# 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:**
Marion Talbi: marion.talbi@gmail.com
Milan Malinsky: 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 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 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 (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). 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 (Úbeda and Wilkins 2010; Latrille et al. 2017; Baker et al. 2023) 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. 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) and there is evidence that species lacking PRDM9 have more conserved recombination landscapes (Lam and Keeney 2015; Singhal et al. 2015). However, the association with these 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; Déserts et al. 2021). 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 (Felsenstein 1981; Barton 2020; Butlin et al. 2021). Consistent with the important role in this context, a recombination landscape, together with natural selection, shape the distribution of genomic regions of divergence along each chromosome (Ravinet et al. 2017; Duranton et al. 2018; Schumer et al. 2018; Martin et al. 2019). The recombination landscape itself is an evolving dynamic parameter (Ritz et al. 2017; Samuk et al. 2020; Déserts et al. 2021), which should be taken into account in genomic studies of speciation (Ortiz-Barrientos et al. 2016; Ortiz-Barrientos and James 2017). In recent years, there has been a growing appreciation of the role of recombination suppression in organismal diversification (Schluter and Rieseberg 2022). These efforts often concentrated on specific large inversions or other low-

87 recombining haplotype blocks (Jay et al. 2018; Faria et al. 2019a; Todesco et al. 2020; Reeve et al.
88 2023). In addition, a recent study based on genome-wide estimates of recombination rates supports
89 the notion that cis-acting recombination modifiers play an important role in promoting adaptive
90 divergence between populations (Venu et al. 2024). Building upon this initial progress, additional
91 comparisons of recombination landscapes in different species and across different levels of
92 divergence will be required to understand where in the genome, how fast, and by what mechanisms
93 recombination rates evolve and, ultimately, the interplay with natural selection and organismal
94 evolution.

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

105 Over the last decade or so, many empirical studies used recombination estimates based on linkage
106 disequilibrium (LD) – patterns of non-random association of genetic variants in a sample of
107 individuals from a population (Auton et al. 2012; Singhal et al. 2015; Shanfelter et al. 2019; Spence
108 and Song 2019). LD-based methods estimate a 'population recombination rate' ($\rho$) – a population
109 genetic parameter which is a product of effective population size ($N_e$) and the historical per-
110 generation recombination rate averaged over the ancestry of the sampled individuals. Differences

111 between LD-based recombination maps can arise from sampling variance, methodological
112 limitations, different ancestries across samples, and factors affecting $N_e$ such as demographic change
113 and natural selection (Coop and Przeworski 2007; Peñalba and Wolf 2020; Samuk et al. 2020).
114 Therefore, it is challenging to use LD-based methods to compare recombination landscapes between
115 populations and species to track their evolution.

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

## 129 Material and Methods

130 **Variant calling and filtering**

131 Genomic DNA from a total of 336 individuals from Lake Masoko has been sequenced on the Illumina
132 HiSeq X Ten platform, obtaining 150bp paired end reads (NCBI Short Read Archive, BioProject ID:
133 PRJEB27804). The reads were aligned to the *Astatotilapia calliptera* reference genome fAstCal1.5
134 (GenBank ID: GCA_900246225.6) using bwa-mem v.0.7.17 (Li 2013), with median depth per-

135 individual of 15.8x (min = 12.4x, max = 22.1x). The reference sequence is based on an *A. calliptera*
136 sample from the Itupi stream which is a close outgroup to Lake Masoko. We used the
137 `MarkDuplicates` tool from the Picard package v.2.26.6 to tag PCR and optical duplicate reads
138 and GATK v.4.2.3 (DePristo et al. 2011) to call variants, using `HaplotypeCaller` in GVCF mode
139 for each individual separately followed by joint genotyping using `GenotypeGVCFs` with the `--`
140 `include-non-variant-sites` option.

141 Next, we generated a *callability mask* to identify and filter out the regions of the genome where we
142 were unable to confidently call variants. The mask included: i) sites with an overall read depth cutoffs
143 based on examining a depth histogram (< 3800 or > 5700); ii) sites where more than 10% of
144 individuals had missing genotypes; iii) sites identified by GATK as low quality (with the `LowQual`
145 tag) and iv) sites with poor mappability. Specifically, to obtain the mappability information, we broke
146 down the genome into overlapping k-mers of 150bp (matching the read length), mapped these k-
147 mers back to the genome, and masked all sites where fewer than 90% of k-mers mapped back to
148 their original location perfectly and uniquely. In total, the callability mask comprised of 311 million
149 bp, or about 35% of the genome. In addition to applying the callability mask, we used several hard
150 filters based on GATK best practices, specifically `MQ<40`, `FS>40`, `QD<2` and `ExcessHet>40`.
151 These additional filters removed fewer than 1% of the remaining SNPs. After discarding 1.4 million
152 of multiallelic sites and indels, the filtered VCF contained 3.86 million SNPs.

153 **Sample selection for recombination analyses**

154 We used the full set of 336 available individuals for variant calling because the inclusion of more
155 samples leads to more accurate genotyping. However, in this study we were specifically interested
156 in differences between the littoral and benthic ecotypes of Lake Masoko. Therefore, we retained 80
157 individuals assigned in the field as benthic and 79 assigned as littoral and excluded 201 other
158 individuals which were not assigned to either ecotype because they were juveniles, females (neither

159 category show the ecotype-distinct male breeding colors), or putative hybrids. To check the validity
160 of these field assignments, we first built a neighbor-joining tree based on a genetic distance matrix,
161 i.e., the average number of single-nucleotide differences between each pair of individuals, using the
162 `stats` command from the evo package v.0.1 r28 and the `--diff-matrix` option. The pairwise
163 difference matrix was divided by the callable genome size to obtain pairwise distances per base pair
164 and this was then used as input into the `nj()` tree-building function implemented in the package
165 *ape* in R (Paradis et al. 2004). Next, for principal component analysis we used smartPCA (Patterson
166 et al. 2006) on data filtered for minor allele frequency >=0.05 using plink v1.9 (Purcell et al. 2007)
167 with the `--maf 0.05` option and LD pruned using the plugin `+prune` from bcftools v.1.16
168 (Danecek et al. 2021) with the `-m 0.8 -w 1000` options. After we identified and removed 20
169 genetically intermediate samples (see **Figure 1C** and **Supplementary Figure 1**), the final VCF with
170 139 individuals was composed of 3.3 million biallelic SNPs.

171 **Genome annotation**

172 Because the fAstCal1.5 assembly was not annotated by NCBI, we 'lifted over' the annotation from
173 the older fAstCal1.2 assembly (GenBank: GCA_900246225.3). We used the UCSC paradigm (Miller
174 et al. 2007) to generate a pairwise whole genome alignment between fAstCal1.2 and fAstCal1.5
175 assembly. Afterwards, we used the UCSC `liftOver` tool to translate the NCBI Annotation Release
176 100 to the new coordinates.

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

178 To estimate split time between the ecotypes and changes in effective population size ($N_e$) through
179 time, we used smc++ v.1.15.4 (Terhorst et al. 2017), using the sequence of smc++ commands:
180 `vcf2smc` -> `estimate` -> `split`, To translate the time axis into number of generations, we used
181 the cichlid-specific mutation rate estimate of $\mu = 3.5\,\mathrm{x}10^{-9}$ per bp per generation with 95%
182 confidence interval $(1.6\,\mathrm{x}10^{-9}, 4.6\,\mathrm{x}10^{-9})$ (Malinsky et al. 2018). Next, we used fastsimcoal2.7
183 (Excoffier et al. 2021) to estimate the level of gene flow between the two ecotypes. In fastsimcoal,

184 we entered the split time and changes in $N_e$ as inferred by smc++ as fixed parameters and estimated
185 continuous asymmetrical migration rates after the ecotype split. We ran 30 simulations with
186 different starting parameter values, which revealed two local peaks in the likelihood surface
187 (**Supplementary Figure 2**). To reduce the confounding effects of selection in these demographic
188 analyses we used only sites from non-coding regions of the genome, masking all annotated exons,
189 introns, and promoters.

190 **Subsampling and bootstrap**

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

198 **Inference of recombination rates**

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

204 Specifically, first, to build likelihood tables between each biallelic sites, we used the `make_table`
205 command with $\mu = 3.5 \times 10^{-9}$, demographic history for each of the ecotypes as inferred by smc++,
206 and the Moran approximation with the `-approx` and `-moran_pop_size N` flags where N
207 equals 1.5x the number of haplotypes in each subset. To determine the best parameter to use for

208 the inference of recombination rates, we processed a set of simulation with evolutionary parameters
209 corresponding to the one of our cichlid species (e.g., $\mu$ , sample size, $N_e$) and chose the value of block
210 penalty that was minimizing the quantity of false negative and false positive with best results
211 obtained for block penalty of 15 and a window size of 50 SNPs. These parameters were then used in
212 all runs of the `optimize` command to infer the recombination maps. The output of `pyrho`
213 contained estimates of recombination rate between each pair of SNPs.

214 **Neutral coalescent simulations**

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

224 To confirm that the conclusions of this manuscript are robust to the specifics of demographic
225 inference, we ran the simulations with an extended set of split time values (1 000, 2 500, 5 000 and
226 10 000 generations ago) and with migration rates corresponding to the two local likelihood peaks in
227 fastsimcoal2 inference ('low migration rates': $11.5 \times 10^{-5}$ for littoral to benthic and $7.01 \times 10^{-5}$ for
228 benthic to littoral; and 'high migration rates': 0.0339 for littoral to benthic and 0.0368 for benthic to
229 littoral).

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

240 $$P_{as} = \frac{(N_{tb} - 1) \ + \ (N_{tl} - 1)}{(N_t - 1)}$$

241 The scripts used to run these simulations are available from GitHub.

242 **Processing and comparisons of recombination maps**

243 We used the `PhysicalWindowAverages` command from the evo package v.0.1 r28 to obtain
244 mean recombination rates in identical 2Kb windows for all datasets, which facilitated easy
245 comparisons between different maps. The correlations, map distances, and other comparisons were
246 then calculated using R scripts, available from GitHub.

247

248 The first set of comparisons among recombination maps was based on correlations, which has
249 allowed us to define a measure of divergence which we call the Population Recombination
250 Divergence Index (PRDI). Briefly, we calculated pairwise Spearman correlations of recombination
251 maps in 2kb windows, binned per 5Mb regions along the genome and obtained the median of these
252 correlations. We first calculated the correlations for the within-ecotype replicate maps (medians
253 denoted $m_{wb}$ and $m_{wl}$ for within-benthic and within-littoral replicates respectively). Then we

calculated correlations for map comparisons between ecotypes, denoting the median of all these comparisons $m_b$. Between-ecotype correlations are generally lower than correlations for within-ecotype replicates and we measure this difference using the medians as:

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

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 non-overlapping 100kb windows (i.e. vectors of 50 values for 2kb each) using the `dist()` function in R. We use *log₁₀* transformed absolute (Manhattan) distances, because these are straightforward to interpret: a *log₁₀* 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 *log₁₀* 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

279 variation (i.e., amount of data available for inference). To understand how our results are affected
280 by these factors, we generated filtered recombination maps where the less reliable regions of the
281 genome were masked. While the main figures of this manuscript report the results for the raw maps,
282 we conducted many of the key analyses also using the filtered maps and present these results as
283 supplementary figures.

284 First, errors in the reference genome can mistakenly place in physical proximity genetic variants that
285 have large genetic distances between them. Therefore, we masked intervals 50kb upstream and
286 50kb downstream of each joint between contiguous sequences (contig joint) in the assembly. There
287 were 514 joints on the 22 chromosomes, leading to masking of 50.46Mb of sequence. Second, we
288 used the callability mask produced for variant filtering (see above). A lack of data can make
289 recombination inference difficult. We reduced this effect by masking each 100kb region within which
290 more than 70% were not callable. Third, to exclude regions where the recombination maps showed
291 especially elevated sampling noise, we took advantage of the bootstrap runs and masked all 100kb
292 regions where the difference in inferred recombination rates between bootstrap runs was greater
293 than one order of magnitude. Overall, the filtered recombination maps had masked a total of 264
294 Mb, or approximately 30.6% of the chromosomes, which is about 5% less than the was filtered out
295 by the callability mask during in genotype filtering.

296 **Hotspot analyses**

297 A recombination hotspot is a narrow region of unusually elevated recombination rate. When
298 searching for hotspots in our data, we required the local recombination rate estimate in any inter-
299 SNPs interval to be at least five times higher than a background rate. For the background
300 recombination rate, we applied three definitions the: (i) mean rate in 40kb around the interval (20kb
301 before and 20kb after), (ii) mean rate in 1Mb around the interval, and (iii) mean recombination rate
302 for the whole chromosome as in (Halldorsson et al. 2019). In most cases, several neighboring

303 intervals were identified as being a part of a hotspot and these intervals were merged using the
304 bedtools v2.29.2 `merge` command if the distance between such intervals was less than 1kb.
305 Proportion of overlap between hotspots from different maps were calculated using the `intersect`
306 command from bedtools v2.29.2 with default parameters, meaning that hotspots were considered
307 overlapping if they shared at least 1bp. When considering the mean recombination rates around
308 hotspots (**Supplementary Figure 5B**), we normalized the highest point in each hotspot to equal 1.0,
309 so that all hotspots were considered equal. Some hotspots were very long and contained implausibly
310 large fractions of recombination, a phenomenon also reported by other studies (Auton et al. 2012;
311 Hoge et al. 2023). We removed hotspots longer than 5kb from the above analyses and from the
312 search for sequence motifs described below.

313 **Measures of genetic differentiation**

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

323 **Distance from CpG islands and transcription start sites**

324 We used two different definitions of CpG islands (GpGi). First, consistent with (Baker et al. 2017), we
325 used the the `maskOutFa` and `cpg_lh` command from UCSC utils. This approach stipulates a
326 minimum of 50% GC content for defining a CpGi. This resulted in 17 000 CpG islands representing a

327 total of 8.5 Mb. Because GC content in the genome is lower in percomorpha that in vertebrates due
328 to less biased gene conversion toward GC (Escobar et al. 2011) and constraining the definition of
329 CpG island by GC content may not be the most appropriate here. We thus also used the `cpgplot`
330 `-minoe 0.6 -minpc 0` command from `EMBOSS` software (Rice et al. 2000), which resulted in
331 a much greater number (228 069) of CpGi. For the Transcription Starting Sites (TSS), we used the
332 genome annotation described before. We then used the `intersect -v` and `closest` command
333 from `bcftools v.2` to obtain the mean recombination rates in 2Kb and 10Kb associated with the
334 distance of the closest TSS non overlapping with a CpGi and vice versa. We finally used the `loess`
335 function in R to obtain the relative recombination rate depending on the distance with the closest
336 TSS or CpGi.

337 **Inference of haplotype blocks**

338 To discover large-scale variation shared by loci along the genome, we used the program `lostruct`
339 that visualizes the local effect of population structure (Li and Ralph 2018). `lostruct` summarize
340 the pattern of relatedness in a local PCA for nonoverlapping windows along the genome and
341 calculate the dissimilarities between each pair of local PCAs. It then uses multidimensional scaling
342 (MDS) to visualize relationships between windows. We ran `lostruct` for each chromosome
343 separately on 100 SNPs windows. We then plot the first and second axis of the MDS against the
344 genome position. We manually identified 47 regions with high MDS values and high difference in the
345 MDS scores (**Supplementary Figure 10**). To visualize population structure in `lostruct` outliers, we
346 used `smartPCA` (Patterson et al. 2006) on data filtered for minor allele frequency >=0.05 using plink
347 v1.9 (Purcell et al. 2007) with the `--maf 0.05` option but did not filter based on LD.

348 **Characterization of haplotype blocks**

349 To better understand the evolutionary history of each of the 47 haplotype blocks we calculated and
350 looked at the relationships of a range of population genetic statistics in these regions. First, to test if
351 the proportion of high-heterozygosity individuals was different between ecotypes, we calculated the
352 proportion of heterozygous sites for each individual ($H_{ind}$), used the `kmeans` function in R with `k=2`
353 to assign these values into two clusters, and then used a binomial distribution to quantify the
354 difference in the number of individuals from each ecotype in the cluster with higher $H_{ind}$. (**Figure 4C**;
355 **Supplementary Table 2**). The inbreeding coefficient $F$ was calculated for each SNP using our R code.
356 To estimate the time to the most recent common ancestor (TMRCA) we calculated pairwise $d_{xy}$ (i.e.
357 the number of differences in the nucleotide sequence) among all individuals. We took the maximum
358 value of pairwise $d_{xy}$ within each of the 47 local PCA outliers to approximate the TMRCA as
359 $\widehat{TMRCA} = \frac{Max(d_{xy})}{2 * \mu}$ (acknowledging that this can be a slight underestimate of the true TMRCA of
360 individual sequences because of using unphased data).

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

366 The highly differentiated regions (HDRs) were defined analogously to ref. (Malinsky et al. 2015). We
367 took the top 1% of the $F_{ST}$ values in 20 SNP windows (calculated as described above) and merged the
368 windows that were within 10kb of each other. This resulted in 352 HDRs which is very similar to the
369 344 HDRs found in (Malinsky et al. 2015). The permutation test used to assess significance of mean
370 $F_{ST}$ per each local PCA outlier was implemented using custom R script. Briefly, for each local PCA
371 outlier we sampled 1,000 random genomic windows of the same size, obtaining a null distribution

372 of $F_{ST}$ values, and considered the $F_{ST}$ significantly elevated when it fell within the 5% of this
373 distribution.

374 **PRDM9 ortholog research and distance from Zinc-Finger binding motif**

375 To find the PRDM9 ortholog in *A. calliptera*, we used the `blastp` command (Altschul et al. 1990)
376 using as query the PRDM9 protein sequence from the Atlantic salmon *Salmo salar* (Gene ID
377 100380788) against the NCBI RefSeq database (O'Leary et al. 2016) for the species *Astatotilapia*
378 *calliptera*. Using the best matching protein sequence (XP_026034002.1), we applied a Zinc-Finger
379 (ZF) motif predictor (Persikov et al. 2009; Persikov and Singh 2014) to obtain the position weight
380 matrix representing the binding site prediction based on 11 identified zinc finger domains. We then
381 used the `FIMO` command from the `MEME Suite` (Bailey et al. 2015) on the reference genome of
382 our species, to localize the binding sites. We then used the `intersect -v` and `closest`
383 command from `bcftools v.2` to obtain the mean recombination rates in 2Kb with the distance
384 of the closest ZF binding DNA motif.

385 **Contribution of genomic features to recombination divergence**

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

392 # Results

393 **Study system and demographic history**

394 We obtained whole genome short read sequences of 159 male individuals, each of which was
395 assigned to either the benthic or littoral ecotype based on a field photograph. The sequencing
396 coverage was $\sim$15x and, after variant calling and filtering, we obtained 3.9 million SNPs that were
397 used for all the following analyses. To check the validity of the field assignment we used the genetic
398 data to run a principal component analysis (PCA; **Figure 1B**) and reconstructed a Neighbor-Joining
399 tree (**Supplementary Figure 1**). While no individuals were misassigned, we identified 20 individuals
400 that appeared to be genetically intermediate which may be the result of recent hybridization.
401 Because our goal was to focus on the differences between the ecotypes, we removed the
402 intermediate individuals (gray in **Figure 1B**) from further analyses.

403 To obtain a more accurate understanding of the historical demographic context of the ecotype
404 divergence, we first used the program SMC++ (Terhorst et al. 2017) to infer the recent changes in
405 effective population size. Consistent with previous results (Malinsky et al. 2015), we find a bottleneck
406 which may be related to the lake colonization, followed by recent demographic expansions in both
407 ecotypes (**Figure 1C**). This approach also allowed us to re-estimate the split time between the two
408 ecotypes, which we now put at $\sim$2,500 generations ago (95% confidence interval: 1902 to 5469
409 generations). While this is considerably older than reported previously (Malinsky et al. 2015), this
410 difference is primarily a result of using the cichlid-specific mutation rate from (Malinsky et al. 2018)
411 in place of the human mutation rate used in the previous study. We also estimated the amount of
412 gene flow between the ecotypes using fastsimcoal2, with best migration rate estimates being
413 $11.5 \times 10^{-5}$ for littoral to benthic and $7.01 \times 10^{-5}$ for benthic to littoral, although higher migration
414 rates of 0.0339 for littoral to benthic and 0.0368 for benthic to littoral had almost equivalent

415 likelihoods and appear more realistic given the number of intermediate individuals we found

416 (**Supplementary Figure 2**; Materials and Methods).

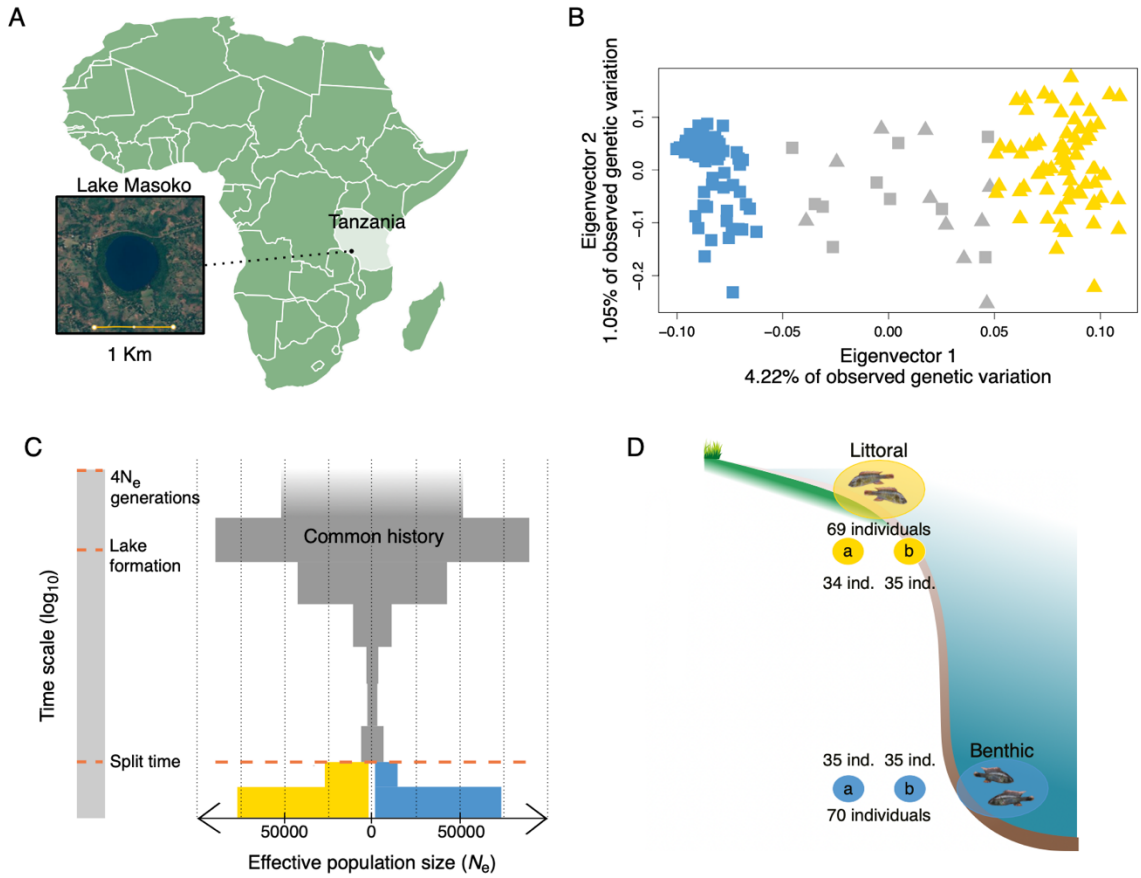



**Figure 1: Study system and demographic history.** (**A**) Lake Masoko is a circular small (~670m diameter) maar-type volcanic crater lake located in the East African rift valley in southern Tanzania. (**B**) A principal component analysis based on 784974 SNPs. Some individuals labelled in the field 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**) 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.

417 **Population recombination landscapes differ between ecotypes**

418 First, to quantify the effects of sampling variability and methodological limitations, we divided the

419 individuals from each ecotype into two independent subsets and reconstructed a separate

420 recombination map for each subset (**Figure 1D**). Thus, we obtained a total of two replicate maps for

421 each ecotype. Spearman correlation between the replicate maps from the same ecotype (within-

422 ecotype) was 0.77 for within-littoral and 0.71 for within-benthic comparisons at 2kb scale (**Figure**

423 **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 between ecotypes (in gray in **Figure 2A**) and found that the correlation coefficients were considerably lower than within ecotypes (**Figure 2A**; 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 Materials and Methods). The key result, *i.e.* that between-ecotype correlations are consistently and considerably lower than for within-ecotype replicates, holds across genomic scales from 2kb to 5Mb (**Supplementary Figure 3**). It is also highly consistent across bootstrap replicates obtained by splitting the ecotypes into different subsamples of individuals (**Supplementary Figure 3**).

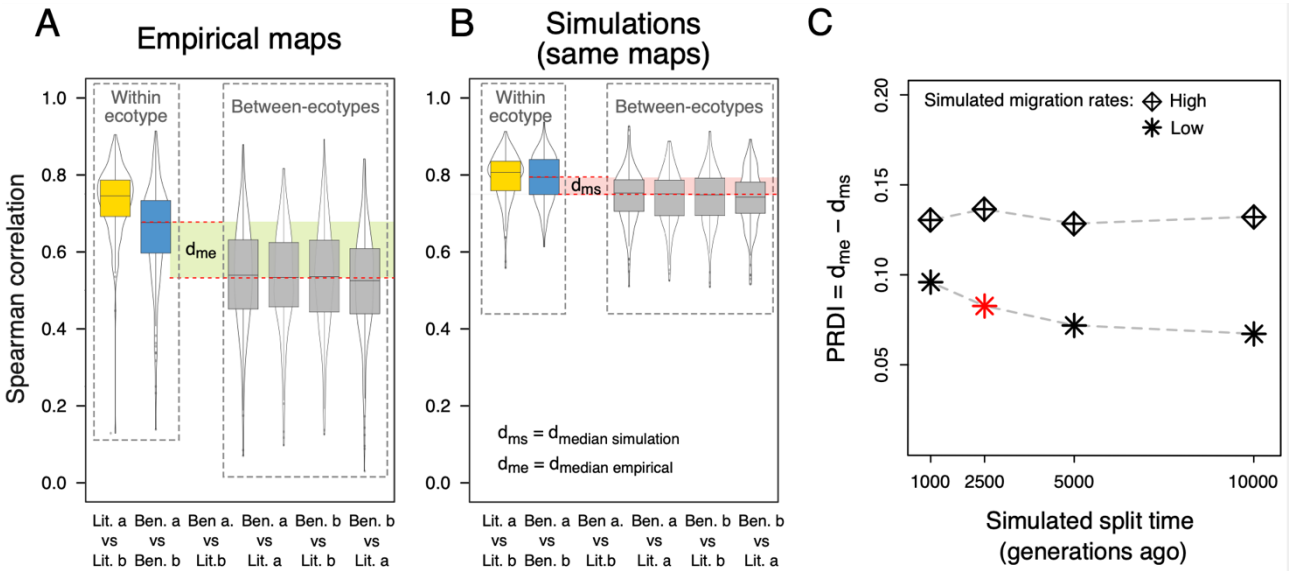



**Figure 2: Rapid evolution of the recombination landscapes between the ecotypes.** (**A**) Spearman correlation between recombination maps on 2kb scale with each datapoint representing a 5Mb genomic interval. The correlations in between-ecotype comparisons are substantially lower than in the within-ecotype replicate comparisons, with the difference measured by $d_{me}$ (Material and Methods). Blue and yellow colors refer to the benthic and littoral ecotypes respectively. (**B**) Correlations among recombination maps inferred from population genetic data simulated under our best-estimate parameters, 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).

To account for the separation of recent benthic vs. littoral ancestry, which leads to non-random 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, $N_e$,

437 and gene flow) that we inferred from empirical data as described above. Based on data simulated
438 under these best-estimate parameters, the within-ecotype vs. between ecotype difference in
439 median recombination map correlations, which we denote $d_{ms}$ was 0.046, was approximately three
440 times lower than what we observed in empirical data (empirical $d_{me}$ = 0.140 vs. simulation $d_{ms}$ =
441 0.046; **Figure 2B**). We consider this difference ($d_{me}$ - $d_{ms}$) to be a meaningful measure of
442 recombination landscape evolution and refer to it as the Population Recombination Divergence
443 Index (PRDI) (Materials and Methods; **Figure 2C**).

444 Because the PRDI estimate depends on the parameters used in the simulations, and because there
445 are uncertainties regarding the correct values for these parameters, we conducted additional
446 simulations across a broad range of split times and using both the higher and lower migration rate
447 estimates from fastsimcoal2. We found that with the higher migration rates, PRDI estimates were
448 consistently high, around 0.14 across the full range of split times, because $d_{ms}$ hovered around zero
449 reflecting the intertwined ancestries of the ecotypes in the presence of high migration (**Figure 2C**).
450 With the low migration rates, PRDI was lower and depended on the simulated split time, reaching a
451 minimum of 0.07 when we simulated ecotype divergence 10,000 generations ago (**Figure 2C**).
452 Overall, these results provide evidence that the population recombination landscapes differ
453 between the ecotypes, with the magnitude of divergence estimated by PRDI to be between 0.07 and
454 0.14.

455 To further confirm that the observed recombination differences between benthic and littoral were
456 not driven by technical artifacts in regions of the genome where inference is especially error prone,
457 we applied a stringent filtering mask. This mask was based on the: (i) location of contig joins in the
458 assembly, (ii) ability to confidently call SNPs across genomic windows, and (iii) consistency of inferred
459 genetic maps across bootstrap replicates (Material and Methods). In total, we masked 30.6% of the
460 sequence, eliminating a substantial proportion of noise from the inferred recombination maps

461 across all genomic scales from 2kb to 5Mb (**Supplementary Figure 3**). We repeated all the analyses
462 above using these filtered maps and found that, despite the stricness of the filtering, the results, and
463 specifically the $d_{me}$ and PRDI estimates were virtually the same as for the raw maps (**Supplementary**
464 **Figures 3, 4A**). We also verified that the $d_{ms}$ estimates and PRDI do not depend on the reference map
465 that is used as input for the coalescent simulations (**Supplementary Figure 4**).

466 In Lake Masoko *A. calliptera*, the ancestry for the sampled individuals extends substantially beyond
467 the ecotype split time (**Figure 1C**), which means that the inferred recombination maps for each
468 ecotype can be interpreted as a mixture of two time periods: (i) recombination events that happened
469 in the common history of the ecotypes and (ii) events that happened after their split. To estimate
470 the contribution of each of these epochs, we used coalescent simulations and found that, across ten
471 simulations with the best-estimate split time of ~2,500 generations ago, an average 34.9% of
472 recombination events that changed the genealogy of the sample occured after the split (min =
473 34.4%; max = 35.7%; Materials and Methods). Thsese are the recombination events that make up
474 the differences between the recombination landscapes between the ecotypes.

475 Recombination landscapes are highly heterogenous, and a large proportion of events tends to occur
476 in so-called 'hotspots' (Coop and Przeworski 2007; Peñalba and Wolf 2020). We quantified the
477 heterogeneity of recombination along the genome in the *A. calliptera* of Lake Masoko and found
478 that 50% of all events were concentrated in less than 9.6% of the genome (**Supplementary Figure**
479 **5A**; Between 8.9% and 10.4% depending on the ecotype and the subsample). Therefore, the
480 concentration of recombination in hotspots can be considered intermediate - substantially lower
481 than for example in humans, but higher than for example in the plant Arabidopsis thaliana
482 (**Supplementary Figure 5A**). Using a definition of hotspots as having at least 5x higher recombination
483 rate than the 500kb of surrounding sequence, we found on average 2322 hotspots in each
484 recombination map (between 2275 and 2345; **Supplementary Table 1**). Only 41.5 % of hotspots

485 were shared when comparing 'a' and 'b' replicate maps within the same ecotype, showing that
486 hotspot detection is particularly sensitive to sampling variance. Nevertheless, we again observed the
487 same pattern as for correlations – the comparisions between benthic and littoral maps showed even
488 less hotspot overlap than expected based on simulations (**Supplementary Figure 5**). Qualitatively
489 similar results were obtained using different hotspot definitions (**Supplementary Figure 5**).

490 **Characterizing recombination rate evolution**

491 The changes in recombination landscapes measured by LD-based maps are based on effective
492 recombination – the recombination events present in the ancestry of the sampled individuals – and
493 are directly influenced by divergent selection between the ecotypes. Therefore, we next investigated
494 the relationship between recombination rates and genetic divergence between the ecotypes in
495 terms of allele frequencies ($F_{ST}$) and levels of nucleotide diversity ($\Delta(\pi)$) along the genome. As
496 expected, we found that PRDI generally increased with $F_{ST}$ and $\Delta(\pi)$, consitent with an effect of
497 divergent selection on genotypes (**Figure 3A**). However, we observed evolution of recombination
498 landscapes (positive PRDI) also at low levels of genetic divergence and were not limited to regions
499 of particularly high $F_{ST}$ or $\Delta(\pi)$ (**Figure 3A**). Therefore, the evolution of recombination landscapes is
500 not driven by selection on genotypes (e.g. selective sweeps), and/or the effect that differences in $\pi$
501 can have on the accuracy of recombination inference (**Supplementary Figure 6;** Raynaud et al, 2023).
502 Moreover, equivalent results were obtained for filtered maps, providing additional confidence
503 regarding the robustness of these conclusions (**Supplementary Figure 7**).

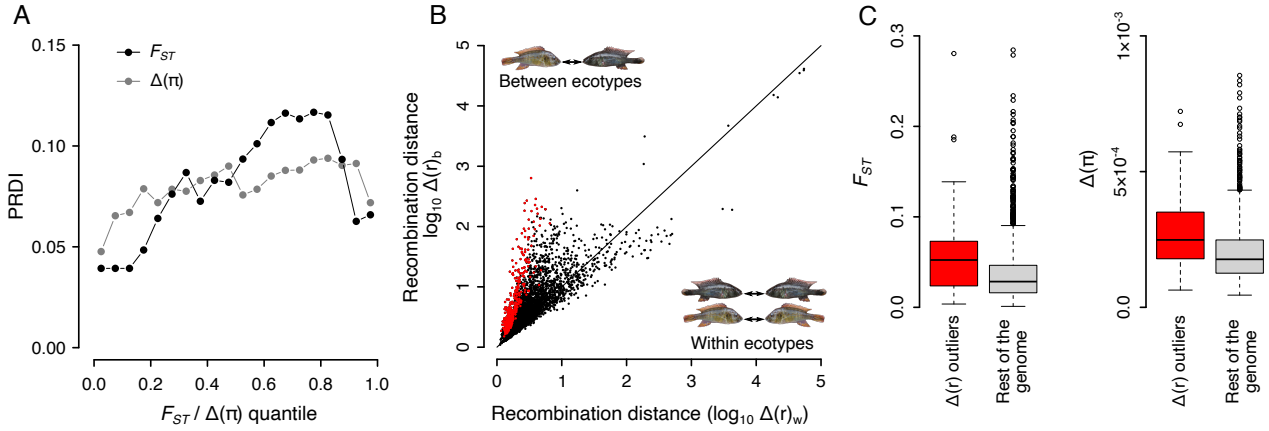



**Figure 3: Interplay between recombination and genomic differentiation.** (**A**) The Population Recombination Divergence Index (PRDI)_as a function of benthic-littoral $F_{ST}$ and of $\Delta(\pi)$. (**B**) A scatterplot of average recombination map distances in 100kb windows, within biological replicates - $\Delta(r)_w$ 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.

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 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 42.7 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 Figure 8A**). 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 8B**), when looking directly at $\Delta(r)$ outliers, we found that both $F_{ST}$ and $\Delta(\pi)$

519 were significantly elevated (Mann–Whitney U test: $p = 1.88 \times 10^{-6}$ for $F_{ST}$ and $p = 5.5 \times 10^{-7}$ for
520 $\Delta(\pi)$), clearly confirming that there is an association between allele frequency divergence and the
521 most rapidly evolving population recombination landscapes (**Figure 3C**).

522 Given the rapid evolution of recombination rates across the genome, we wanted to verify whether
523 the PRDM9 mechanism may be active in Lake Masoko *A. calliptera*. As in several other percomorph
524 species (Cavassim et al. 2022), we found one incomplete PRDM9 ortholog missing the KRAB and
525 SSXRD domains that appear to be necessary for PRDM9 to direct recombination (Baker et al. 2017).
526 Consistent with this, recombination rates were elevated at and near CpG islands (~1.2x higher;
527 **Supplementary Figure 9**) and transcription start sites (TSS; ~1.3x higher; **Supplementary Figure 9**),
528 a pattern that is similar to that reported previously for swordtail fish (Baker et al. 2017). Because the
529 PRDM9 zinc finger array in *A. calliptera* was intact, we predicted its binding sites across the genome
530 and found no increase in recombination rates at or near the binding sites (**Supplementary Figure 9**).
531 Overall, these results confirm that PRDM9 does not direct recombination in *A. calliptera* and,
532 therefore, cannot contribute to the rapid evolution of recombination rates in this system.

533 **Large haplotype blocks contribute to evolution of recombination rates**

534 Ecotype divergence and speciation in the face of gene-flow are often facilitated by regions of
535 suppressed recombination, which allow a buildup of linkage between multiple loci under divergent
536 selection (Faria et al. 2019b). Non-recombining haplotype blocks can be revealed as extended
537 regions of the genome with distinct population structure, substantially different from the genome-
538 wide average (Ma and Amos 2018; Mérot 2020; Todesco et al. 2020). To look for such regions, we
539 used a local PCA approach (Li and Ralph 2018) and identified a total of 47 outliers (**Supplementary**
540 **Figures 10, 11**), ranging in size between 550kb and 25.7Mb (mean 3.9Mb) and covering in total
541 21.4% of the genome. Importantly, these regions contain 31.6% of $\Delta(r)$ outliers, and thus contribute
542 disproportionately to the observed differences in recombination rates between the ecotypes (**Table**

543 **1**). Consistent with this, we found that the signal of recombination rate evolution measured by PRDI
544 is about 10% stronger in the local PCA outliers than in the rest of the genome (**Supplementary Figure**
545 **12**). At the same time, it should be emphasized that there is also a strong PRDI signal of
546 recombination rate evolution outside of these blocks (**Supplementary Figure 12**).

547 **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 |

548

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

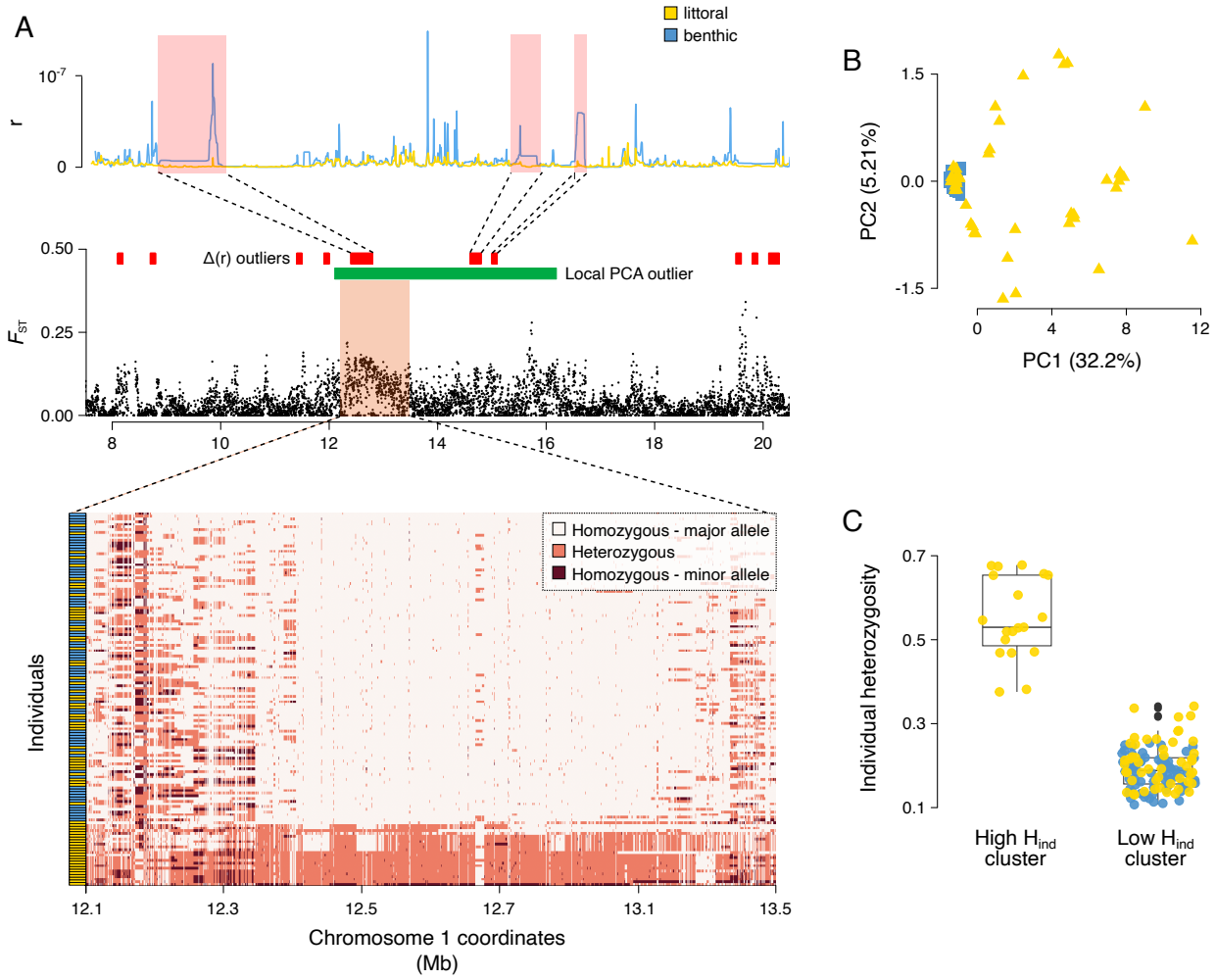



**Figure 4: A local PCA outlier associated with difference in recombination.** (**A**) An example region of chromosome 1 (LS420019.2) illustrating the link between a large haplotype block, $F_{ST,}$ and recombination rate evolution. Top: average recombination rates in the two ecotypes. There is lower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

As predicted, across the 47 haplotype blocks we found 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 12**). In contrast, the association of $\Delta(r)$ with allele frequency divergence (mean $F_{ST}$ per haplotype block) with was less pronounced and not statistically significant (**Figure 5B**).

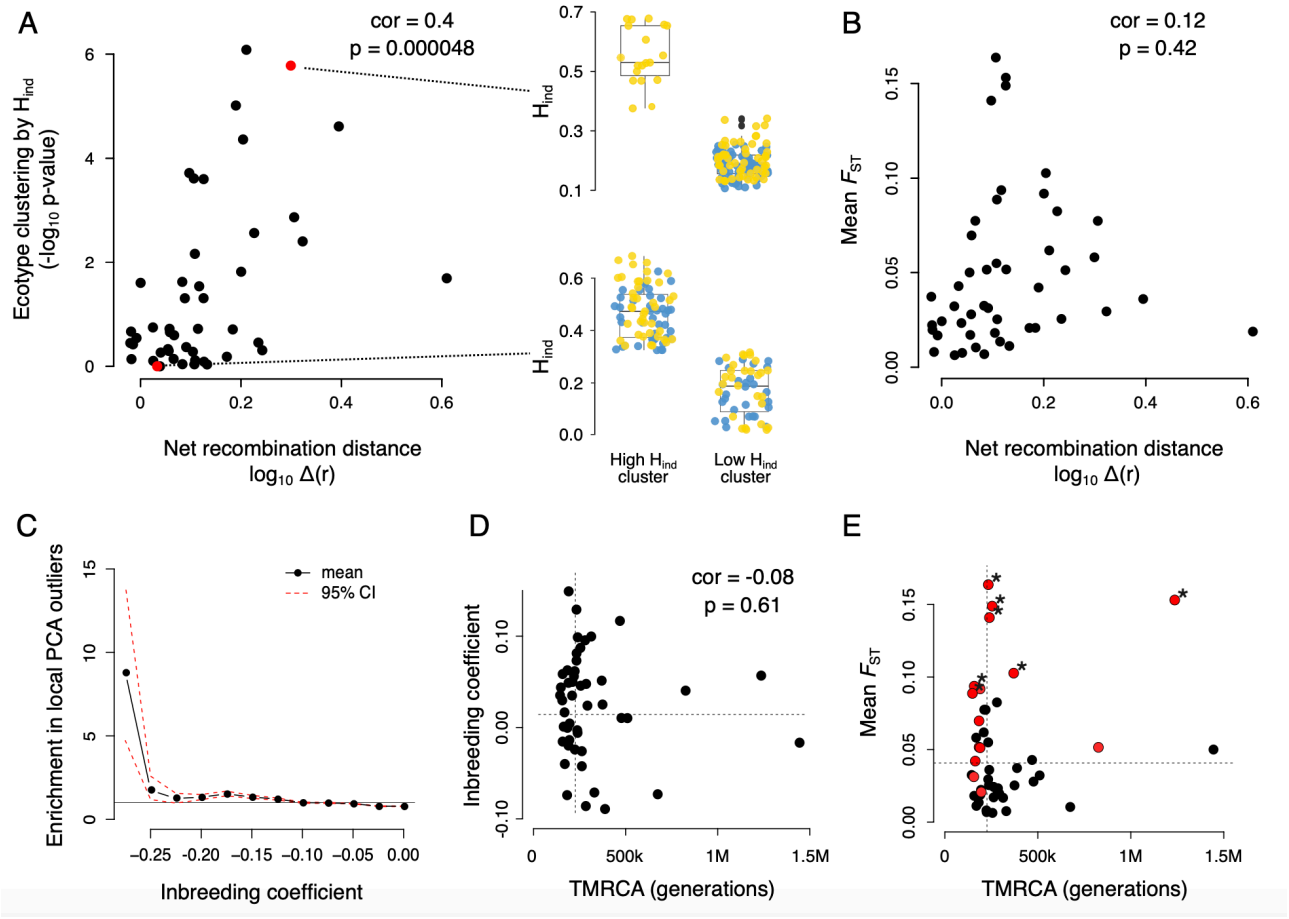



**Figure 5: Characterizing haplotype blocks and their link with recombination in Lake Masoko.** (A) Relationship, across the 47 local PCA outliers, between the degree of ecotype clustering by $H_{ind}$ and net between-ecotype recombination map distances. We highlighted examples of a strong (top) and 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 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).

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 13, 14**).

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

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

## 591 Discussion

592 The landscape of recombination across the genome is not static but evolves through time. In this
593 study, we undertook a holistic investigation of recombination rate evolution between two ecotypes

594 that diverged very recently, in sympatry with gene flow, and adapted rapidly across multiple
595 phenotypic traits to new lake environments (Malinsky et al. 2015). Given ecotype divergence at
596 many – at least a hundred or so – genomic loci (Malinsky et al. 2015), we could expect that
597 recombination, or the lack thereof, would be important in bringing together and keeping together
598 the alleles that are beneficial in each environment (Ortiz-Barrientos et al. 2016; Todesco et al. 2020;
599 Schluter and Rieseberg 2022; Battlay et al. 2023). Our findings reveal, characterize, and quantify
600 substantial differences in population recombination rates between the ecotypes, complementing
601 previous studies of divergence in ecology, mate choice, and allele frequencies (Malinsky et al. 2015),
602 methylation (Vernaz et al. 2022), gene expression (Carruthers et al. 2022), and in sex determination
603 (Munby et al. 2021).

604 The use of LD-based maps, integrating over a large number of recombination events throughout the
605 ancestry of the ecotypes, allowed us to analyze recombination landscape evolution at a fine scale.
606 Despite the challenges inherent in such comparisons, we used subsampling (replicate maps from the
607 same ecotype) and extensive simulations to derive meaningful measures of evolution in population
608 landscapes. We show how the recombination differences are distributed across the genome, how
609 they are linked to allele frequency divergence ($F_{ST;}\Delta(\pi)$) and to large haplotype blocks identified by
610 local PCA.

611 As we expected, we found a link between recombination rate divergence and regions of high allele
612 frequency divergence. This link can be expected for a variety of reasons – for example, local
613 recombination rates can impact inference of $F_{ST}$ (Booker et al. 2020), they partly determine the
614 extent of background selection which can affect $F_{ST}$ (Burri et al. 2015; Matthey-Doret and Whitlock
615 2019), and they play a major role in in the formation of genomic islands of divergence during
616 speciation with gene-flow (Feder et al. 2012; Cruickshank and Hahn 2014). Moreover, selection on
617 genotypes (e.g., incomplete selective sweeps) is directly reflected in LD-based recombination maps

618 (in contrast to gamete-based or pedigree-based maps) (Coop and Przeworski 2007; Peñalba and Wolf
619 2020). Given these considerations, it is surprising that allele frequency divergence does not appear
620 to be the main driver of the observed recombination rate differences. For example, only 13.7% of
621 significant differences in recombination, i.e. $\Delta(r)$ outliers, are co-located with the top 10% of $F_{ST}$
622 (**Table 1**), and PRDI is substantially positive in regions where $F_{ST}$ and $\Delta(\pi)$ are zero.

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

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

642 analyses of structural variants and their roles in evolution of recombination landscapes (Coster et al.
643 2021).

644 We found substantial recombination landscape evolution where the ecotypes cluster by individual
645 heterozygosity levels (**Figure 5A**, **Supplementary Figure 12**). This could be because inversions
646 prevent crossover formation only in the gametes of heterozygous individuals (Faria et al. 2019b). A
647 recent study using pedigree-based recombination maps raises an additional possibility. Venu et al.
648 (2024) found that regions with higher heterozygosity have lower recombination rates due to
649 haplotype incompatibilities between diverging ecotypes. Therefore, the same mechanism could play
650 a role in the divergence of recombination landscapes seen in Lake Masoko.

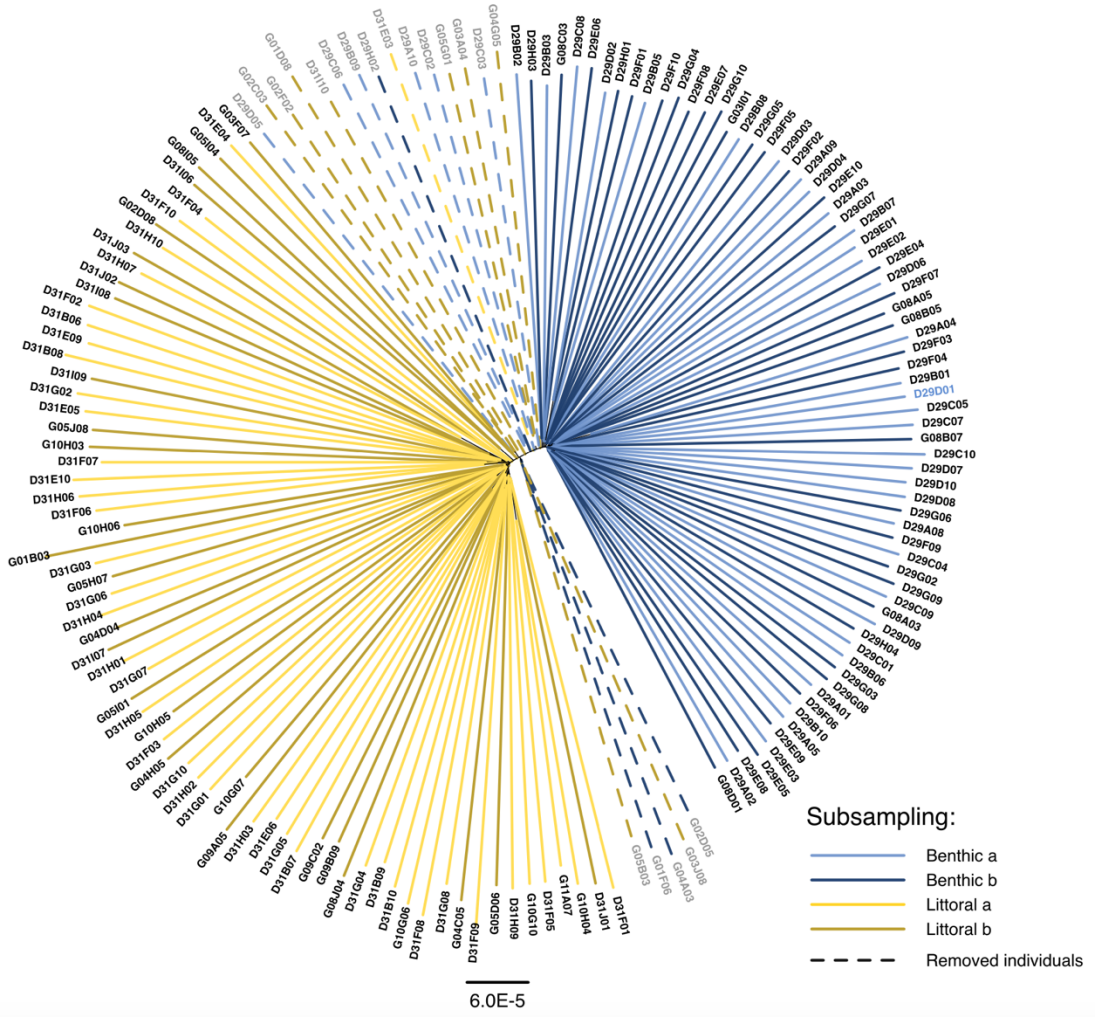
651 More general and important open questions concern the nature of selection on recombination.
652 Which of the observed differences in population recombination rates are a result of changes in the
653 distribution of crossovers during gamete formation? And which changes are an indirect effect of
654 subsequent selection for or against specific genetic variants and recombinant haplotypes? These
655 questions are not possible to answer conclusively with LD-based estimates alone. A future
656 comparison of our LD-based maps against recombination landscapes obtained by sequencing of
657 gametes and / or individuals related by pedigrees will shed further light on this question (Peñalba
658 and Wolf 2020).

659 Our finding of rapid recombination rate evolution, while consistent with some previous studies
660 (Shanfelter et al. 2019; Déserts et al. 2021), seems to be in conflict with the current paradigm of
661 evolutionary stability of recombination landscapes in species lacking the PRDM9 mechanism (Lam
662 and Keeney 2015; Singhal et al. 2015). However, this is not as surprising as it may appear because it
663 is common that different tempos of molecular evolution are observed between micro- and macro-
664 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).

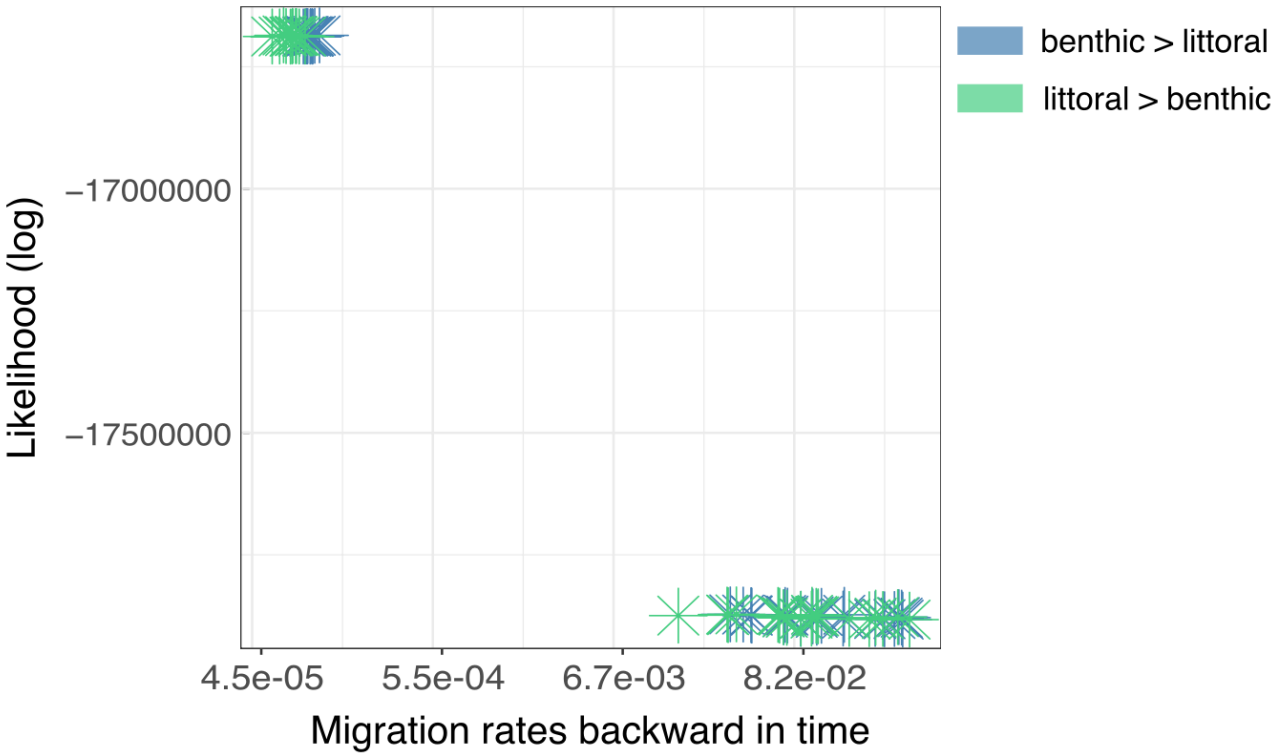
Adaptation and organismal diversification are increasingly seen as multidimensional and combinatorial, typically with involvement of multiple polygenic traits and epistasis (Marques et al. 2019; Barton 2022; Yeaman 2022), and the relative genetic distances between the loci involved constitute key parameters. 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. However, this is still typically done at a rough Megabase-scale resolution (Haenel et al. 2018; Brazier and Glémin 2022). 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 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 Materials

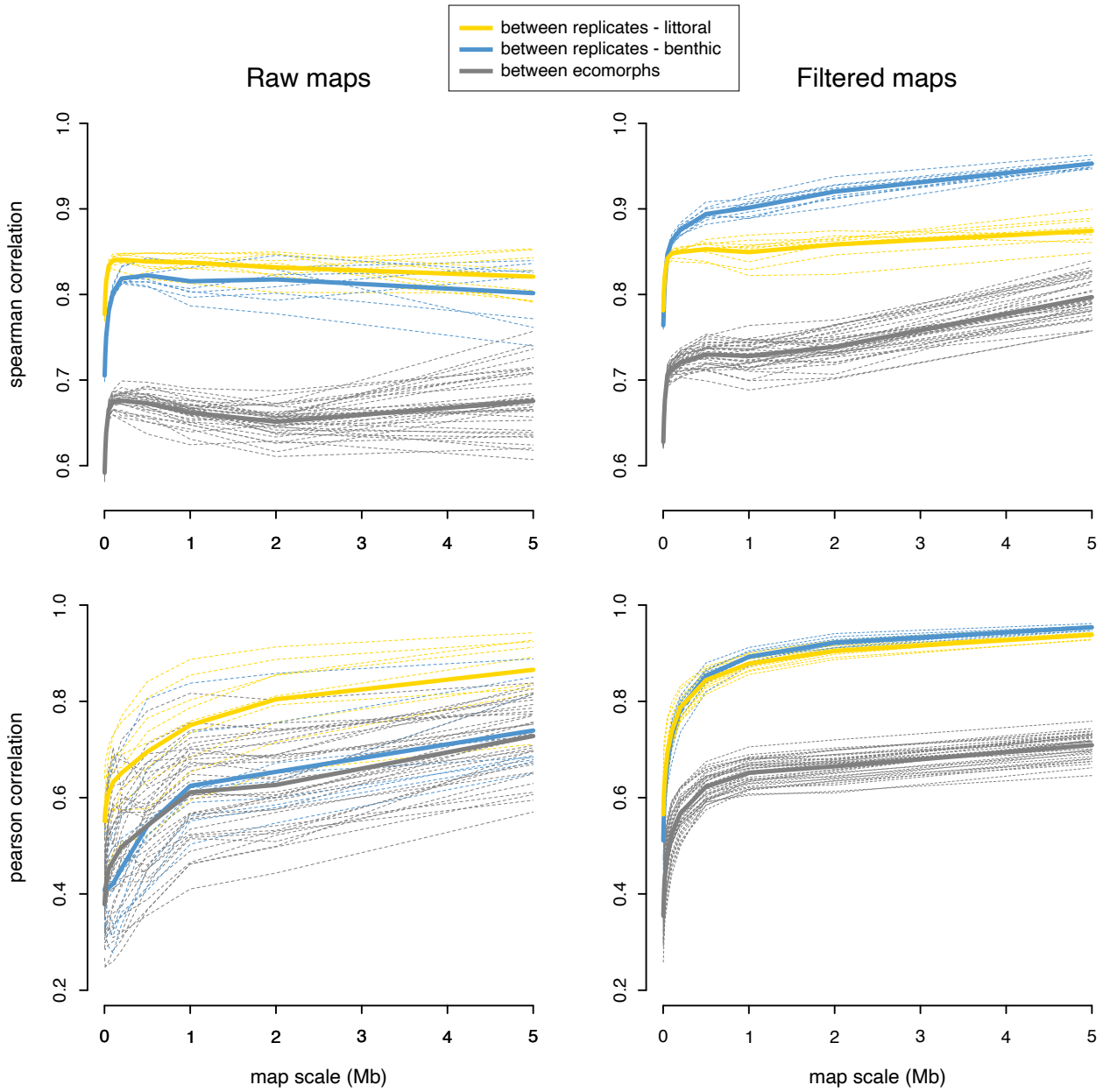



**Supplementary Figure 1: Neighbor joining tree based on the number of differences in polymorphic sites between individuals.** Each tip represents one individual. We used this tree and the PCA showed in Figure 1 to determine which intermediate individuals to discard from downstream analyses. After discarding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

691

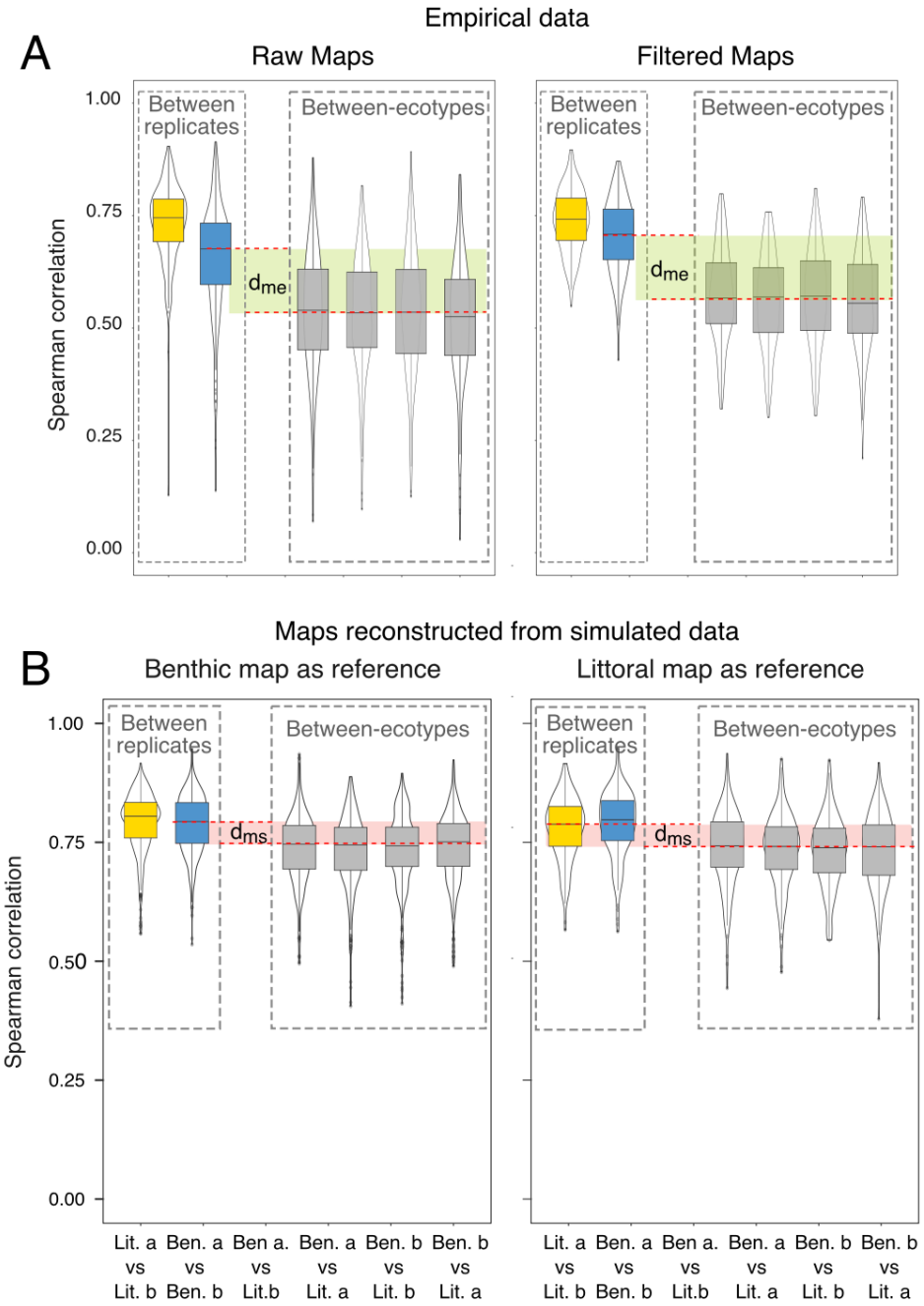



**Supplementary Figure 2: Estimating migration rates with fastsimcoal2.** Each point represents an estimate from an independent run with different starting parameters. The estimates with lower migration rates (top left corner) have superior likelihoods to the estimates in the bottom right. In the main 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.

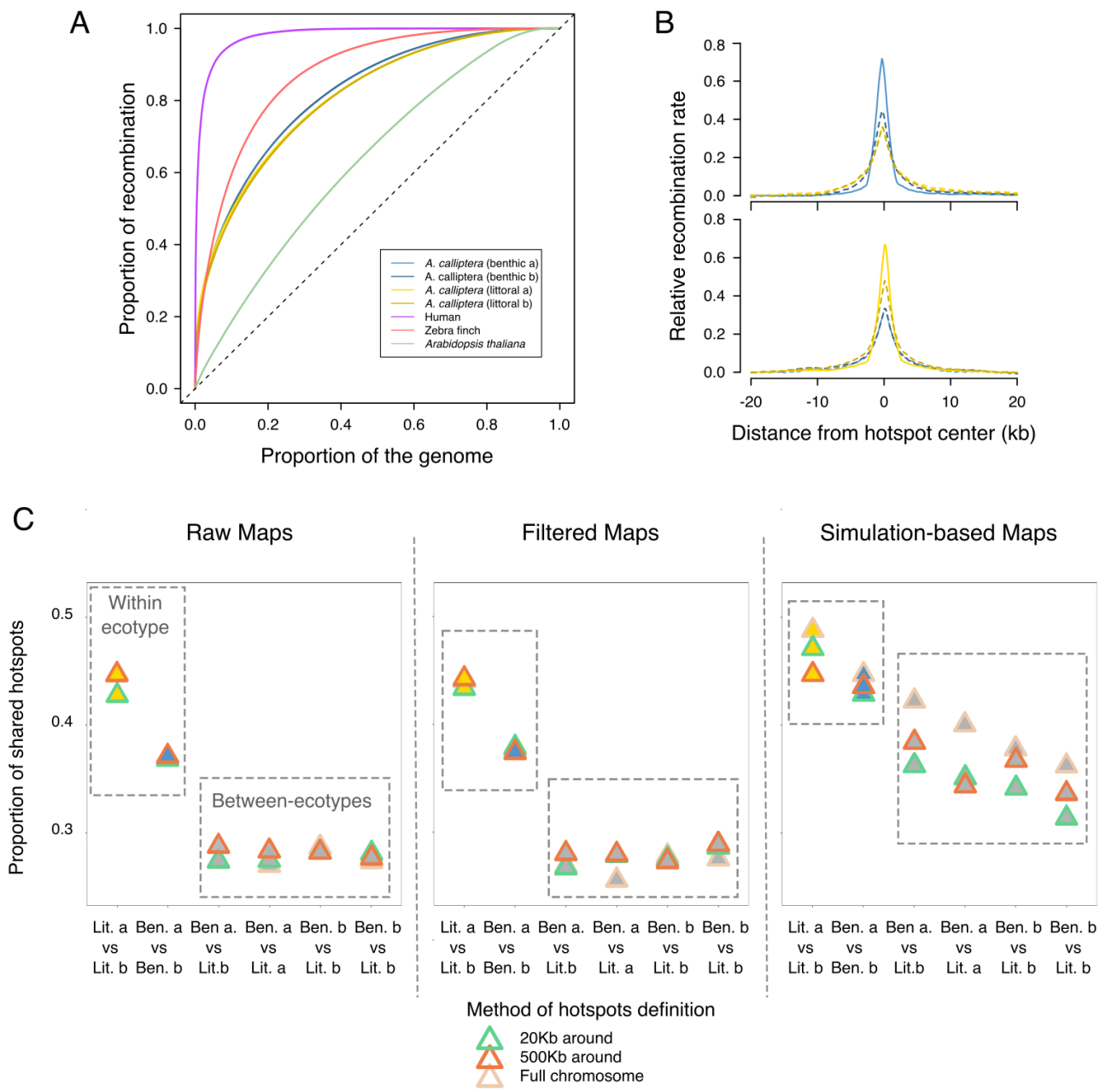



**Supplementary Figure 3: Correlation between recombination maps at different scales.** We see that correlations in the between ecotype comparisons are considerably lower than correlations for the within-ecotype replicates. This holds across all genomic scales (2kb to 5Mb) for Spearman correlation, across almost all genomic scales for Pearson correlation, and across bootstrap replicates of genetic maps obtained by randomly sub-sampling individuals for each ecotype (dashed lines). 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 even clearer.

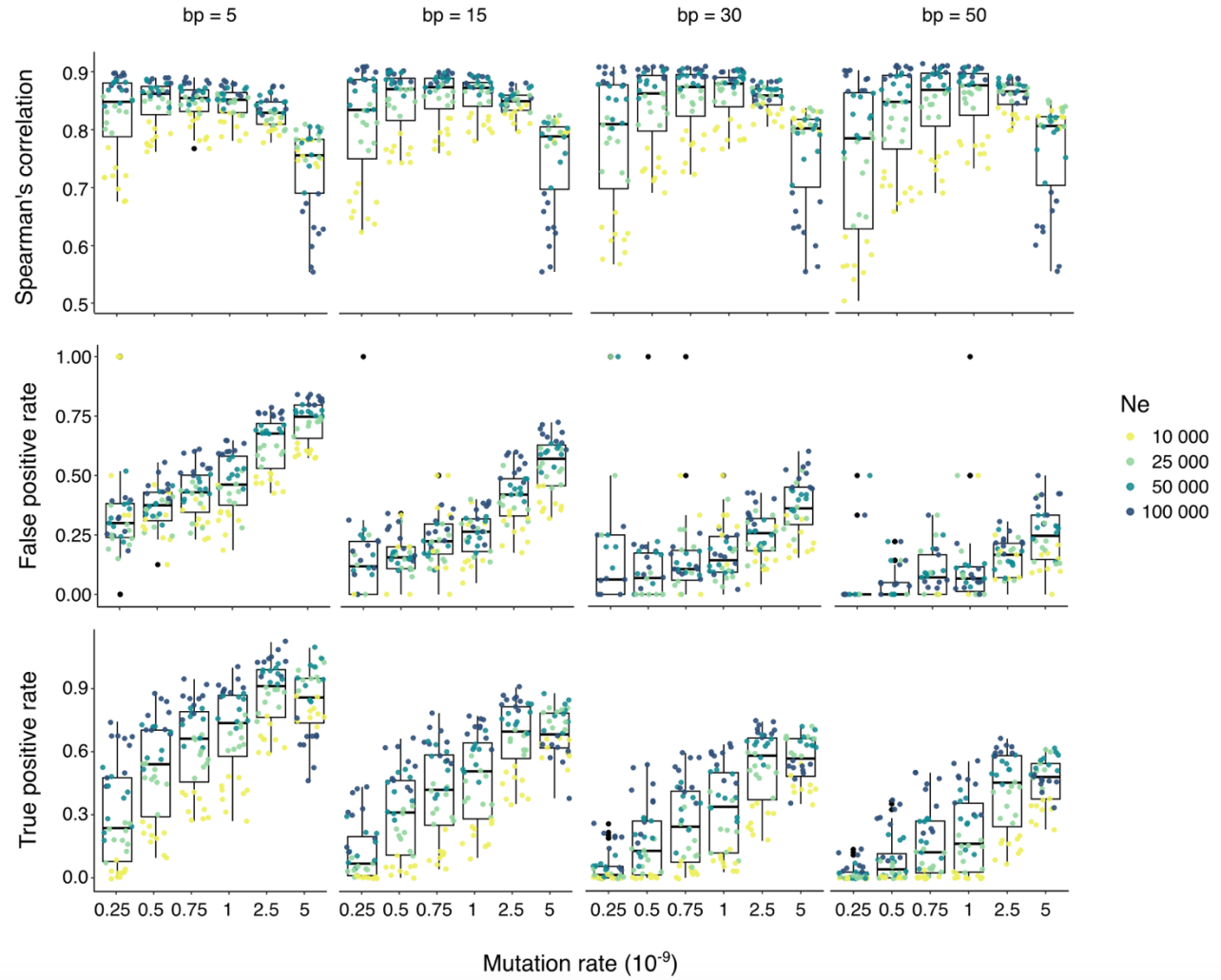
692



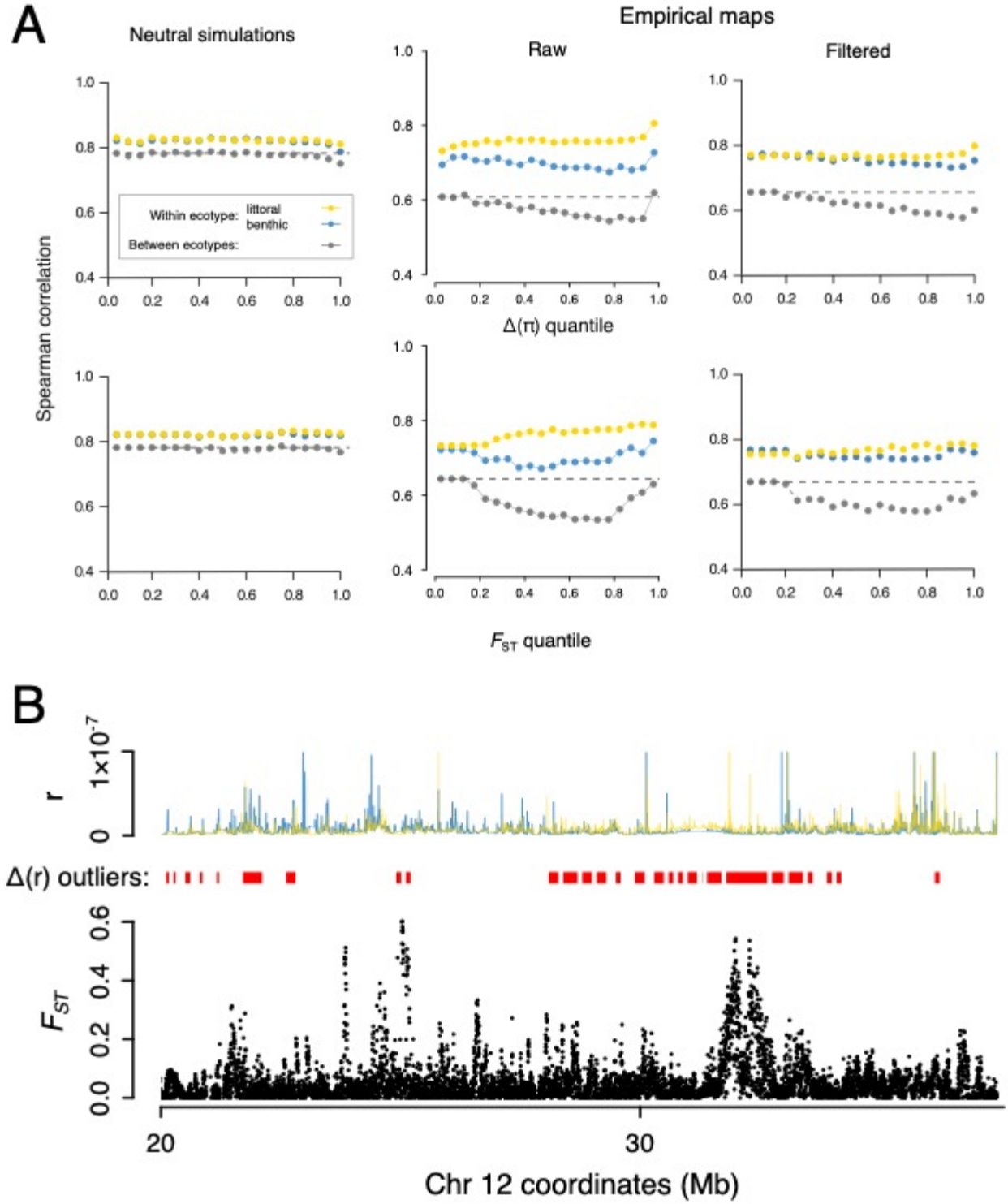

**Supplementary Figure 4: Confirming the robustness of the PRDI estimate of recombination landscape evolution. (A)** A comparison of recombination map correlations derived from the raw maps (equivalent to Figure 2A) and maps that have been carefully filtered to exclude regions where recombination inference might have been error prone (Filtered Maps; see main text for details). Importantly, the $d_{me}$ estimate, which is the first term in the PRDI calculation, is virtually identical. **(B)** A comparison of recombination map correlations for maps derived from simulated data. Regardless of which empirical map was used as an input/reference for the simulations, we got virtually identical estimates of $d_{ms}$, the second term in the PRDI calculations.




**Supplementary Figure 5: Characterizing recombination hotspots and their evolution in Lake Masoko.** Blue and yellow colors refer to the benthic and littoral ecotypes respectively. **(A)** Empirical cumulative distribution function of for recombination fraction as a function of proportion of genome covered. **(B)** Scaled recombination rates around hotspots defined in the benthic a (top) and littoral a (bottom) subsamples (solid lines) and the scaled rates at these position in the other subsamples (dashed lines). Hotspots in (B) are defined against a 1Mb background. The light color (blue and yellow) corresponds to the subsamples a (in benthic and littoral) and the dark color to the subsamples b. **(C)** The proportion of overlapping hotspots between recombination maps, based on three different background sizes - these three methods differ on the size of the genomic background used as baseline (e.g. 20kb, 500kb and full chromosome). We observe very similar proportion of overlapping hotspots with different the different

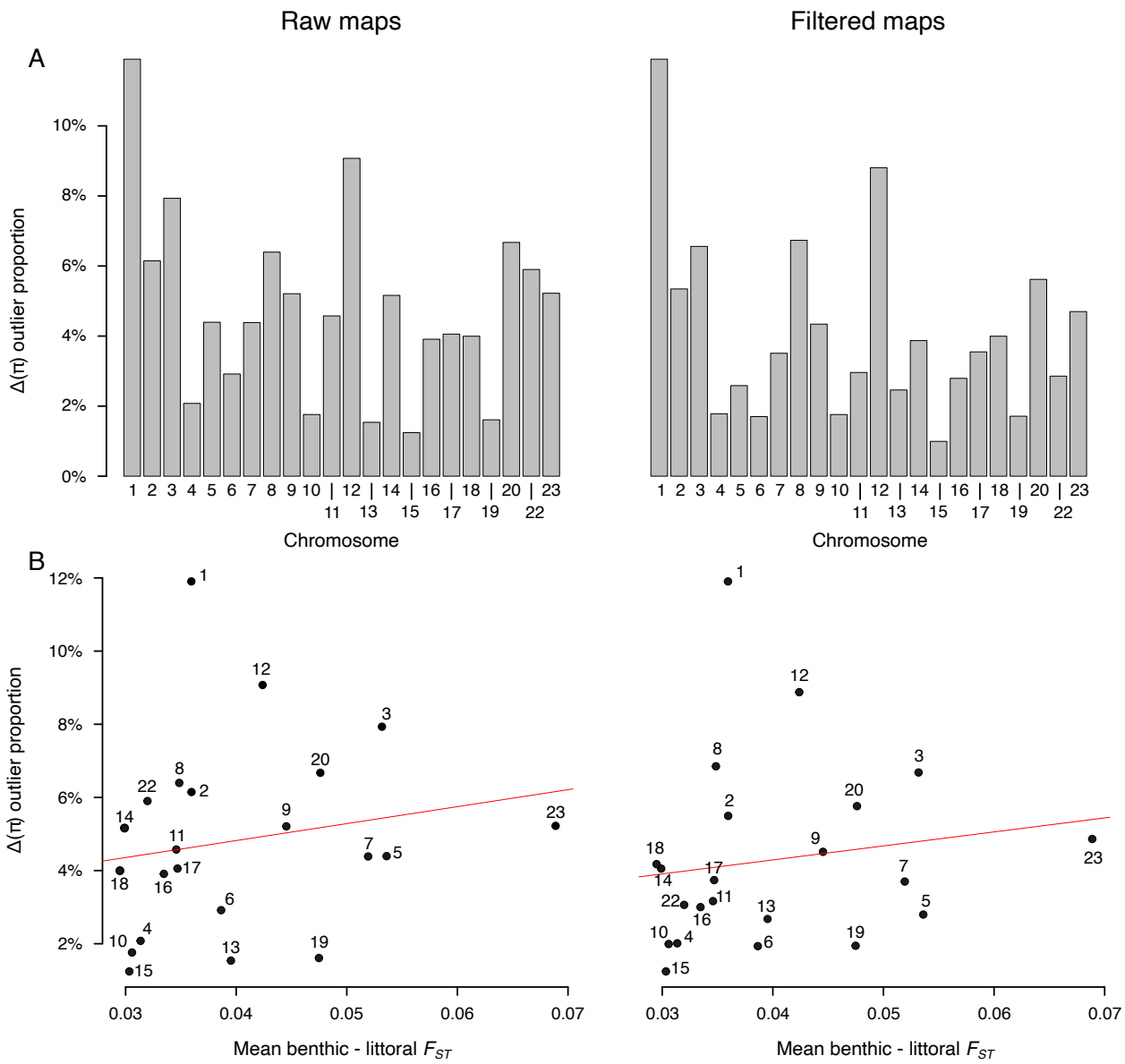



**Supplementary Figure 6: Inference of recombination maps and hotspots from coalescent simulations.** We used `msprime` to simulate ancestral history with mutations based on different mutation rates ($\mu$) and effective population sizes ($N_e$). We inferred for each VCF the recombination landscape with *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 $\pi$ is higher.



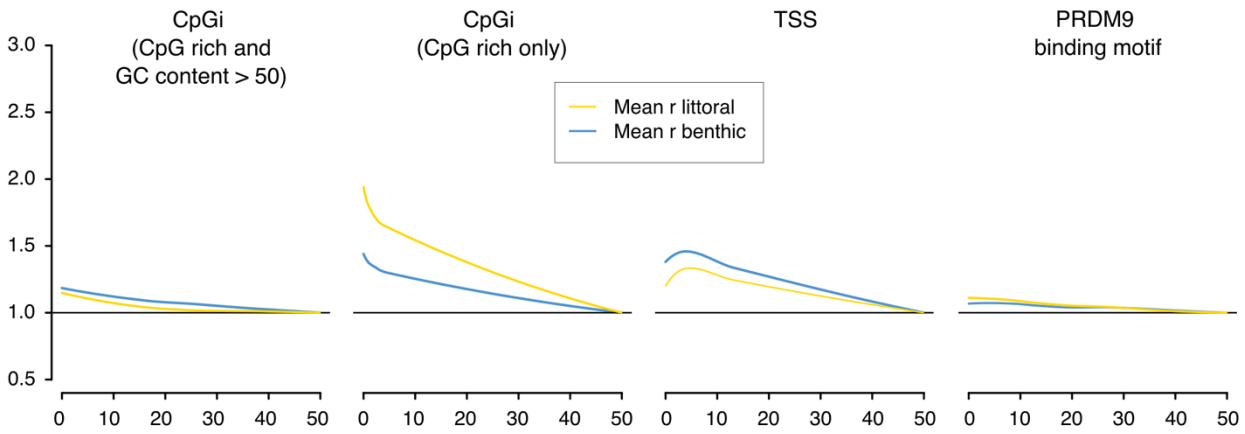

**Supplementary Figure 7: The relationship between allele frequency divergence and recombination divergence.** (**A**) Spearman correlation measured between recombination maps as a function of benthic-littoral $F_{ST}$ and of Δ(π). There is no relationship in neutral coalescent simulations. Empirical maps, both raw and filtered, show a consistent result of always lower-than-expected correlations between ecotypes and a consistent trend of lower benthic vs. littoral correlations in regions of greater $F_{ST}$ and of Δ(π). (**B**) An illustrative example. While some Δ(r) outliers – regions of significantly divergent recombination – co-locate 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.

701

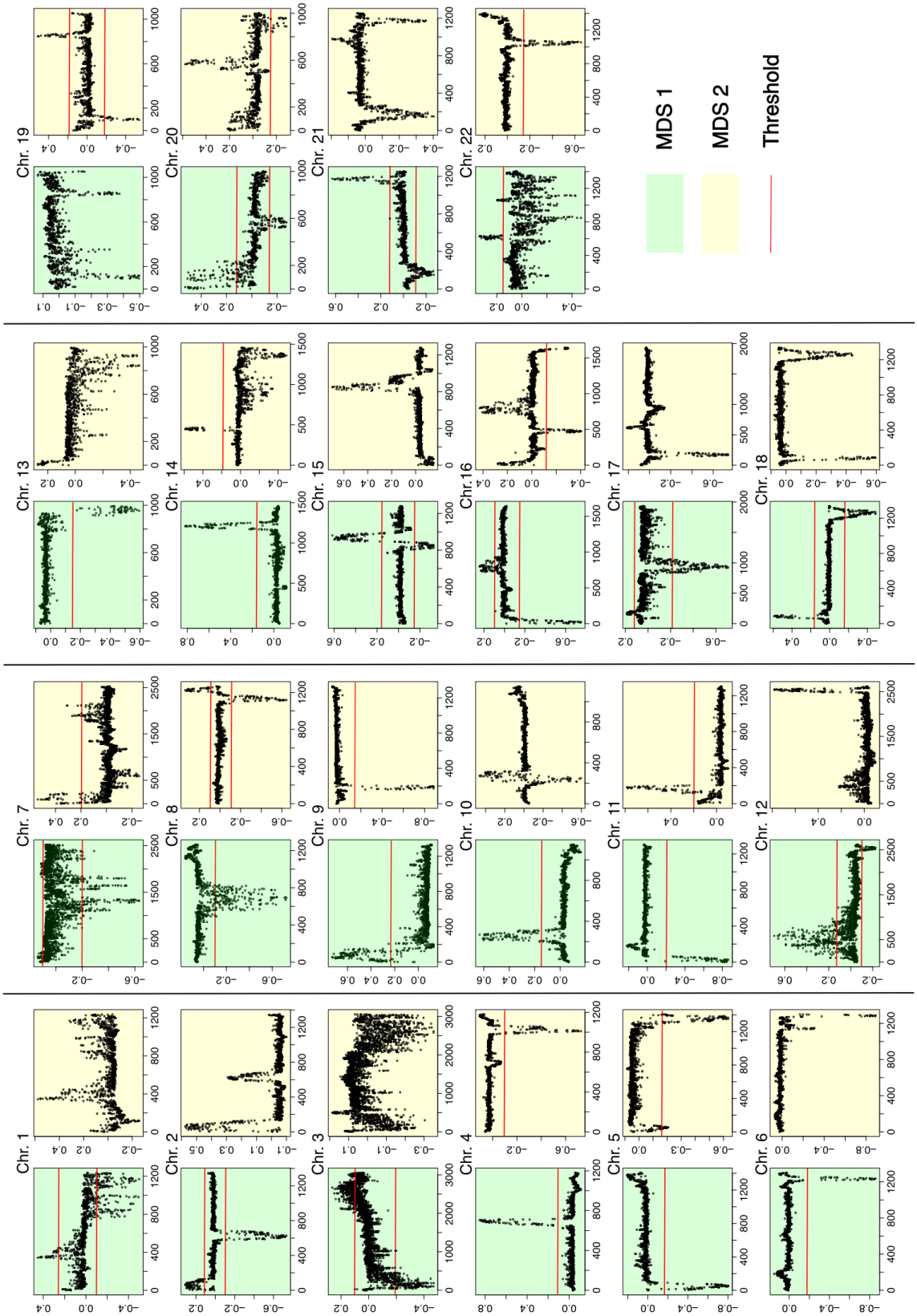



**Supplementary Figure 8: The distribution of $\Delta(r)$ outliers across chromosomes.** (**A**) The proportion of each chromosome assigned as $\Delta(r)$ outliers, i.e., regions with rapidly changing recombination landscapes. (**B**) The relationship between $\Delta(r)$ outlier proportion and mean $F_{ST}$ per chromosome.

703




**Supplementary Figure 9: Recombination landscapes in Lake Masoko have PRDM9-independent profiles.** Relative recombination rates in 2kb windows are higher near CpG islands (GpGi) and transcription start sites (TSS) but are independent of the binding motif associated with the incomplete PRDM9 ortholog presents in this species. CpG islands were defined with and without condition of 50% minimal GC content (see Materials and Methods).

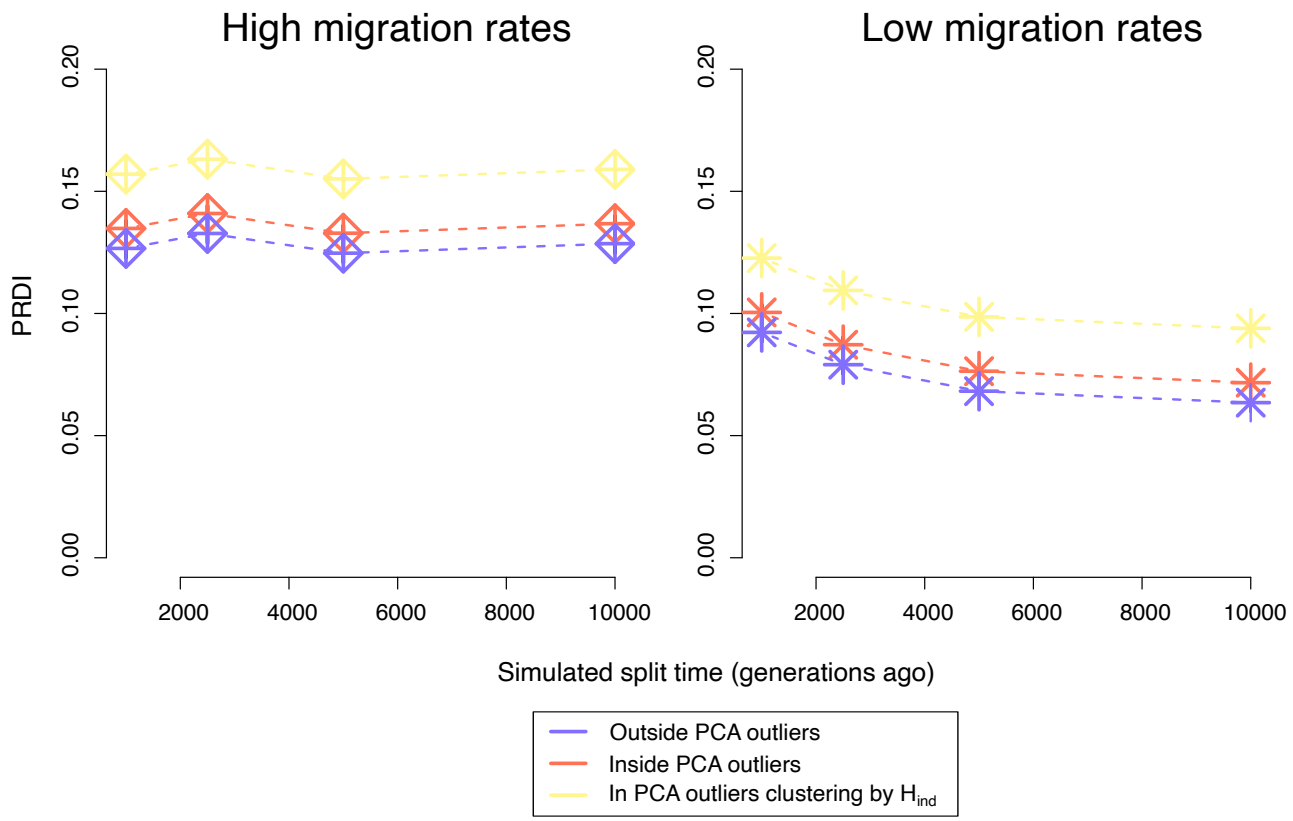



**Supplementary Figure 10: Detection of genomic regions showing clearly distinct population structure.** Each datapoint corresponds to a local PCA for a 100 SNP genomic window. The (dis-)similarity between population structure shown in different local PCA windows is measured by multidimensional scaling (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.

706

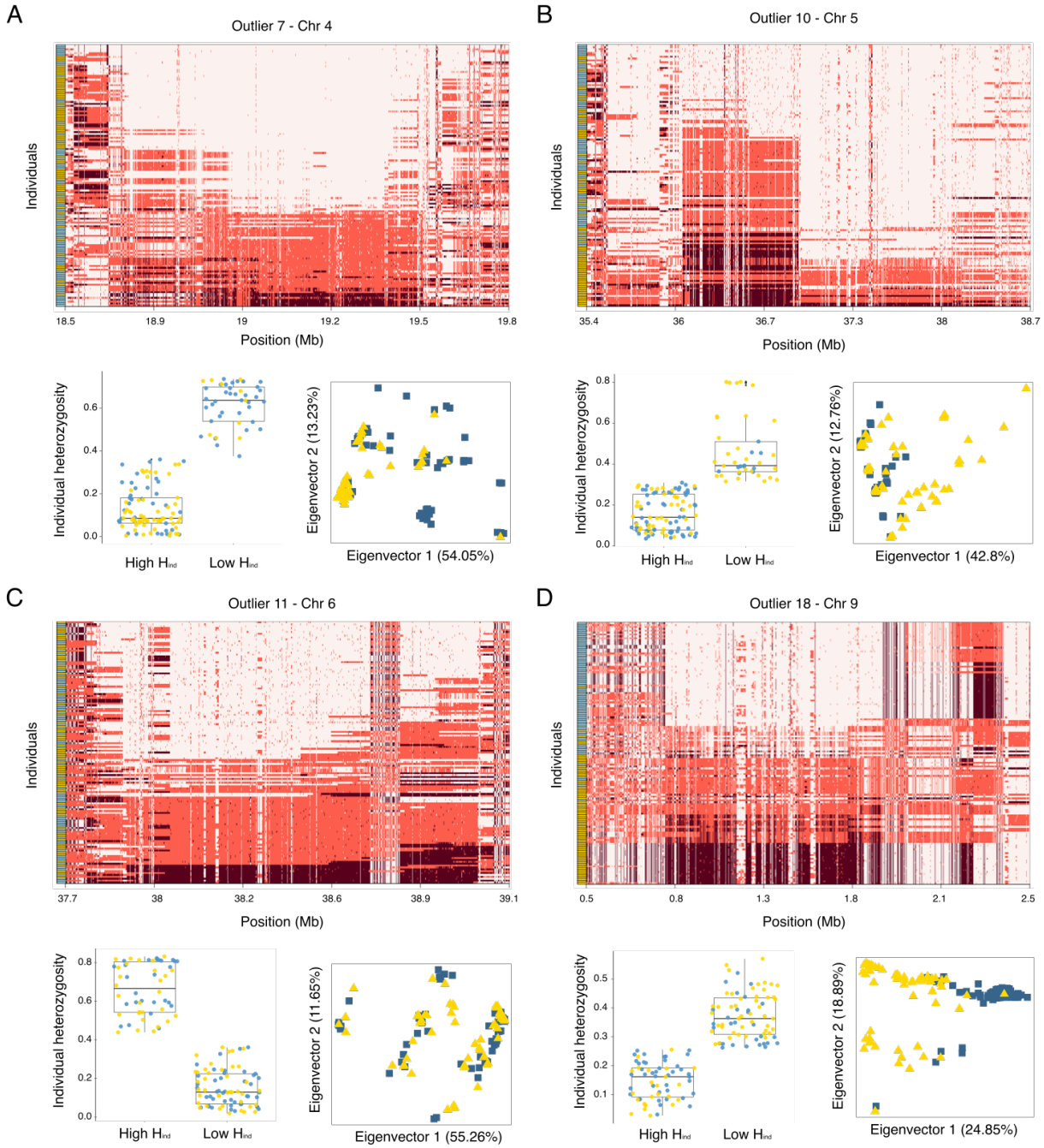
**Supplementary Figure 11: Local PCAs within the 47 haplotype blocks identified with lostruct**. Each blue square represents a benthic individual, and each yellow triangle a littoral individual. The patterns of population structure are very diverse, with some showing clear differences between the ecotypes. 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.

707

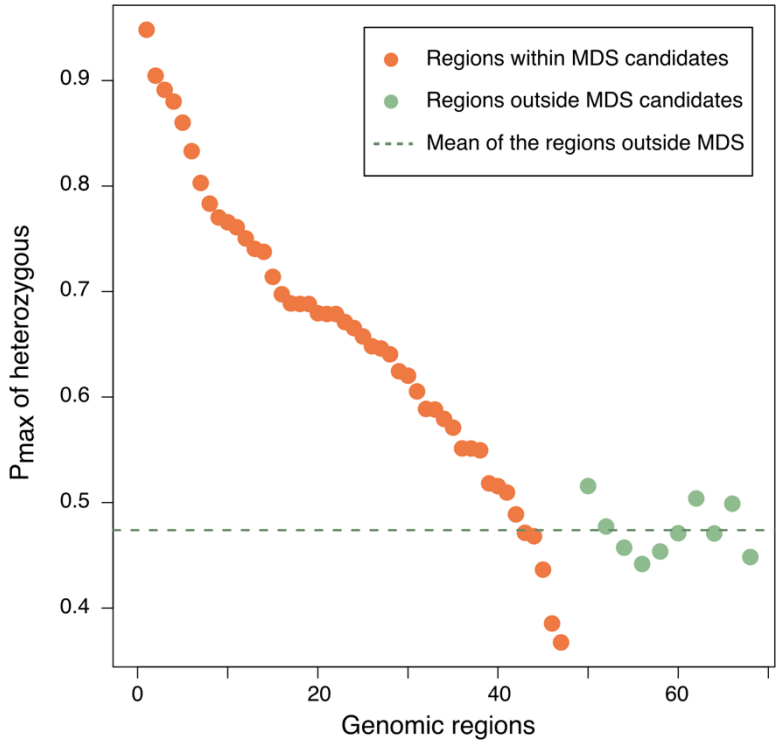



**Supplementary Figure 12**: **The relationship between ecotype divergence in recombination and local PCA outliers.** The Population Recombination Divergence Index (PRDI) is about 10% higher in local PCA outliers than outside of these blocks. PRDI is further elevated in the local PCA outliers where the ecotypes cluster based on individual heterozygosity ($H_{ind}$). The same conclusions are obtained regardless of the split times and/or migration rates used in the neutral simulations that form a part of PRDI estimation.

708




**Supplementary Figure 13: 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 **Figure 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.




**Supplementary Figure 14: 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.

712
713
714
715
716

***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 1: 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 2: 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 Figure 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, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410.

Auton, A., A. Fledel-Alon, S. Pfeifer, O. Venn, L. Ségurel, T. Street, E. M. Leffler, R. Bowden, I. Aneas, J. Broxholme, P. Humburg, Z. Iqbal, G. Lunter, J. Maller, R. D. Hernandez, C. Melton, A. Venkat, M. A. Nobrega, R. Bontrop, S. Myers, P. Donnelly, M. Przeworski, and G. McVean. 2012. A Fine-Scale Chimpanzee Genetic Map from Population Sequencing. Science (New York, N.Y.) 336:193–198.

Bailey, T. L., J. Johnson, C. E. Grant, and W. S. Noble. 2015. The MEME Suite. Nucleic Acids Res. 43:W39–W49.

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

Baker, Z., M. Schumer, Y. Haba, L. Bashkirova, C. H. Elife, and 2017. 2017. Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates. cdn.elifesciences.org, doi: 10.7554/elife.24133.

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

Barton, N. H. 2020. On the completion of speciation. Philos. Trans. R. Soc. B: Biol. Sci. 375:20190530.

Barton, N. H. 2022. The “New Synthesis.” Proc. Natl. Acad. Sci. 119:e2122147119.

Battlay, P., J. Wilson, V. C. Bieker, C. Lee, D. Prapas, B. Petersen, S. Craig, L. van Boheemen, R. Scalone, N. P. de Silva, A. Sharma, B. Konstantinović, K. A. Nurkowski, L. H. Rieseberg, T. Connallon, M. D. Martin, and K. A. Hodgins. 2023. Large haploblocks underlie rapid adaptation in the invasive weed Ambrosia artemisiifolia. Nat. Commun. 14:1717.

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

Baumdicker, F., G. Bisschop, D. Goldstein, G. Gower, A. P. Ragsdale, G. Tsambos, S. Zhu, B. Eldon, E. C. Ellerman, J. G. Galloway, A. L. Gladstein, G. Gorjanc, B. Guo, B. Jeffery, W. W. Kretzschumar, K. Lohse, M. Matschiner, D. Nelson, N. S. Pope, C. D. Quinto-Cortés, M. F. Rodrigues, K. Saunack, T. Sellinger, K. Thornton, H. van Kemenade, A. W. Wohns, Y. Wong, S. Gravel, A. D. Kern, J. Koskela, P. L. Ralph, and J. Kelleher. 2021. Efficient ancestry and mutation simulation with msprime 1.0. Genetics 220:iyab229.

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

Blumer, L. M., V. Burskaia, I. Artiushin, J. Saha, J. C. Garcia, J. Elkin, B. Fischer, C. Zhou, S. Gresham, M. Malinsky, T. Linderoth, W. Sawasawa, I. Bista, A. Hickey, M. Kucka, S. Louzada,

775 R. Zatha, F. Yang, B. Rusuwa, M. E. Santos, Y. F. C. F, D. A. J. DA, A. Böhne, E. A. Miska, M.
776 Ngochera, G. F. Turner, R. Durbin, and H. Svardal. 2024. Introgression dynamics of sex-linked
777 chromosomal inversions shape the Malawi cichlid adaptive radiation. bioRxiv
778 2024.07.28.605452.

779 Booker, T. R., S. Yeaman, and M. C. Whitlock. 2020. Variation in recombination rate affects
780 detection of outliers in genome scans under neutrality. Mol Ecol 29:4274–4279.

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

783 Burri, R., A. Nater, T. Kawakami, C. F. Mugal, P. I. Olason, L. Smeds, A. Suh, L. Dutoit, S. Bures,
784 L. Z. Garamszegi, S. Hogner, J. Moreno, A. Qvarnström, M. Ružić, S.-A. Sæther, G.-P. Saetre,
785 J. Török, and H. Ellegren. 2015. Linked selection and recombination rate variation drive the
786 evolution of the genomic landscape of differentiation across the speciation continuum of
787 Ficedula flycatchers. Genome Research 25:1656–1665.

788 Butlin, R. K., M. R. Servedio, C. M. Smadja, C. Bank, N. H. Barton, S. M. Flaxman, T. Giraud, R.
789 Hopkins, E. L. Larson, M. E. Maan, J. Meier, R. Merrill, M. A. F. Noor, D. Ortiz-Barrientos, and
790 A. Qvarnström. 2021. Homage to Felsenstein 1981, or why are there so few/many species?
791 Evolution 75:978–988.

792 Carruthers, M., D. E. Edgley, A. D. Saxon, N. P. Gabagambi, A. Shechonge, E. A. Miska, R.
793 Durbin, J. R. Bridle, G. F. Turner, and M. J. Genner. 2022. Ecological Speciation Promoted by
794 Divergent Regulation of Functional Genes Within African Cichlid Fishes. Mol. Biol. Evol.
795 39:msac251.

796 Cavassim, M. I. A., Z. Baker, C. Hoge, M. H. Schierup, M. Schumer, and M. Przeworski. 2022.
797 PRDM9 losses in vertebrates are coupled to those of paralogs ZCWPW1 and ZCWPW2. Proc.
798 Natl. Acad. Sci. 119:e2114401119.

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

801 Coster, W. D., M. H. Weissensteiner, and F. J. Sedlazeck. 2021. Towards population-scale long-
802 read sequencing. Nat. Rev. Genet. 22:572–587.

803 Cruickshank, T. E., and M. W. Hahn. 2014. Reanalysis suggests that genomic islands of speciation
804 are due to reduced diversity, not reduced gene flow. Molecular Ecology 23:3133–3157.

805 Danecek, P., J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T.
806 Keane, S. A. McCarthy, R. M. Davies, and H. Li. 2021. Twelve years of SAMtools and
807 BCFtools. Gigascience 10:giab008.

808 DePristo, M. A. M., E. E. Banks, R. R. Poplin, K. V. K. Garimella, J. R. J. Maguire, C. C. Hartl, A.
809 A. A. Philippakis, G. G. del Angel, M. A. M. Rivas, M. M. Hanna, A. A. McKenna, T. J. T.
810 Fennell, A. M. A. Kernytsky, A. Y. A. Sivachenko, K. K. Cibulskis, S. B. S. Gabriel, D. D.
811 Altshuler, and M. J. M. Daly. 2011. A framework for variation discovery and genotyping using
812 next-generation DNA sequencing data. Nature Genetics 43:491–498.

813 Déserts, A. D. des, S. Bouchet, P. Sourdille, and B. Servin. 2021. Evolution of recombination
814 landscapes in diverging populations of bread wheat. Genome Biol. Evol. 13:evab152-.

Duranton, M., F. Allal, C. Fraïsse, N. Bierne, F. Bonhomme, and P.-A. Gagnaire. 2018. The origin and remolding of genomic islands of differentiation in the European sea bass. Nat. Commun. 9:2518.

Escobar, J. S., S. Glémin, and N. Galtier. 2011. GC-Biased Gene Conversion Impacts Ribosomal DNA Evolution in Vertebrates, Angiosperms, and Other Eukaryotes. Mol. Biol. Evol. 28:2561–2575.

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

Faria, R., P. Chaube, H. E. Morales, T. Larsson, A. R. Lemmon, E. M. Lemmon, M. Rafajlović, M. Panova, M. Ravinet, K. Johannesson, A. M. Westram, and R. K. Butlin. 2019a. Multiple chromosomal rearrangements in a hybrid zone between Littorina saxatilis ecotypes. Mol. Ecol. 28:1375–1393.

Faria, R., K. Johannesson, R. K. Butlin, and A. M. Westram. 2019b. Evolving Inversions. Trends Ecol. Evol. 34:239–248.

Feder, J. L., S. P. Egan, and P. Nosil. 2012. The genomics of speciation-with-gene-flow. Trends in Genetics 28:342–350.

Feldman, M. W., S. P. Otto, and F. B. Christiansen. 1996. Population genetic perspectives on the evolution of recombination. Annu Rev Genet 30:261–295.

Felsenstein, J. 1981. Skepticism Towards Santa Rosalia, or Why are There so Few Kinds of Animals? Evolution; international journal of organic evolution 35:124.

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

Genestier, A., L. Duret, and N. Lartillot. 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., T. G. Laurentino, M. Roesti, and D. Berner. 2018. Meta-analysis of chromosome-scale crossover rate variation in eukaryotes and its significance to evolutionary genomics. Mol. Ecol. 27:2477–2497.

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

Harringmeyer, O. S., and H. E. Hoekstra. 2022. Chromosomal inversion polymorphisms shape the genomic landscape of deer mice. Nat. Ecol. Evol. 6:1965–1979.

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

854 Hoge, C., M. de Manuel, M. Mahgoub, N. Okami, Z. Fuller, S. Banerjee, Z. Baker, M. McNulty, P.
855 Andolfatto, T. S. Macfarlan, M. Schumer, A. C. Tzika, and M. Przeworski. 2023. Patterns of
856 recombination in snakes reveal a tug of war between PRDM9 and promoter-like features.
857 bioRxiv 2023.07.11.548536.

858 Jay, P., A. Whibley, L. Frézal, M. Á. R. de Cara, R. W. Nowell, J. Mallet, K. K. Dasmahapatra, and
859 M. Joron. 2018. Supergene Evolution Triggered by the Introgression of a Chromosomal
860 Inversion. Curr. Biol. 28:1839-1845.e3.

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

863 Kent, T. V., J. Uzunović, and S. I. Wright. 2017. Coevolution between transposable elements and
864 recombination. Philos. Trans. R. Soc. B: Biol. Sci. 372:20160458.

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

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

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

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

873 Ma, J., and C. I. Amos. 2018. Investigation of Inversion Polymorphisms in the Human Genome
874 Using Principal Components Analysis. PLoS ONE 7:e40224.

875 Malinsky, M., R. J. Challis, A. M. Tyers, S. Schiffels, Y. Terai, B. P. Ngatunga, E. A. Miska, R.
876 Durbin, M. J. Genner, and G. F. Turner. 2015. Genomic islands of speciation separate cichlid
877 ecomorphs in an East African crater lake. Science 350:1493–1498.

878 Malinsky, M., H. Svardal, A. M. Tyers, E. A. Miska, M. J. Genner, G. F. Turner, and R. Durbin.
879 2018. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by
880 gene flow. Nature Ecology & Evolution 2:1940–1955.

881 Marques, D. A., J. I. Meier, and O. Seehausen. 2019. A Combinatorial View on Speciation and
882 Adaptive Radiation. Trends Ecol. Evol. 34:531–544.

883 Martin, S. H., J. W. Davey, C. Salazar, and C. D. Jiggins. 2019. Recombination rate variation
884 shapes barriers to introgression across butterfly genomes. PLoS Biol. 17:e2006288.

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

887 McDonald, M. J., D. P. Rice, and M. M. Desai. 2016. Sex speeds adaptation by altering the
888 dynamics of molecular evolution. Nature 531:233–236.

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

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

Miller, W., K. Rosenbloom, R. C. Hardison, M. Hou, J. Taylor, B. Raney, R. Burhans, D. C. King, R. Baertsch, D. Blankenberg, S. L. K. Pond, A. Nekrutenko, B. Giardine, R. S. Harris, S. Tyekucheva, M. Diekhans, T. H. Pringle, W. J. Murphy, A. Lesk, G. M. Weinstock, K. Lindblad-Toh, R. A. Gibbs, E. S. Lander, A. Siepel, D. Haussler, and W. J. Kent. 2007. 28-way vertebrate alignment and conservation track in the UCSC Genome Browser. Genome Research 17:1797–1808.

Munby, H., T. Linderoth, B. Fischer, M. Du, G. Vernaz, A. M. Tyers, B. P. Ngatunga, A. Shechonge, H. Denise, S. A. McCarthy, I. Bista, E. A. Miska, M. E. Santos, M. J. Genner, G. F. Turner, and R. Durbin. 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., R. Bowden, A. Tumian, R. E. Bontrop, C. Freeman, T. S. MacFie, G. McVean, and P. Donnelly. 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, N. A., M. W. Wright, J. R. Brister, S. Ciufo, D. Haddad, R. McVeigh, B. Rajput, B. Robbertse, B. Smith-White, D. Ako-Adjei, A. Astashyn, A. Badretdin, Y. Bao, O. Blinkova, V. Brover, V. Chetvernin, J. Choi, E. Cox, O. Ermolaeva, C. M. Farrell, T. Goldfarb, T. Gupta, D. Haft, E. Hatcher, W. Hlavina, V. S. Joardar, V. K. Kodali, W. Li, D. Maglott, P. Masterson, K. M. McGarvey, M. R. Murphy, K. O'Neill, S. Pujar, S. H. Rangwala, D. Rausch, L. D. Riddick, C. Schoch, A. Shkeda, S. S. Storz, H. Sun, F. Thibaud-Nissen, I. Tolstoy, R. E. Tully, A. R. Vatsan, C. Wallin, D. Webb, W. Wu, M. J. Landrum, A. Kimchi, T. Tatusova, M. DiCuccio, P. Kitts, T. D. Murphy, and K. D. Pruitt. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 44:D733–D745.

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

Ortiz-Barrientos, D., and M. E. James. 2017. Evolution of recombination rates and the genomic landscape of speciation. J. Evol. Biol. 30:1519–1521.

Otto, S. P., and B. A. Payseur. 2019. Crossover Interference: Shedding Light on the Evolution of Recombination. Annu. Rev. Genet. 53:1–26.

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

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

Peñalba, J. V., and J. B. W. Wolf. 2020. From molecules to populations: appreciating and estimating recombination rate variation. Nat Rev Genet 21:476–492.

930 Persikov, A. V., R. Osada, and M. Singh. 2009. Predicting DNA recognition by Cys2His2 zinc
931 finger proteins. Bioinformatics 25:22–29.

932 Persikov, A. V., and M. Singh. 2014. De novo prediction of DNA-binding specificities for
933 Cys2His2 zinc finger proteins. Nucleic Acids Res. 42:97–108.

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

936 Ponticelli, A. S., E. P. Sena, and G. R. Smith. 1988. Genetic and physical analysis of the M26
937 recombination hotspot of Schizosaccharomyces pombe. Genetics 119:491–497.

938 Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar,
939 P. I. W. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: A Tool Set for Whole-Genome
940 Association and Population-Based Linkage Analyses. The American Journal of Human Genetics
941 81:559–575.

942 Ravinet, M., R. Faria, R. K. Butlin, J. Galindo, N. Bierne, M. Rafajlović, M. A. F. Noor, B. Mehlig,
943 and A. M. Westram. 2017. Interpreting the genomic landscape of speciation: a road map for
944 finding barriers to gene flow. J. Evol. Biol. 30:1450–1477.

945 Reeve, J., R. K. Butlin, E. L. Koch, S. Stankowski, and R. Faria. 2023. Chromosomal inversion
946 polymorphisms are widespread across the species ranges of rough periwinkles (Littorina saxatilis
947 and L. arcana). Mol. Ecol., doi: 10.1111/mec.17160.

948 Rice, P., I. Longden, and A. Bleasby. 2000. EMBOSS: The European Molecular Biology Open
949 Software Suite. Trends Genet. 16:276–277.

950 Rice, W. R., and A. K. Chippindale. 2001. Sexual Recombination and the Power of Natural
951 Selection. Science 294:555–559.

952 Ritz, K. R., M. A. F. Noor, and N. D. Singh. 2017. Variation in Recombination Rate: Adaptive or
953 Not? Trends Genet 33:364–374.

954 Rolland, J., L. F. Henao-Diaz, M. Doebeli, R. Germain, L. J. Harmon, L. L. Knowles, L. H. Liow,
955 J. E. Mank, A. Machac, S. P. Otto, M. Pennell, N. Salamin, D. Silvestro, M. Sugawara, J. Uyeda,
956 C. E. Wagner, and D. Schluter. 2023. Conceptual and empirical bridges between micro- and
957 macroevolution. Nat. Ecol. Evol. 7:1181–1193.

958 Rowan, B. A., D. Heavens, T. R. Feuerborn, A. J. Tock, I. R. Henderson, and D. Weigel. 2019. An
959 Ultra High-Density Arabidopsis thaliana Crossover Map That Refines the Influences of
960 Structural Variation and Epigenetic Features. Genetics 213:771–787.

961 Samuk, K., B. Manzano-Winkler, K. R. Ritz, and M. A. F. Noor. 2020. Natural Selection Shapes
962 Variation in Genome-wide Recombination Rate in Drosophila pseudoobscura. Curr Biol
963 30:1517-1528.e6.

964 Schluter, D., and L. H. Rieseberg. 2022. Three problems in the genetics of speciation by selection.
965 Proc. Natl. Acad. Sci. 119:e2122153119.

966 Schumer, M., C. Xu, D. L. Powell, A. Durvasula, L. Skov, C. Holland, J. C. Blazier, S.
967 Sankararaman, P. Andolfatto, G. G. Rosenthal, and M. Przeworski. 2018. Natural selection

interacts with recombination to shape the evolution of hybrid genomes. Science (New York, N.Y.) 360:656–660.

Shanfelter, A. F., S. L. Archambeault, and M. A. White. 2019. Divergent Fine-Scale Recombination Landscapes between a Freshwater and Marine Population of Threespine Stickleback Fish. Genome Biology and Evolution 11:1573–1585.

Singhal, S., E. M. Leffler, K. Sannareddy, I. Turner, O. Venn, D. M. Hooper, A. I. Strand, Q. Li, B. Raney, C. N. Balakrishnan, S. C. Griffith, G. McVean, and M. Przeworski. 2015. Stable recombination hotspots in birds. Science 350:928–932.

Spence, J. P., and Y. S. Song. 2019. Inference and analysis of population-specific fine-scale recombination maps across 26 diverse human populations. Sci Adv 5:eaaw9206.

Szankasi, P., W.-D. Heyer, P. Schuchert, and J. Kohli. 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., J. A. Kamm, and Y. S. Song. 2017. Robust and scalable inference of population history from hundreds of unphased whole genomes. Nat. Genet. 49:303–309.

Todesco, M., G. L. Owens, N. Bercovich, J.-S. Légaré, S. Soudi, D. O. Burge, K. Huang, K. L. Ostevik, E. B. M. Drummond, I. Imerovski, K. Lande, M. A. Pascual-Robles, M. Nanavati, M. Jahani, W. Cheung, S. E. Staton, S. Muños, R. Nielsen, L. A. Donovan, J. M. Burke, S. Yeaman, and L. H. Rieseberg. 2020. Massive haplotypes underlie ecotypic differentiation in sunflowers. Nature 584:602–607.

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

Venu, V., E. Harjunmaa, A. Dreau, S. Brady, D. Absher, D. M. Kingsley, and F. C. Jones. 2024. Fine-scale contemporary recombination variation and its fitness consequences in adaptively diverging stickleback fish. Nat. Ecol. Evol. 1–16.

Vernaz, G., A. G. Hudson, M. E. Santos, B. Fischer, M. Carruthers, A. H. Shechonge, N. P. Gabagambi, A. M. Tyers, B. P. Ngatunga, M. Malinsky, R. Durbin, G. F. Turner, M. J. Genner, and E. A. Miska. 2022. Epigenetic divergence during early stages of speciation in an African crater lake cichlid fish. Nat Ecol Evol 1–12.

Wellenreuther, M., and L. Bernatchez. 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-.