Recombination rate variation shapes barriers to introgression across butterfly genomes

Simon H. Martin[1], John W. Davey[2], Camilo Salazar[3], Chris D. Jiggins[1]

[1]Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, United Kingdom

[2]Department of Biology, University of York, YO10 5DD, United Kingdom

[3]Biology Program, Faculty of Natural Sciences and Mathematics, Universidad del Rosario, Carrera. 24 No. 63C-69, Bogota, D.C. 111221, Colombia

ABSTRACT

Hybridisation and introgression can dramatically alter the relationships among groups of species, leading to phylogenetic discordance across the genome and between populations. Introgression can also erode species differences over time, but selection against introgression at certain loci acts to maintain post-mating species barriers. Theory predicts that species barriers made up of many loci throughout the genome should lead to a broad correlation between introgression and recombination rate, which determines the extent to which selection on deleterious foreign alleles will affect neutral alleles at physically linked loci. Here we describe the variation in genealogical relationships across the genome among three species of *Heliconius* butterflies: *H. melpomene*, *H. cydno* and *H. timareta,* using whole genomes of 92 individuals, and ask whether this variation can be explained by heterogeneous barriers to introgression. We find that species relationships vary predictably at the chromosomal scale. By quantifying recombination rate and admixture proportions, we then show that rates of introgression are predicted by variation in recombination rate. This implies that species barriers are highly polygenic, with selection acting against introgressed alleles across most of the genome. In addition, long chromosomes, which have lower recombination rates, produce stronger barriers on average than short chromosomes. Finally, we find a consistent difference between two species pairs on either side of the Andes, which suggests differences in the architecture of the species barriers. Our findings illustrate how the combined effects of hybridisation, recombination and natural selection, acting at multitudes of loci over long periods, can dramatically sculpt the phylogenetic relationships among species.

INTRODUCTION

The genealogical relationships among closely-related species can be complex, varying across the genome and among individuals. This heterogeneity is most prevalent in the presence of introgressive hybridisation, which can alter species' relationships in certain parts of the genome. Genome-scale studies have revealed that particular genomic regions such as sex chromosomes and chromosomal inversions can have distinct phylogenetic histories [1–3], indicating heterogeneity in introgression across the genome. Indeed, the establishment of barriers to introgression in certain parts of the genome through selection against hybrids or admixed individuals is a key part of the speciation process [4–7]. Selection against foreign genetic variation can be driven by both extrinsic factors, such as adaptation to distinct environments, and intrinsic factors, such as genetic incompatibilities [6,8]. The heterogeneous landscape of species relationships therefore carries information about the 'barrier loci' that contribute to the origin and maintenance of species.

The barriers between closely-related subspecies or ecotypes that interbreed frequently are often restricted to just a few loci that contribute to local adaptation, resulting in narrow 'islands' of genetic differentiation between populations [1,9–11]. Recently, it has become evident that patterns of genomic differentiation between more strongly isolated species are often complex and result from an interaction of genomic processes including localised selective sweeps and background selection within species that reduce variation at linked sites [12–14]. This is particularly the case for conventional measures of genetic differentiation, such as F_{ST}, which provide a poor proxy for the strength of a local barrier. However, it is possible to largely avoid the confounding effects of positive and background selection by directly estimating the effective migration rate and how it varies across the genome, either using summary statistics that are less prone to artefacts [15], or through model-based inference [16]. Provided there has been sufficient migration between the species, regions of the genome where admixture is reduced can be inferred to have experienced selection against foreign genetic variation.

Once beyond the very earliest stages of divergence, post-mating species barriers could involve many loci ('polygenic' barriers) [17]. If species barriers are highly polygenic and each locus has only a weak effect on fitness, their individual localised effects on levels of admixture might be difficult to detect, analogous to the difficulties in studying polygenic adaptation more generally [18–20]. While it may not be possible to identify all barrier loci in such a situation, we can test hypotheses about the architecture of barriers by studying genome-wide patterns of admixture. In particular, barriers made up of many loci of small effect are expected to be weaker where recombination rates are higher. Foreign chromosomes that enter a population through hybridisation and backcrossing will be more rapidly broken down over subsequent generations in regions with higher recombination rates. This creates more opportunities for neutral (or mutually beneficial) foreign alleles to be separated from detrimental foreign alleles at other loci, and thus avoid being removed by selection [16,21–23]. A correlation between the recombination rate and the inferred rate of effective migration has been observed between subspecies of house mice [24], subspecies of *Mimulus* monkeyflowers [16] and even between humans and Neanderthals [25], suggesting that loci experiencing selection against introgression among close relatives can be widespread in the genome. To date, it has not been investigated whether such widespread barrier loci could also explain large-scale heterogeneity in phylogenetic relationships across the genome.

We explored species relationships and barriers to introgression among species of *Heliconius* butterflies. Many *Heliconius* species are divided into geographically distinct 'races' with distinct warning patterns, which signal their distastefulness to local predators. Selection favouring locally recognised warning patterns maintains narrow islands of divergence at a few wing patterning loci between otherwise genetically similar races [1,9,26]. However, there are also more strongly differentiated pairs of sympatric species that hybridise rarely and have strong post-zygotic barriers, leading to higher genome-wide genetic differentiation [1]. We studied three such species: *Heliconius melpomene* ('*mel*'), *Heliconius cydno* ('*cyd*') and *Heliconius timareta* ('*tim*'), which form at least two independent zones of sympatry separated by the Andes mountains. While *mel* is found throughout much of South and Central America, *cyd* is largely restricted to the west of the Andes and the inter-Andean valleys, where it overlaps with the western populations of *mel*, whereas *tim* occurs only on the eastern slopes of the Andes, where it co-occurs with the eastern populations of *mel*. In addition to strong assortative mating based on chemical cues and wing patterns in the case of *cyd* and *mel* [27–30], and mainly chemical cues between *mel* and *tim* [29,31,32], both species pairs show ecological differences as well as partial hybrid sterility [29,30,33–37] (and see [29] for a review). Nevertheless, previous studies have revealed surprisingly pervasive admixture between these species in sympatry, most likely explained by a low rate of ongoing hybridisation over an extended period of time [1,38,39]. There is also considerable heterogeneity in the relationships among these populations across the genome [1]. Adaptive introgression in *Heliconius* is well documented. Mimicry between sympatric races of *mel* and *tim* has been facilitated by exchange of multiple wing patterning alleles [40,41], and at least one case of introgression between *mel* and *cyd* has allowed the latter to mimic other unpalatable species [42]. However, the extent to which introgression among these species might be selected against remains unclear.

Using 92 whole genome sequences, we asked whether the heterogeneous relationships observed among these species reflect the influence of polygenic barriers to introgression that vary in their strength across the genome. Then, taking advantage of high-resolution linkage maps for these species [43], we show that admixture is correlated with recombination rate, consistent with polygenic species barriers leading to widespread selection against introgression. This selection also explains broader variation in admixture at the chromosomal scale. Overall, our results highlight the pervasive role of natural selection in shaping the ancestry of hybridising species.

RESULTS

Population Structure

We analysed whole genome sequence data from 90 butterflies representing nine populations of *H. melpomene* ('*mel*', 50 samples from five populations or 'races'), *H. cydno* ('*cyd*', 20 samples from two races) and *H. timareta* ('*tim*', 20 samples from two races), along with two samples from an outgroup species *Heliconius numata* ('*num*') (Table S1). Our sampling included four regions of sympatry: two on the west of the Andes where *cyd* co-occurs with *mel-W*, and two on the eastern slopes of the Andes where *tim* co-occurs with *mel-E*, as well as an allopatric population, *mel-G*, from French Guiana (Figure 1A). Principal components analysis (PCA) based on whole-genome SNP data shows clear distinctions between the three species, and also between *mel-W*, *mel-E* and *mel-G* (Figure 1B). By contrast, pairs of races of the same species from the same broad geographic area (i.e. 'West', 'East' and 'Guiana') are not clearly distinct in the PCA, indicating virtually panmictic populations in each species in each area, despite variation at a few wing patterning loci, as shown previously [44,45]. These results therefore highlight the contrast between the clear barriers that exist between sympatric species, even in sympatry, and the continuity that

exists within species, with the Andes mountains and wide Amazon basin presenting the only major sources of discontinuity among sampled populations of the same species [44,45].

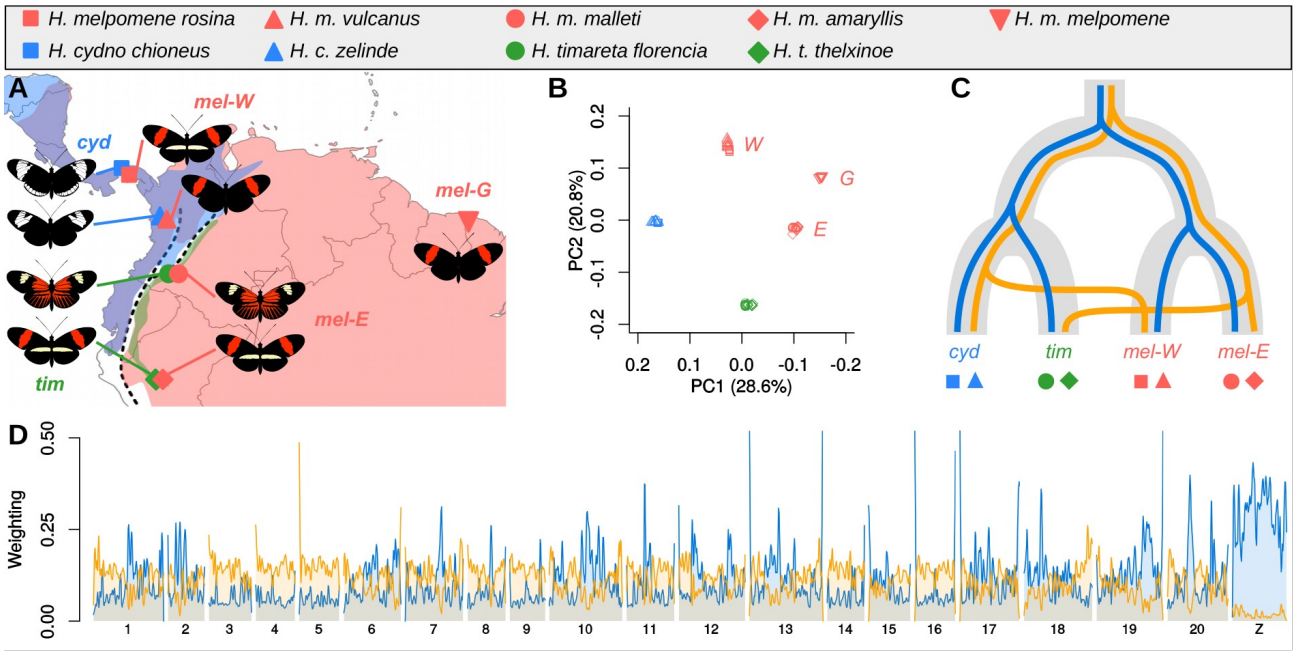


Figure 1. The three species are clearly distinct but their relationships vary across the genome

A. Sampling locations of the nine races from three species included in this study. Species ranges for *H. melpomene*, *H. cydno* and *H. timareta* are indicated by red, blue and green shading, respectively. The dashed line indicates the central part of the Andean mountains **B.** Principal components 1 and 2 differentiate both *cyd* and *tim* from three *mel* populations, but do not separate the two sampled races of each species on either side of the Andes. **C.** The two topologies with the highest weightings, the species topology ('T3', blue) and geography topology ('T6', yellow), shown here as lineages within the hypothesized population branching history. While the geography can arise through both introgression and lineage sorting effects, here it is represented as arising through recent introgression, with the direction (i.e. from *mel-E* into *tim* and from *cyd* into *mel-W*) representing the inferred prevailing direction of introgression, as described below. **D.** Weightings for the 'species' and 'geography' topologies plotted across the 21 chromosomes and smoothed as a locally-weighted average (loess span = 1 Mb). See Figure S4 for a detailed plot without smoothing.

Topology weighting reveals both introgression and incomplete linage sorting

We explored species relationships across the genome using *Twisst* [46], which quantifies the frequency (or 'weighting') of alternative topological relationships among all sampled individuals in narrow windows of 50 single nucleotide polymorphisms (SNPs) each. Consistent with previous results, topology weighting indicates that both large-scale introgression and stochasticity in lineage sorting have shaped the relationships among these species. All 15 possible rooted topologies that describe the relationship between *cyd*, *tim*, *mel-W* and *mel-E* (rooted with *H. numata* as the outgroup) (Figure S1) have non-zero weightings (Figure S2). Despite the strong clustering of distinct species in the PCA, less than 0.5% of windows have completely-sorted genealogies (i.e. in which all groups cluster according to a single topology, resulting in a weighting of one) (Figure S2). This low level of lineage sorting is not surprising given the large effective population sizes (> 2 million individuals [45]), fairly recent divergence times, and large sample sizes used. The two most common topologies are 3 and 6, which differ entirely in the relationships among the ingroup taxa (Figure 1C). Topology 3 matches the expected species branching order, in which *cyd* and *tim* are sister species and *mel*-W groups with *mel*-E ((*cyd*, *tim*), (*mel-W*, *mel-E*)) [44,47]. We refer to this as the 'species topology'. Topology 6, by contrast, groups populations by geography: *cyd* with *mel-W*, and *tim* with *mel-E* ((*cyd, mel-W*), (*tim, mel-E*)). We refer to this as the 'geography topology'. We

therefore hypothesise that the history of these species can be modelled as a branching process following the species topology, but with considerable variation in lineage sorting as well as introgression that increases the rate of coalescence between sympatric populations from distinct species, as in the geography topology. Across the autosomes, the next two most highly weighted topologies (14 and 5) further support this hypothesis, as they each group one pair of sympatric taxa but otherwise match the species topology (Figure 2A).

Relative coalescence times in each genealogy provide further support for the hypothesis that introgression has led to increased clustering by geography. Topologies grouping sympatric non-sister taxa tend to have shallower splits, consistent with recent coalescence resulting from post-speciation gene flow (Figure S3). Topologies grouping allopatric non-sister taxa tend to have deeper splits, consistent with coalescence in the ancestral population and incomplete lineage sorting (Figure S3). These average branch lengths reflect a combination of both recent and ancient coalescence events and depend on population size, and should therefore be interpreted with caution. Nonetheless, the differences between topologies are in agreement with our model of recent or ongoing gene flow between both sympatric species pairs.

Our sampling design allows us to make inferences about biases in the direction of gene flow. The third most abundant topology genome-wide (Topology 14) has *tim* nested within the *mel* clade, suggesting introgression from *mel-E* into *tim* (Figure 2A). This would be expected given the much smaller range and lower effective population size of *tim*, which also has the lowest nucleotide diversity of all the taxa studied [1]. Likewise, Topology 5, which is the fourth most abundant across autosomes suggests that most introgression in the west of the Andes has been from *cyd* into *mel-W* (Figure 2A). This direction was also inferred to be the most likely in a previous study using coalescent modelling [39], and is consistent with the fact that F1 hybrids show mate preference for *H. melpomene* in experiments with Panama populations [27].

Species relationships vary across the genome

Topology weightings vary considerably across the genome. To highlight this heterogeneity, we first focus on the two most abundant patterns of relatedness: the species and geography topologies (Figure 1 C,D). The species topology has the highest weighting in narrow peaks on some of the autosomes and across the Z chromosome, while the geography topology has a higher weighting than the species topology throughout the rest of the genome. In other words, throughout large parts of the genome, samples of *mel-W* and *mel-E* tend to be more closely related to their respective sympatric counterparts, *cyd* and *tim*, than to one another. However, the species topology tends to occur in sharp peaks, which frequently have a weighting approaching 1, indicating complete lineage sorting (Figure S4, note that these narrow peaks are not visible in Figure 1D due to smoothing). By contrast, all other topologies, including the geography topology, seldom approach a weighting of 1, indicating that they tend to be incompletely sorted, with individuals from each group being more dispersed in each genealogy. The much higher occurrence of complete sorting in the species topology is consistent with selection maintaining strong barriers between species in certain parts of the genome.

There are also strong trends in the abundance of the species and geography topologies across the 20 autosomes. All species in this clade have ten short and ten long autosomes. The latter formed through ten independent fusions in the ancestor of *Heliconius*, which had 30 autosomes [48,49].

The species topology is less abundant on the ten short autosomes as compared to the ten long, fused autosomes, and there is a fairly linear increase in its weighting with chromosome length (Figure 2A). By contrast, the geography topology shows decreasing abundance with chromosome length, and tends to be far more abundant on the short chromosomes. There is also a fairly consistent within-chromosome trend, with higher weightings for the geography topology and lower weightings for the species topology toward the outer third of the chromosomes compared to the chromosome centres (Figure 2C).

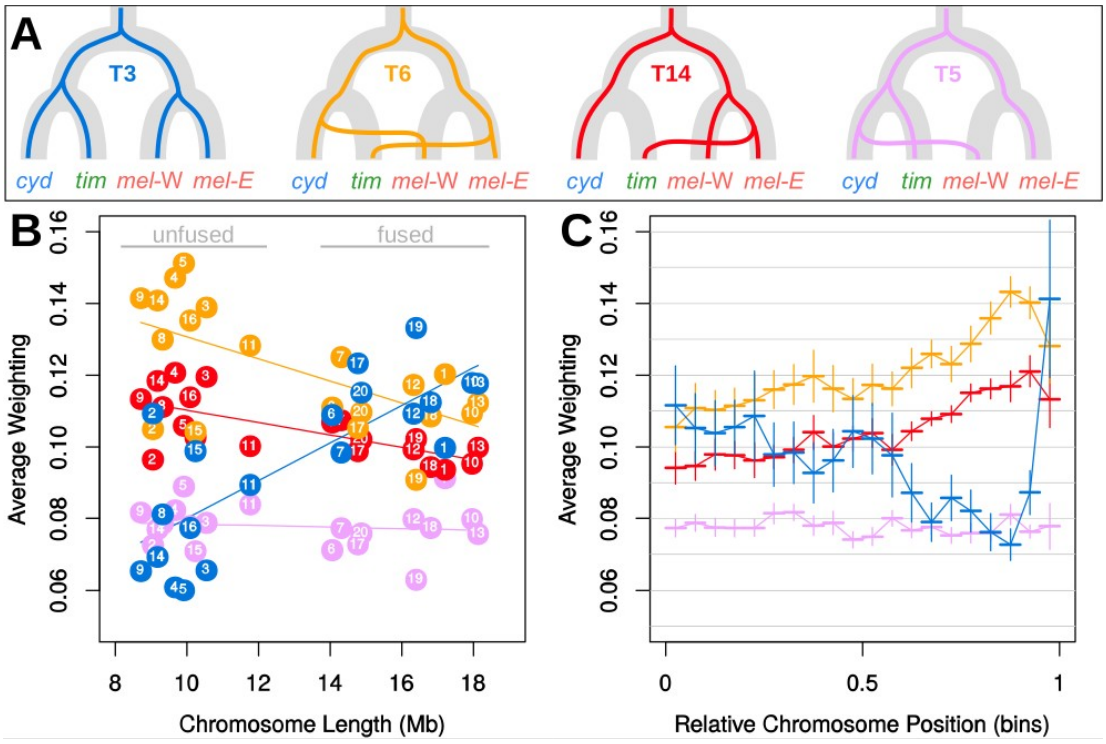


Figure 2. Species relationships vary consistently both among and within chromosomes

A. The four topologies that are most abundant across autosomes, the species topology (3, blue), the geography topology (6, yellow), and two other topologies consistent with introgression (14, red; and 5, pink). Note that topologies 14 and 5 suggest a likely predominant direction of introgression from *cyd* into *mel-W* and from *mel-E* into *tim*, and therefore the same directions are indicated in the illustration of topology 6. **B.** The average weighting for the same four topologies (colours as in **A**) for each of the 20 autosomes, plotted against the physical length of the chromosome. **C.** Average weightings for the same four topologies (colours as in **A** and **B**) binned according to their relative chromosome position, from the centre (0) to the periphery (1). Each bin represents 5% of the chromosome arm, with the range indicated by a horizontal line. Vertical lines indicate +/- one standard error.

The above trends might be partly explained if the variance in lineage sorting is greatest on short chromosomes and away from chromosome centres. Indeed, we previously found a negative correlation between chromosome length and effective population size in *H. melpomene* [45]. However, several observations suggest that these patterns also reflect variation in the extent of introgression between genomic regions. Topology 14, which is consistent with introgression between *mel-E* and *tim* is less abundant than the species topology on long chromosomes and at chromosome centres, but is more abundant on short chromosomes and in chromosome peripheries (Figure 2B,C). Such a switch in rank is not expected if the short chromosomes and peripheries simply experience more variation in lineage sorting, but is consistent with differences in the extent of introgression between short and long chromosomes and between centres and peripheries. Interestingly topology 5, which is consistent with introgression between *cyd* and *mel-W*, does not show any clear relationship with chromosome length or relative chromosome position

(Figure 2B,C). This implies that there may be less consistent variation in the extent of introgression between *cyd* and *mel-W* in different regions, and also further strengthens the argument that differences in the level of variation in lineage sorting alone cannot explain the differences in relatedness among chromosome regions. The chromosome-level trends for all 15 topologies are shown in Figure S5. In summary, topology weighting reveals quantitative variation in species relationships both within and among chromosomes consistent with heterogeneity in the level of introgression. However, topology weighting does not explicitly distinguish between introgression and shared ancestral variation. We therefore set out to explicitly test the hypotheses that (1) there is heterogeneity in the level of admixture across the genome, and (2) that this heterogeneity can be explained by variation in the strength of selection against introgression.

Heterogeneous admixture suggests variable selection against introgression

We used the summary statistic f_d [15] to quantify admixture separately between *cyd* and *mel-W* and between *tim* and *mel-E*. This approach also measures an excess of genealogical clustering of sympatric non-sister taxa. However, f_d provides a normalised measure that is approximately proportional to the effective migration rate [15]. Building on previous work, we first investigated the degree to which f_d might be influenced by variation in effective population size (N_e) across the genome. N_e tends to be reduced in regions of reduced recombination rate due to linked selection. By means of simulations, we find that, across a large range of realistic population sizes, f_d is a reliable estimator of admixture. Furthermore, f_d outperforms the commonly-used divergence statistics F_{ST} and d_{XY}, which are both highly sensitive to N_e (Figure S6). When population sizes are very large, f_d tends to underestimate the true level of admixture. This is caused by a loss of information when population sizes are large relative to the split times: the lack of lineage sorting means that there is insufficient information available to accurately quantify admixture. The population sizes for which this is relevant are at the upper end of estimates for these species [45]. Moreover, this error would cause a conservative bias in our results, as we expect reduced admixture in low-recombination regions, where N_e is expected to be the smallest. Most important for our subsequent analysis, high background selection in regions of low recombination, which is known to influence measures such as F_{ST} is not likely to strongly bias our estimates using f_d. We therefore conclude that f_d provides a suitable, albeit conservative, measure to test the hypothesis that species barriers are enhanced in regions of reduced recombination rate.

Computation of f_d requires the use of a 'control' population that is ideally allopatric and unaffected by introgression. To confirm the robustness of our results, we computed f_d with several different sets of populations, varying the control population, as well as splitting or joining each of *cyd*, *tim*, *mel-W* and *mel-E* into their two constituent sub-populations (Figure S7).

Patterns of admixture estimated by f_d show considerable heterogeneity across the genome (Figure 3A). As expected, admixture is minimal across the Z chromosome in both pairs, indicating a strong barrier to introgression. There is also heterogeneity in admixture proportion across the autosomes. This is most striking between *tim* and *mel*-E, where some regions exhibit deep troughs, implying strong, localised species barriers. Some of this heterogeneity likely reflects individual barrier loci of large effect. Indeed, the known wing-patterning loci provide a useful example. The pattern differences between *cyd* and *mel-W* are determined by regulatory modules around three major genes: *wnt-A* (chromosome 10), *cortex* (chromosome 15) and *optix* (chromosome 18) [42,50–54]. These probably act as strong barriers to introgression between *cyd* and *mel-W,* due to increased

predation against hybrids with intermediate wing patterns [36]. By contrast, the shared wing patterns of *tim* and *mel*-E are thought to result from adaptive-introgression of wing patterning alleles. As expected, there is a strong reduction in admixture between *cyd* and *mel*-W in the vicinity of all three genes, while there are peaks of admixture between the co-mimetic *tim* and *mel-E* populations in the corresponding regions (Figure S8).

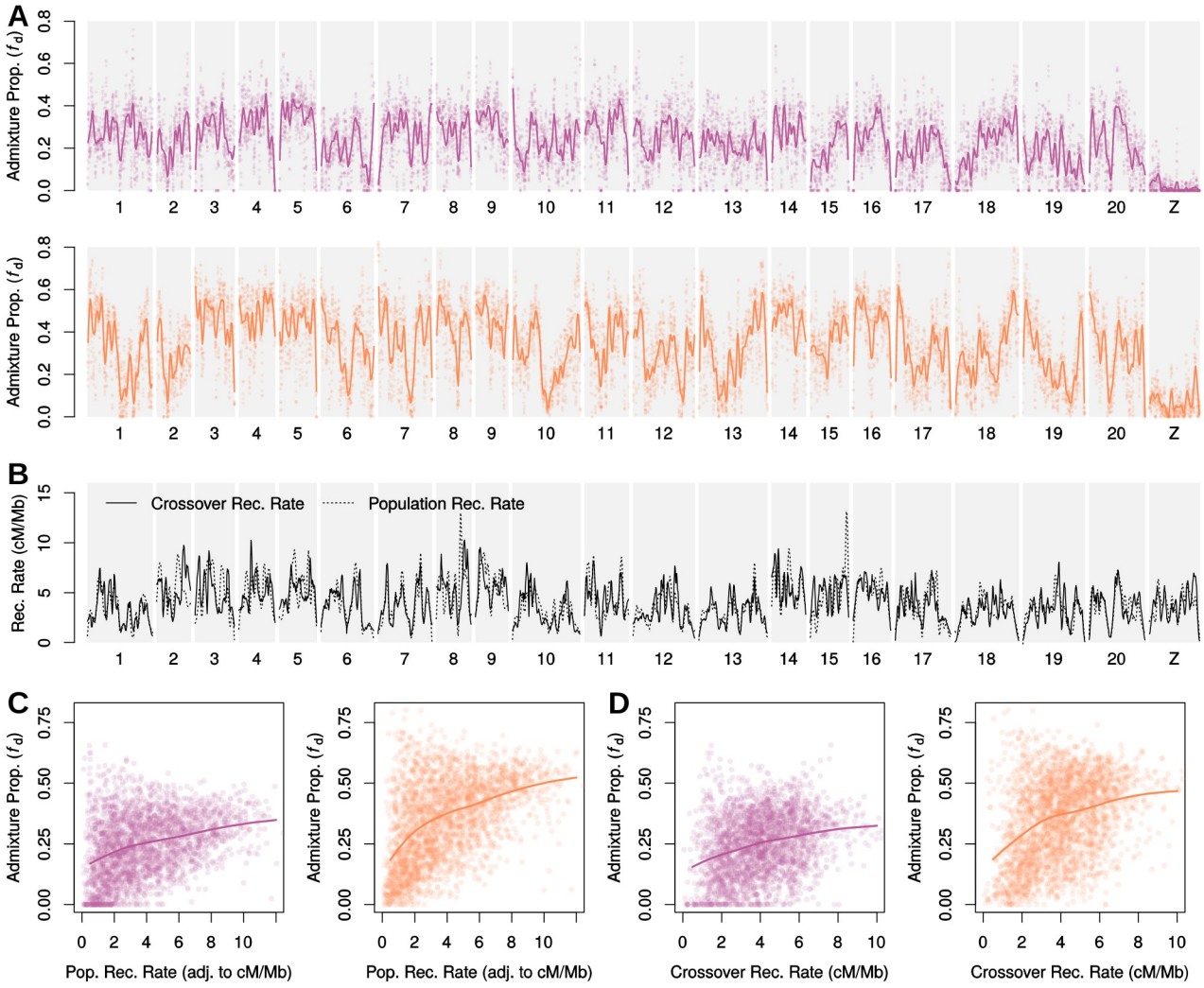


Figure 3. Admixture proportions are correlated with recombination rate

A. Estimated admixture proportions (f_d) between *cyd* and *mel-W* (upper) and between *tim* and *mel-E* (lower) plotted across all 21 chromosomes in 100 kb windows, sliding in increments of 20 kb. A locally-weighted average (loess span=2 Mb) is included. Results shown are for population Sets 1 and 4 of Figure S7. **B.** Recombination rate estimated from the crossover rate in linkage maps (solid line) and the population recombination rate averaged across the four populations considered and plotted as a locally-weighted average (loess span = 2 Mb) (dashed line). See Figure S9 for a more detailed plot. **C.** Admixture proportions for *cyd* and *mel-W* (left) and *tim* and *mel-E* (right) for non-overlapping 100 kb windows plotted against the population recombination rate. Solid lines indicate the locally-weighted average (loess span = 0.75). Dashed lines indicate the same, but when windows in the outer third of each chromosome are excluded. **D.** As in C, except that the x-axis is the crossover recombination rate inferred using linkage maps.

Admixture proportions are correlated with recombination rate

We hypothesized that many loci across the genome contribute to the species barriers, which leads to the expectation that the level of admixture will be correlated with the recombination rate [16]. We quantified variation in recombination rate across the genome using high-resolution linkage maps

[43] as well as using LDHelmet, which estimates the population recombination rate (ρ) based on linkage-disequilibrium (LD) in the genomic data from natural populations. At a broad scale, the map-based estimates are highly concordant with the population-based estimates, and the latter are also strongly conserved across the different species (Figure 3B, S9). There is considerable variation in recombination rate across the genome, allowing us to investigate whether admixture proportions are correlated with recombination rate.

There is a strong positive and non-linear relationship between admixture proportion and recombination rate in both species pairs (Figure 3C, 3D, S11). Strong reductions in admixture, implying barriers to introgression, are concentrated in genomic regions where recombination rates are below 2 cM/Mb. However, there is also more variability in admixture proportions in these low-recombination regions, with some showing high levels of admixture (Figure 3C, 3D). This might imply that some regions do not harbour loci that contribute to the species barrier, although the variance in admixture proportions may also be increased in low-recombination regions due to increased genetic drift resulting from enhanced linked selection. In addition, the relationship between admixture and recombination rate is clearest in the *tim* and *mel-E* pair implying that the more heterogeneous pattern of admixture across the genome between this pair is more consistent with a model in which barrier loci are widespread and recombination rate modulates the strength of the barrier to introgression.

Admixture proportions are less well predicted by the map-based estimates of crossover recombination rate (Figure 3D) compared to the inferred population recombination rate (ρ) (Figure 3C,). This probably partly reflects inaccuracy in fine-scale recombination rate estimated from the linkage maps. However, it may also be that ρ (=$4N_er$) provides a more meaningful predictor for the admixture proportion, as it is a composite of the per-generation recombination rate (r) and local effective population size (N_e). Due to linked selection, parts of the genome with a low recombination rate and a high density of selected sites are expected to have locally reduced N_e and therefore reduced ρ. Indeed, ρ is strongly negatively correlated with gene density (linear regression, r^2=0.368, p=$9.67\text{x}10^{-273}$, Figure S10). However, there is also a weaker but significant negative relationship between gene density and the crossover recombination rate (linear regression, r^2=0.064, p=$7.93\text{x}10^{-41}$, Figure S10). This implies that linked selection in regions of low recombination rate may be further enhanced by a higher density of selected loci. As the conditions that enhance linked selection are the same as those expected to strengthen barriers to introgression (i.e. a high ratio of selected loci relative to the recombination rate, also called the 'selection density' [16]) it is to be expected that ρ would provide a better predictor of barrier strength and therefore admixture proportion. As expected, there is a negative relationship between admixture proportion and the proportion of coding sequence per window (referred to as 'gene density' below) (Figure S12). However, the fact that regions with a high gene density also tend to have lower recombination rates makes it difficult to determine whether such regions harbour a higher physical density of barrier loci, but this seems likely given the arguments above.

The above trends are robust to using different allopatric 'control' populations when estimating admixture proportions (Figure S11, S12), with the exception that using very closely related control populations lead to very low estimated rates of admixture, for which the relationships with recombination rate and gene density are not clear. See Figure S11 and Figure S12 for details.

Large scale trends in patterns of admixture

Average chromosomal admixture proportions are negatively correlated with chromosome length (Figure 4A, 4B). This is expected given the extremely strong negative correlation between physical chromosome length (in base pairs) and average recombination rate (Figure 4C). By contrast, there is no clear relationship between chromosome length and gene density (Figure 4D). The broadly enhanced barrier to introgression on long chromosomes is therefore more consistent with an effect of increased linkage, rather than an increased density of barrier loci. As in the trends above, the relationship with chromosome length is stronger for admixture between *tim* and *mel-E* (correlation coefficient = -0.76, p=8e-05, df=18), than for admixture between *cyd* and *mel-W* (correlation coefficient = -0.52, p=0.018, df=18). Between *tim* and *mel-E,* the shortest chromosomes experience about 50% more admixture than the longest chromosomes, with the exception of chromosome 2, which has strongly reduced admixture compared to other short chromosomes with similarly high recombination rates. This might reflect a higher density of barrier loci on this chromosome, which seems possible as it also has the highest gene density of all chromosomes (Figure 4D).

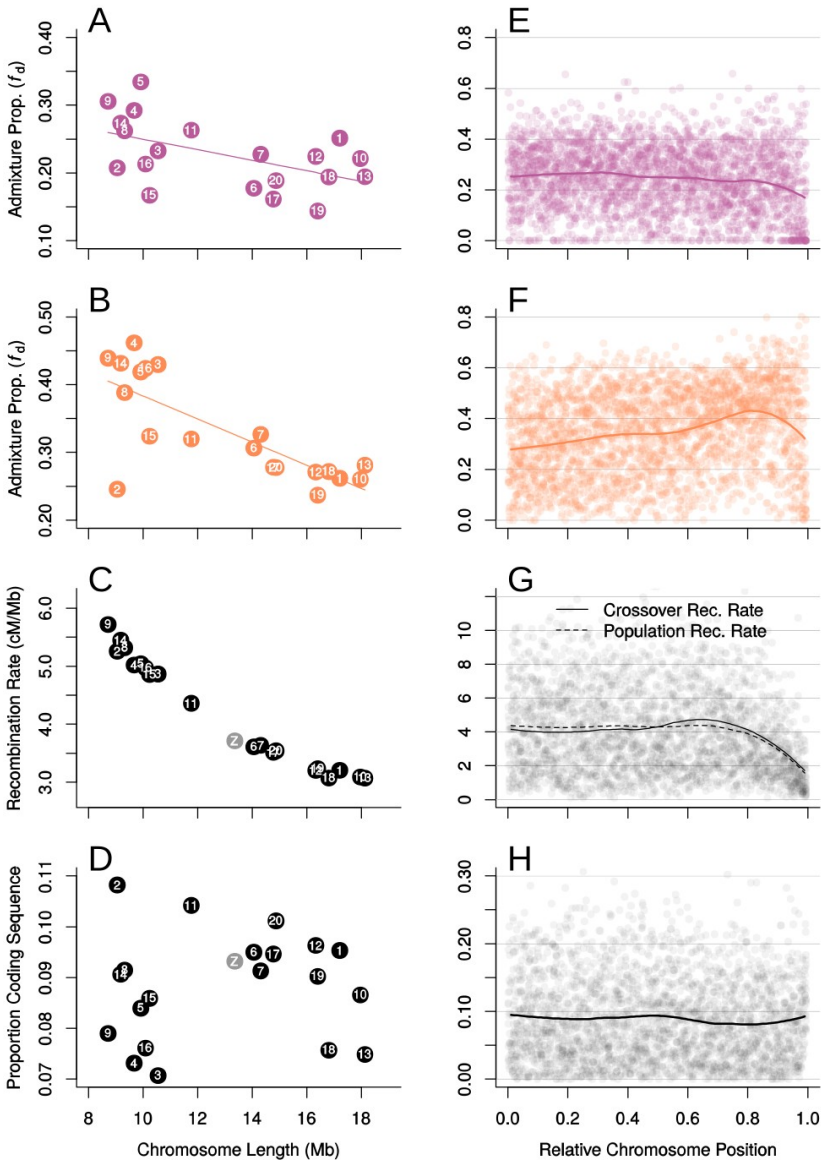


Figure 4. Variation in admixture proportions among and within chromosomes is explained by recombination rate and proximity to chromosome edges

A. & **B.** Estimated admixture proportions (f_d) between *cyd* and *mel-W* (**A**) and between *tim* and *mel-E* (**B**) for each chromosome plotted against chromosome length. **C.** Crossover recombination rate for each chromosome plotted against chromosome length. **D**. Proportion of coding sequence for each chromosome plotted against chromosome length. **E.** & **F.** Estimated admixture proportions (f_d) between *cyd* and *mel-W* (**E**) and between *tim* and *mel-E* (**F**) for 100 kb windows plotted against relative chromosome position. A locally-weighted average (loess span=0.25) is included. **G.** Crossover recombination rate plotted against relative chromosome position. A locally weighted average is included, along with the corresponding line for the population recombination rate (ρ). **H.** Proportion of coding sequence per 100 Kb window plotted against average chromosome position, again with a locally-weighted average shown.

As indicated by topology weighting, there is an effect of position along the chromosome on the proportion of admixture between *mel-E* and *tim,* where admixture increases on average towards the distal region of the chromosome, but decreases again at chromosome ends (Figure 4F). This is not seen in the proportion of admixture between *cyd* and *mel-W* (Figure 4E). Unlike in many other taxa, there is no consistent decrease in recombination rate toward chromosome centres. By contrast, both the crossover recombination rate and ρ show a sharp decrease at the chromosome ends (Figure 4G). Gene density is roughly uniform across chromosomes on average (Figure 4H). Theory predicts that, given a uniform recombination rate and distribution of selected loci, species barriers should weaken toward chromosome ends, leading to increased admixture [22]. Therefore, the different trends seen in the two species pairs might imply a different balance between this edge effect, which should weaken species barriers, and reduced recombination, which should strengthen them.

DISCUSSION

Introgression effectively acts to rewrite the evolutionary history of the genome. Genome-scale data have revealed that the extent of introgression in some species may be far greater than previously imagined [1,2,55,56]. Despite the strong behavioural isolation among the three species studied here, we find that relationships among them vary dramatically across the genome, and that in some parts introgression has overwhelmed the genealogical footprints of the original population branching pattern. Similar dramatic heterogeneity in species relationships has been described in several other taxa [2,57]. For example, introgression among some *Anopheles* mosquitos has almost entirely eliminated the signal of the original species branching events on autosomes [2]. In the present study, by analysing genomes from multiple samples per population, we show that *Heliconius* species relationships vary quantitatively within and among chromosomes. Our main finding is that this variation in species relationships is predictable, and can be explained by quantitative variation in the strength of the selective barrier to introgression, which depends on the local recombination rate. Our findings therefore show how hybridisation and natural selection act in combination to shape the tree of life.

There has been considerable interest in making inferences about species barriers from the genomic landscape of divergence between hybridising species, based on the idea that selection should resist genetic homogenisation by gene flow at barrier loci. However there has also been an increasing realisation that patterns of differentiation and divergence can be influenced by unrelated factors, such as linked selection acting within species [13,58–60]. One effect of these confounding factors is that relative measures of divergence, such as F_{ST}, can show elevated values in regions of the genome that experience stronger linked selection, even if there is no reduction in effective migration in such regions, and indeed even when there is no gene flow at all [14]. In other words,

the observation of increased F_{ST} in regions of the genome with reduced recombination rate is not informative about the architecture of species barriers. This is particularly problematic when considering species that hybridise rarely, as the contribution of gene flow to patterns of differentiation may be small compared to that of within-species linked selection. Previous analyses of these *Heliconius* species revealed a highly heterogeneous pattern of F_{ST}, in which even known wing patterning loci that have a major impact on hybrid fitness are not particularly prominent [1]. Until now, it has not been known whether barrier loci are indeed widespread throughout the genome in these species.

Our results suggest that there are highly polygenic barriers that maintain these *Heliconius* species. Genome windows in regions of high recombination rate (>5 cM/Mb) almost invariably show increased levels of admixture, whereas windows showing reduced admixture are concentrated in parts of the genome with low recombination rates (<2 cM/Mb). This is consistent with theoretical expectations that barrier loci will cause a stronger localised reduction in introgression in regions of lower recombination rate [16,21–23]. Interestingly, some windows in regions of low recombination nevertheless show high levels of admixture. This may indicate that barrier loci, although abundant, are not ubiquitously distributed across the genome. However, we also expect increased variance in levels of admixture in these low-recombination regions due to increased genetic drift, which will be compounded by the reduced independence among sites in the 100 kb windows. This increased variance does not explain the positive relationship between admixture and recombination, however. Selection against introgression also produces a global pattern of decreasing admixture with chromosome length. Long chromosomes, which have similar gene density but lower per-base recombination rates than short chromosomes, form stronger barriers to introgression on average. It is likely that species barriers are also stronger in gene-rich regions, due to an increased density of barrier loci. While we did find a weak trend of reduced admixture in gene-rich regions, this is difficult to interpret, as the recombination rate is also lower in gene-rich regions. A final factor that could influence our conclusions is the fact that barrier loci may not be expected to accumulate randomly across the genome. Some models predict that, under a scenario of ecological divergence in the face of gene flow, the accumulation of barrier loci may be clustered [23,61]. This could increase the correlation between admixture and recombination, and perhaps lead to overestimation of the density of barrier loci. Clustering could theoretically be further enhanced by genomic rearrangements between species that physically suppress recombination in hybrids. However, we have previously found that there are no major inversions maintaining barriers between *H. melpomene* and *H. cydno* [43]. Nonetheless, the lack of information about the distribution of barrier loci and their effect sizes means that it is currently not possible to estimate the number of loci involved, except that it is probably very large.

In agreement with previous findings, we find that barriers to introgression are far stronger across the Z chromosome compared to autosomes. Enhanced barriers to introgression on sex chromosomes have been observed in genomic studies of a range of taxa with both XY and ZW systems [1–3,62] . This has been attributed to a more rapid build-up of incompatibilities due to hemizygozity and a key role played by sex chromosomes in reproduction and fertility. Comparing genetic differentiation on sex chromosomes to autosomes can be complicated by their reduced effective population size, and this can be further confounded by changes in population size, which can affect sex chromosomes differently [63]. However, our simulations show that these factors cannot explain the reduction in admixture detected here using f_d. In these *Heliconius* species, hybrid female sterility is associated with one or more loci on the Z chromosome [35,64,65]. The

observed reduction of admixture across the Z chromosome must result from selection against foreign Z chromosome alleles in backcross progeny and their descendants, such that there are opportunities for independent assortment of chromosomes prior to selection. Segregation of sterility in backcrosses has indeed been observed in crossing experiments [35,65,66]. This also means that there are opportunities for recombination before selection. The fairly even reduction in admixture we observe across most of the chromosome is therefore perhaps surprising, and implies that there are multiple barrier loci spread across the Z chromosome. A similar pattern of widespread incompatibilities throughout much of the sex chromosome has been shown experimentally between *Drosophila* species [67].

The model proposed here of a highly polygenic species barrier between *H. melpomene* and its relatives contrasts with previous studies that identified a few major effect loci that control differences in wing pattern and mate preference between *H. cydno* and *H. melpomene* [36,66]. However, multiple additional behavioural and ecological differences are known to distinguish these species [29,68], and each of these may have a more polygenic basis. Interestingly, unlike *H. cydno*, *H. timareta* races have wing patterns that commonly match those of the local *H. melpomene* races. Hence, the large effect wing patterning loci do not contribute to the barrier between *H. timareta* and *H. melpomene*. This difference might explain why admixture between this pair is more strongly correlated with recombination rate. When barrier loci are weak and dispersed across the genome, admixture proportions should be more strongly predicted by recombination rate than when there are few large-effect barrier loci. Hence, perhaps counter-intuitively, the more heterogeneous pattern of admixture between *H. timareta* and *H. melpomene* is in fact more consistent with the model of small-effect barrier loci evenly distributed across the genome. The heterogeneity reflects the underlying heterogenous recombination landscape, rather than a patchy distribution of large-effect barrier loci.

The cause of the more even pattern of admixture between *H. cydno* and western *H. melpomene* is less clear, but it may be explained by epistatic selection on the patterning loci. Lindtke and Buerkle [69] distinguish between two types of epistatic barrier loci: “DMI-type” incompatibilities that cause reduced fitness in hybrids but can be broken down by recombination in backcross progeny, following the formulation by Dobzhansky [4] and Muller [7], and “pathway-type” incompatibilities that reduce fitness in recombinant hybrids in which co-adapted alleles become separated. Simulations show that pathway-type incompatibilities can produce pronounced localised barriers to introgression, while DMI-type incompatibilities can have more even, genome-wide effects, if selection is strong enough [69]. Hybrids between *H. cydno* and *H. melpomene* have intermediate wing patterns that do not resemble the recognisable warning patterns of either species, making them roughly twice as vulnerable to predation [36]. The patterning loci may act as DMI-like incompatibilities, as backcrossing restores a recognisable warning pattern to a proportion of the progeny. It is therefore possible that the presence of a few large-effect incompatibilities could in fact explain the less heterogeneous landscape of introgression between *H. cydno* and *H. melpomene*, although this requires further investigation.

In conclusion, our findings imply that barrier loci have accumulated rapidly in the 1-1.5 million years over which these butterfly species have diverged. This joins a growing number of examples showing that selection against introgression between fairly young species can be pervasive across the genome [16,24,25]. Further work is still required to determine the generality of these trends, and also to account for complications such as clustered barrier loci and epistasis. These new

insights into the polygenic nature of species boundaries highlight the dangers of assuming strictly neutral evolution when modelling speciation. Models that incorporate variable selection pressures among sites [55,70] are likely to be more realistic. Our results here are intriguing in that they show that, despite the widely distributed barriers across the genome, introgression has nonetheless dramatically reshaped species relationships. A few recent examples have shown how introgression can lead to dramatically different topologies across genome regions, but our data goes further in showing (1) how this phylogenetic heterogeneity can be predicted by recombination rate and (2) how relationships can vary across a species range. The focal species *H. melpomene* has very different relationships to its sister taxa both depending on the genome region and on the population sampled. These patterns raise questions about how we view the species as an entity and the degree to which animal life can be accurately viewed as a bifurcating tree.

METHODS AND MATERIALS

Samples and genotyping

We used whole genome resequencing data from 92 wild-caught butterflies (Table S1) [1,42,45,52,63]. Reads were mapped to the *H. melpomene* genome assembly v2 [49] using BWA mem v0.7.2, using default parameters. Read depth was computed using GATK v3.4 DepthOfCoverage [71]. Average read depth across all 92 samples was 29.22 (Table S1). Genotyping was performed using the GATK v3.4 HaplotypeCaller and GenotypeGVCFs tools [71], using default parameters except that heterozygosity was set to 0.02. Each geographic population (10 samples each) was genotyped separately. Variant sites were accepted only if the quality (QUAL) value was ≥ 30, and individual genotype calls were accepted only where the sample depth of coverage for the position was ≥ 8. Scaffold positions in the Hmel2 assembly were converted to chromosome positions based on the most recent scaffolding [43], now released as Hmel2.5. Two sets of filtered SNPs were generated for the analyses below. In addition to the requirement of ≥ 8X depth of coverage, SNPs were required to be biallelic and sites at which more than 75% of samples were heterozygous, or where the minor allele was present in only a single haploid copy, were discarded. SNP Set 1 had the further requirement that at least nine out of the ten samples representing each of the nine ingroup populations, and one of the two outgroup samples, had an accepted genotype call, resulting in 14,406,386 SNPs. SNP Set 2 had the less stringent requirement that at least four of the samples from each ingroup population, and one of the outgroup samples, had an accepted genotype call, resulting in 23,084,596 SNPs.

Principal Components Analysis

We used Eigenstrat SmartPCA [72] to investigate population structure and confirm sample identity. SNP Set 1 was used for this analysis.

Topology Weighting

To quantify genealogical relationships among taxa, we used topology weighting by iterative sampling of subtrees (*Twisst*) [46] (github.com/simonhmartin/twisst). This also made use of SNP Set 1. Genotypes for all samples were first phased and imputed using SHAPEIT v2 [73,74]. Neighbour-joining phylogenies were inferred for windows of 50 SNPs, following extensive simulations by Martin et al. [46]. Exact weightings were computed for all inferred genealogies that could be simplified to ≤ 2000 remaining sample combinations (see reference [46] for details). In cases where this was not possible, approximate weightings were computed by randomly sampling combinations of haplotypes until estimated weightings for all 15 possible topologies had a 95%

binomial confidence interval of <0.05. Confidence intervals were computed according to the Wilson method, as implemented by the package binom [75] in R [76].

Admixture proportions

We estimated admixture proportions for 100 kb windows using f_d [15], which is based on the so-called ABBA-BABA test [77,78]. This analysis was implemented using the python script fourPopWindows.py, available from github.com/simonhmartin/genomics_general. To ensure that f_d is not affected by confounding factors such as effective population size and selective sweeps, we first tested its performance in quantifying the proportion of admixture using coalescent simulations, and compared it to other methods used to study admixture and genomic divergence. We used msms [79] to simulate the evolution of independent windows of 50 kb, with a population recombination at rate of 1%. The models used, along with the range of effective population sizes and rates of gene flow tested are shown in Figure S6.

Analyses of real data focused on quantifying admixture between the two sympatric species pairs: *cyd* and *mel-W*, and *tim* and *mel-E*. f_d was computed using a range of combinations, including different allopatric control populations and either combining the two races that represent each broad geographic area ('East', 'West' and 'Guiana') or keeping them separate (Figure S7). SNP Set 2 was used for these analyses, with the added requirement that for the given run, at least 50% of samples in each population were genotyped and the outgroup was fixed for the ancestral state.

Recombination rate estimation

Recombination rates were estimated in two different ways. First, we used the high-resolution linkage maps recently produced for *H. melpomene*, *H. cydno* and hybrids [43] to estimate the local crossover rate. The recombination rate was computed as the slope of the locally weighted regression (loess span = 2 Mb) between physical position and map position along each chromosome [45,80]. Note that because recombination is male-limited in Lepidoptera, the values presented here represent the male-specific recombination rate. Conversion to an effective recombination rate at the population level would require knowledge of the effective sex ratio, which we don not have, so we chose here to use the male-specific rate. Second, we computed the population recombination rate (ρ) for 100 kb windows using the maximum likelihood method implemented in LDHelmet [81]. This analysis was run separately for each population of 20 samples (i.e. combining races from the same area following the results of the PCA in Figure 1), using SNP Set 1, phased as described above. A window size of 50 SNPs was used, along with the recommended range of pre-computed pairwise likelihoods. For convenience, ρ values were converted to cM/Mb by scaling values for each chromosome according to the map length of each chromosome, averaged across the three linkage maps used. As above, these values are therefore scaled to the male-specific recombination rate.

ACKNOWLEDGEMENTS

This work was funded by ERC grant SpeciationGenetics (339873) to CDJ. SHM was funded by a research fellowship from St John's College, Cambridge. CS was funded by Fondos Concursables Universidad del Rosario 2016-PIN-2017-001.This work made use of the Darwin Supercomputer of the University of Cambridge High Performance Computing Service (http://www.hpc.cam.ac.uk/), provided by Dell Inc. using Strategic Research Infrastructure Funding from the Higher Education Funding Council for England and funding from the Science and Technology Facilities Council. We

thank Dorothea Lindtke for contributing to extensive discussions and exploration of our results, and Sarah Barker for technical assistance. We are also grateful to Richard Merrill, Claire Mérot, Markus Möst, Steven Van Belleghem and Konrad Lohse for useful discussions that shaped this paper.

REFERENCES

1. Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, et al. Genome-wide evidence for speciation with gene flow in Heliconius butterflies. Genome Res. 2013;23: 1817–1828. doi:10.1101/gr.159426.113

2. Fontaine MC, Pease JB, Steele A, Waterhouse RM, Neafsey DE, Sharakhov I V, et al. Extensive introgression in a malaria vector species complex revealed by phylogenomics. Science. 2015;347: 1258524. doi:10.1126/science.1258524

3. Garrigan D, Kingan S. Genome sequencing reveals complex speciation in the Drosophila simulans clade. Genome Res. 2012;22: 1499–511. doi:10.1101/gr.130922.111

4. Dobzhansky TG. Genetics and the Origin of Species. Columbia University Press; 1937.

5. Gavrilets S. Fitness landscapes and the origin of species. Monographs in Population Biology. Princeton University Press; 2004.

6. Coyne JA, Orr HA. Speciation. Sunderland, MA: Sinauer; 2004.

7. Muller HJ. Isolating mechanisms, evolution and temperature. Biol. Symp. 1942. pp. 71–125.

8. Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, et al. Genomics and the origin of species. Nat Rev Genet. Nature Publishing Group; 2014;15: 176–92. doi:10.1038/nrg3644

9. Nadeau NJ, Whibley A, Jones RT, Davey JW, Dasmahapatra KK, Baxter SW, et al. Genomic islands of divergence in hybridizing Heliconius butterflies identified by large-scale targeted sequencing. Philos Trans R Soc B Biol Sci. 2012;367: 343–353. doi:10.1098/rstb.2011.0198

10. Poelstra JW, Vijay N, Bossu CM, Lantz H, Ryll B, Müller I, et al. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science. 2014;344: 1410–4. doi:10.1126/science.1253226

11. Malinsky M, Challis RJ, Tyers AM, Schiffels S, Terai Y, Ngatunga BP, et al. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. Science. 2015;350: 1493–8. doi:10.1126/science.aac9927

12. Charlesworth B. Measures of divergence between populations and the effect of forces that reduce variability. Mol Biol Evol. 1998;15: 538–43.

13. Cruickshank TE, Hahn MW. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol Ecol. 2014;23: 3133–3157. doi:10.1111/mec.12796

14. Burri R, Nater A, Kawakami T, Mugal CF, Olason PI, Smeds L, et al. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. Genome Res. 2015;25: 1656–1665. doi:10.1101/gr.196485.115

15. Martin SH, Davey JW, Jiggins CD. Evaluating the Use of ABBA-BABA Statistics to Locate Introgressed Loci. Mol Biol Evol. 2015;32: 244–257. doi:10.1093/molbev/msu269

16. Aeschbacher S, Selby JP, Willis JH, Coop G. Population-genomic inference of the strength and timing of selection against gene flow. Proc Natl Acad Sci. 2017;114: 7061–7066.

17. Wu C, Palopoli MF. Genetics of Postmating Reproductive Isolation in Animals. Annu Rev Genet. 1994;28: 283–308. doi:10.1146/annurev.ge.28.120194.001435

18. Rockman M V. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. Evolution. Society for the Study of Evolution; 2012;66: 1–17. doi:10.1111/j.1558-5646.2011.01486.x

19. Jiggins CD, Martin SH. Glittering gold and the quest for Isla de Muerta. J Evol Biol. 2017;30: 1509–1511. doi:10.1111/jeb.13110

20. Baird SJE. The impact of high-throughput sequencing technology on speciation research: maintaining perspective. J Evol Biol. 2017;30: 1482–1487. doi:10.1111/jeb.13099

21. Barton N. Multilocus Clines Author(s): N. H. Barton Source: Evolution. 1983;37: 454–471.

22. Barton N, Bengtsson BO. The barrier to genetic exchange between hybridising populations. Heredity. 1986;57 (Pt 3): 357–76.

23. Aeschbacher S, Bürger R. The effect of linkage on establishment and survival of locally beneficial mutations. Genetics. 2014;197: 317–336. doi:10.1534/genetics.114.163477

24. Janoušek V, Munclinger P, Wang L, Teeter KC, Tucker PK. Functional organization of the genome may shape the species boundary in the house mouse. Mol Biol Evol. 2015;32: 1208–1220. doi:10.1093/molbev/msv011

25. Juric I, Aeschbacher S, Coop G. The Strength of Selection against Neanderthal Introgression. PLoS Genet. 2016;12: 1–25. doi:10.1371/journal.pgen.1006340

26. Van Belleghem SM, Rastas P, Papanicolaou A, Martin SH, Arias CF, Supple MA, et al. Complex modular architecture around a simple toolkit of wing pattern genes. Nat Ecol Evol. Nature Publishing Group; 2017;1: 52. doi:10.1038/s41559-016-0052

27. Merrill RM, Gompert Z, Dembeck LM, Kronforst MR, McMillan WO, Jiggins CD. Mate preference across the speciation continuum in a clade of mimetic butterflies. Evolution. 2011;65: 1489–500. doi:10.1111/j.1558-5646.2010.01216.x

28. Jiggins CD, Naisbit RE, Coe RL, Mallet J. Reproductive isolation caused by colour pattern mimicry. Nature. 2001;411: 302–5. doi:10.1038/35077075

29. Mérot C, Salazar C, Merrill RM, Jiggins C, Joron M. What shapes the continuum of reproductive isolation? Lessons from Heliconius butterflies. Proc R Soc B Biol Sci. 2017;284: 20170335. doi:https://doi.org/10.1101/107011

30. Jiggins C. Ecological speciation in mimetic butterflies. Bioscience. 2008;58: 541–548.

31. Merot C, Frerot B, Leppik E, Joron M. Beyond magic traits: Multimodal mating cues in Heliconius butterflies. Evolution. 2015; 2891–2904. doi:10.1111/evo.12789

32. Darragh K, Vanjari S, Mann F, Gonzalez-Rojas MF, Morrison CR, Salazar C, et al. Male sex pheromone components in *Heliconius* butterflies released by the androconia affect female choice. PeerJ. 2017;5: e3953. doi:10.7717/peerj.3953

33. Jiggins CD, Estrada C, Rodrigues a. Mimicry and the evolution of premating isolation in Heliconius melpomene Linnaeus. J Evol Biol. 2004;17: 680–91. doi:10.1111/j.1420-9101.2004.00675.x

34. Naisbit RE, Jiggins CD, Mallet J. Disruptive sexual selection against hybrids contributes to speciation between Heliconius cydno and Heliconius melpomene. Proc Biol Sci. 2001;268: 1849–54. doi:10.1098/rspb.2001.1753

35. Jiggins CD, Linares M, Naisbit RE, Salazar C, Yang ZH, Mallet J. Sex-linked hybrid sterility in a butterfly. Evolution. SOC STUDY EVOLUTION; 2001;55: 1631–1638.

36. Merrill RM, Wallbank RWR, Bull V, Salazar PC a, Mallet J, Stevens M, et al. Disruptive ecological selection on a mating cue. Proc Biol Sci. 2012;279: 4907–13. doi:10.1098/rspb.2012.1968

37. Sánchez AP, Pardo-Diaz C, Enciso-Romero J, Muñoz A, Jiggins CD, Salazar C, et al. An introgressed wing pattern acts as a mating cue. Evolution. 2015;69: 1619–1629. doi:10.1111/evo.12679

38. Kronforst MRR, Hansen MEB, Crawford NGG, Gallant JRR, Zhang W, Kulathinal RJJ, et al. Hybridization reveals the evolving genomic architecture of speciation. Cell Rep. The Authors; 2013;5: 666–77. doi:10.1016/j.celrep.2013.09.042

39. Lohse K, Chmelik M, Martin SH, Barton NH. Efficient strategies for calculating blockwise likelihoods under the coalescent. Genetics. 2016;202: 775–786. doi:10.1534/genetics.115.183814

40. Pardo-Diaz C, Salazar C, Baxter SW, Merot C, Figueiredo-Ready W, Joron M, et al. Adaptive introgression across species boundaries in Heliconius butterflies. PLoS Genet. 2012;8: e1002752. doi:10.1371/journal.pgen.1002752

41. Heliconius Genome Consortium T. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature. 2012;487: 94–8. doi:10.1038/nature11041

42. Enciso-Romero J, Pardo-Díaz C, Martin SH, Arias CF, Linares M, McMillan WO, et al. Evolution of novel mimicry rings facilitated by adaptive introgression in tropical butterflies. Mol Ecol. 2017;26: 5160–5172. doi:10.1111/mec.14277

43. Davey JW, Barker SL, Rastas PM, Pinharanda A, Martin SH, Durbin R, et al. No evidence for maintenance of a sympatric Heliconius species barrier by chromosomal inversions. Evol Lett. 2017;1: 138–154. doi:10.1002/evl3.12

44. Nadeau NJ, Martin SH, Kozak KM, Salazar C, Dasmahapatra KK, Davey JW, et al. Genome-wide patterns of divergence and gene flow across a butterfly radiation. Mol Ecol. 2013;22: 814–26. doi:10.1111/j.1365-294X.2012.05730.x

45. Martin SH, Moest M, Palmer WJ, Salazar C, McMillan WO, Jiggins FM, et al. Natural selection and genetic diversity in the butterfly Heliconius melpomene. Genetics. 2016;203: 525–541. doi:10.1101/042796

46. Martin SH, Van Belleghem SM. Exploring evolutionary relationships across the genome using topology weighting. Genetics. 2017;206. doi:10.1534/genetics.116.194720

47. Kozak KM, Wahlberg N, Neild AFE, Dasmahapatra KK, Mallet J, Jiggins CD. Multilocus species trees show the recent adaptive radiation of the mimetic heliconius butterflies. Syst Biol. Cold Spring Harbor Labs Journals; 2015;64: 505–524. doi:10.1093/sysbio/syv007

48. Ahola V, Lehtonen R, Somervuo P, Salmela L, Koskinen P, Rastas P, et al. The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. Nat Commun. 2014;5: 1–9. doi:10.1038/ncomms5737

49. Davey JW, Chouteau M, Barker SL, Maroja L, Baxter SW, Simpson F, et al. Major Improvements to the Heliconius melpomene Genome Assembly Used to Confirm 10 Chromosome Fusion Events in 6 Million Years of Butterfly Evolution. G3. Genetics Society of America; 2016;6: 695–708. doi:10.1534/g3.115.023655

50. Baxter SW, Nadeau NJ, Maroja LS, Wilkinson P, Counterman BA, Beltran M, et al. Genomic Hotspots for Adaptation : The Population ¨ llerian Mimicry in the Heliconius Genetics of Mu melpomene Clade. Nachman MW, editor. PLoS Genet. Public Library of Science; 2010;6: e1000794. doi:10.1371/journal.pgen.1000794

51. Wallbank RWR, Baxter SW, Pardo-Diaz C, Hanly JJ, Martin SH, Mallet J, et al. Evolutionary Novelty in a Butterfly Wing Pattern through Enhancer Shuffling. PLoS Biol. PLoS; 2016;14: 1–16. doi:10.1371/journal.pbio.1002353

52. Nadeau NJ, Pardo-diaz C, Whibley A, Supple MA, Suzanne V, Richard W, et al. The gene cortex controls mimicry and crypsis in butterflies and moths. Nature. Nature Publishing Group; 2016;534: 106–110. doi:10.1038/nature17961

53. Reed RD, Papa R, Martin A, Hines HM, Kronforst MR, Chen R, et al. optix Drives the Repeated Convergent Evolution of Butterfly Wing Pattern Mimicry. Science. 2011;333: 1137–1142.

54. Martin A, Papa R, Nadeau NJ, Hill RI, Counterman BA, Halder G. Diversification of complex butterfly wing patterns by repeated regulatory evolution of a Wnt ligand. Proc Natl Acad Sci U S A. 2012;109: 12632–12637. doi:10.1073/pnas.1204800109/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1204800109

55. Roux C, Fraïsse C, Romiguier J, Anciaux Y, Galtier N, Bierne N. Shedding Light on the Grey Zone of Speciation along a Continuum of Genomic Divergence. Moritz C, editor. PLOS Biol. Oxford University Press; 2016;14: e2000234. doi:10.1371/journal.pbio.2000234

56. Mallet J, Besansky N, Hahn MW. How reticulated are species? BioEssays. 2016;38: 140–149. doi:10.1002/bies.201500149

57. Gante HF, Matschiner M, Malmstrøm M, Jakobsen KS, Jentoft S, Salzburger W. Genomics of speciation and introgression in Princess cichlid fishes from Lake Tanganyika. Mol Ecol. 2016;25: 6143–6161. doi:10.1111/mec.13767

58. W Wolf JB, Ellegren H. Making sense of genomic islands of differentiation in light of speciation. Nat Publ Gr. Nature Publishing Group; 2016;18: 87–100. doi:10.1038/nrg.2016.133

59. Ravinet M, Faria R, Butlin RK, Galindo J, Bierne N, Rafajlović M, et al. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. J Evol Biol. 2017;in press: 1450–1477. doi:10.1111/jeb.13047

60. Burri R. Interpreting differentiation landscapes in the light of long-term linked selection. 2017; 118–131. doi:10.1002/evl3.14

61. Yeaman S, Whitlock MC. The genetic architecture of adaptation under migration-selection balance. Evolution. 2011;65: 1897–1911. doi:10.1111/j.1558-5646.2011.01269.x

62. Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, et al. The genomic landscape of Neanderthal ancestry in present-day humans. Nature. 2014;507: 354–7. doi:10.1038/nature12961

63. Van Belleghem SM, Baquero M, Papa R, Salazar C, McMillan WO, Counterman BA, et al. Patterns of Z chromosome divergence among Heliconius species highlight the importance of historical demography. Mol Ecol. Cold Spring Harbor Laboratory; 2018;In Press. doi:10.1101/222430

64. Salazar CA, Jiggins CD, Arias CF, Tobler A, Bermingham E, Linares M. Hybrid incompatibility is consistent with a hybrid origin of Heliconius heurippa Hewitson from its close relatives, Heliconius cydno Doubleday and Heliconius melpomene Linnaeus. J Evol Biol. Blackwell Science Ltd; 2004;18: 247–256. doi:10.1111/j.1420-9101.2004.00839.x

65. Naisbit RE, Jiggins CD, Linares M, Salazar C, Mallet J. Hybrid Sterility, Haldane's Rule and Speciation in. Race. 2002;1526: 1517–1526.

66. Merrill RM, Van Schooten B, Scott J a, Jiggins CD. Pervasive genetic associations between traits causing reproductive isolation in Heliconius butterflies. Proc Biol Sci. 2011;278: 511–8. doi:10.1098/rspb.2010.1493

67. Masly JP, Presgraves DC. High-Resolution Genome-Wide Dissection of the Two Rules of Speciation in Drosophila. 2007;5. doi:10.1371/journal.pbio.0050243

68. Estrada C, Jiggins CD. Patterns of pollen feeding and habitat preference among Heliconius species. Ecol Entomol. 2002;27: 448–456. doi:10.1046/j.1365-2311.2002.00434.x

69. Lindtke D, Buerkle CA. The genetic architecture of hybrid incompatibilities and their effect on barriers to introgression in secondary contact. Evolution. 2015;69: 1987–2004. doi:10.1111/evo.12725

70. Roux C, Fraïsse C, Castric V, Vekemans X, Pogson GH, Bierne N. Can we continue to neglect genomic variation in introgression rates when inferring the history of speciation? A case study in a *Mytilus* hybrid zone. J Evol Biol. 2014;27: 1662–1675. doi:10.1111/jeb.12425

71. DePristo M a, Banks E, Poplin R, Garimella K V, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43: 491–8. doi:10.1038/ng.806

72. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38: 904–909. doi:10.1038/ng1847

73. Delaneau O, Zagury J-F, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods. 2013;10: 5–6. doi:10.1038/nmeth.2307

74. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nat Methods. 2012;9: 179–181. doi:10.1038/nmeth.1785

75. Dorai-Raj S. binom: Binomial Confidence Intervals For Several Parameterizations. 2014.

76. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2015.

77. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A draft sequence of the Neandertal genome. Science. 2010;328: 710–22. doi:10.1126/science.1188021

78. Durand EY, Patterson N, Reich D, Slatkin M. Testing for ancient admixture between closely related populations. Mol Biol Evol. 2011;28: 2239–2252. doi:10.1093/molbev/msr048

79. Ewing G, Hermisson J. MSMS: A coalescent simulation program including recombination, demographic structure and selection at a single locus. Bioinformatics. 2010;26: 2064–2065. doi:10.1093/bioinformatics/btq322

80. Corbett-Detig RB, Hartl DL, Sackton TB. Natural Selection Constrains Neutral Diversity across A Wide Range of Species. PLOS Biol. 2015;13: e1002112. doi:10.1371/journal.pbio.1002112

81. Chan AH, Jenkins PA, Song YS. Genome-Wide Fine-Scale Recombination Rate Variation in Drosophila melanogaster. 2012;8. doi:10.1371/journal.pgen.1003090

SUPPLEMENTARY INFORMATION

Table S1. Sample and sequencing information

sample ID	Sequence ID	Taxon	Country	Sex	Longitude	Latitude	Accession (ENA, NCBI)	Ref	Total Gbp	Mean depth
CAM025091	chi.CAM25091	H. cydno chioneus	Panama	f	9.1200	-79.7020	SAMEA104585050	[63]	7.05	25.71
CAM025137	chi.CAM25137	H. cydno chioneus	Panama	f	9.1200	-79.7020	SAMEA104585051	[63]	7.54	27.5
CAM000580	chi.CAM580	H. cydno chioneus	Panama	m	9.1200	-79.7020	SAMEA104585044	[63]	6.16	22.45
CAM000582	chi.CAM582	H. cydno chioneus	Panama	m	9.1200	-79.7020	SAMEA104585045	[63]	6.68	24.36
CAM000585	chi.CAM585	H. cydno chioneus	Panama	m	9.1200	-79.7020	SAMEA104585047	[63]	6.30	22.98
CAM000586	chi.CAM586	H. cydno chioneus	Panama	m	9.1200	-79.7020	SAMEA104585048	[63]	6.37	23.24
CAM000553	chi.CJ553	H. cydno chioneus	Panama	m	9.1714	-79.7573	SAMEA1919256	[1]	9.69	35.33
CAM000560	chi.CJ560	H. cydno chioneus	Panama	m	9.1714	-79.7573	SAMEA1919265	[1]	9.55	34.84
CAM000564	chi.CJ564	H. cydno chioneus	Panama	m	9.1714	-79.7573	SAMEA1919278	[1]	10.63	38.75
CAM000565	chi.CJ565	H. cydno chioneus	Panama	m	9.1714	-79.7573	SAMEA1919262	[1]	12.50	45.58
CS002242	zel.CS1	H. cydno zelinde	Colombia	m	3.9394	-77.3689	SAMEA104106540	[42]	9.88	36.04
CS001028	zel.CS1028	H. cydno zelinde	Colombia	m	3.9583	-77.3733	SAMEA104585054	[63]	8.94	32.59
CS001029	zel.CS1029	H. cydno zelinde	Colombia	m	3.9394	-77.3689	SAMEA104585055	[63]	6.70	24.44
CS001030	zel.CS1030	H. cydno zelinde	Colombia	m	3.9394	-77.3689	SAMEA104585056	[63]	7.35	26.8
CS001033	zel.CS1033	H. cydno zelinde	Colombia	m	3.9583	-77.3733	SAMEA104585057	[63]	7.73	28.18
CS001035	zel.CS1035	H. cydno zelinde	Colombia	m	3.9583	-77.3733	SAMEA104585058	[63]	8.32	30.34
CS002261	zel.CS2	H. cydno zelinde	Colombia	m	3.9394	-77.3689	SAMEA104106542	[42]	9.10	33.2
CS002262	zel.CS2262	H. cydno zelinde	Colombia	f	3.9583	-77.3733	SAMEA3670517	[42]	4.91	17.91
CS000273	zel.CS273	H. cydno zelinde	Colombia	m	3.9583	-77.3733	SAMEA104585059	[63]	7.11	25.93
CS002260	zel.CS30	H. cydno zelinde	Colombia	f	3.9394	-77.3689	SAMEA104106543	[42]	11.35	41.38
JM-09-313	thxn.JM313	H. timareta thelxinoe	Peru	m	-6.4584	-76.2877	SAMEA1919266	[1]	9.84	35.89
JM-09-57	thxn.JM57	H. timareta thelxinoe	Peru	m	-6.4528	-76.2987	SAMEA1919254	[1]	12.52	45.64
JM-09-84	thxn.JM84	H. timareta thelxinoe	Peru	m	-6.4528	-76.2987	SAMEA1919273	[1]	8.72	31.8
JM-09-86	thxn.JM86	H. timareta thelxinoe	Peru	m	-6.4528	-76.2987	SAMEA1919263	[1]	11.06	40.33
MJ12-3221	thxn.MJ12-3221	H. timareta thelxinoe	Peru	m	-5.6546	-77.6938	SAMEA104585110	[63]	6.74	24.58
MJ12-3233	thxn.MJ12-3233	H. timareta thelxinoe	Peru	m	-6.4519	-76.2985	SAMEA104585111	[63]	6.80	24.79
MJ12-3308	thxn.MJ12-3308	H. timareta thelxinoe	Peru	m	-5.6546	-77.6938	SAMEA104585112	[63]	6.57	23.95

MJ11-3339	txn.MJ11-3339	H. timareta thelxinoe	Peru	m	-5.6546	-77.6938	SAMEA104585113	[63]	7.25	26.43
MJ11-3340	txn.MJ11-3340	H. timareta thelxinoe	Peru	m	-5.6546	-77.6938	SAMEA104585114	[63]	6.34	23.12
MJ11-3460	txn.MJ12-3460	H. timareta thelxinoe	Peru	m	-5.6546	-77.6938	SAMEA104585115	[63]	10.55	38.46
CS002395	flo.CS12	H. timareta florencia	Colombia	m	1.7097	-75.6976	SAMEA104585100	[63]	7.77	28.35
CS002402	flo.CS13	H. timareta florencia	Colombia	m	1.7097	-75.6976	SAMEA104585101	[63]	7.41	27.03
CS002403	flo.CS14	H. timareta florencia	Colombia	m	1.7097	-75.6976	SAMEA104585102	[63]	8.34	30.4
CS002406	flo.CS15	H. timareta florencia	Colombia	m	1.7097	-75.6976	SAMEA104585103	[63]	8.30	30.27
CS002337	flo.CS2337	H. timareta florencia	Colombia	m	1.7108	-75.7089	SAMEA104585104	[63]	11.99	43.74
CS002338	flo.CS2338	H. timareta florencia	Colombia	m	1.7108	-75.7089	SAMEA104585105	[63]	6.48	23.63
CS002341	flo.CS2341	H. timareta florencia	Colombia	m	1.8136	-75.6686	SAMEA104585106	[63]	7.90	28.81
CS002350	flo.CS2350	H. timareta florencia	Colombia	m	1.7108	-75.7089	SAMEA104585107	[63]	6.73	24.53
CS002358	flo.CS2358	H. timareta florencia	Colombia	m	1.7108	-75.7089	SAMEA104585108	[63]	7.16	26.12
CS002359	flo.CS2359	H. timareta florencia	Colombia	m	1.7108	-75.7089	SAMEA104585109	[63]	6.33	23.07
CAM001841	ros.CAM1841	H. melpomene rosina	Panama	m	9.0760	-79.6590	SAMEA104585083	[63]	7.53	27.47
CAM001880	ros.CAM1880	H. melpomene rosina	Panama	m	9.0760	-79.6590	SAMEA104585084	[63]	8.39	30.59
CAM002045	ros.CAM2045	H. melpomene rosina	Panama	m	9.1103	-79.6907	SAMEA104585085	[63]	6.24	22.76
CAM002059	ros.CAM2059	H. melpomene rosina	Panama	m	9.1103	-79.6907	SAMEA104585086	[63]	7.05	25.7
CAM002519	ros.CAM2519	H. melpomene rosina	Panama	m	9.0109	-79.5477	SAMEA104585087	[63]	7.68	28
CAM002552	ros.CAM2552	H. melpomene rosina	Panama	m	9.0109	-79.5477	SAMEA104585088	[63]	6.53	23.83
CAM002071	ros.CJ2071	H. melpomene rosina	Panama	m	9.1206	-79.6969	SAMEA1919257	[1]	10.08	36.77
CAM000531	ros.CJ531	H. melpomene rosina	Panama	m	9.1206	-79.6969	SAMEA1919271	[1]	7.36	26.83
CAM000533	ros.CJ533	H. melpomene rosina	Panama	m	9.1206	-79.6969	SAMEA1919260	[1]	7.31	26.68
CAM000546	ros.CJ546	H. melpomene rosina	Panama	m	9.1206	-79.6969	SAMEA1919279	[1]	7.25	26.43
CS000710	vul.CS10	H. melpomene vulcanus	Colombia	m	3.9000	-76.6325	SAMEA3723391	[42]	9.76	35.59
CS003603	vul.CS3603	H. melpomene vulcanus	Colombia	m	3.5175	-76.7572	SAMEA104585091	[63]	8.30	30.25
CS003605	vul.CS3605	H. melpomene vulcanus	Colombia	m	3.5175	-76.7572	SAMEA104585092	[63]	7.55	27.55
CS003606	vul.CS3606	H. melpomene vulcanus	Colombia	m	3.5175	-76.7572	SAMEA104585093	[63]	7.31	26.66
CS003612	vul.CS3612	H. melpomene vulcanus	Colombia	m	3.5175	-76.7572	SAMEA104585094	[63]	9.07	33.06
CS003614	vul.CS3614	H. melpomene vulcanus	Colombia	m	3.5175	-76.7572	SAMEA104585095	[63]	7.49	27.33
CS003615	vul.CS3615	H. melpomene vulcanus	Colombia	m	3.5175	-76.7572	SAMEA104585096	[63]	6.44	23.5

CS003617	vul.CS3617	H. melpomene vulcanus	Colombia	m	3.5175	-76.7572	SAMEA104585097	[63]	6.92	25.22
CS003618	vul.CS3618	H. melpomene vulcanus	Colombia	m	3.5175	-76.7572	SAMEA104585098	[63]	6.92	25.24
CS003621	vul.CS3621	H. melpomene vulcanus	Colombia	m	3.5175	-76.7572	SAMEA104585099	[63]	6.25	22.8
CS001002	mal.CS1002	H. melpomene malleti	Colombia	m	1.8033	-75.6553	SAMEA104585067	[63]	6.27	22.85
CS001011	mal.CS1011	H. melpomene malleti	Colombia	m	1.8033	-75.6553	SAMEA104585068	[63]	7.76	28.29
CS001815	mal.CS1815	H. melpomene malleti	Colombia	m	1.8033	-75.6553	SAMEA104585069	[63]	5.82	21.21
CS002311	mal.CS21	H. melpomene malleti	Colombia	m	1.8136	-75.6686	SAMEA3723397	[42]	9.75	35.56
CS001286	mal.CS22	H. melpomene malleti	Colombia	m	1.6097	-75.6669	SAMEA3723398	[42]	8.78	32.01
CS001321	mal.CS24	H. melpomene malleti	Colombia	m	1.7506	-75.6319	SAMEA3723399	[42]	7.47	27.25
CS000586	mal.CS586	H. melpomene malleti	Colombia	m	1.8033	-75.6553	SAMEA104585071	[63]	6.55	23.87
CS000594	mal.CS594	H. melpomene malleti	Colombia	f	1.8033	-75.6553	SAMEA104585072	[63]	9.36	34.12
CS000604	mal.CS604	H. melpomene malleti	Colombia	m	1.8033	-75.6553	SAMEA104585073	[63]	6.99	25.48
CS000615	mal.CS615	H. melpomene malleti	Colombia	m	1.8033	-75.6553	SAMEA104585074	[63]	5.93	21.62
JM-11-160	ama.JM160	H. melpomene amaryllis	Peru	f	-5.6756	-77.6747	SAMEA1919261	[1]	12.08	44.05
JM-09-216	ama.JM216	H. melpomene amaryllis	Peru	m	-6.4685	-76.3533	SAMEA1919261	[1]	8.89	32.41
JM-11-293	ama.JM293	H. melpomene amaryllis	Peru	f	-6.4703	-76.3473	SAMEA1919277	[1]	14.67	53.5
JM-11-48	ama.JM48	H. melpomene amaryllis	Peru	m	-6.0960	-76.9774	SAMEA1919269	[1]	15.30	55.81
MJ11-3188	ama.MJ11-3188	H. melpomene amaryllis	Peru	m	-5.6728	-77.7195	SAMEA104585061	[63]	6.87	25.05
MJ11-3189	ama.MJ11-3189	H. melpomene amaryllis	Peru	m	-5.6728	-77.7195	SAMEA104585062	[63]	7.66	27.95
MJ11-3202	ama.MJ11-3202	H. melpomene amaryllis	Peru	m	-5.6745	-77.6711	SAMEA104585063	[63]	6.19	22.58
MJ12-3217	ama.MJ12-3217	H. melpomene amaryllis	Peru	m	-6.4547	-76.2994	SAMEA104585064	[63]	7.04	25.67
MJ12-3258	ama.MJ12-3258	H. melpomene amaryllis	Peru	m	-6.4530	-76.2876	SAMEA104585065	[63]	6.79	24.77
MJ12-3301	ama.MJ12-3301	H. melpomene amaryllis	Peru	m	-6.4528	-76.2862	SAMEA104585066	[63]	6.39	23.29
CAM001349	melG.CAM1349	H. melpomene melpomene	French Guiana	f	2.5222	-51.1934	SAMEA104585075	[63]	6.44	23.49
CAM001422	melG.CAM1422	H. melpomene melpomene	French Guiana	m	2.5222	-51.1934	SAMEA104585076	[63]	10.00	36.47
CAM002035	melG.CAM2035	H. melpomene melpomene	French Guiana	m	2.5222	-51.1934	SAMEA104585077	[63]	7.19	26.24
CAM008171	melG.CAM8171	H. melpomene melpomene	French Guiana	f	2.5222	-51.1934	SAMEA104585078	[63]	7.01	25.57
CAM008216	melG.CAM8216	H. melpomene melpomene	French Guiana	m	2.5222	-51.1934	SAMEA104585080	[63]	7.84	28.57
CAM008218	melG.CAM8218	H. melpomene melpomene	French Guiana	m	2.5222	-51.1934	SAMEA104585081	[63]	6.24	22.74
CAM013435	melG.CJ13435	H. melpomene melpomene	French Guiana	m	2.5222	-51.1934	SAMEA1919276	[1]	9.82	35.81

CAM009315	melG.CJ9315	H. melpomene melpomene	French Guiana	m	2.5222	-51.1934	SAMEA1919270	[1]	6.65	24.26
CAM009316	melG.CJ9316	H. melpomene melpomene	French Guiana	m	2.5222	-51.1934	SAMEA1919252	[1]	6.35	23.15
CAM009317	melG.CJ9317	H. melpomene melpomene	French Guiana	m	2.5222	-51.1934	SAMEA1919267	[1]	9.64	35.14
MJ09-4125	MJ09.4125	H. numata numata	French Guiana		4.0833	-52.6753	SAMEA3888884	[52]	6.31	22
MJ09-4184	MJ09.4184	H. numata silvana	French Guiana		4.0834	-52.6753	SAMEA3888889	[52]	7.71	27.11

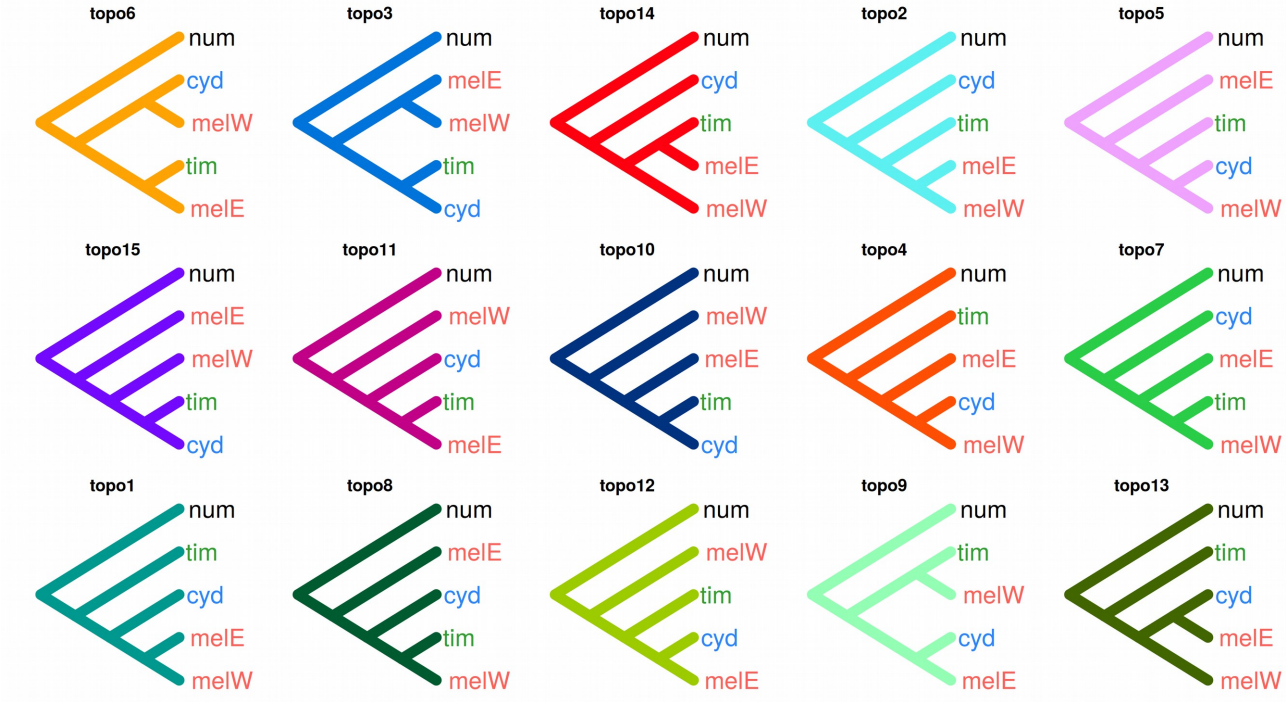


Figure S1. The fifteen possible topologies describing genealogical realtionships among *H. cydno* (*cyd*), *H. timareta* (*tim*), and *H melpomene* from the west (*mel-W*) and east (*mel-E*) of the Andes

The outgroup, *H. numata* (*num*), was used to polarize the genealogies. Topologies are arranged in order of their average weighting across the whole genome, from highest to lowest.

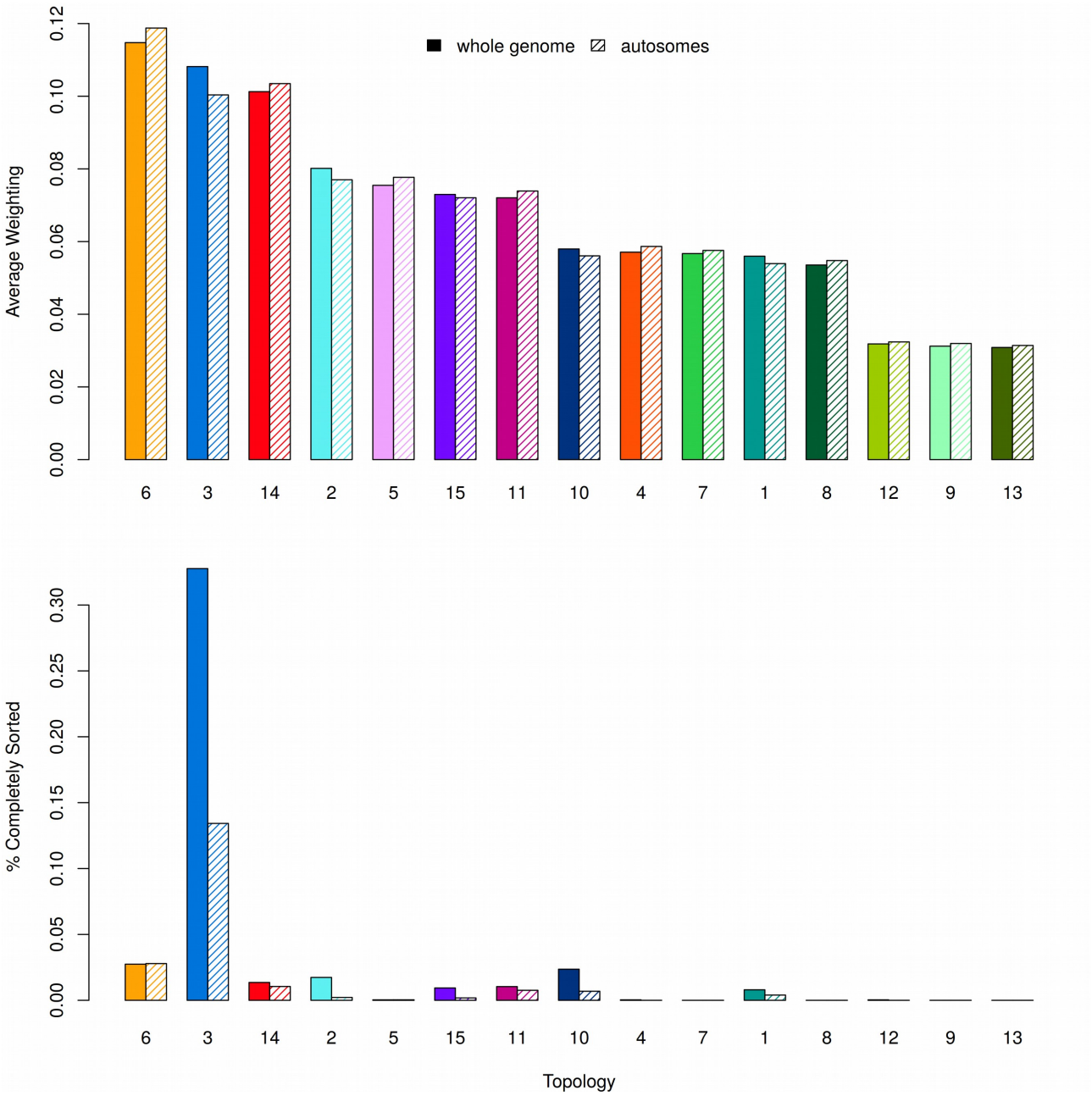


Figure S2. Average weightings and levels of sorting
Upper: Average genome-wide weightings for the 15 topologies in Figure S1, ordered accordingly. Corresponding weightings for autosomes only are also indicated. **Lower:** The percentage of windows in which the genealogy is completely sorted (i.e. has a weighting of 1).

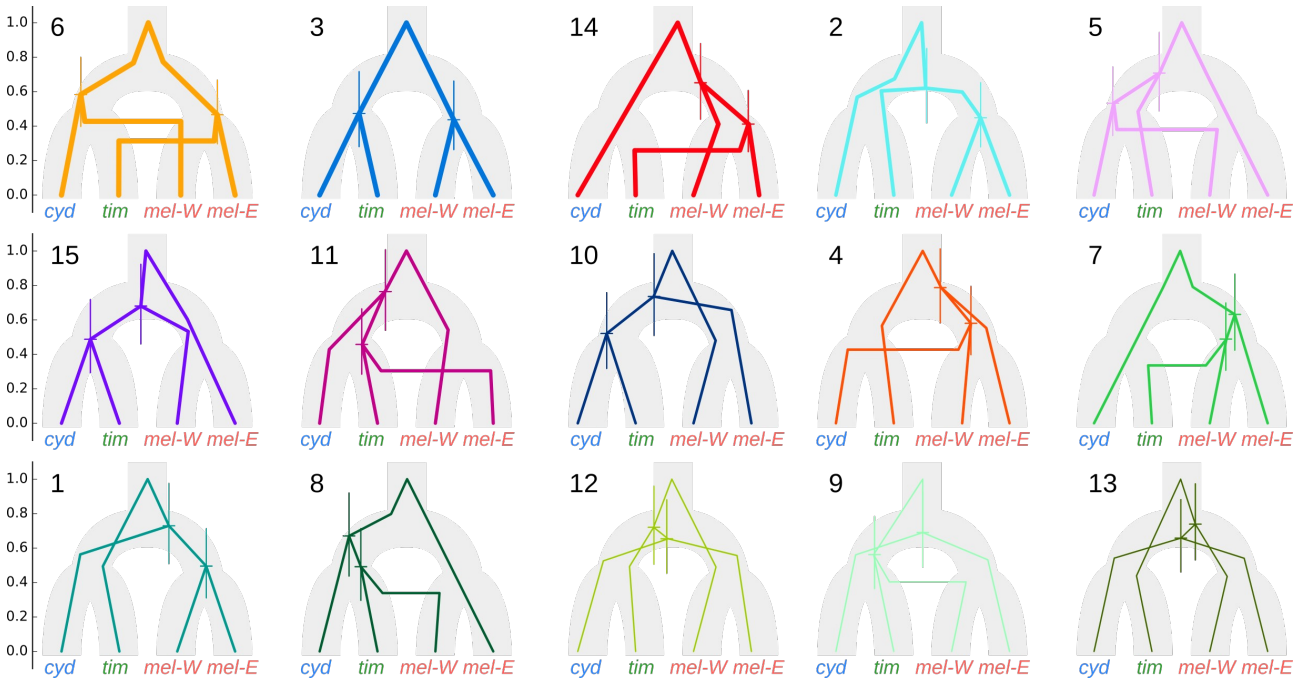


Figure S3. Average relative coalescence times for each topology
The fifteen possible topologies as in Figure S1, here illustrated as lineages within the presumed population branching tree. Topologies are ordered according to their average genome-wide weighting, and line widths are drawn proportionally. Coalescence times were calculated as the median branch length separating daughter taxa in each genealogy, scaled for each window by the branch length to the outgroup, to control for substitution rate variation among windows. The 25% and 75% quantiles for each split are indicated by vertical lines. Gene flow causing introgression is indicated where lineages cross between populations. Note that the each scenario shown here represents a hypothesis, and is one of several possible scenarios that can give rise to the same genealogy. These hypotheses were guided by the split times: more ancient split times are consistent with lineage sorting effects in the ancestral population (e.g. topology 13), whereas more recent split times are consistent with introgression (e.g. topology 14). It is important to note that the split times shown here represent the average across the whole genome and across multiple samples, and therefore represent the average over a range of different histories. Finally, because coalescence time will always pre-date the time of introgression, an arbitrary lag time is added prior to each introgression event. In reality, the length of this period depends on population size, and can therefore not be estimated with this technique. For this reason, these relative times of introgression between different taxa must be interpreted with caution.

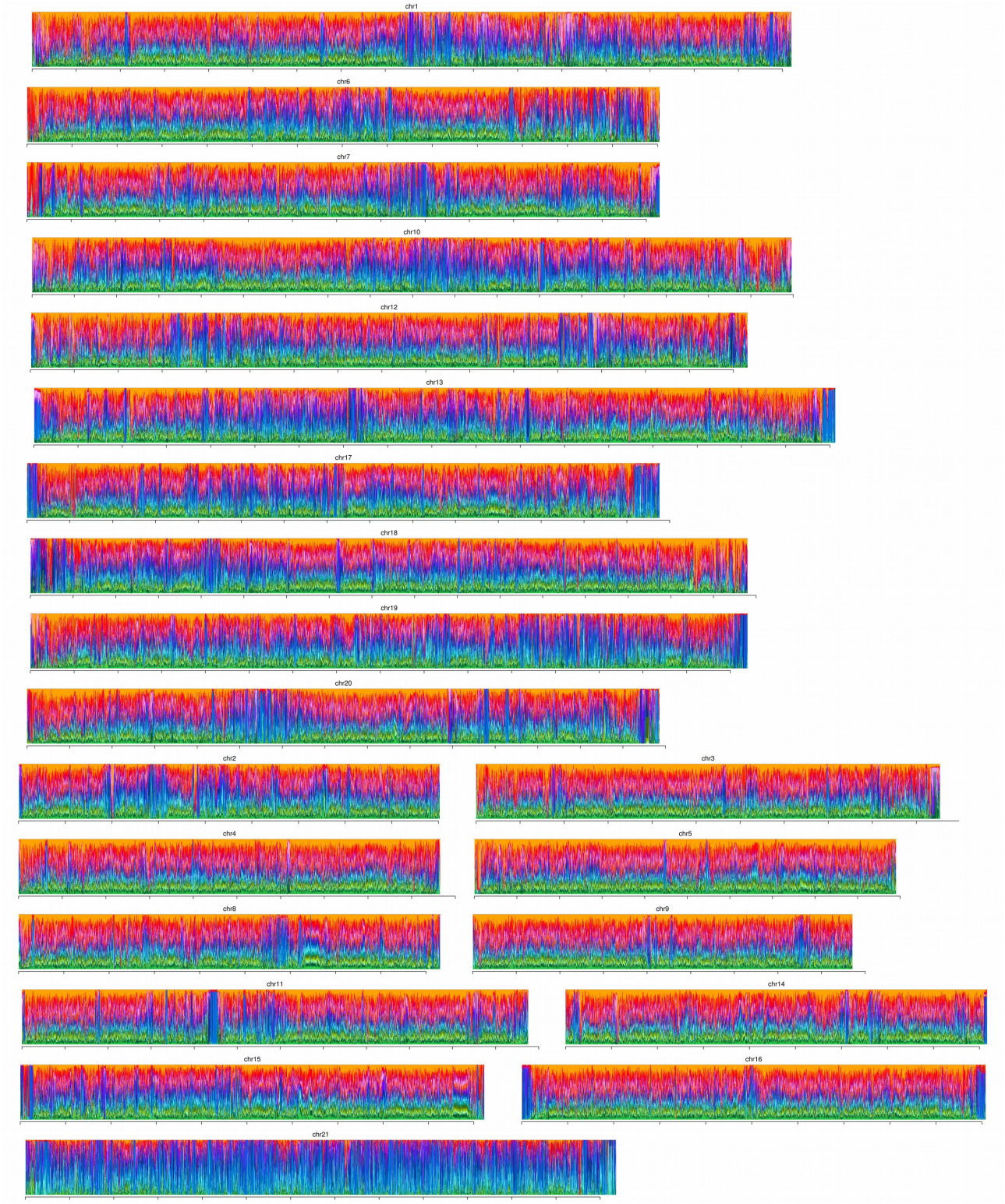


Figure S4. Raw weightings for all topologies across all chromosomes
See Figure S1 for the colour legend. Raw values without smoothing are plotted here, unlike in Figure 1 of the main paper. Weightings are stacked so that all 15 topologies can be distinguished. X-axis tick marks are spaced by 1 Mb.

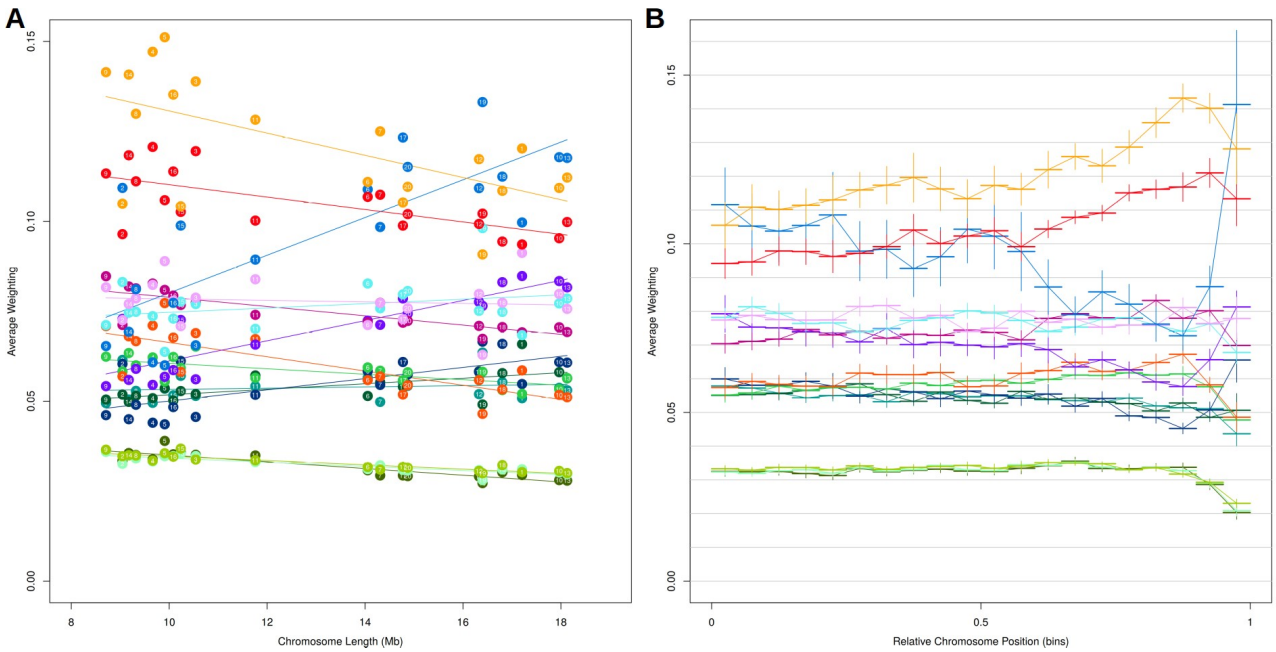


Figure S5. Heterogeneity in topology weightings among and within chromosomes
A. The average weighting for all 15 topologies (colours as in Figure S1) for each of the 20 autosomes, plotted against the physical length of the chromosome. Fitted linear regressions are shown for reference. **B.** Average weightings for all 15 topologies (coloured as in Figure S1) binned according to their relative chromosome position, from the centre (0) to the periphery (1). Each bin represents 5% of the chromosome arm, with the range indicated by a horizontal line. Vertical lines indicate +/- one standard error.

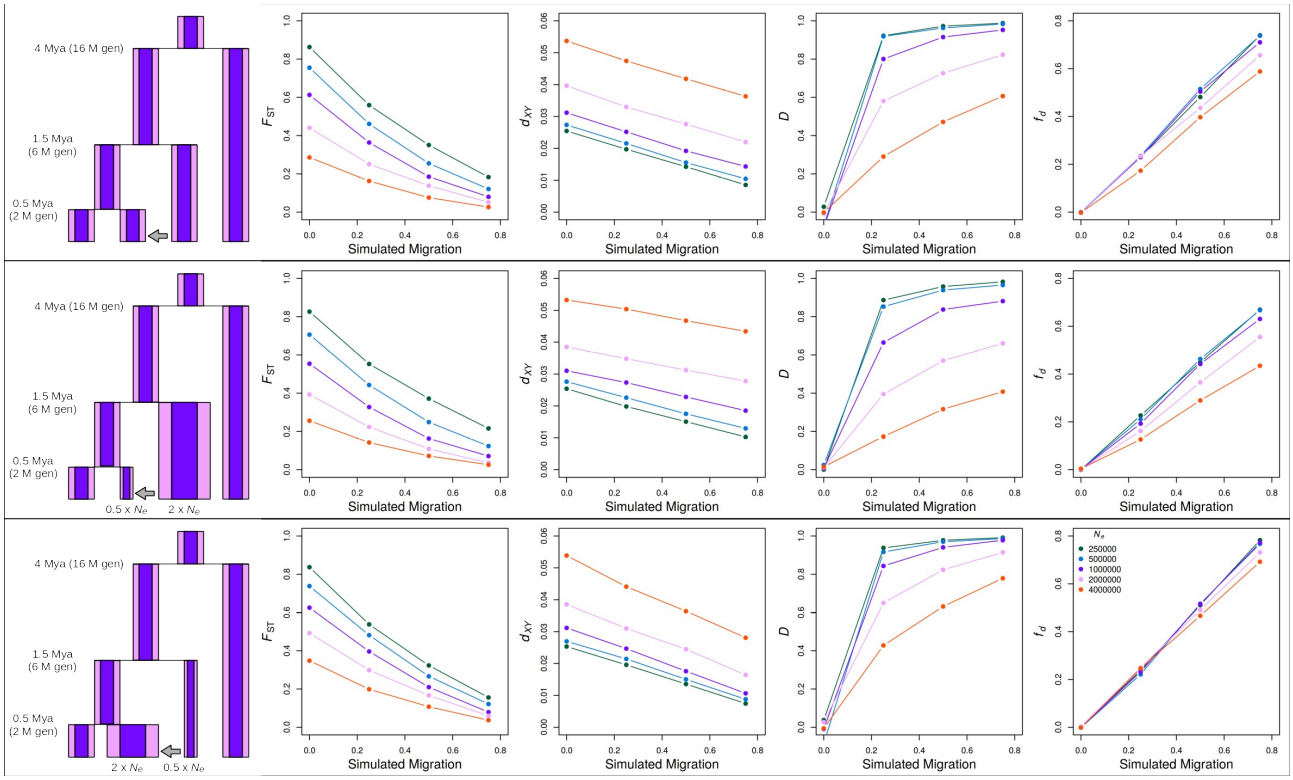


Figure S6. Testing the robustness of f_d to estimate the admixture proportion
Simulations show that f_d is largely robust to effective population size (N_e). Sequences were simulated following the model on the left, with a range of different population sizes, indicated by different colours (two are shown in the model for example). Simulated divergence times were chosen to approximate the splits between the outgroup silvaniform clade and the clade of *H. melpomene*, *H. cydno* and *H. timareta* (~4 million years ago, [47]), and the divergence between *H. melpomene* and the ancestor of *H. cydno* and *H. timareta* (1-1.5 Mya, [38,39]). Population sizes ranging from 250,000 to 4,000,000 were tested. For comparison, other divergence and admixture statistics are included. Relative and absolute divergence statistics F_{ST} and d_{XY} are both strongly dependent on N_e. Patterson's D statistic is strongly affected by N_e and is non-linear. By contrast, f_d is approximately proportional to the simulated level of migration, and is largely unaffected by N_e, except when N_e is large in which case f_d tends to underestimate the simulated admixture proportion. This is consistent with a loss of power with reduced lineage sorting in large populations. N_e for *H. melpomene* was estimated to be 2-3 Million [45], suggesting that admixture would indeed be weakly underestimated. However, as we are primarily interested in testing for reduced admixture in parts of the genome with reduced recombination rate, which usually corresponds to reduced N_e due to enhanced linked selection, the observed bias would have a conservative influence on our main analysis. We also tested simulated histories in which the donor and recipient populations undergo an expansion and contraction, respectively (second row), or the inverse (third row). Expansion of the donor population causes an exaggeration of the underestimate of admixture when N_e is large, but otherwise these changes don't have a significant effect on the performance of f_d.

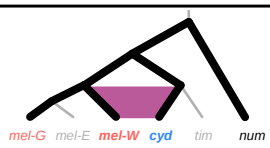
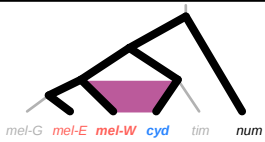
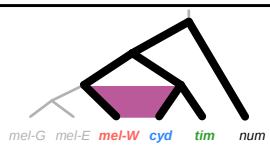
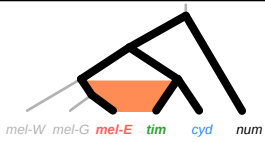
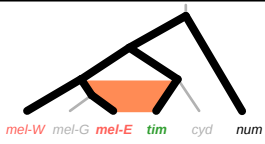
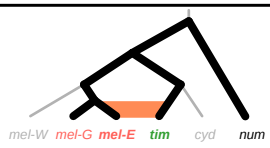
	P1 (allopatric)	P2 (sympatric)	P3 (sympatric)	
Set 1	*mel-G*	***mel-W*** (*ros* + *vul*)	***cyd*** (*chi* + *zel*)	
Set 1a	*mel-G*	***ros***	***chi***	
Set 1b	*mel-G*	***vul***	***zel***	
Set 2	*mel-E*	***mel-W*** (*ros* + *vul*)	***cyd*** (*chi* + *zel*)	
Set 3	***tim*** (*flo* + *txn*)	***cyd*** (*chi* + *zel*)	***mel-W*** (*ros* + *vul*)	
Set 4	*cyd* (*chi* + *zel*)	***tim*** (*flo* + *txn*)	***mel-E*** (*mal* + *ama*)	
Set4a	*cyd* (*chi* + *zel*)	***flo***	***mal***	
Set4b	*cyd* (*chi* + *zel*)	***txn***	***ama***	
Set 5	*mel-W* (*ros* + *vul*)	***mel-E*** (*mal* + *ama*)	***tim*** (*flo* + *txn*)	
Set 6	*mel-G*	***mel-E*** (*mal* + *ama*)	***tim*** (*flo* + *txn*)	

Figure S7. Sets of taxa used to estimate admixture proportions using f_d
Sets 1-3 were used to estimate admixture between *cyd* and *mel-W*. Sets 4-6 were used to estimate admixture between *tim* and *mel-E*. In each set, P2 and P3 represent the two sympatric populations between which the level of admixture is to be measured. P1 represents the allopatric 'control' population that is closely related to P2, but thought not to be subject to contemporary hybridisation with P3. The figure on the right shows, for each set, the relationships among the three populations considered (bold lines), as well as the period over which admixture between P2 and P3 can be detected given P1 (shaded). In all sets, H. numata (*num*) was used as the outgroup. '*ros*' = *H. m. rosina*, '*vul*' = *H. m. vulcanus*, '*mal*' = *H. m. malleti*, '*ama*' = *H. m. amaryllis*, '*mel-G*' = *H. m. melpomene* from French Guiana, '*chi*' = *H. c. chioneus*, '*zel*' = *H. c. zelinde*, '*flo*' = *H. t. florencia*, '*txn*' = *H. t. thelxinoe*.

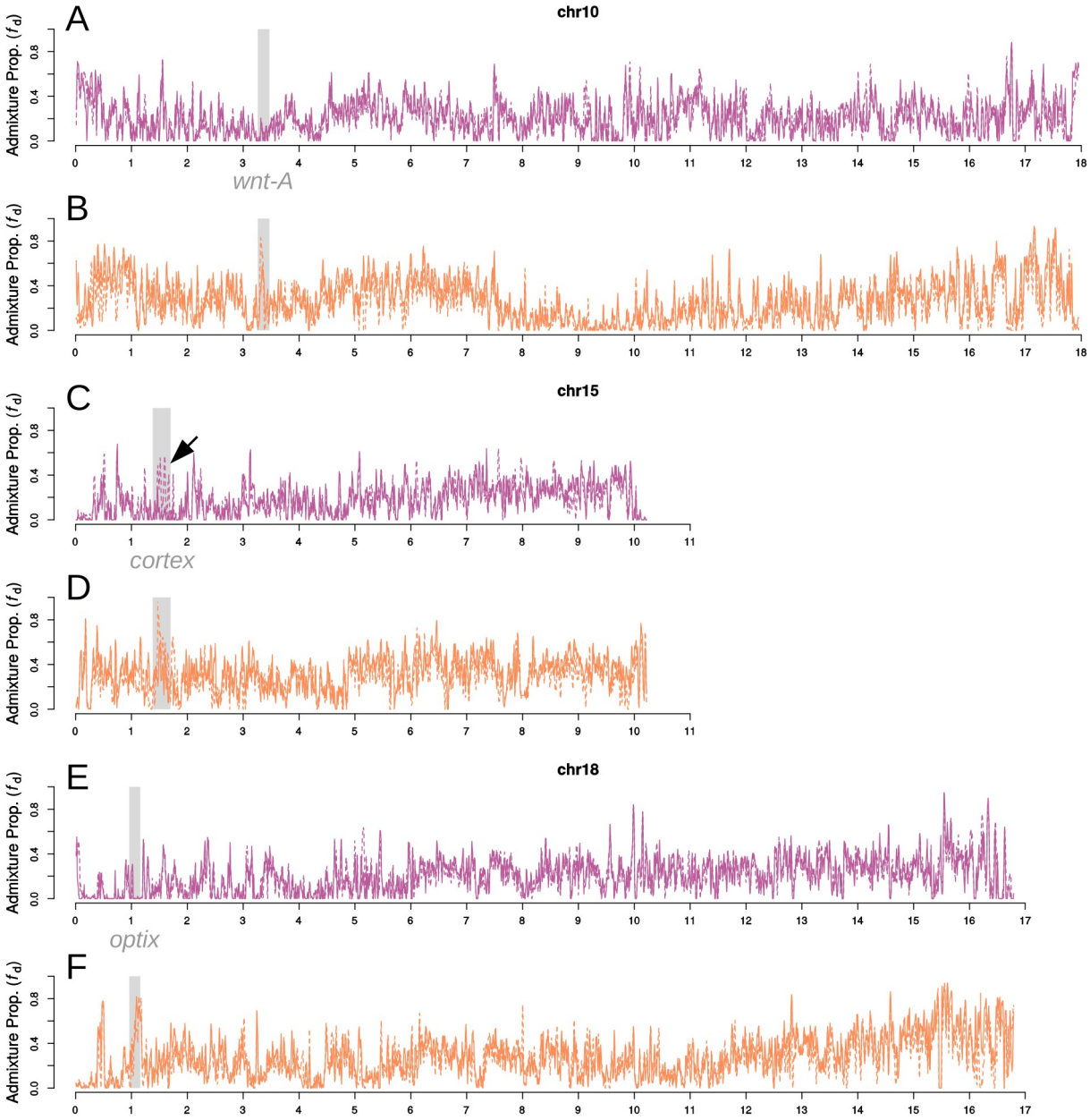


Figure S8. **Fine scale patterns of admixture across four chromosomes containing wing patterning loci**

Estimated admixture proportion (f_d) computed in 20 Kb slinding windows, sliding in increments of 5 Kb, plotted across three chromosomes carrying known wing patterning genes: *wnt-A* (chromosome 10), *cortex* (chromosome 15) and *optix* (chromosome 18). For A, C, and E, f_d was computed between cyd and mel-W using Set 1a (solid purple line) and Set 1b (dashed purple) (see Figure S7). For B, D, and F, f_d was computed between *tim* and *mel-E* using Set 4a (solid purple line) and Set 4b (dashed purple). In all cases, there is reduced admixture between *cyd* and *mel-W*, and elevated admixture between *tim* and *mel-E* in the vicinity of the patterning loci, consistent with a barrier to introgression in the former, but not the latter. The one exception is the cortex locus on chromosome 15, at which there is elevated admixture for Set 1b (i.e. between *H. cydno zelinde* and *H. melpomene vulcanus*, indicated by an arrow). This has in fact been previously recorded as a probable rare instance of introgression of a wing patterning allele between *H. cydno* and *H. melpomene* [42] . This allele appears to be responsible for the dorsal melanisation of the hindwing yellow bar in *H. m. vulcanus*. Therefore, these loci provide robust support for the use of f_d to quantify admixture between these taxa.

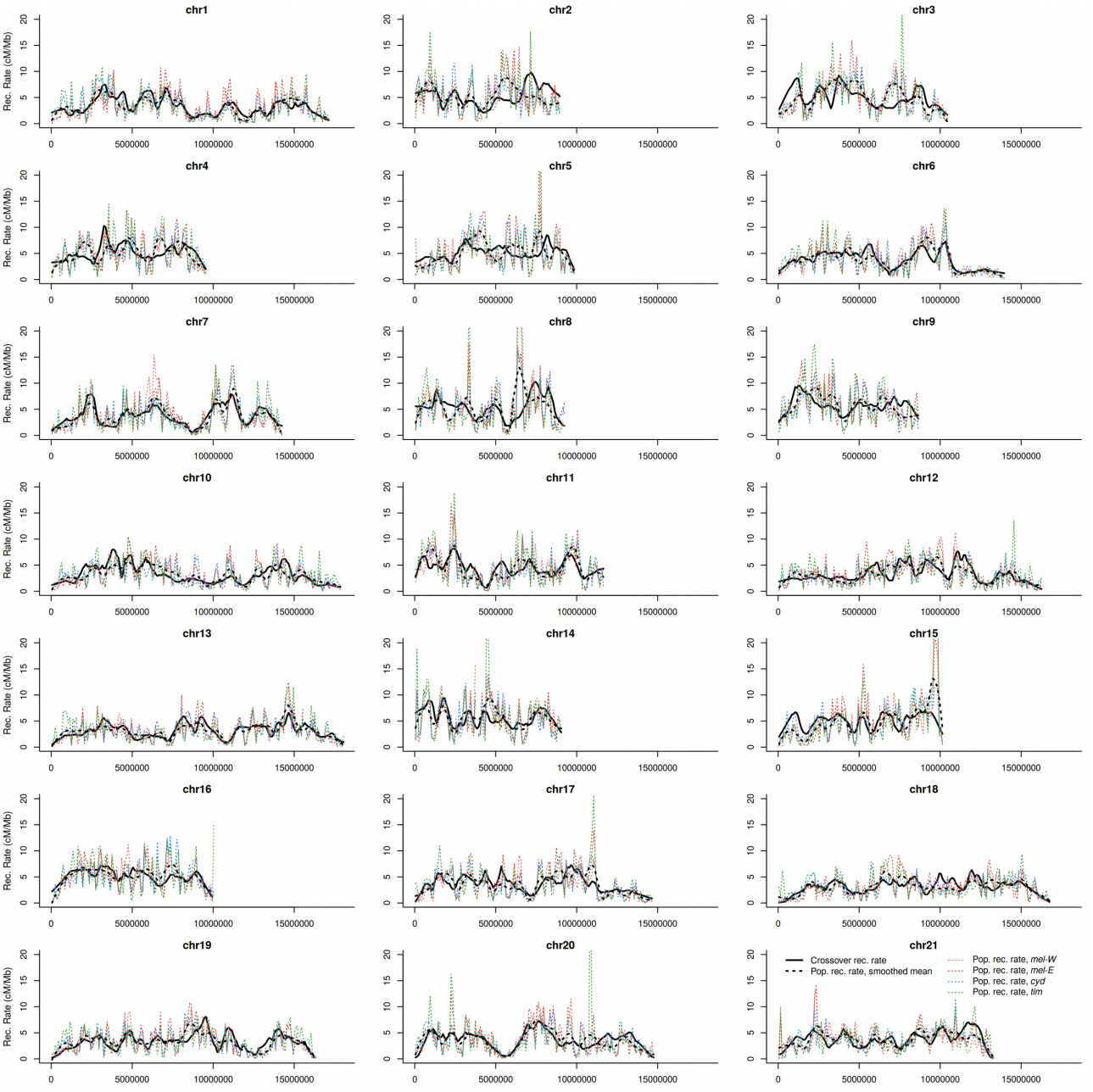


Figure S9. Recombination rates plotted across chromosomes
Solid lines show the crossover recombination rate estimated from linkage maps [43]. Dashed lines show the maximum likelihood estimate for the population recombination rate (ρ), computed for 100 Kb windows separately for *cyd*, *tim*, *mel-W* and *mel-E* (indicated by colours). The black dashed line indicates the mean ρ across the four populations, plotted as a locally-weighted average (loess span = 2 Mb).

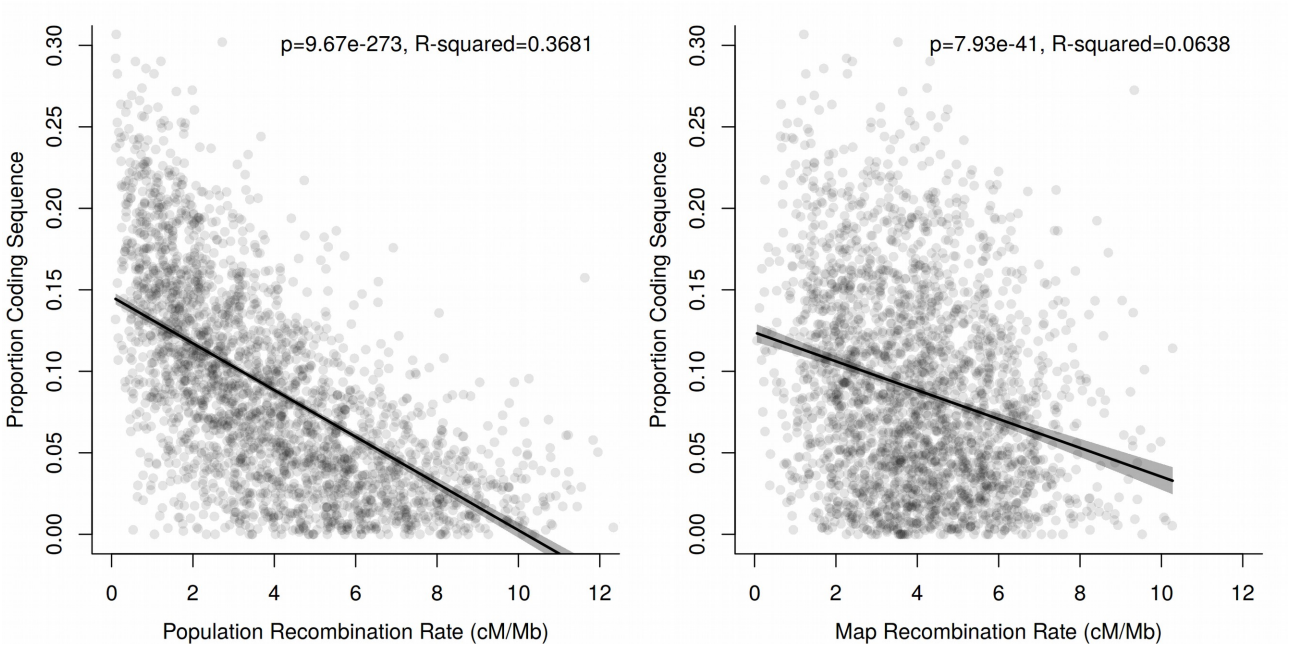


Figure S10. Relationship between recombination rate and gene density
Gene density (i.e. the proportion of coding sequence) in 100 Kb windows plotted against the population recombination rate (ρ) (left) and crossover recombination rate (right). The line shows a fitted linear regression. While there is a strong negative relationship between ρ and gene density, this may partly reflect the fact that ρ represents a composite of recombination and local effective population size, which will tend to be lower in regions of high gene density, due to linked selection [45]. Nevertheless, there is also a negative relationship between the crossover recombination rate and gene density, indicating that regions of lower recombination do indeed tend to harbour more coding sequence. The relationship is fairly weak, but it is unclear to what extent this might reflect the inaccuracies of measuring local recombination rates based on linkage mapping [43].

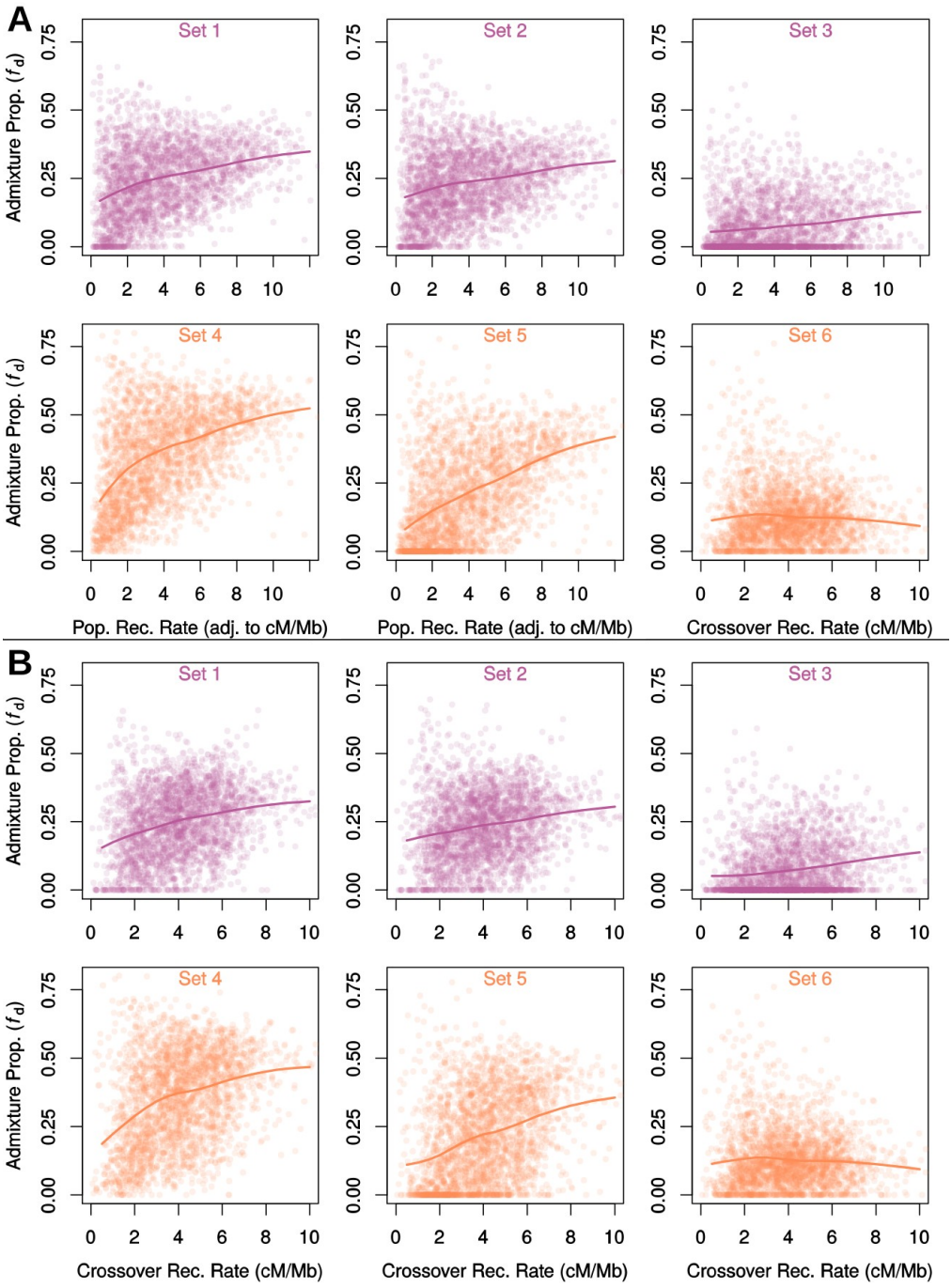


Figure S11. Admixture is positively correlated with recombination rate
Admixture proportions estimated for non-overlapping 100 Kb windows, plotted against the local recombination rate, either computed as the population recombination rate and rescaled to cM/Mb (**A**) or estimated directly from linkage maps (**B**). Solid lines indicate the locally-weighted average (loess span = 0.75). Dashed lines indicate the same average when windows in the outer third of each chromosome are excluded. Admixture between *cyd* and *mel-W* (Sets 1-3, see Figure S7) as well as that between *tim* and *mel-E* (Sets 4-6) increases non-linearly with increasing recombination rate, with the exception of Set 6, for which admixture proportions are low, and there is only evidence for a weak positive relationship in windows of low recombination rate. This may be driven by the close relationship and likely ongoing migration between *mel-E* and *mel-G* (see Figures 1 and S7), which could limit our ability to detect admixture between *tim* and *mel-E*. The estimated admixture proportion between *cyd* and *mel-W* using Set 3 is also much lower than for sets 1 and 2. This may be driven by strongly direction introgression from *cyd* into *mel-W*, which is also indicated by topology weightings, as described in the main paper. If introgression is largely in the direction from P2 into P3, f_d tends to underestimate the trough admixture proportion [15].

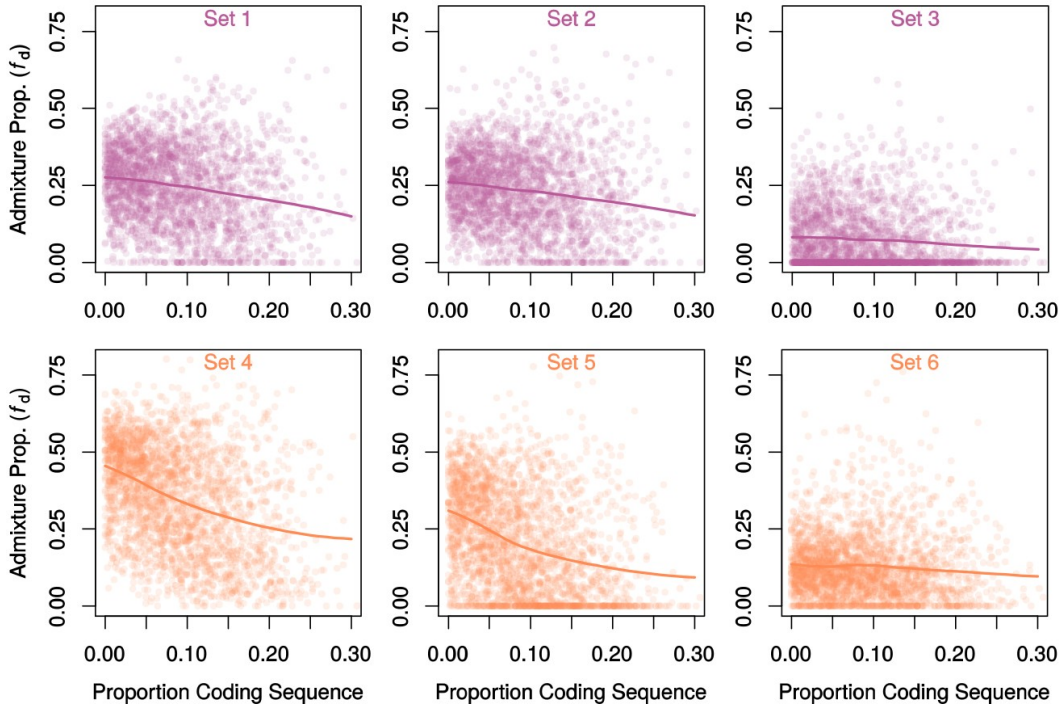


Figure S12. Admixture is negatively correlated with the proportion of coding sequence Admixture proportions estimated for non-overlapping 100 Kb windows, plotted against the proportion of coding sequence. Solid lines indicate the locally-weighted average (loess span = 0.75). Explanations for the lower average levels of admixture in Sets 3 and 6 are discussed in the legend of Figure S11 above.