancing selection, for most, the age of ancestry is close to the genome-wide average. Among haplotype blocks, there is a strong and clear association between the degree of recombination divergence and ecotype clustering by individual heterozygosity. Overall, our work provides a holistic view of changes in recombination landscapes during early stages of speciation with gene flow and advances our understanding of the combinatorial basis of evolution.

Keywords (three to ten keywords that represent the content of the article):
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 (Rice and Chippindale 2001; Nielsen 2006; McDonald et al. 2016), but in some contexts can also hinder adaptation by breaking up locally adapted haplotypes (Ortiz-Barrientos et al. 2016; Schluter and Rieseberg 2022). In recent years, there has been a growing appreciation of the role of recombination suppression in organismal diversification (Schluter and Rieseberg 2022). 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 (Nei 1967; Feldman et al. 1996; Coop and Przeworski 2007). Many genetic variants are known to affect the recombination rates and the positioning of recombination events (Halldorsson et al. 2019; Rowan et al. 2019). 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 and 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 15x 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 fulfil 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 and Bomblies 2021) and contributes to limiting average recombination rates to a relatively narrow range above this minimum via the mechanism of 'crossover interference' (Otto and Payseur 2019). Second, in some vertebrate species, principally in mammals, recombination is directed towards binding sites of the zinc-finger protein PRDM9 (Baudat

et al. 2010; Myers et al. 2010; Baker et al. 2017; Cavassim et al. 2022). In these species, rapid evolution of recombination landscapes is mediated by intra-genomic conflict. Specifically, degradation and extinction of PRDM9 binding sites due to biased gene conversion induces positive selection on PRDM9 zinc finger sequences to maintain recombination (Úbeda and Wilkins 2010; Latrille et al. 2017) and / or efficient double strand break repair and homolog pairing (Baker et al. 2023). The 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. 2023).

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 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). These genomic features are relatively stable in comparison with the rapid turnover of PRDM9 binding sites, and there is some evidence that species lacking PRDM9 have more conserved recombination landscapes, at least with regards to the location of recombination hotspots (Lam and Keeney 2015; Singhal et al. 2015). However, the association with genomic features is only partial (Singhal et al. 2015) and recombination rates do evolve also in species without PRDM9 (Ritz et al. 2017; Samuk et al. 2020). It has even been suggested that in the PRDM9-lacking stickleback fish, hotspots may evolve at similar rates to those observed in species with PRDM9 (Shanfelter et al. 2019). Despite the growing interest in this topic, many questions remain with regards to recombination rate evolution across different species, and particularly in species without PRDM9. 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 (~650m in diameter) crater lake in Southern Tanzania (**Fig 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).

In this study, we reconstruct genetic maps from patterns of linkage disequilibrium (LD) in whole genome population genetic data of 70 benthic and 69 littoral individuals to investigate the evolution of recombination landscapes in Lake Masoko. We first demonstrate that the population recombination landscapes are considerably different despite the recent split time between these ecotypes. 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 provides new evidence of and insights into rapid recombination rate evolution in the absence of PRDM9 dynamics.

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 morph based on a field photograph. The sequencing coverage was $\sim$15x and, after variant calling and filtering, we obtained 3.9 million SNPs that were used for all the following analyses. To check the validity of the field assignment we used the genetic data to run a principal component analysis (PCA; **Fig. 1B**) and reconstructed a Neighbor-Joining tree (**Supplementary Fig. S1**). 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 **Fig. 1B**) from further 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 which may be related to the lake colonization, followed by recent demographic expansions in both ecotypes (**Fig. 1D**). This approach also allowed us to re-estimate the split time between the two ecotypes, which we now put at $\sim$2,500 generations ago (95% confidence interval: 1902 to 5469 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 morphs 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 (**Supplementary Fig. S2**; Materials and Methods).

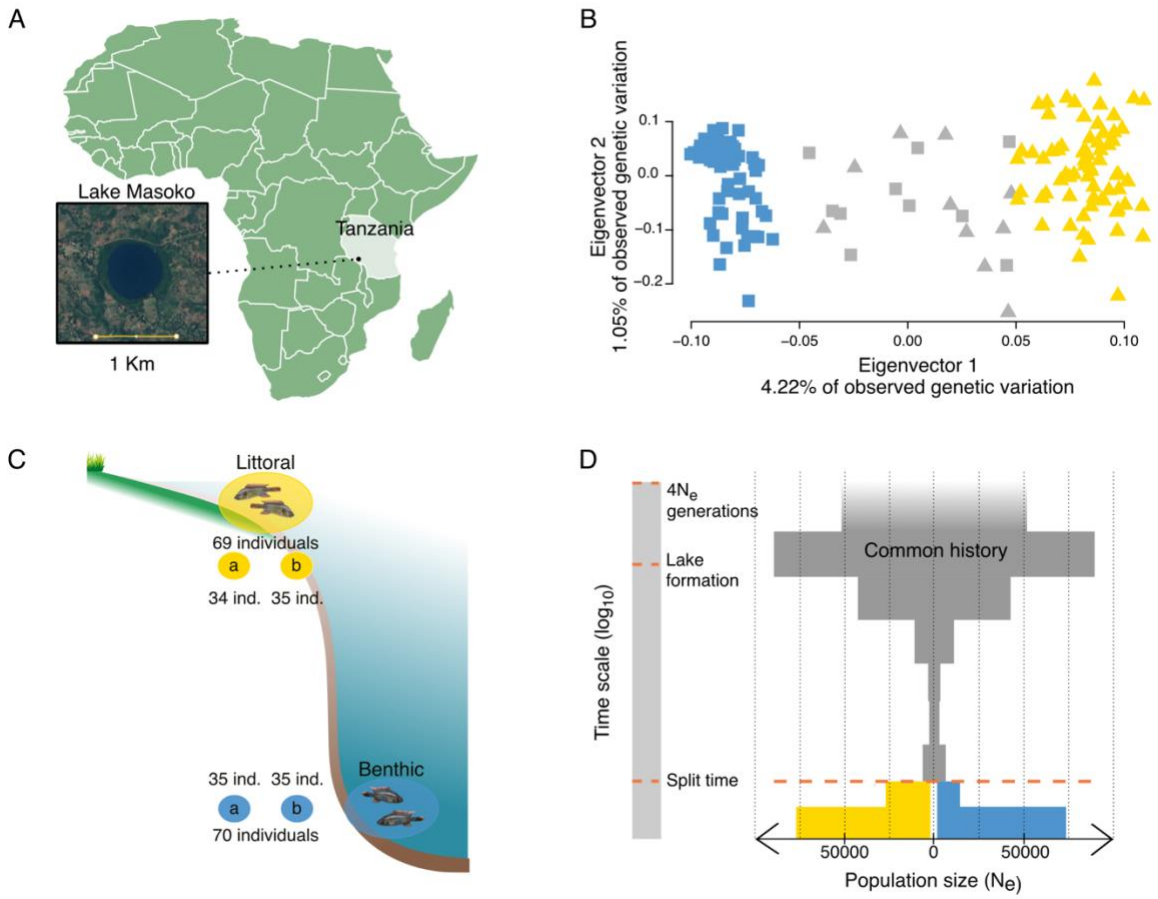

Fig. 1: Study system and demographic history. (**A**) Lake Masoko is a circular small (~670m diameter) maar-type volcanic crater lake located in the East African rift valley in southern Tanzania. (**B**) A principal component analysis based on SNPs. Some individuals labelled in the field as 'benthic' 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**) 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. (**D**) 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 morph are shown in yellow, and for the benthic morph in blue.

Underlying recombination landscapes differ between ecotypes

To quantify sampling variability, we divided the individuals from each morph into two independent subsets and reconstructed a separate recombination map for each subset (**Fig. 1C**). 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 2kb scale (**Fig. 2A**; Materials and Methods). 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 for the between-ecotype replicates (in gray in **Fig. 2A**) and found that the correlation coefficients were considerably lower

than within-ecotype (**Fig. 2A**; mean Spearman correlation = 0.57). This key result, *i.e.* that between-ecotype correlations are consistently lower than for within-ecotype replicates, holds across genomic scales from 2kb to 5Mb (**Supplementary Fig. S3**).

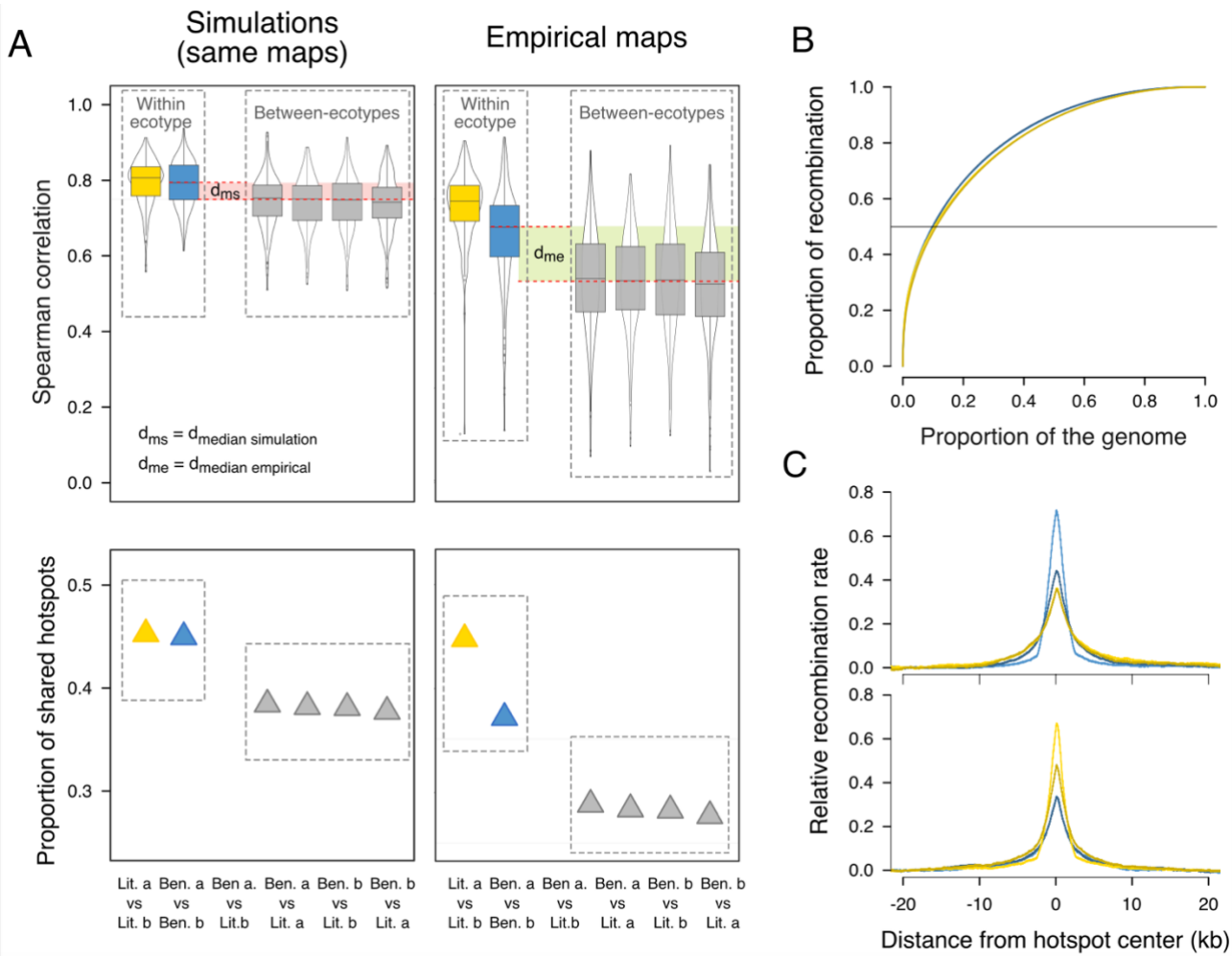

Fig. 2: Rapid evolution of the recombination landscapes between the ecotypes. (**A**) Top: Spearman correlation between recombination maps on 2 kb scale with each datapoint representing a 5Mb genomic interval. In the empirical data, we see that the correlations in between-ecotypes comparisons are substantially lower than expected based on the simulations – more than 75% of the values are below the level defined by d_{ms} (see Material and Methods). Bottom: The proportion of overlapping hotspots between recombination maps. The higher similarity for the within-littoral compared to the within-benthic replicates is a result of the higher N_e and greater genetic diversity in the littoral ecotype leading to more accurate genetic map reconstruction. (**B**) Empirical cumulative distribution function of for recombination fraction as a function of proportion of genome covered. The black line corresponds to 50% of recombination events. (**C**) Scaled recombination rates around hotspots defined in the benthic a (top) and littoral a (bottom) subsamples. Hotspots in (**B**) and (**C**) are defined against a 1Mb background. Blue and yellow colors refer to the benthic and littoral morphs respectively.

Moreover, simulations show that the recombination map differences between the ecotypes cannot be fully accounted for by non-random sampling of coalescent histories arising from the separation of recent benthic vs. littoral ancestry. Specifically, the difference in median correlations was more than three times greater in empirical data than in simulations without recombination rate evolution

(**Fig. 2A**; empirical d_{me} = 0.140 vs. simulation d_{ms} = 0.046; Materials and Methods). These results provide evidence that the population recombination landscapes differ between the ecotypes.

Recombination landscapes are highly heterogenous, and a large proportion of events tends to occur in so-called 'hotspots' (Coop and Przeworski 2007; Peñalba and 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 (**Fig. 2B**; Between 8.9% and 10.4% depending on the morph and the subsample). Using a definition of hotspots as having at least 5x higher recombination rate than the 500kb of surrounding sequence, we found on average 2322 hotspots in each recombination map (between 2275 and 2345; **Supplementary Table S1**). Only 41.5 % of hotspots were shared when comparing 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 comparisions between benthic and littoral maps showed even less hotspot overlap than expected based on simulations (**Fig. 2A,C**). Qualitatively similar results were obtained using different hotspot definitions (**Supplementary Fig. S4**).

To verify that the observed recombination differences between benthic and littoral were not driven by technical artifacts in regions of the genome where inference is especially error prone, we applied a stringent filtering mask. This mask was based on the: (i) location of contig joins in the assembly, (ii) ability to confidently call SNPs across genomic windows, and (iii) consistency of inferred genetic maps across bootstrap replicates (Material and Methods). In total, we masked 30.6% of the sequence, eliminating a substantial proportion of noise from the inferred recombination maps across all genomic scales from 2kb to 5Mb (**Supplementary Fig. S3**). We repeated all the analyses above using these filtered maps and found that, despite the stricness of the filtering, the results were qualitatively the same as in the raw maps (**Supplementary Figs. S3, S5**).

LD-based genetic maps reflect recombination events that have happened throughout the ancestry, i.e., the coalescence history, of the sampled individuals. In Lake Masoko *A. calliptera*, the ancestry for the sampled individuals extends substantially beyond the ecotype split time of ~2,500 generations ago (**Fig. 1C**), which means that the inferred recombination maps for each morph can be interpreted as a mixture of two time periods: (i) recombination events that happened in the common history of the morphs 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, an average 34.9% of recombination events that changed the genealogy of the sample occured after the split (min = 34.4%; max = 35.7%; Materials and Methods).

Characterizing recombination rate evolution

We next investigated the relationship between recombination rates and genetic divergence between the morphs in terms of allele frequencies (F_{ST}) and levels of nucleotide diversity ($\Delta(\pi)$) along the genome. First, we found that the greater-than-expected difference in recombination rates between the morphs is present consistently at all levels of genetic divergence and not limited to regions of particularly high F_{ST} or $\Delta(\pi)$ (**Fig. 3A**). This result provides further evidence that the benthic vs. littoral map differences reflect divergence in the true realized recombination histories of these ecotypes and are not driven by the effect that differences in π can have on the accuracy of recombination inference (**Supplementary Fig. S6**; Raynaud et al, 2023). Second, we found that the between-ecotype map correlations tended to be lower in regions of greater genetic divergence, a trend that is particularly clear when looking at the filtered maps, but not seen in simulations in the absence of selection (**Supplementary Fig. S7A**). In some analyses we also observed a reversal of this trend at the highest levels of F_{ST} and $\Delta(\pi)$; thus, the exact relationship between recombination divergence and levels of genetic differentiation appears to be complex (**Supplementary Fig. S7**).

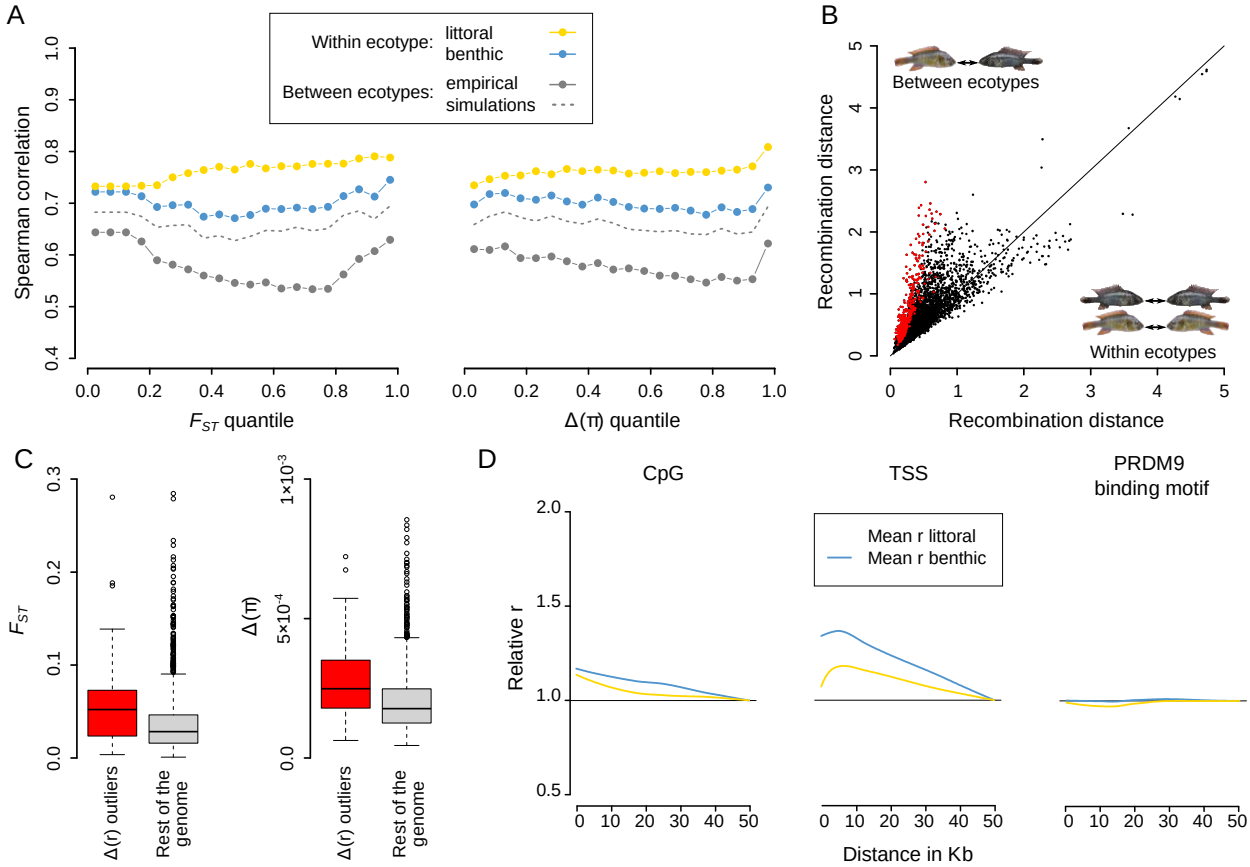

Fig. 3: Interplay between recombination and genomic differentiation. (**A**) Spearman correlation measured between recombination maps as a function of benthic-littoral F_{ST} and of $\Delta(\pi)$. (**B**) A scatterplot of average recombination map distances in 100kb windows, within biological replicates - $\Delta(r)_w$ 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. (**D**) Relative recombination rates in 2kb windows are higher near CpG islands and TSS but are independent of the binding motif associated with the incomplete PRDM9 ortholog presents in this species.

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 100kb windows (see **Methods**). The average recombination distance beween benthic and littoral maps (denoted $\Delta(r)_b$) was greater than the analogous distance for within-ecotype replicates (denoted $\Delta(r)_w$) in 83.1% of the windows (**Fig 3B**). This metric allowed us to identify genomic regions with rapidly changing recombination rates. In the following, we 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 42.7 Mb 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 Fig. S8A**). 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 Fig. S8B**), 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 (**Fig. 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 (~1.2x higher; Fig 3D) and transcription start sites (TSS; ~1.3x higher; **Fig 3D**), a pattern that is similar to that reported previously for swordtail fish (Baker et al. 2017). Because the PRDM9 zinc finger 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 (**Fig 3D**). 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, Johannesson, et al. 2019). Non-recombining haplotype blocks can be revealed as

extended regions of the genome with distinct population structure, substantially different from the genome-wide average (Ma and Amos 2018; Mérot 2020; Todesco et al. 2020). To look for such regions, we used a local PCA approach (Li and Ralph 2018) and identified a total of 47 outliers (**Supplementary Figs. S9, S10**), ranging in size between 550kb and 25.7Mb (mean 3.9Mb) and covering in 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**).

Table 1: Genomic regions underlying recombination rate evolution.

Feature	**$\Delta(r)$ outlier proportion**	**$\Delta(r)$ outlier excess**	**Permutation p-value**	**Mean $\Delta(r)$ (log10)**
High F_{ST} (top 10%)	13.7%	+58.3%	< 0.001	0.109
High $\Delta(\pi)$ (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

Large haplotype blocks can be the result of lack of recombination between alternative haplotypes segregating in Lake Masoko. We hypothesized that, on average, recombination would be reduced in the morph with the higher proportion of individuals who are heterozygous for such non-recombining 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 local PCA outlier region associated with differences in recombination is shown in **Fig. 4,** with strong benthic vs. littoral clustering by H_{ind} illustrated in **Fig. 4C**.

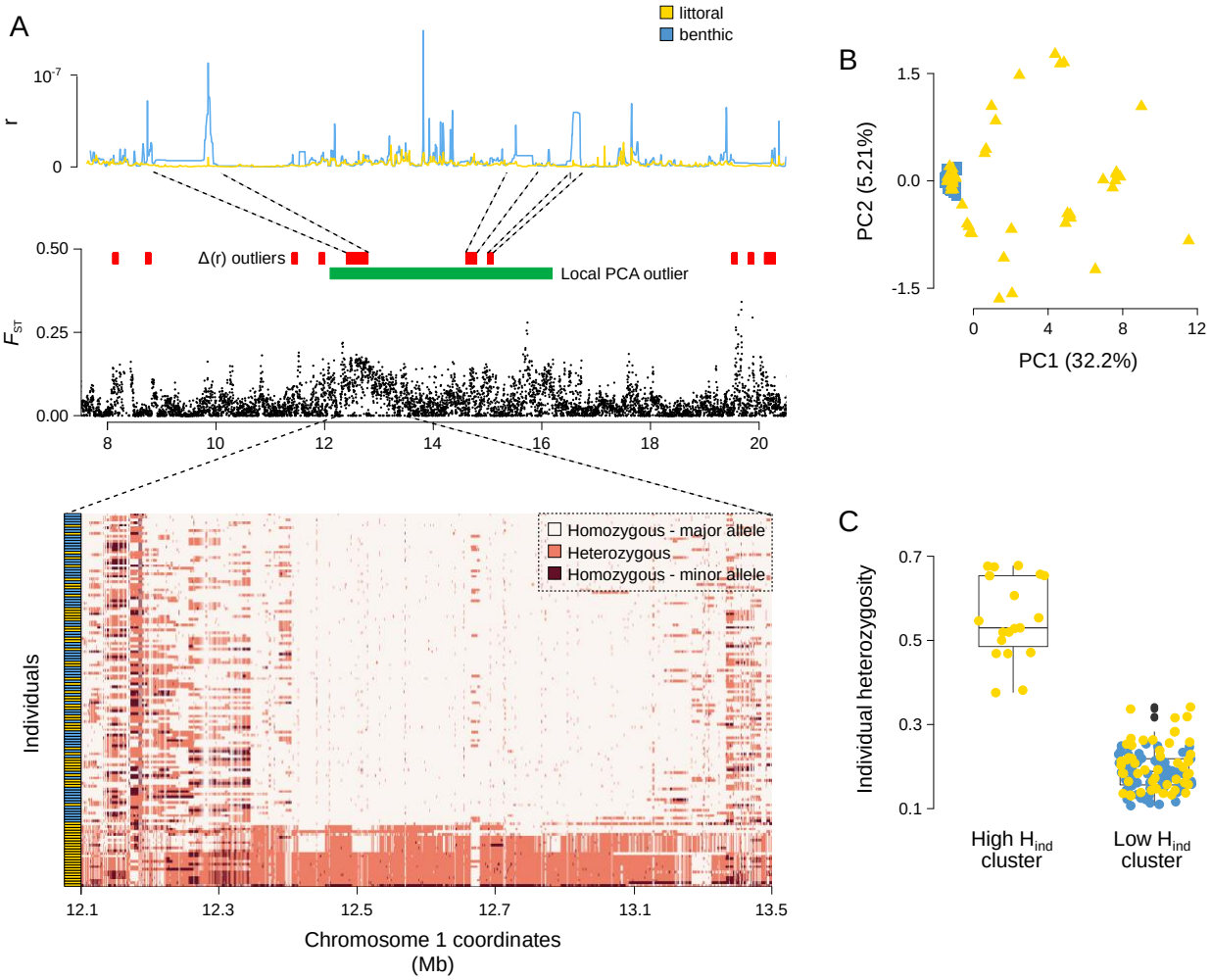

Figure 4: A local PCA outlier associated with difference in recombination. (**A**) An example region of chromosome 1 (LS420019.2) illustrating the link between a large haplotype block, F_{ST}, and recombination rate evolution. Top: average recombination rates in the two ecotypes. There is lower recombination in the littoral morph 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). (**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).

As predicted, across the 47 haplotype blocks we found significant positive association between the degree of ecotype clustering by H_{diff} and the divergence in recombination rates between ecotypes (**Fig. 5A**). Interestingly, the association of allele frequency divergence (mean F_{ST} per haplotype block) with recombination divergence was less pronounced and not statistically significant (**Fig. 5B**). This suggests that, at least in some cases, the heterozygous state may be overrepresented, which would be consistent with a form of balancing selection. Therefore, we compared the distribution of inbreeding coefficient (*F*) per SNP within local PCA outliers against SNPs from a set of control regions (**Methods**). 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* (**Fig. 5C**).

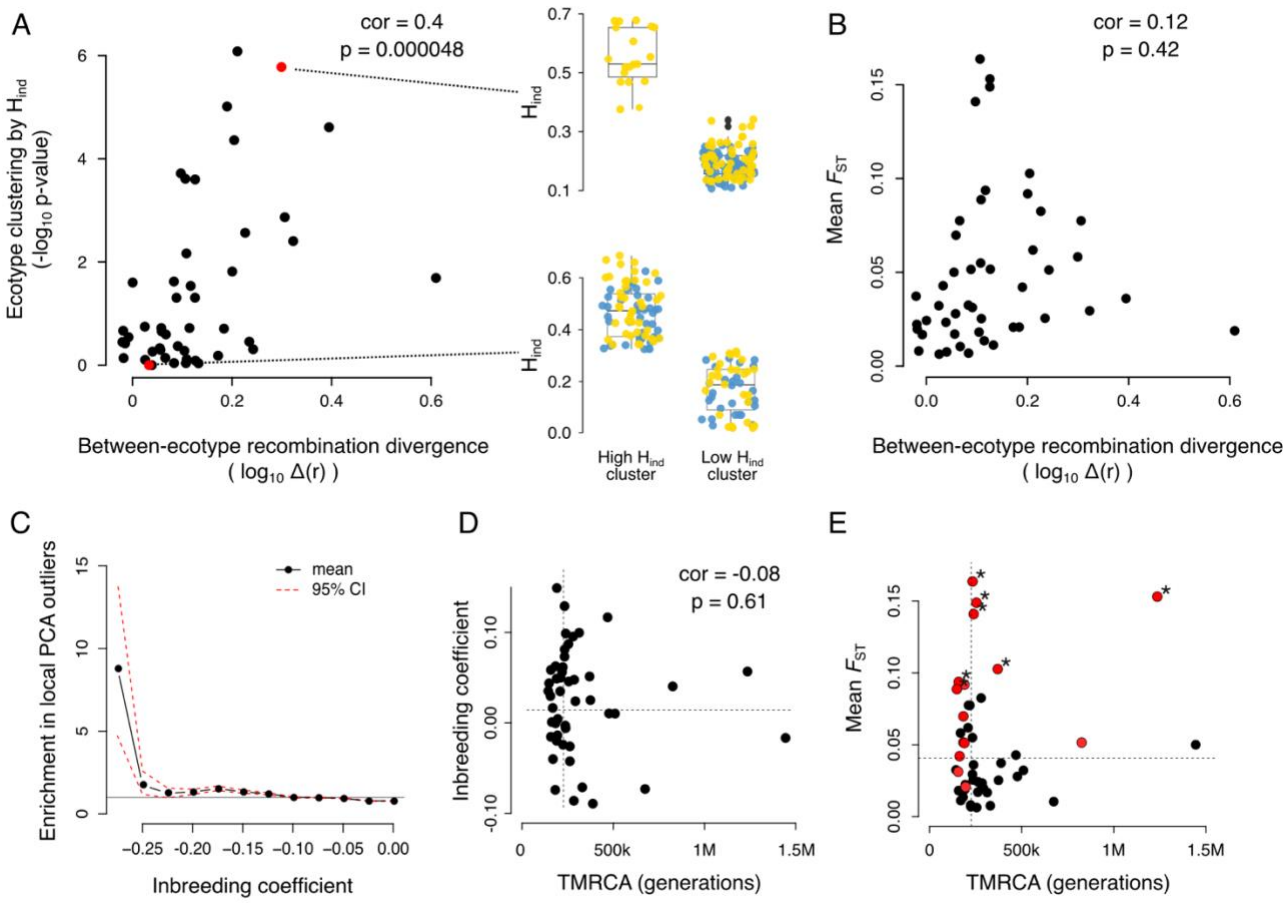

Figure 5: Characterizing haplotype blocks and their link with recombination in Lake Masoko. (A) Relationship, across the 47 local PCA outliers, between the degree of ecotype clustering by H_{ind} and the between-ecotype differences in recombination maps. We highlighted examples of a strong (top) and weak (bottom) of ecotype clustering by H_{ind}. (B) Relationship between the mean F_{ST} and the between-ecotype differences in recombination maps. (C) For each value of inbreeding coefficient, we show the relative enrichment of SNPs in the PCA outlier regions in comparison with the control regions. (D) Relationship between the inbreeding coefficient and the 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).

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 Figs. S11, S12**).

We expected that the haplotype blocks with the lower value of inbreeding coefficient may have been maintained as polymorphism for a long time through the effect of balancing selection. Therefore, for each PCA outlier we estimated the time to the most recent common ancestor (TMRCA) of all our samples. Perhaps surprisingly, we found that the ancestry of these regions was of a similar age as the genome wide average, with a few clear much older outliers (**Fig. 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.

Finally, we investigated in more details the link between allele frequency divergence and the haploblocks in the 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 (**Fig. 5E**). Islands of differentiation – also referred to as highly diverged regions (HDRs), defined as in 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 (**Fig. 5E**; **Supplementary table S2**). Interestingly, the haplotype blocks which contain HDRs are not necessarily among the ones with high mean F_{ST}, highlighting the different perspective that we can obtain when considering recombination and linkage.

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 along multiple axes 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 (Ortiz-Barrientos et al. 2016; Todesco et al. 2020; Schluter and Rieseberg 2022; Battlay et al. 2023). Our findings reveal and characterize 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).

We found a link between recombination rate divergence and islands of high allele frequency divergence. While such a link can be expected for a variety of reasons (Cruickshank and Hahn 2014; Burri et al. 2015; Matthey-Doret and Whitlock 2019; Booker et al. 2020) it 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 co-located with the top 10% of F_{ST} (**Table 1**).

A link was also identified between recombination rate evolution and large haplotype blocks, which comprise more than a fifth of the genome. There is growing evidence that such haplotype blocks, typically caused by large inversions, are present in many species (Wellenreuther and Bernatchez 2018), often comprise considerable proportions of the genome, and have clear links to adaptation and diversification in both animals (Faria, Chaube, et al. 2019; Harringmeyer and Hoekstra 2022; Reeve et al. 2023) and plants (Todesco et al. 2020; Battlay et al. 2023). 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; Rowan et al. 2019; Mérot et al. 2020). 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 evolution of recombination landscapes (Coster et al. 2021).

We can infer that inversions will prevent crossover formation in the gametes of heterozygous individuals (Faria, Johannesson, et al. 2019). However, an important open question concerns the degree to which the observed differences in population recombination rates are the result of changes in the distribution of crossovers during gamete formation vs. an indirect effect of subsequent selection for or against specific recombinant haplotypes. This question is not possible to answer with our LD-based estimates, which reflect recombination events realized in the ancestry of the sampled individuals. A future comparison of our LD-based maps against recombination landscapes obtained by sequencing of gametes and / or individuals related by pedigrees will shed further light on this question, although these approaches are still typically limited in terms of genomic resolution (Peñalba and Wolf 2020).

Our finding of rapid recombination rate evolution, while consistent with one previous study (Shanfelter et al. 2019), stands in apparent conflict with the current paradigm of evolutionary stability of recombination landscapes in species lacking the PRDM9 mechanism (Lam and Keeney 2015; Singhal et al. 2015). This is not as surprising as it may appear, as 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. 2023; Joseph et al. 2023).

Adaptation and organismal diversification are increasingly seen as multidimensional and combinatorial, typically with involvement of multiple polygenic traits and epistasis (Marques et al. 2019; Barton 2022; 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, albeit typically at a rough resolution where constraints related to the role of recombination in meiosis and in chromosome segregation seem to prevail (Haenel et al. 2018; Brazier and Glémin 2022). We adopted the LD-based approach, enabling us to see that fine-scale rates can evolve rapidly. We envisage that the large and growing amount of population genomic data available will enable 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.

Material and Methods

Variant calling and filtering

Genomic DNA from a total of 336 individuals from Lake Masoko has been sequenced on the Illumina HiSeq X Ten platform, obtaining 150bp paired end reads (NCBI Short Read Archive, BioProject ID: PRJEB27804). The reads were aligned to an *Astatotilapia calliptera* reference genome using bwa-mem v.0.7.17 (Li 2013), with median depth per-individual of 15.8x (min = 12.4x, max = 22.1x). 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 (< 3800 or > 5700); 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, the callability mask comprised of 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 `MQ<40`, `FS>40`, `QD<2` and `ExcessHet>40`. These additional filters removed fewer than 1% of the remaining SNPs. After discarding 1.4 million of multiallelic sites and indels, the filtered VCF contained 3.86 million 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 which 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 numbers of single-nucleotide differences between haplotypes for 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 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. After we identified and removed 20 genetically intermediate samples (see Figs. 1C and S1), the final VCF with 139 individuals was composed of 3.3 million biallelic SNPs.

Genome annotation

We used the UCSC paradigm (Miller et al. 2007) to generate a pairwise whole genome alignment between the fAstCal12 assembly (GenBank: GCA_900246225.3) and the fAstCal14 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 number of generations, we used the cichlid-specific mutation rate estimate of $\mu = 3.5\,\mathrm{x}10^{-9}$ per bp per generation with 95% confidence interval $(1.6\,\mathrm{x}10^{-9}, 4.6\,\mathrm{x}10^{-9})$ (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 but with the likelihood for the migration rate estimates reported here being clearly superior (Fig S2). To reduce the confounding effects of selection in these demographic analyses we used only sites from non-coding regions of the genome, masking all annotated exons, introns, and promoters.

Subsampling and bootstrap

We first used the `shuf -n` command to randomly draw a set of 35 individuals from each morph to for the first subset. The remaining individuals (35 littoral and 34 benthic) then formed the second subset. Therefore, these (a) and (b) subsets (Fig. 1B) are independent, in the sense that they are composed of non-overlapping 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 bootstrap replicates over individuals.

Inference of recombination rates

We used the `pyrho` software (Spence and Song 2019) to infer recombination rates along the genome based on patterns of linkage disequilibrium. 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 and Song 2019) and confirmed by our own simulations (data not shown)].

Specifically, first, to build likelihood tables between each biallelic sites, we used the `make_table` command with $\mu = 3.5 \times 10^{-9}$, demographic history for each of the morphs as inferred by smc++, and the Moran approximation with the `-approx` and `-moran_pop_size N` flags where N equals 1.5x the number of haplotypes in each subset. To determine the best parameter to use for the inference of recombination rates, we processed a set of simulation with evolutionary parameters corresponding to the one of our cichlid species (e.g., μ , sample size, N_e) and chose the value of block penalty that was minimizing the quantity of false negative and false positive with best results obtained for block penalty of 15 and a window size of 50 SNPs. These parameters were then used in all runs of the `optimize` command 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 rates were constant and natural selection 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 one of the empirical recombination maps as input. From each simulation, we sampled 70 individuals from each population, labelled 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.

We also used msprime to estimate how our recombination maps reflect the relative contributions recombination events that happened in the common history of the morphs vs. events that happened

after their split. To do this, we counted the recombination events that changed the local genealogy of the sample, i.e. the local tree. Specifically, we used the `end_time` option in the `sim_ancestry()` function of msprime to stop the simulation at the split time, counted the distinct genealogies at that time point in benthic (N_{tb}) and in littoral (N_{tl}) morphs, and then we continued the simulation all the way to the common ancestor of all samples and counted the total number of genealogies (N_t). Then we calculated the proportion of genealogy-changing events that happened in the ecotypes after their split (P_{as}) as:

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

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

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. For the background recombination rate, we applied three definitions the: (i) mean rate in 40kb around the interval (20kb before and 20kb after), (ii) mean rate in 1Mb around the interval, and (iii) mean recombination rate for the whole chromosome as in (Halldorsson et al. 2019). 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 1kb. Proportion of overlap between hotspots from different maps were 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 (Fig. 2C), 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. 2023). We removed hotspots longer than 5kb from the above analyses and from the search for sequence motifs described below.

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 2Kb 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). Map distances were calculated in non-overlapping 100kb windows (i.e. vectors of 50 values for 2kb each) using the `dist()` function in R and represent the average of absolute (Manhattan) distances between these vectors. We use the log_{10} transformed values for these distances, which 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-morph replicate maps (denoted $\Delta(r)_w$). Then we calculated the analogous measure for map comparisons between morphs (denoted $\Delta(r)_b$). Finally, we calculated the standard deviation of the log_{10} transformed $\Delta(r)_w$ measure across the bootstrap replicates (we denote this standard deviation as sd_w). We then refer to any 100kb interval of the genome as a $\Delta(r)$ outlier if it satisfies the following inequality:

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

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 50kb upstream and 50kb downstream of each joint between contiguous sequences (contig joint) in the assembly. There were 514 joints on the 22 chromosomes, leading to masking of 50.46Mb 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 100kb 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 bootstrap runs and masked all 100kb regions where the difference in inferred recombination rates between bootstrap runs 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 the was filtered out by the callability mask during in genotype filtering.

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 morph, 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 and TSS

We use the `maskOutFa` and `cpg_lh` command from the *cpg_lh* program (Jim Kent reference – UCSC utils) to define 17 000 CpG islands representing a total of 8.5 Mb. For the 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 with the distance of the closest TSS non overlapping with a CpGi and vice versa.

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 and 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 Fig. S9**). 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 and looked at the relationships of a range of population genetic statistics in these regions. First, to test if the proportion of high-heterozygosity individuals was different between ecotypes, we calculated the proportion of heterozygous sites for each individual (H_{ind}), used the `kmeans` function in R with `k=2` to assign these values into two clusters, and then used a binomial distribution to quantify the difference in the number of individuals from each ecotype in the cluster with higher H_{ind}. (**Fig. 4C**; **Supplementary Table 2**). 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 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 10kb 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 Zinc-Finger binding motif

To find the PRDM9 orthologs, 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*. From this protein sequence, we used a Zinc-Finger (ZF) motif predictor (Persikov et al. 2009; Persikov and Singh 2014) to obtain the nucleotide sequence of the binding site. We finally used the `FIMO` command from the `MEME Suite` (Bailey et al. 2015) on the reference genome of our species, to localize the binding site along the genome. 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 ($\Delta(r)$ outlier regions) coincide with different genomic regions (regions of high F_{ST}, regions of 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 – that is the association of with these genomic regions – was calculated using 1,000 permutations with the R package `regioneR v. 1.34.0` (Gel et al. 2016).

Acknowledgements

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). 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.

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.

Competing interests

The authors declare that they have no competing interests.

Data Availability

Genomic DNA from all 336 individuals used in this study is available from the NCBI Short Read Archive (BioProject ID: PRJEB27804). Recombination maps and code used to analyze the data are available on GitHub (https://github.com/MarionTalbi/MasokoPaper).

Supplementary Materials

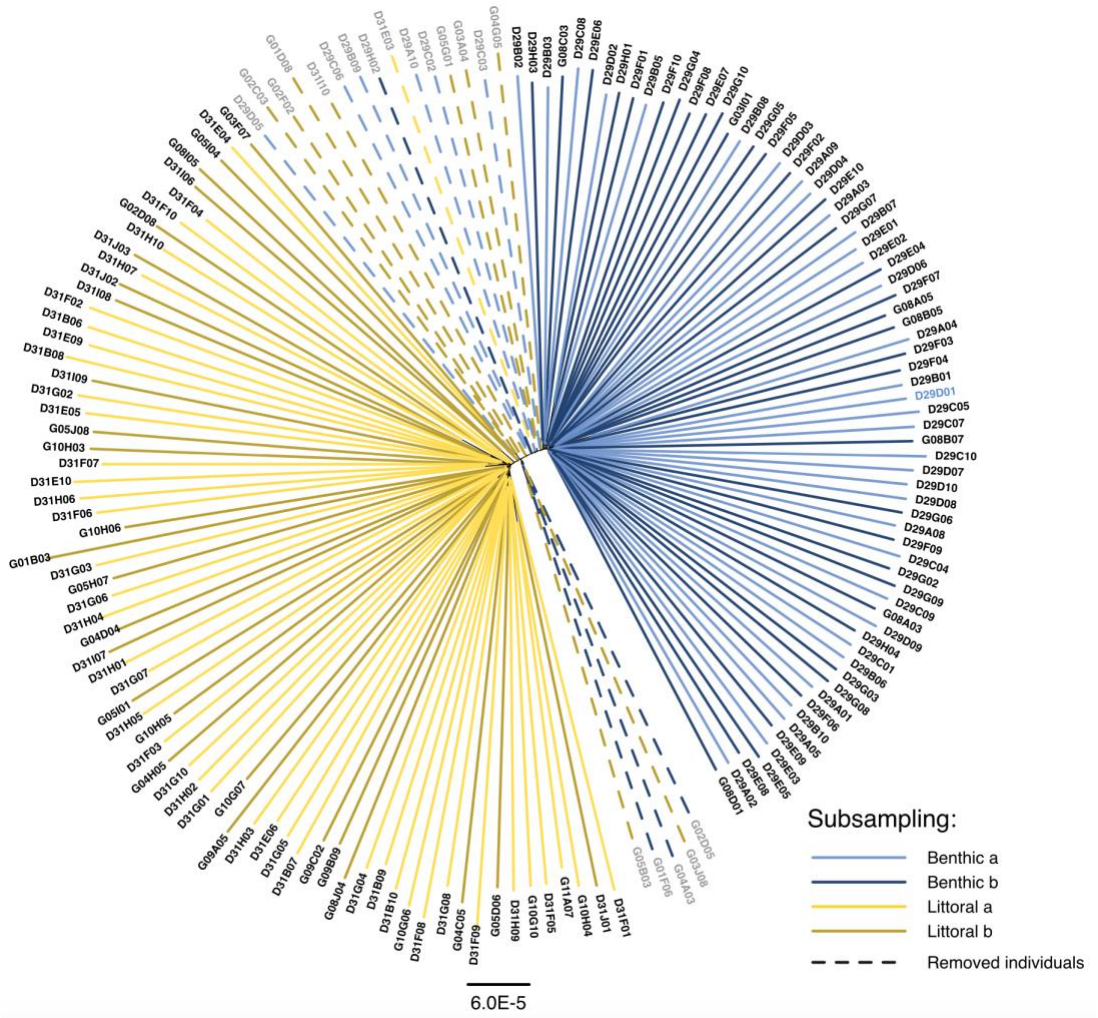

Fig. S1: Neighbor joining tree based on the number of differences in polymorphic sites between individuals. Each tip represents one individual. We used this tree and the PCA showed in Fig. 1 to determine which intermediate individuals to discard from downstream analyses. After discarding the intermediate individuals (indicated by dashed lines), the remaining samples constitute two monophyletic groups in this tree.

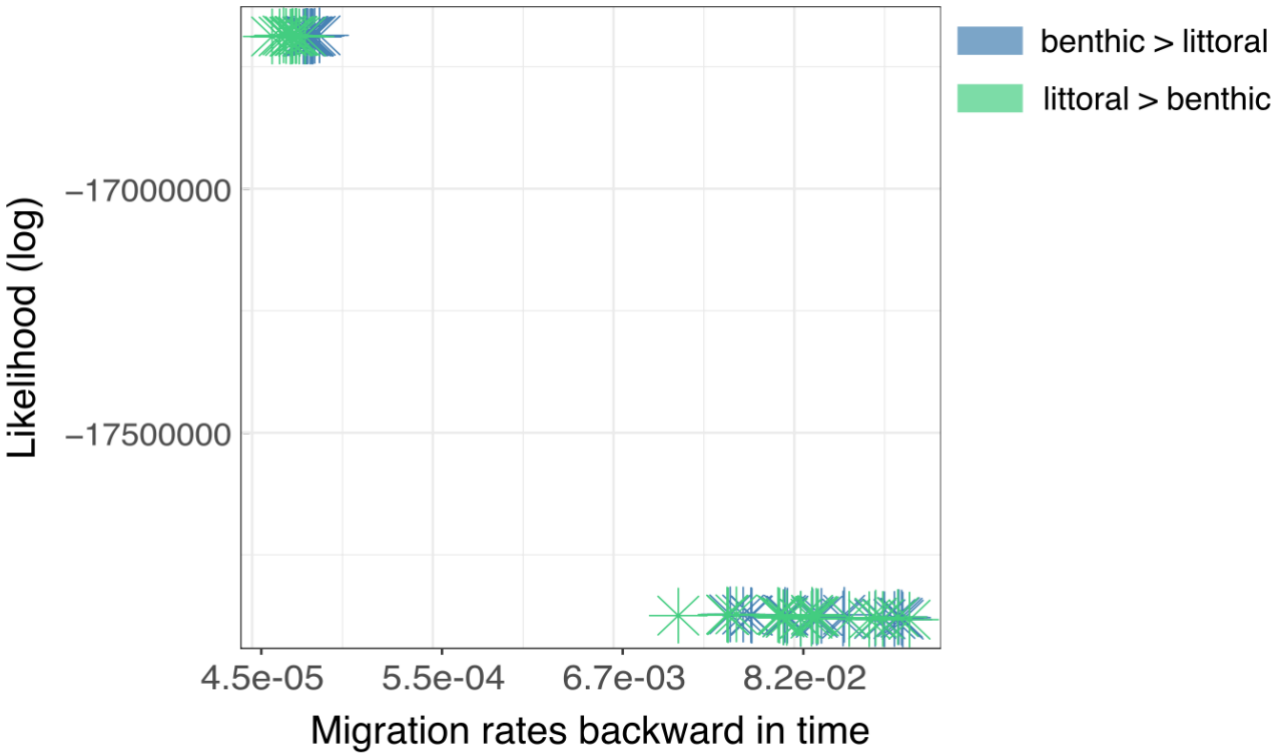

Fig. S2: Estimating migration rates with fastsimcoal2. Each point represents an estimate from an independent run with different starting parameters. The estimates with lower migration rates (top left corner) have superior likelihoods to the estimates in the bottom right. In the main text, we report the overall maximum likelihood estimates. However, we note that given the number of intermediate individuals in our sample (presumably recent hybrids), the maximum likelihood rates are lower that we would expect and the results with the lower likelihood should not be completely discounted.

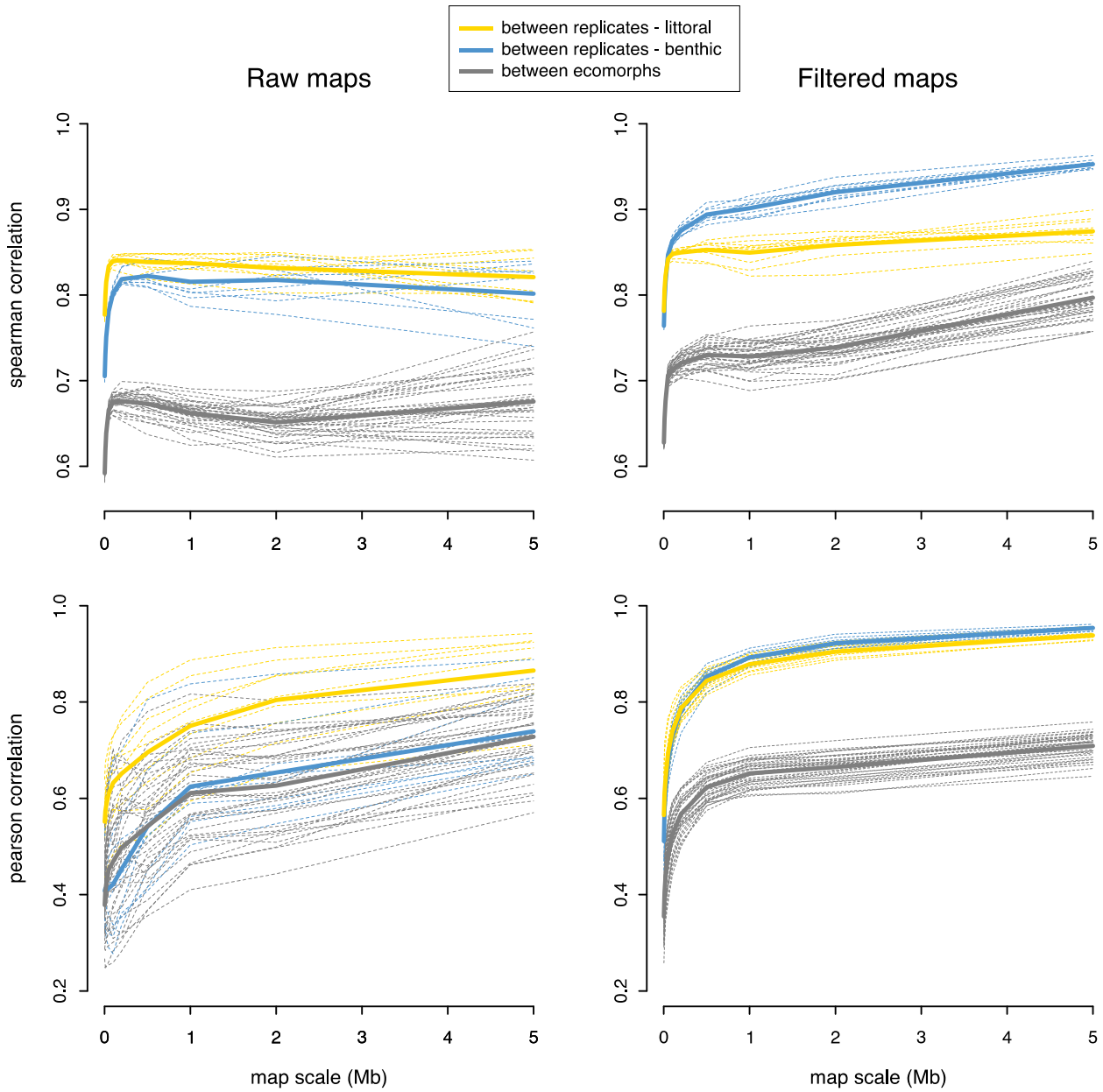

Fig. S3: Correlation between recombination maps at different scales. We see that correlations in the between ecotype comparisons are considerably lower than correlations for the within-ecotype replicates. This holds across all genomic scales (2kb to 5Mb) for Spearman correlation, and across almost all genomic scales for Pearson correlation. We also see that the filtered maps (after masking 30.6% of the genome; see Materials and Methods) show considerably less inference noise than the raw maps, making the results clearer.

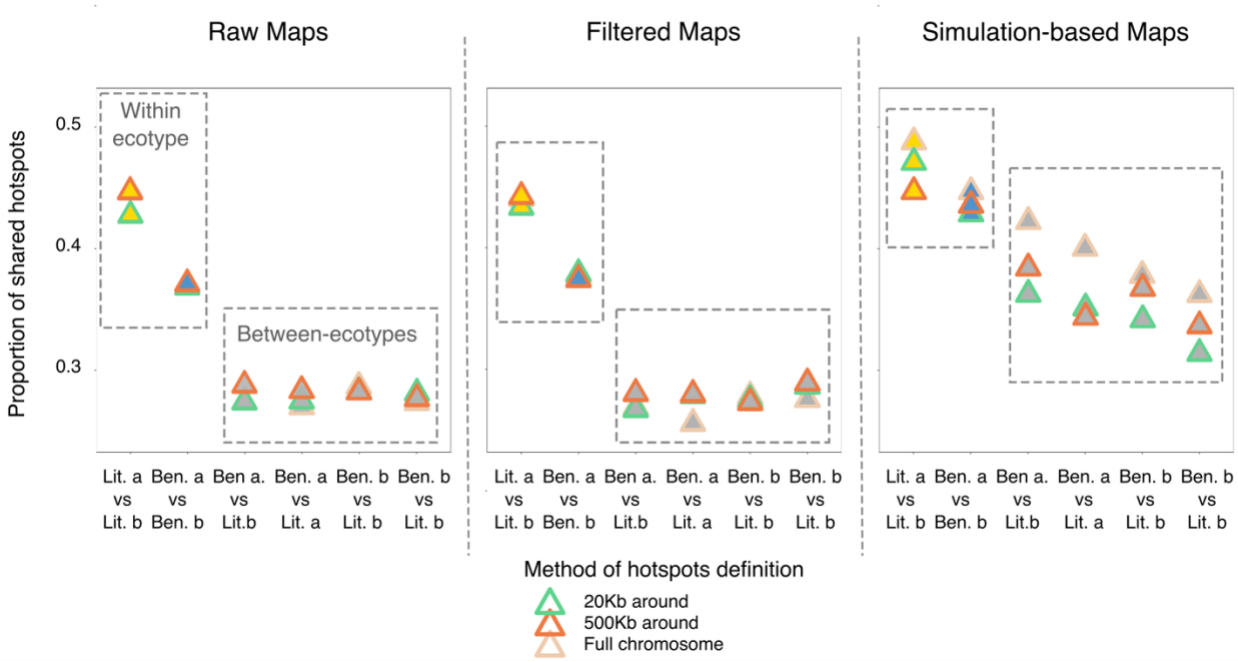

Fig. S4: The proportion of overlapping hotspots between recombination maps based on three different background size. We observe qualitatively identic and quantitatively very similar proportion of overlapping hotspots with different methods of hotspots definition. These three methods differ on the size of the genomic background used as baseline (e.g. 20kb, 500kb and full chromosome).

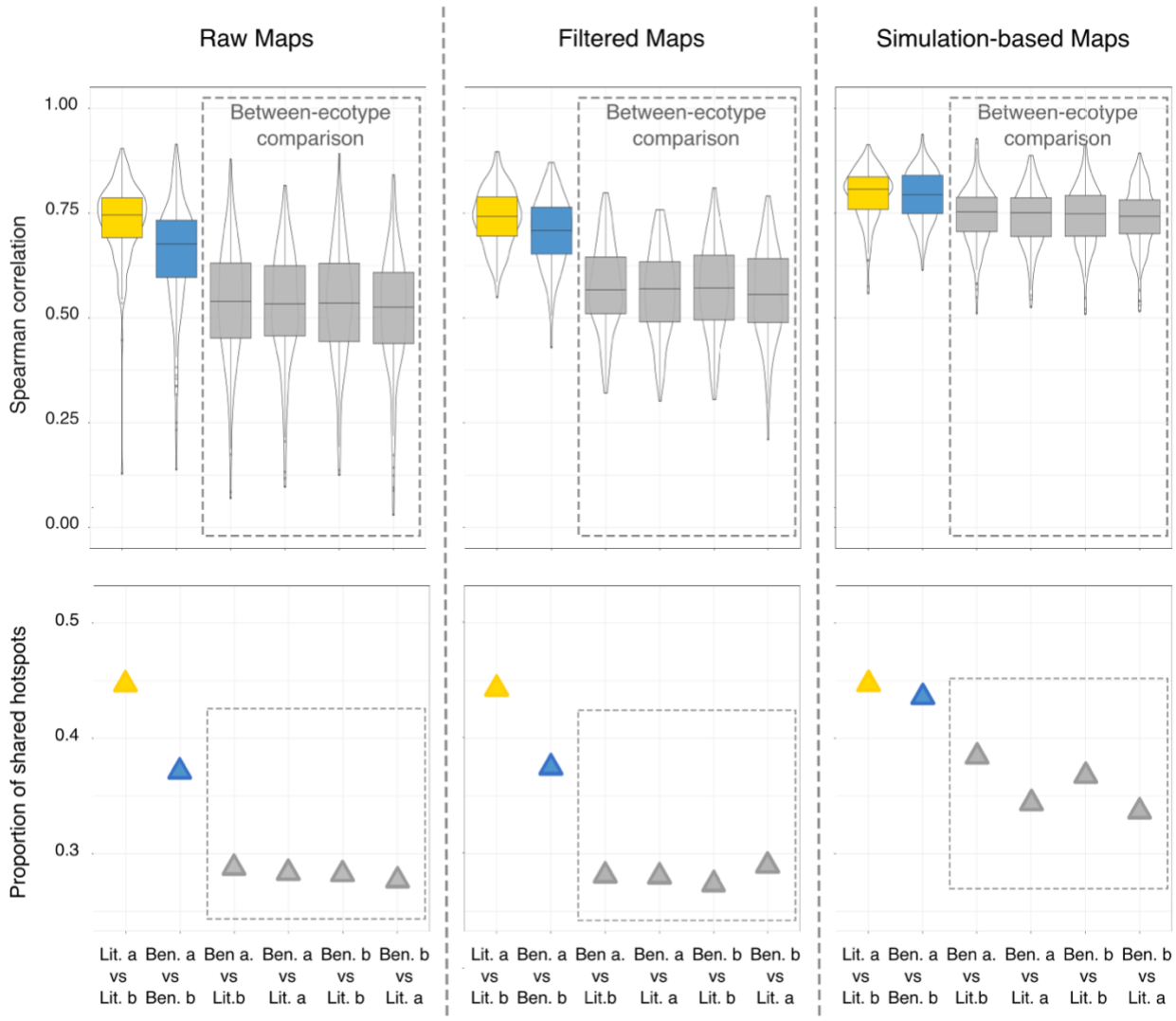

Fig. S5: Equivalent of Fig. 2A with results from filtered recombination maps. Top: Spearman correlation between recombination maps on 2 kb scale with each datapoint representing a 5Mb genomic interval. Bottom: The proportion of overlapping hotspots between recombination maps. The results based on the filtered maps are very similar to the results based on the raw maps.

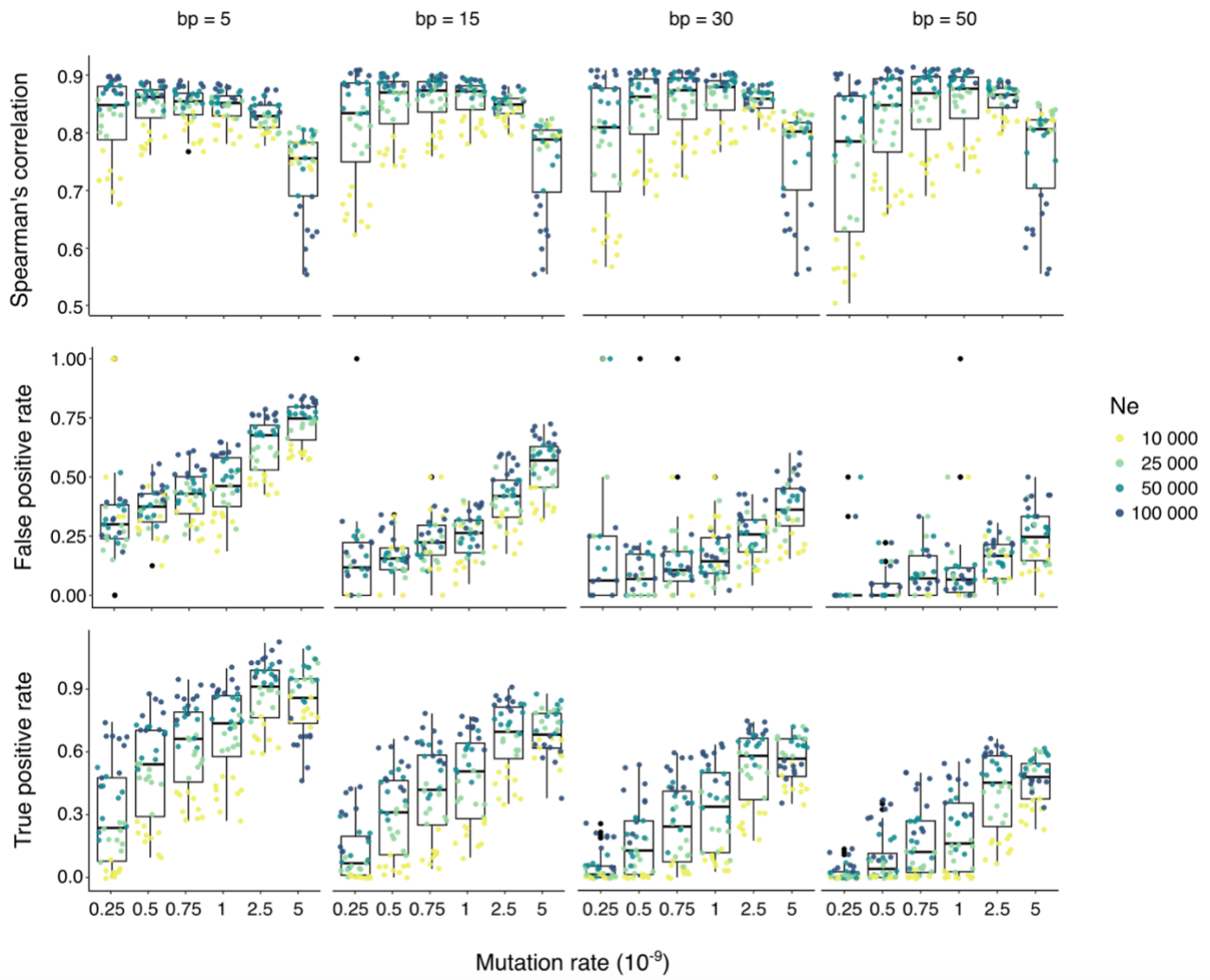

Fig. S6: Inference of recombination maps and hotspots from coalescent simulations. We used `msprime` to simulate ancestral history with mutations based on different mutation rates (μ), effective population sizes (N_e) and sample sizes. We inferred for each VCF the recombination landscape with *pyrho* using different block penalty values (bp = 5 to bp = 50). Here, we plot (i) the Spearman correlation, (ii) the false positive rate in hotspot discovery and (iii) the true positive rate in hotspot discovery between the recombination maps inferred from `msprime` simulation and the reference map. Recombination inference clearly depends on genetic diversity $\pi = 4N_e\mu$, and performs generally better when the level of π is higher.

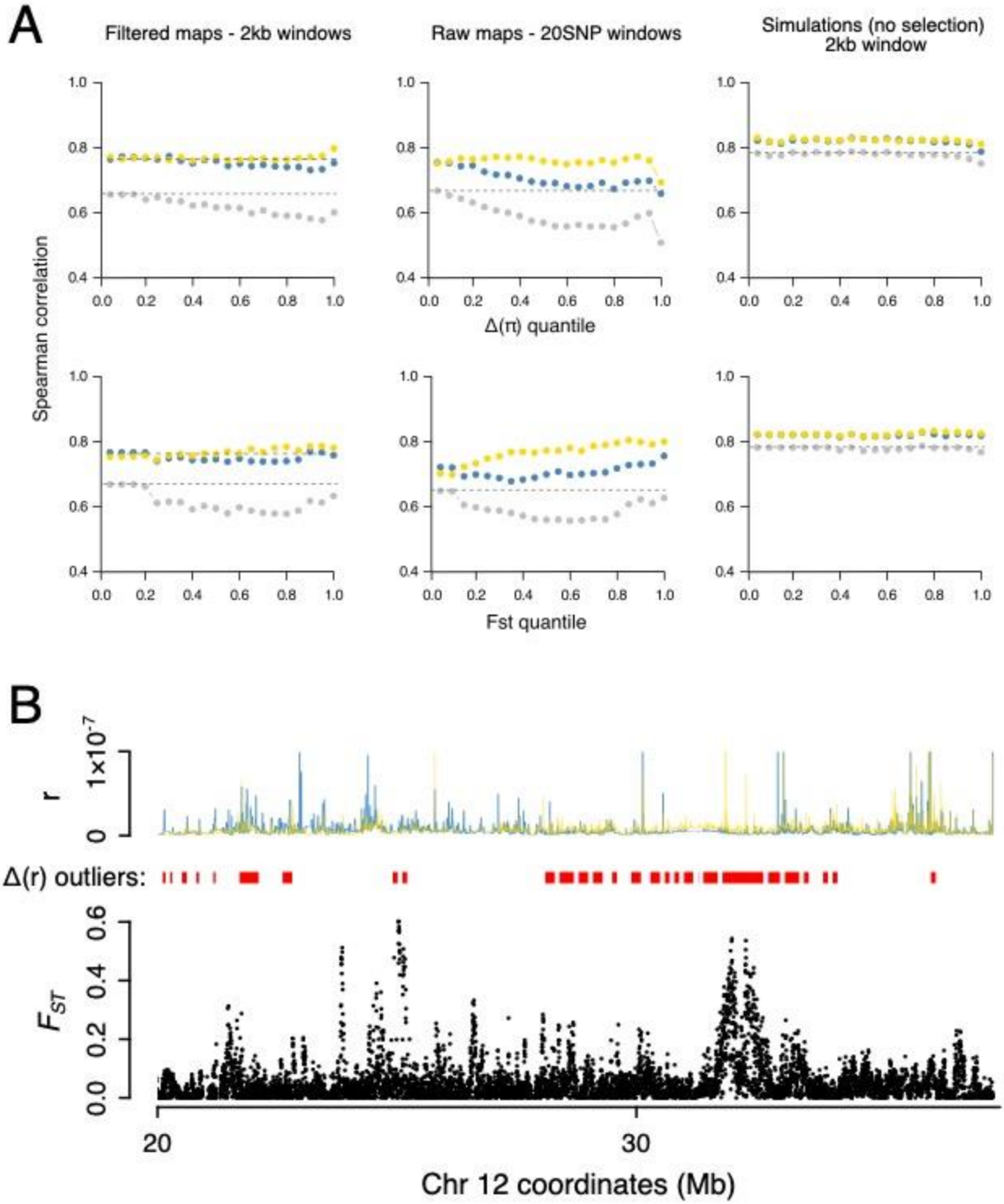

Fig. S7: The complex relationship between allele frequency divergence and recombination divergence. (**A**) Spearman correlation measured between recombination maps as a function of benthic-littoral F_{ST} and of Δ(π). For filtered maps, there is a clear trend of lower benthic vs. littoral correlations in regions of greater F_{ST} and of Δ(π). Changing the scale of the analysis from 2kb to 20SNP windows (i.e. equal number of variants but varying physical window size) has a noticeable effect on the results. Finally, there is no relationship between correlations of recombination maps in neutral coalescent simulations. (**B**) An illustrative example. While some Δ(r) outliers – regions of significantly divergent recombination – collocate clearly with elevated F_{ST} others do not and there are also clear F_{ST} peaks in regions where benthic and littoral recombination rates appear to be similar.

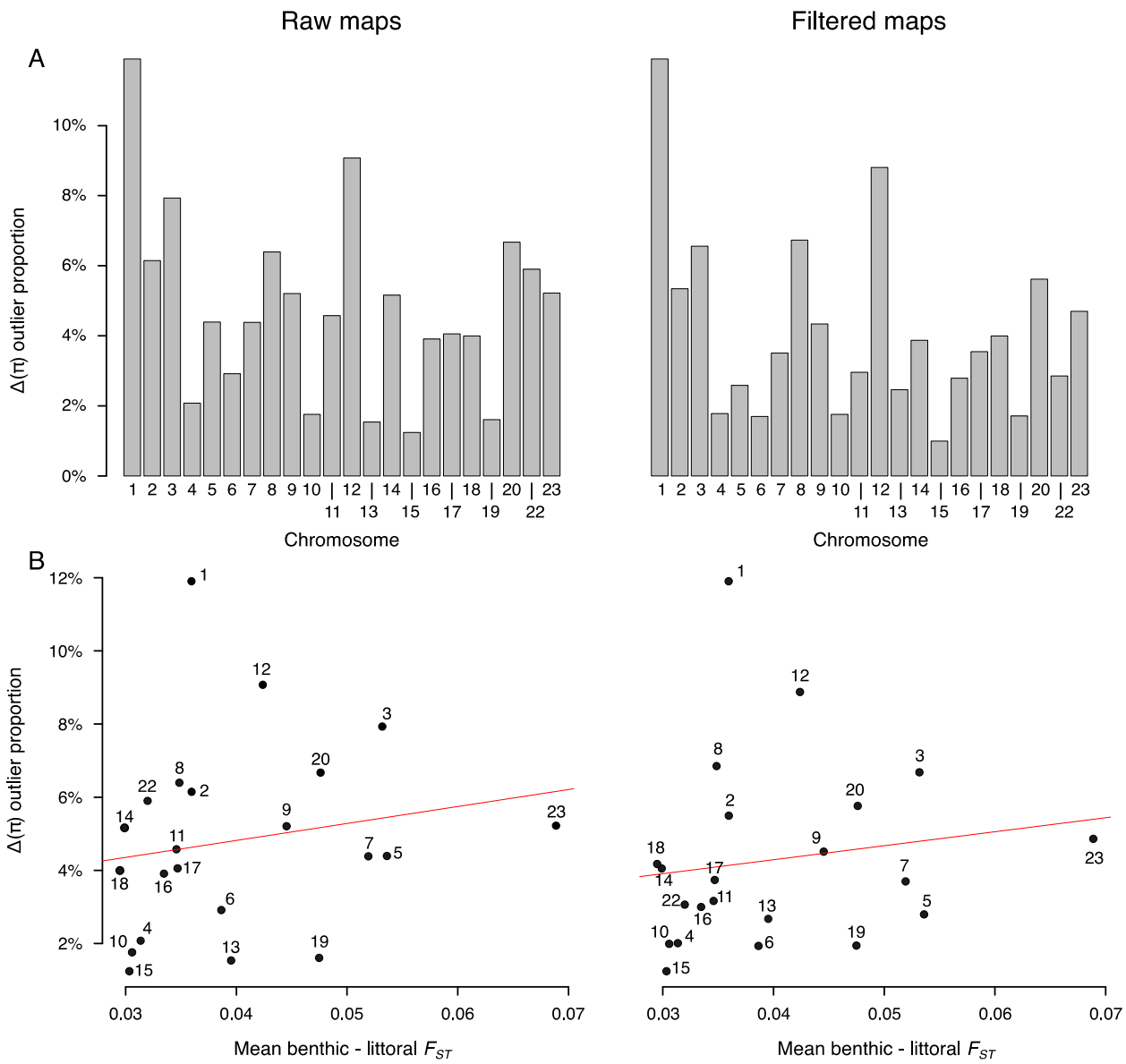

Fig. S8: The distribution of $\Delta(r)$ outliers across chromosomes. (**A**) The proportion of each chromosome assigned as $\Delta(r)$ outliers, i.e., regions with rapidly changing recombination landscapes. (**B**) The relationship between $\Delta(r)$ outlier proportion and mean F_{ST} per chromosome.

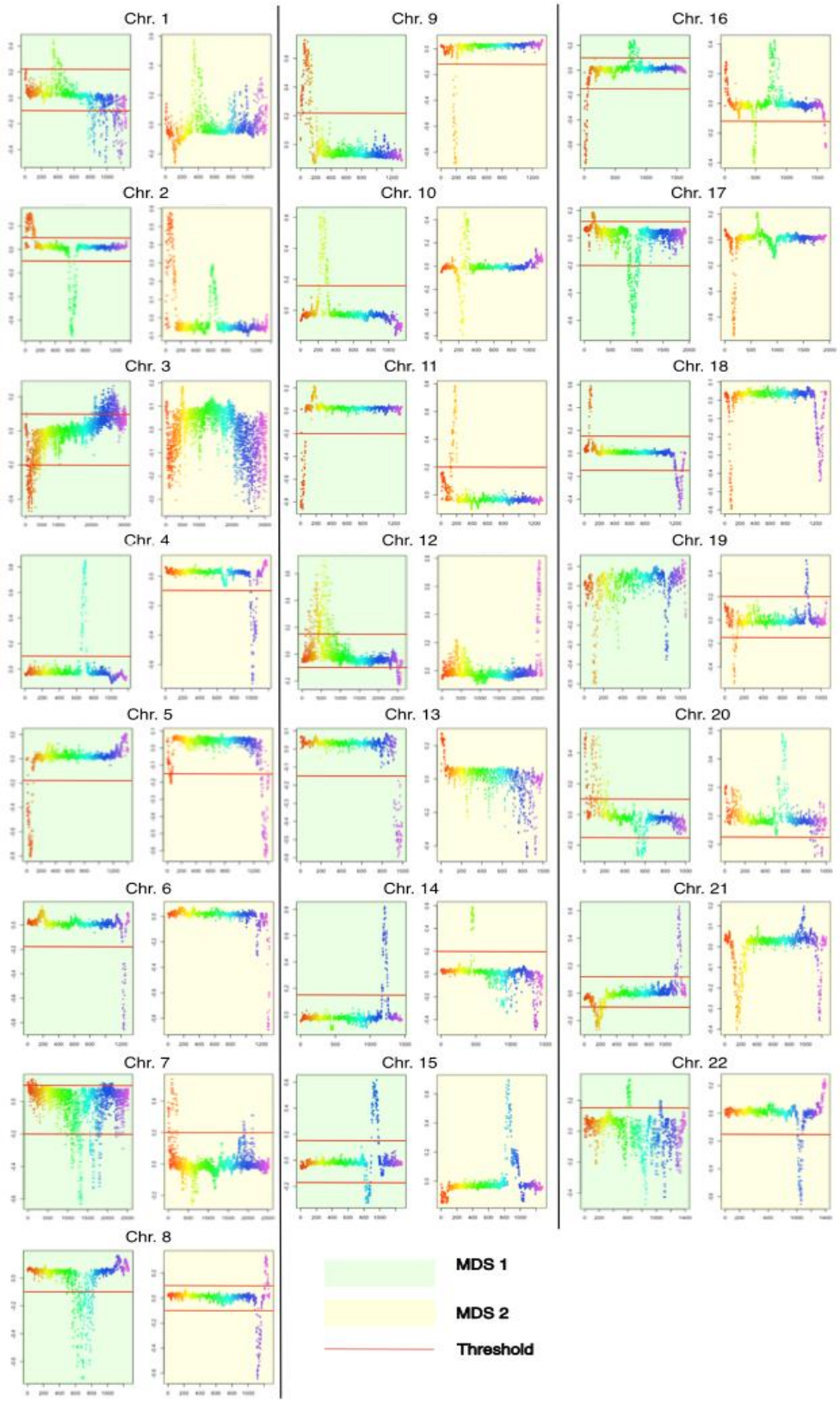

Fig. S9: Detection of genomic regions showing clearly distinct population structure. Each datapoint corresponds to a local PCA for a 100 SNP genomic window. The (dis-)similarity between population structure shown in different local PCA windows is measured by multidimensional scaling (MDS): the greater the MDS deviation from the chromosome average, the more distinct the local population structure. We visually inspected two MDS coordinates (MDS 1 and MDS 2) to select thresholds for each chromosome, identifying 47 distinct regions.

Fig. S10: Local PCAs within the 47 haplotype blocks identified with lostruct. Each blue square represents a benthic individual, and each yellow triangle a littoral individual. The patterns of population structure are very diverse, with some showing clear differences between the ecotypes. In all cases, the first two principal components explain substantially greater proportion of total genetic variation than in the whole genome PCA shown in Fig 1B.

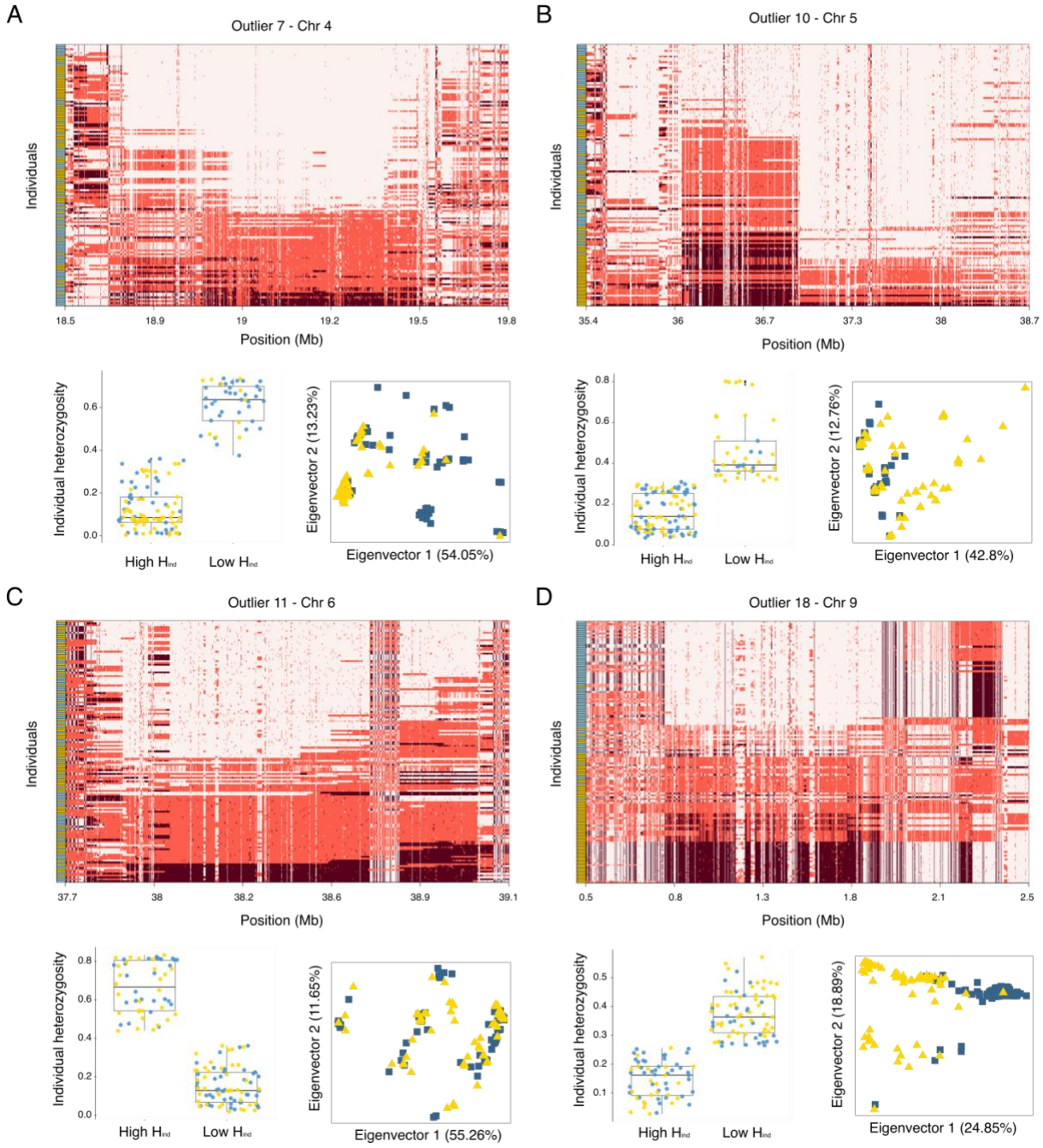

Fig. S11: Examples of large haplotype blocks with typical inversion signatures. For four `lostruct` outliers we show individual genotypes (with minor allele frequency > 5%; color scheme as in **Fig. 4A**), results of k-means clustering (K=2) of individuals based on their heterozygosity, and the local PCA for that region. We see sharp edges of haplotype blocks, individuals who are consistently heterozygous for long stretches, and individuals falling clear clusters in the PCA, all characteristic signatures of inversion. However, each of these `lostruct` outliers contains several different haplotype blocks. A more detailed study of history and mechanisms behind these haplotype blocks will require careful delineation of the boundaries of each of these haplotypes.

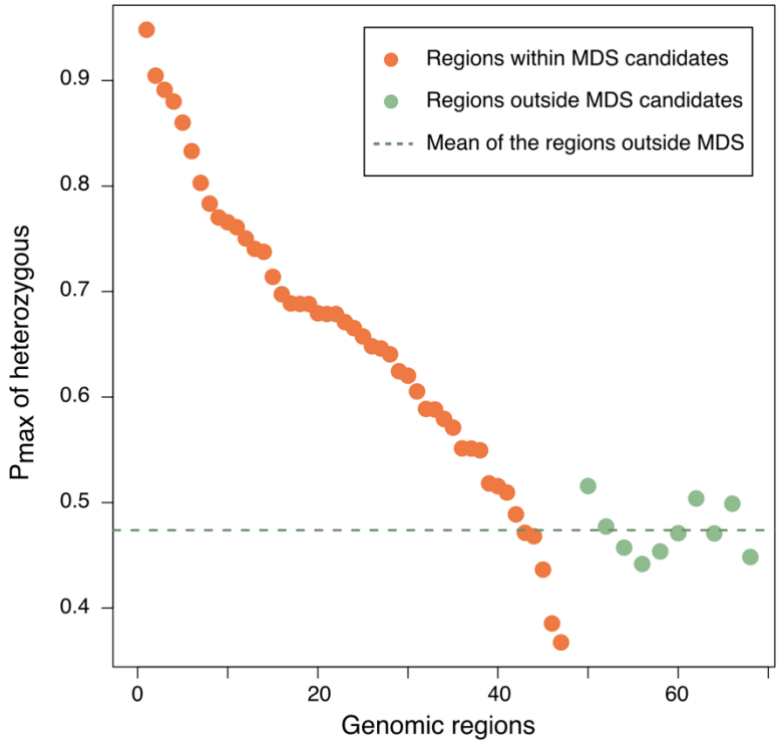

Fig. S12: High levels of individual heterozygosity in regions identified as distinct by local PCA. We found that for 90% of the local PCA outlier regions, the maximum of individual heterozygosity (H_{ind}) was higher than the mean of the equivalent measure in length-matched control regions.

Population	*Number of hotspots*		
	Define on 20kb around	**Define on 500kb around**	**Define on the chromosome mean**
Benthic a	2350	2275	2132
Benthic b	2488	2341	2249
Littoral a	2086	2345	2378
Littoral b	2086	2328	2433

Supplementary Table S1: The number of hotspot defined relative to the 3 different background that we used. We defined a hotspot of recombination in a given genomic region if the recombination rate was at least 5 times higher than the recombination rate in the background region. The genomic area longer than 5kb were not considered as hotspots.

Supplementary Table S2: The genomic coordinates and key statistics for the 47 regions identified as local PCA outliers. The "cluster p-value" column refers to the ecotype clustering $\log_{10}$ p-value shown in Fig. 5A. HDRs refers to the number of highly diverged regions (HDRs; defined as in Malinsky et al. (2015)) that overlap the local PCA outlier.

chr	start	end	F	F_{ST}	H_{ind}	cluster p-value	$\log_{10} \Delta(r)$	TMRCA	HDRs
1	12102312	16189411	0.0166	0.0582	0.6882	5.78	0.2992	168146	0
1	26036869	41119330	9e-04	0.0421	0.5495	5.01	0.1899	163307	1
2	19465	1831057	0.0732	0.055	0.6711	0	0.1073	232967	0
2	14469831	18569566	-0.04	0.0112	0.6787	0.04	0.1326	169577	0
3	1119826	10621518	0.0513	0.1027	0.3854	4.36	0.2042	369977	15
3	37891964	63560498	0.0404	0.0515	0.3675	1.35	0.0882	825661	3
4	18591787	19609472	0.0568	0.1531	0.7405	3.6	0.1257	1235990	1
4	27040234	29900969	-0.0241	0.0069	0.6787	0.08	0.0835	224469	0
5	61769	1324715	0.0616	0.008	0.8801	0.42	-0.015	223426	0
5	35372227	38655485	0.0955	0.0825	0.803	2.57	0.2263	279811	0
6	37886749	39005213	0.0457	0.0063	0.8331	0.1	0.0254	255967	0
7	44298	5486542	0.0239	0.0197	0.5156	0.14	-0.018	292390	0
7	23358419	45879790	0.0585	0.0937	0.4364	1.54	0.1168	157760	44
7	45461192	55025348	0.0297	0.0312	0.5182	0.37	0.091	155663	4
8	13442845	19170725	0.035	0.0619	0.6053	6.08	0.2107	209024	0
8	26904667	27684980	0.0996	0.0168	0.6654	0.54	-0.008	313103	0
8	28217567	29212295	-0.0138	0.0222	0.7377	0.68	-0.0188	194304	0
9	194816	5233966	0.0872	0.1489	0.571	1.31	0.1255	254345	11
9	3624571	5069209	-0.003	0.0255	0.7702	0.46	0.2347	236930	0
10	4970531	7454652	-0.0154	0.0181	0.5512	0.28	0.1042	157832	0
11	42902	989171	-0.0259	0.0243	0.8601	1.6	2e-04	261982	0
11	1717094	2887045	0.1291	0.0295	0.7833	2.4	0.3228	232926	0
12	3835486	14509449	0.0352	0.0325	0.4888	1.54	0.0831	143924	0
12	35703715	36304361	-0.0166	0.05	0.6245	0.38	0.0546	1444439	0
13	30006077	32432700	0.0561	0.0774	0.6975	2.87	0.3059	219150	0
14	13470079	14575006	-0.086	0.0208	0.761	0.71	0.1837	283673	0
14	34884867	35895818	-0.0424	0.017	0.7657	0.3	0.0561	262086	0
15	25408840	28073014	-0.074	0.0135	0.7139	0.72	0.1144	181994	0
15	28507085	31697055	0.05	0.0775	0.6882	0.14	0.0659	211765	0
16	51475	1382176	0.0103	0.0322	0.646	0.84	0.0247	509643	0
16	6127996	6968254	-0.0712	0.0076	0.6481	0.26	0.0402	330029	0
16	10128603	14367042	0.0043	0.0207	0.4714	0.37	0.1719	195626	1
16	34796693	35691328	-0.0892	0.0372	0.9482	0.45	-0.0203	387988	0
17	3768696	4458439	0.1491	0.0188	0.9046	1.69	0.6097	190618	0
17	11894487	14516568	0.0987	0.141	0.5793	3.72	0.0971	239485	22
18	1680439	2232789	-0.0731	0.0104	0.8912	0.6	0.0669	674088	0
18	32534225	34007640	0.0477	0.0233	0.6794	0	0.0387	285507	0
19	2765772	4034315	0.0489	0.0918	0.7504	1.82	0.2004	189647	1
19	24725346	25415225	0.0812	0.1637	0.6887	3.61	0.1061	233503	7
20	57079	5144292	-0.0058	0.036	0.5514	4.61	0.3948	238749	0
20	14695766	18898982	0.0438	0.0887	0.6405	2.17	0.1082	148164	2
20	25953482	30007504	0.0626	0.0516	0.5095	0.14	0.1265	183742	3
22	2331584	5114600	0.0251	0.0253	0.4681	0.17	0.1088	374464	0
22	30331726	32901550	-3e-04	0.0698	0.6574	0.66	0.0585	183682	3
23	16708893	17785992	0.0104	0.0279	0.6202	0.51	0.0578	476754	0
23	32884884	34749208	-0.0198	0.0512	0.5888	0.31	0.2424	189584	1
23	43724278	44592772	0.1167	0.0428	0.5882	0	0.0332	468786	0

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–410.

Auton A, Fledel-Alon A, Pfeifer S, Venn O, Ségurel L, Street T, Leffler EM, Bowden R, Aneas I, Broxholme J, et al. 2012. A Fine-Scale Chimpanzee Genetic Map from Population Sequencing. *Science (New York, N.Y.)* 336:193–198.

Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME Suite. *Nucleic Acids Res.* 43:W39–W49.

Baker Z, Przeworski M, Sella G. 2023. Down the Penrose stairs, or how selection for fewer recombination hotspots maintains their existence. *eLife* 12:e83769.

Baker Z, Schumer M, Haba Y, Bashkirova L, Elife CH, 2017. 2017. Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates. *cdn.elifesciences.org*.

Barker, Williamson, Gasse, Gibert. 2003. Climatic and volcanic forcing revealed in a 50,000-year diatom record from Lake Massoko, Tanzania. *Quaternary Research* 60:9–9.

Barton NH. 2022. The "New Synthesis." *Proc. Natl. Acad. Sci.* 119:e2122147119.

Battlay P, Wilson J, Bieker VC, Lee C, Prapas D, Petersen B, Craig S, Boheemen L van, Scalone R, Silva NP de, et al. 2023. Large haploblocks underlie rapid adaptation in the invasive weed Ambrosia artemisiifolia. *Nat. Commun.* 14:1717.

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.

Baumdicker F, Bisschop G, Goldstein D, Gower G, Ragsdale AP, Tsambos G, Zhu S, Eldon B, Ellerman EC, Galloway JG, et al. 2021. Efficient ancestry and mutation simulation with msprime 1.0. *Genetics* 220:iyab229.

Bhatia G, Patterson N, Sankararaman S, Price AL. 2013. Estimating and interpreting FST: The impact of rare variants. *Genome Research* 23:1514–1521.

Booker TR, Yeaman S, Whitlock MC. 2020. Variation in recombination rate affects detection of outliers in genome scans under neutrality. *Mol Ecol* 29:4274–4279.

Brazier T, Glémin S. 2022. Diversity and determinants of recombination landscapes in flowering plants. *PLoS Genet.* 18:e1010141.

Burri R, Nater A, Kawakami T, Mugal CF, Olason PI, Smeds L, Suh A, Dutoit L, Bures S, Garamszegi LZ, et al. 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.

Carruthers M, Edgley DE, Saxon AD, Gabagambi NP, Shechonge A, Miska EA, Durbin R, Bridle JR, Turner GF, Genner MJ. 2022. Ecological Speciation Promoted by Divergent Regulation of Functional Genes Within African Cichlid Fishes. *Mol. Biol. Evol.* 39:msac251.

Cavassim MIA, Baker Z, Hoge C, Schierup MH, Schumer M, Przeworski M. 2022. PRDM9 losses in vertebrates are coupled to those of paralogs ZCWPW1 and ZCWPW2. *Proc. Natl. Acad. Sci.* 119:e2114401119.

Coop G, Przeworski M. 2007. An evolutionary view of human recombination. *Nature reviews. Genetics* 8:23–34.

Coster WD, Weissensteiner MH, Sedlazeck FJ. 2021. Towards population-scale long-read sequencing. *Nat. Rev. Genet.* 22:572–587.

Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology* 23:3133–3157.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* 10:giab008.

DePristo MAM, Banks EE, Poplin RR, Garimella KVK, Maguire JRJ, Hartl CC, Philippakis AAA, Angel GG del, Rivas MAM, Hanna MM, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* 43:491–498.

Excoffier L, Marchi N, Marques DA, Matthey-Doret R, Gouy A, Sousa VC. 2021. fastsimcoal2 : demographic inference under complex evolutionary scenarios. *Bioinformatics* 37:4882–4885.

Faria R, Chaube P, Morales HE, Larsson T, Lemmon AR, Lemmon EM, Rafajlović M, Panova M, Ravinet M, Johannesson K, et al. 2019. Multiple chromosomal rearrangements in a hybrid zone between Littorina saxatilis ecotypes. *Mol. Ecol.* 28:1375–1393.

Faria R, Johannesson K, Butlin RK, Westram AM. 2019. Evolving Inversions. *Trends Ecol. Evol.* 34:239–248.

Feldman MW, Otto SP, Christiansen FB. 1996. POPULATION GENETIC PERSPECTIVES ON THE EVOLUTION OF RECOMBINATION. *Annu Rev Genet* 30:261–295.

Gel B, Díez-Villanueva A, Serra E, Buschbeck M, Peinado MA, Malinverni R. 2016. regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics* 32:289–291.

Genestier A, Duret L, Lartillot N. 2023. Bridging the gap between the evolutionary dynamics and the molecular mechanisms of meiosis : a model based exploration of the PRDM9 intra-genomic Red Queen. *bioRxiv*:2023.03.08.531712.

Haenel Q, Laurentino TG, Roesti M, Berner D. 2018. Meta-analysis of chromosome-scale crossover rate variation in eukaryotes and its significance to evolutionary genomics. *Mol. Ecol.* 27:2477–2497.

Halldorsson BV, Palsson G, Stefansson OA, Jonsson H, Hardarson MT, Eggertsson HP, Gunnarsson B, Oddsson A, Halldorsson GH, Zink F, et al. 2019. Characterizing mutagenic

effects of recombination through a sequence-level genetic map. *Science (New York, N.Y.)* 363:eaau1043.

Harringmeyer OS, Hoekstra HE. 2022. Chromosomal inversion polymorphisms shape the genomic landscape of deer mice. *Nat. Ecol. Evol.* 6:1965–1979.

Henderson IR, Bomblies K. 2021. Evolution and Plasticity of Genome-Wide Meiotic Recombination Rates. *Annu Rev Genet* 55:1–21.

Hoge C, Manuel M de, Mahgoub M, Okami N, Fuller Z, Banerjee S, Baker Z, McNulty M, Andolfatto P, Macfarlan TS, et al. 2023. Patterns of recombination in snakes reveal a tug of war between PRDM9 and promoter-like features. *bioRxiv*:2023.07.11.548536.

Jay P, Whibley A, Frézal L, Cara MÁR de, Nowell RW, Mallet J, Dasmahapatra KK, Joron M. 2018. Supergene Evolution Triggered by the Introgression of a Chromosomal Inversion. *Curr. Biol.* 28:1839-1845.e3.

Joseph J, Prentout D, Laverré A, Tricou T, Duret L. 2023. High prevalence of Prdm9-independent recombination hotspots in placental mammals. *bioRxiv*:2023.11.17.567540.

Kent TV, Uzunović J, Wright SI. 2017. Coevolution between transposable elements and recombination. *Philos. Trans. R. Soc. B: Biol. Sci.* 372:20160458.

Lam I, Keeney S. 2015. Nonparadoxical evolutionary stability of the recombination initiation landscape in yeast. *Science (New York, N.Y.)* 350:932–937.

Latrille T, Duret L, Lartillot N. 2017. The Red Queen model of recombination hot-spot evolution: a theoretical investigation. *Philos. Trans. R. Soc. B: Biol. Sci.* 372:20160463.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv.org* [Internet] q-bio.GN. Available from: arXiv.org

Li H, Ralph P. 2018. Local PCA Shows How the Effect of Population Structure Differs Along the Genome. *Genetics* 211:289–304.

Ma J, Amos CI. 2018. Investigation of Inversion Polymorphisms in the Human Genome Using Principal Components Analysis. *PLoS ONE* 7:e40224.

Malinsky M, Challis RJ, Tyers AM, Schiffels S, Terai Y, Ngatunga BP, Miska EA, Durbin R, Genner MJ, Turner GF. 2015. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science* 350:1493–1498.

Malinsky M, Svardal H, Tyers AM, Miska EA, Genner MJ, Turner GF, Durbin R. 2018. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology & Evolution* 2:1940–1955.

Marques DA, Meier JI, Seehausen O. 2019. A Combinatorial View on Speciation and Adaptive Radiation. *Trends Ecol. Evol.* 34:531–544.

Matthey-Doret R, Whitlock MC. 2019. Background selection and FST : consequences for detecting local adaptation. *Molecular Ecology* 28:mec.15197-3914.

McDonald MJ, Rice DP, Desai MM. 2016. Sex speeds adaptation by altering the dynamics of molecular evolution. *Nature* 531:233–236.

Mérot C. 2020. Making the most of population genomic data to understand the importance of chromosomal inversions for adaptation and speciation. *Mol. Ecol.* 29:2513–2516.

Mérot C, Oomen RA, Tigano A, Wellenreuther M. 2020. A Roadmap for Understanding the Evolutionary Significance of Structural Genomic Variation. *Trends Ecol. Evol.* 35:561–572.

Miller W, Rosenbloom K, Hardison RC, Hou M, Taylor J, Raney B, Burhans R, King DC, Baertsch R, Blankenberg D, et al. 2007. 28-way vertebrate alignment and conservation track in the UCSC Genome Browser. *Genome Research* 17:1797–1808.

Munby H, Linderoth T, Fischer B, Du M, Vernaz G, Tyers AM, Ngatunga BP, Shechonge A, Denise H, McCarthy SA, et al. 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 RE, Freeman C, MacFie TS, McVean G, Donnelly P. 2010. Drive Against Hotspot Motifs in Primates Implicates the PRDM9 Gene in Meiotic Recombination. *Science (New York, N.Y.)* 327:876–879.

Nei M. 1967. MODIFICATION OF LINKAGE INTENSITY BY NATURAL SELECTION. *Genetics* 57:625–641.

Nielsen R. 2006. Why Sex? *Science* 311:960–961.

O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44:D733–D745.

Ortiz-Barrientos D, Engelstädter J, Rieseberg LH. 2016. Recombination Rate Evolution and the Origin of Species. *Trends in Ecology & Evolution* 31:226–236.

Otto SP, Payseur BA. 2019. Crossover Interference: Shedding Light on the Evolution of Recombination. *Annu. Rev. Genet.* 53:1–26.

Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics (Oxford, England)* 20:289–290.

Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLoS genetics* 2:e190–e190.

Peñalba JV, Wolf JBW. 2020. From molecules to populations: appreciating and estimating recombination rate variation. *Nat Rev Genet* 21:476–492.

Persikov AV, Osada R, Singh M. 2009. Predicting DNA recognition by Cys2His2 zinc finger proteins. *Bioinformatics* 25:22–29.

Persikov AV, Singh M. 2014. De novo prediction of DNA-binding specificities for Cys2His2 zinc finger proteins. *Nucleic Acids Res.* 42:97–108.

Petronczki M, Siomos MF, Nasmyth K. 2003. Un Ménage à Quatre The Molecular Biology of Chromosome Segregation in Meiosis. *Cell* 112:423–440.

Ponticelli AS, Sena EP, Smith GR. 1988. Genetic and physical analysis of the M26 recombination hotspot of Schizosaccharomyces pombe. *Genetics* 119:491–497.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, Bakker PIW de, Daly MJ, et al. 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* 81:559–575.

Reeve J, Butlin RK, Koch EL, Stankowski S, Faria R. 2023. Chromosomal inversion polymorphisms are widespread across the species ranges of rough periwinkles (Littorina saxatilis and L. arcana). *Mol. Ecol.*

Rice WR, Chippindale AK. 2001. Sexual Recombination and the Power of Natural Selection. *Science* 294:555–559.

Ritz KR, Noor MAF, Singh ND. 2017. Variation in Recombination Rate: Adaptive or Not? *Trends Genet* 33:364–374.

Rolland J, Henao-Diaz LF, Doebeli M, Germain R, Harmon LJ, Knowles LL, Liow LH, Mank JE, Machac A, Otto SP, et al. 2023. Conceptual and empirical bridges between micro- and macroevolution. *Nat. Ecol. Evol.* 7:1181–1193.

Rowan BA, Heavens D, Feuerborn TR, Tock AJ, Henderson IR, Weigel D. 2019. An Ultra High-Density Arabidopsis thaliana Crossover Map That Refines the Influences of Structural Variation and Epigenetic Features. *Genetics* 213:771–787.

Samuk K, Manzano-Winkler B, Ritz KR, Noor MAF. 2020. Natural Selection Shapes Variation in Genome-wide Recombination Rate in Drosophila pseudoobscura. *Curr Biol* 30:1517-1528.e6.

Schluter D, Rieseberg LH. 2022. Three problems in the genetics of speciation by selection. *Proc. Natl. Acad. Sci.* 119:e2122153119.

Shanfelter AF, Archambeault SL, White MA. 2019. Divergent Fine-Scale Recombination Landscapes between a Freshwater and Marine Population of Threespine Stickleback Fish.Hurst L, editor. *Genome Biology and Evolution* 11:1573–1585.

Singhal S, Leffler EM, Sannareddy K, Turner I, Venn O, Hooper DM, Strand AI, Li Q, Raney B, Balakrishnan CN, et al. 2015. Stable recombination hotspots in birds. *Science* 350:928–932.

Spence JP, Song YS. 2019. Inference and analysis of population-specific fine-scale recombination maps across 26 diverse human populations. *Sci Adv* 5:eaaw9206.

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. *J. Mol. Biol.* 204:917–925.

Terhorst J, Kamm JA, Song YS. 2017. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* 49:303–309.

Todesco M, Owens GL, Bercovich N, Légaré J-S, Soudi S, Burge DO, Huang K, Ostevik KL, Drummond EBM, Imerovski I, et al. 2020. Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature* 584:602–607.

Úbeda F, Wilkins JF. 2010. The Red Queen theory of recombination hotspots. *J. Evol. Biol.* 24:541–553.

Vernaz G, Hudson AG, Santos ME, Fischer B, Carruthers M, Shechonge AH, Gabagambi NP, Tyers AM, Ngatunga BP, Malinsky M, et al. 2022. Epigenetic divergence during early stages of speciation in an African crater lake cichlid fish. *Nat Ecol Evol*:1–12.

Wellenreuther M, Bernatchez L. 2018. Eco-Evolutionary Genomics of Chromosomal Inversions. *Trends Ecol Evol* 33:427–440.

Yeaman S. 2022. Evolution of polygenic traits under global vs local adaptation. *Genetics* 220:iyab134-.