Genomic regions of current low hybridisation mark long-term barriers to gene flow in scarce swallowtail butterflies

Sam Ebdon*, [1], Dominik R. Laetsch[1], Roger Vila[2], Stuart J.E. Baird[3, †], and Konrad Lohse[1, †]

[1]Institute of Ecology and Evolution, University of Edinburgh, Edinburgh, EH9 3FL, UK

[2]Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Passeig Marítim de la Barceloneta 37, ESP-08003, Barcelona, Spain

[3]Institute of Vertebrate Biology, Academy of Sciences of the Czech Republic, Brno, Czech Republic

∗ Corresponding author: se13@sanger.ac.uk

† These authors contributed equally to this work

Abstract

Many closely related species continue to hybridise after millions of generations of divergence. However, the extent to which current patterning in hybrid zones connects back to the speciation process remains unclear: does evidence for current multilocus barriers support the hypothesis of speciation due to multilocus divergence? We analyse whole-genome sequencing data to investigate the speciation history of the scarce swallowtails *Iphiclides podalirius* and *I. feisthamelii*, which abut at a narrow (∼25 km) contact zone north of the Pyrenees. We first quantify the heterogeneity of effective migration rate under a model of isolation with migration, using genomes sampled across the range to identify long-term barriers to gene flow. Secondly, we investigate the recent ancestry of individuals from the hybrid zone using genome polarisation and estimate the coupling coefficient under a model of a multilocus barrier. We infer a low rate of long-term gene flow from *I. feisthamelii* into *I. podalirius* the direction of which matches the admixture across the hybrid zone and complete reproductive isolation across ≈ 33% of the genome. Our contrast of recent and long-term gene flow shows that regions of low recent hybridisation are indeed enriched for long-term barriers which maintain divergence between these hybridising sister

species. This paves the way for future analysis of the evolution of reproductive isolation along the speciation continuum.

Author summary

Efforts to understand how new species evolve typically approach the problem through either: 1) investigating patterns of genetic exchange across 'hybrid zones' — where closely related species interbreed — or 2) modelling the demographic history of closely related species using geneological trees. Both approaches are capable of quantifying variation in genetic exchange, or 'gene flow', along the genome to identify regions of reproductive isolation; yet they leverage genetic signatures across vastly different timescales. The former exploits very recent signatures, while the latter averages long-term signatures over the history of divergence. Hence, we can compare and contrast each approach to understand how patterns of gene flow change across the speciation continuum. Here we use this strategy to capture the speciation dynamics of a pair of hybridising papilionid butterflies. Our results show that not only are these species continuing to produce hybrids after more than a million years since the onset of divergence, but there is a significant degree of concordance between patterns of gene flow observed along the genome across time scales.

Introduction

A fundamental goal of speciation research is to understand the genetic basis of reproductive isolation (RI) between diverging species and quantify the demographic and selective processes that lead to a build-up of RI (Campbell et al. 2018). We now know that speciation in the face of gene flow is not only possible (Bush 1994, Dieckmann & Doebeli 1999, Nosil 2008, Feder et al. 2012) but frequent (Taylor et al. 2001, Wang et al. 1997, Martin et al. 2013, Roux et al. 2016, Taylor & Larson 2019): closely related species often continue to hybridise after millions of generations of divergence (Mallet 1986, Harrison et al. 1990, Rieseberg et al. 1999, Sankararaman et al. 2014), yet remain distinct despite low levels of gene flow (Stankowski et al. 2021). However, given the time scales over which speciation events develop, the processes that contribute to RI are likely to vary over time (Drès & Mallet 2002). Stankowski & Ravinet (2021) define the speciation continuum as a continuum of reproductive isolation, from incipience to complete hybrid inviability, and highlight that the species we observe today are at different stages along this continuum. For example, sister species in many temperate plant and animal taxa form secondary contact zones (Hewitt 1988, Remington 1968) which may be substantially younger than the onset of speciation. In other words, contemporary contact zones are likely one of many instances of secondary contact generated by drastic environmental changes, such as glacial cycles over the last 800,000 years (Ebdon et al.

2021). Hybrid zones (HZs) for such taxa are stable exactly because there is strong selection against admixed ancestry. Whilst it is evident that genetic divergence and differentiation varies along the genome (Grainger Hunt & Selander 1973, Nosil et al. 2009), understanding the extent to which this reflects variation in the rate of gene flow requires explicit modelling of the interplay between migration, selection, genetic drift, and recombination.

Locally beneficial alleles may be selected against if they migrate into unfavourable environments and/or genetic backgrounds, reducing effective gene flow (Lenormand 2002). Over time, gene flow may vary as a consequence of range shifts and/or other changes in demography caused by glacial cycles, and – as a consequence – the selective forces and targets underpinning RI may also vary over time. Periods of complete allopatry during which gene flow is interrupted facilitate the build-up of strong endogeneous (Kruuk et al. 1999) or 'intrinsic' barriers (Bank et al. 2012). Genetic studies of HZs show that recent introgression may vary considerably along the genome (Payseur & Rieseberg 2016): some genome regions harbouring strong incompatibilities act as strong barriers and show steep clines (Barton & Bengtsson 1986, Yang et al. 2020). Steep clines in nature predicted (Macholán et al. 2011) the approximate location of the second mammalian hybrid sterility gene Hstx2 years before its identification in the lab (Bhattacharyya et al. 2014). In other regions introgression may be unimpeded by selection or even advantageous (Hedrick 2013, Arnold et al. 2008, Racimo et al. 2015). Importantly, the barriers associated with RI since the onset of divergence may differ from those acting during secondary contact, many of which may have arisen recently (Abbott et al. 2013, Ravinet et al. 2017). To date, very few studies have compared the targets of short-term selection against heterospecific ancestry in hybrids with long-term barriers to gene flow. Previous attempts to investigate the overlap of long-term barriers and hybrid-zone (HZ) introgression relied on comparing outliers of genetic differentiation with cline-based analyses and yielded mixed results: while several systems show strong associations between divergence and introgression (Gompert & Buerkle 2012, Gosset & Bierne 2013, Taylor et al. 2014, Yang et al. 2020), numerous other case studies revealed considerable heterogeneity (Teeter et al. 2008, Parchman et al. 2013, Kingston et al. 2017, Westram et al. 2021). The power of these approaches has been limited for two reasons: firstly, summary statistics of the SFS (such as F_{ST}) confound the effect of long-term barriers with other population genetic processes and, by focusing on the most extreme outliers, limit sample size and power. Similarly the power of clinal analyses has been limited for two reasons: extensive geographic sampling requirements and insufficient summary statistics; clines often being merely centered and scaled without allowing for variation in their shape and asymmetry.

The lack of comparisons of barriers to gene flow across recent and long-term timescales has been

highlighted in a recent review on speciation genomics (Ravinet et al. 2017). However, filling this gap requires quantitative inference frameworks that can distinguish barriers on both time scales from other evolutionary processes; frameworks that have only become available recently.

Speciation in *Iphiclides* butterflies

The southern European 'scarce swallowtails' *Iphiclides podalirius* and *I. feisthamelii* are large papilionid butterflies, typically associated with various species of *Prunus* and other Rosaceae bushes and trees. While *I. podalirius* ranges across the north of the Mediterranean from France to East Asia, *I. feisthamelii* is restricted to the Iberian peninsula and the northwest of Africa (Fig 1). Despite a long appreciation of the phenotypic differences between the two species (Godart 1842), including genital morphology (Coutsis & Van Oorschot 2011) and male ultraviolet wing patterns (Gaunet et al. 2019), their taxonomic status has been disputed after DNA barcoding at the mitochondrial (*mt*) *COX1* locus revealed that the two taxa share *mt* haplotypes (Wiemers 2003, Wiemers & Fiedler 2007). However, it has been suggested recently that the *mt* genealogy reflects an introgression sweep most likely linked to infection by *Wolbachia* (Gaunet et al. 2019). The two taxa diverged approximately 1.2 million years ago (Wiemers & Gottsberger 2010, Ebdon et al. 2021) and today, both species abut at a narrow (∼25 km) contact zone north of the Pyrenees (Descimon & Mallet 2009) (Fig 1). While potential hybrids have been diagnosed based on morphology in a small set of museum specimens from France (Lafranchis et al. 2015), the putative hybrid zone has not yet been characterised genetically. Here, we quantify introgression across the hybrid zone and estimate long-term effective migration rates (m_{e}) for this pair of sister species.

Aims and objectives

We investigate the history of speciation and quantify both long and short-term barriers between *I. podalirius* and *I. feisthamelii*. We use two recently developed minimal-assumption inference methods `gIMble` (Laetsch et al. 2023) and `diem` (Baird et al. 2023) on whole-genome sequencing (WGS) data to quantify long-term barriers and selection against recent introgression in the HZ, respectively. Firstly, we infer variation in the rate of long-term effective migration along the genome under an explicit demographic model and locate putative long-term barriers to gene flow with `gIMble`. Secondly, we use the genome polarisation framework `diem` to quantify recent introgression barriers using six putative hybrid individuals sampled from the *Iphiclides* hybrid zone (Fig 1) and characterize the multilocus barrier to ongoing hybridisation between these species. Finally,

we investigate the overlap between long-term barriers to gene flow and genomic regions that are depleted for recent introgression. We address the following questions:

1. What is the direction and rate of gene flow between *I. podalirius* and *I. feisthamelii* since their initial divergence?
2. What fraction of the *Iphiclides* genome acts as a long-term barrier to gene flow, and what are the properties of barrier regions?
3. What is the evidence for recent gene flow between the species across the hybrid zone and the strength of the multilocus barrier acting against it?
4. Is introgression across the hybrid zone impeded by long-term barriers to gene flow?

Results

Interspecific variation is consistent with hybridisation in *Iphiclides*

We generated WGS data for six individuals from the *Iphiclides* hybrid zone (we will refer to these as "the HZ set") in southern France (Fig 2 B) and 14 individuals sampled throughout the ranges of both species ("the non-HZ set", Fig 2 A). The non-HZ set includes six samples of *I. podalirius* and eight of *I. feisthamelii*.

Visualising genetic variation in a PCA reveals distinct clusters both within and between species. The first principle component (PC1) captures interspecies differentiation between *I. podalirius* and *I. feisthamelii* ($\approx$ 23% of the variation, Fig 2). Samples assigned morphologically to each species form two clusters along PC1 with individuals from the HZ (Fig 2) falling between the two parental species clusters (Fig 2). The North African *I. feisthamelii* samples are separated from the European *I. feisthamelii* samples along PC2 ($\approx$ 14% of the variation, Fig 2). Genetic diversity is slightly larger in *I. feisthamelii* compared to *I. podalirius* (Table 1).

We identified runs of homozygosity (ROH) using PLINK (v1.9) (Purcell et al. 2007) to estimate inbreeding via F_{roh}. One *I. podalirius* sample from Sicily (RVcoll12R048) was particularly inbred ($F_{roh} \approx 0.25$), and two other *I. podalirius* individuals, one from Romania (RVcoll14E561) and one from the hybrid zone (1325) were somewhat inbred ($F_{roh} \approx 0.076$ and 0.061 respectively). F_{roh} in the remaining samples (including all *I. feisthamelii*) was negligible (see Supplementary Table 1).

Table 1 – Estimates of genetic diversity (H, π, and θ), divergence (d_{xy}), and differentiation (F_{ST}) at intergenic (I) and fourfold degenerate (4D) sites between *I. feisthamelii* and *I. podalirius*. Estimates for 4D sites are taken from (Ebdon et al. 2021).

	H_{I}	H_{4D}	π_{I}	θ_{I}	$d_{\mathrm{xy,I}}$	$d_{\mathrm{xy,4D}}$	$F_{\mathrm{ST,I}}$	$F_{\mathrm{ST,4D}}$
I. feisthamelii	0.00676	0.00794	0.00682	0.00692	0.0244	0.0275	0.594	0.575
I. podalirius	0.00585	0.00521	0.00652	0.00659				

An IM model supports a history of speciation with gene flow from *I. feisthamelii* into *I. podalirius*

We use the composite-likelihood method `gIMble` (Laetsch et al. 2023) to infer long-term barriers. This analytic framework co-estimates variation in long-term m_e along the genome. Contrasting the support for a background IM model and a history without gene flow in genomic windows gives a measure for the strength of putative barriers. Since this analysis of long-term barriers relies on the assumption of a two population history, HZ and North African samples were excluded. We summarise genetic variation in short 64 base 'blocks', which are assumed to be non-recombining, under no direct selection, and to evolve with a constant mutation rate. These assumptions allow modelling the shared genealogical history of closely linked variants in the composite likelihood framework implemented in `gIMble`. To maximise the density of neutrally evolving sites we follow Laetsch et al. (2023) and restrict this analysis to intergenic sequence which have similar per site diversity and divergence to fourfold degenerate (4D) sites in coding regions (Table 1). After applying coverage based filters (see Methods), our analyses include 39% of the genome ($\approx$ 160 out of 408 Mb).

We fit a series of demographic models to the blockwise site-frequency spectrum (bSFS) of the whole genome: a model of strict divergence (DIV) and an IM model with migration in either direction with three N_{e} parameters ($IM_{3,\rightarrow\mathrm{pod}}$ and $IM_{3,\rightarrow\mathrm{fei}}$). Out of these three scenarios (Table 2), the best fitting history is an IM_3 model with unidirectional gene flow (forwards in time) from *I. feisthamelii* into *I. podalirius* ($IM_{3,\rightarrow\mathrm{pod}}$). Since an IM model will always fit at least as well as the (nested) DIV model, we compared the observed improvement in model fit ($\ln CL$) relative to the null distribution of $\ln CL$ which we obtained by simulating 100 data sets under the best fitting DIV history. This parametric bootstrap confirms that the $IM_{3,\rightarrow\mathrm{pod}}$ model does indeed fit significantly better than a DIV_3 history (Fig S4).

Under the background/global IM model, we infer a split time (T) 2,180,000 generations ago

(Fig 2, Table 2) and N_e estimates of 483,000, 377,000 and 1,140,000 for *I. feisthamelii* and *I. podalirius* and the ancestral population, respectively. This split time is close to previous estimates for the species pair (Ebdon et al. 2021). The long-term genome-wide rate of gene flow m_e from *I. feisthamelii* into *I. podalirius* is 4.73 $\times 10^{-8}$ which corresponds to $M/2 = 2N_e m_e = 0.046$ migrants per generation.

Extensive genome-wide reproductive isolation between *I. feisthamelii* and *I. podalirius*

We infer the effective migration rate m_e along the genome in sliding windows. For each window, we estimate parameters under the best fitting $IM_{3,\rightarrow \mathrm{pod}}$ history using a pre-computed-grid. Note that whilst the model and split time T are fixed globally, the remaining parameters (N_e for each population and m_e) are estimated locally, i.e. per window. The estimates of local N_e are approximately normally distributed (Fig S1) while local m_e estimates have a strongly leptokurtic distribution with a long tail up to the maximum value in the grid (Fig S1). While the large number of windows in the largest m_e bin reflects the fact that our grid of m_e estimates truncates the long tail of the m_e distribution, this does not affect our analyses of barriers.

Following Laetsch et al. (2023) we label windows as barriers to gene flow if a DIV history ($m_e = 0$) has greater marginal support than an IM history assuming the best fitting genome-wide value of m_e (Table 2), i.e. $\Delta_{B0} > 0$. We find that barriers to gene flow are widespread in the *Iphiclides* genome, making up 20% of windows across all chromosomes. Combining overlapping barrier windows, defines 555 barrier regions across all autosomes which cover 35% of the genome ($\approx$ 143Mb). The number of barrier regions varies widely between chromosomes: e.g. chromosomes 29 and 31 harbour no and one barrier region respectively. In contrast, chromosomes 19 and 17 have 53 (29.3% of windows) and 49 (29.5% of windows) barrier regions respectively. The average length of barrier region is $\approx$ 257 kb (median $\approx$ 200 kb) with a maximum of 1.4 Mb.

The genomic correlates of barrier regions

Models of local adaptation under migration selection balance predict a concentration of barriers in regions of low recombination and a positive correlation between m_e and recombination (Barton & Hewitt 1985). In line with this prediction we find that barriers and m_e are correlated with recombination rates over two different scales. First, the proportion of barrier sequence is positively correlated with chromosome length (Pearson's $\rho = 0.193$). Second, we observe that, on average,

barrier regions are closer to the centre of chromosomes (which generally have a reduced rate of recombination) than non-barrier regions (t-test, p = 0.0513). Interestingly, this is the opposite of the pattern observed in *Heliconius melpomene* and *H. cydno* (Laetsch et al. 2023). Contrary to expectation, we find that both the density of coding sequence (CDS) and repeats are reduced in barrier regions (Fig S3, see discussion).

Genome polarisation reveals complex hybrids

Chromosome-painting of individual genomes via diagnostic markers provides a direct way to visualize the mosaic ancestry of hybrid zone samples (Meier et al. 2021, Iasi et al. 2024, Saarman et al. 2019). While many have relied on assigning hybrid genotypes based on a reference panel which is assumed to reflect 'pure' ancestry, this assumption is both unnecessary and biases inference against introgression (Gompert et al. 2017). We labeled the ancestry of alleles at all non-singleton SNPs with respect to the focal barrier using the EM algorithm implemented in `diem` (Baird et al. 2023). This approach requires no *a priori* defined sets of reference individuals or candidate sites and assigns genotypes of 0/0, 1/1 or 0/1 at each variant position. All samples were included in the polarisation.

Genome polarisation shows that the HZ individuals (Fig 1 B) have an alternating pattern of *podalirius*—*feisthamelii* ancestry along all autosomes (Fig 3) which can only be generated by hybridisation after divergence. We find substantial heterogeneity in the degree of hybridisation, as quantified by the hybrid index (HI which is the average across genotype assignments in an individual) both between individuals and chromosomes (Fig 5). The two most intermediate HZ samples, IF_1325 ($HI = 0.712$) and IF_1322 ($HI = 0.420$), show large stretches of all three possible ancestries, that is, tracts that are heterozygous by source (interspecific H) or homozygous for each species respectively, a pattern that can only be generated by a complex history of hybridisation: the expected HI of a simple n-th generation backcross hybrid is $(1/2)^n$ with homozygous tracts for one species only. While we expect substantial variation around this expectation, successive generations of simple backcrosses move the expected HI towards the parental values of 0 or 1 along the sides of the De Finetti triangle, interspecific H (the proportion of genotypes with an allele from each side of the barrier) remains maximal. Thus, the relationship between the HI and interspecific H, i.e. the De Finetti diagram, shows that samples 1325 and 1330 are complex backcrosses (i.e. their ancestry does not reflect the simple successive crossing of hybrids with parental species), and suggests that in both cases introgression is towards *I. podalirius* (Fig 5A). All samples except 1322 show a deficit in

interspecific heterozygosity, the genomic Wahlund effect (Fig 5). Interestingly, we find that the six HZ individuals only exhibit one Z chromosome of mixed ancestry (1325), which suggests that the Z is a substantially greater barrier than autosomes of equivalent length, as predicted by numerous theoretical models (Charlesworth et al. 1987, Mank et al. 2009, Bachtrog et al. 2019, Turelli & Orr 1995).

Modelling the history of introgressed blocks in complex hybrids

If we make the simplifying assumption that introgression across the HZ into either species started at time T, occurs at a constant rate m, and only involves backcrosses (i.e. the recipient population is infinitely large), the lengths of admixture tracts depend only on the coupling coefficient $\theta = S/R$ (Barton & Hewitt 1985, Baird 1995). More precisely, the equilibrium solution for the gradient of the admixture tract length distribution on a log-log scale is $-\frac{3+\theta}{1+\theta}$ (see Fig 4, Appendix 1). Thus, to learn about the direction and dynamics of recent gene flow across the HZ, we fit the distribution of admixture tract length of colinear introgressed ancestry to the analytic expectation under this model (Baird 1995). In the absence of a recombination map, we measured the length of admixture tracts (x) relative to chromosome length, i.e. we assumed that each autosome has a map length of 0.25M, which corresponds to an average of one cross-over event per male meiosis (female butterflies are achiasmatic). To avoid block lengths being fragmented by rare errors we kernel smoothed `diem` labeling along chromosomes at a scale of $10^{-4}\times$ chromosome length.

Considering long tracts ($l > 0.04$), neutrality provides a good fit for the distribution of tracts of *I. feisthamelii* ancestry in *I. podalirius* (i.e. a gradient of -3). In contrast, the size distribution of *I. podalirius* admixture tracts in *I. feisthamelii* is best explained by $\theta = 0.44$ (Fig 4). This asymmetry mirrors the asymmetry of long-term gene flow under the IM history inferred by `gIMble` and suggests that gene flow has been stable for much of the last 150 - 275 generations (75 - 140 years). In contrast, over the same timescale *I. podalirius* admixture into *I. feisthamelii* has been strongly selected against. Interestingly, we find that the admixture tract length distribution in neither direction involves a simple asymptote for short tracts as predicted under a constant m secondary contact initiated at time T. This may be due to several factors: first, gene flow in the more distant past may have been genuinely lower, which — given the dependence of *Iphiclides* on orchards and other anthropogenic habitats — may reflect changes in landscape use. Secondly, our simplistic measure of x as a proportion of chromosome length is inadequate for short blocks; given that the recombination is known to vary substantially along butterfly chromosomes (Martin

Table 2 – Maximum composite likelihood parameters for three demographic models of species divergence. We use the $IM_{3,\rightarrow\text{pod}}$ model for barrier inference.

model	N_{fei}	N_{pod}	N_{anc}	T	m	$\ln CL$
DIV	459,000	416,000	1,240,000	1,810,000	-	-310,884,113
$IM_{3,\rightarrow\text{pod}}$	483,000	377,000	1,150,000	2,180,000	4.73E-08	-310,494,739
$IM_{3,\rightarrow\text{fei}}$	438,000	429,000	1,190,000	2,020,000	2.80E-08	-310,733,475

et al. 2019). Thirdly, our model is unrealistically simplistic in that it ignores space: it assumes introgression into an infinitely large panimictic population. In a spatially continuous population we expect shorter tracts with increasing distance from the HZ which is indeed what we observe in *Iphiclides*. Fourthly, violating our infinite population size assumption, hybrids in finite populations can interact, producing complex admixture tracts. This may inflate the distribution of small tracts in both admixture directions and explain the asymmetric deviation from expectation we see between fits. Finally, it is also plausible that the polygenic model of barrier architecture assumed by Baird (1995) breaks down over short scales, because it does not include the possibility that increasingly short blocks may have increasingly variable fitness effects (including the possibility of adaptive introgression).

Long-term barriers to gene flow are associated with regions of low hybridity

To estimate current barrier strength along the genome, we estimated D (Barton & Gale 1993), the multi-site mean pairwise LD in windows defined in the `gIMble` analysis. This is numerically equivalent to the variance in HI and captures the strong LD seen along stretches of co-introgressing variation: when minimal for polarised data ($D = 0$) states of sites are uncorrelated, and when maximal ($D = 0.25$) each sample reflects pure ancestry, three from either species, as expected under a strong central barrier.

We find that — at the scale of `gIMble` windows ($\approx$ 100 kb) — D is negatively correlated with $m_{\text{e,i}}$ (Pearson's ρ = -0.14). Furthermore, we found that D is increased for barrier windows (ANOVA, p $<$ 2e-16) compared to non-barrier windows. This suggests ongoing selection against hybrid ancestry within the HZ at long-term barrier loci (Fig 6).

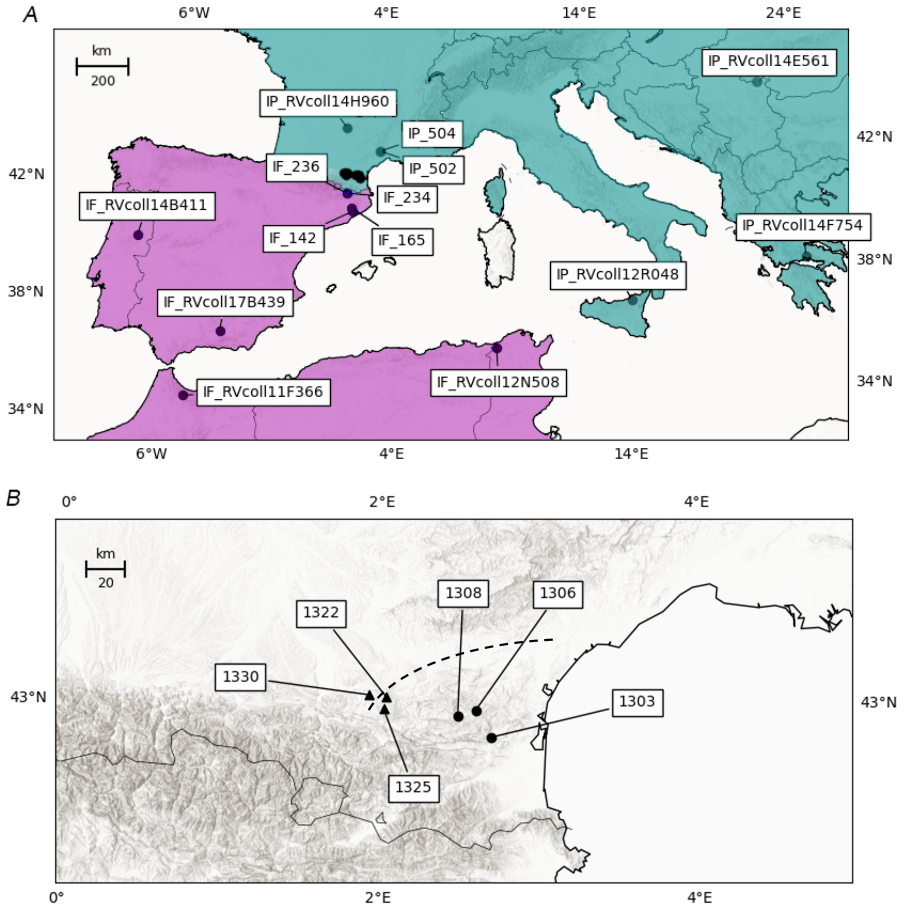

Figure 1 – (A) Sampling locations and ranges of *I. feisthamelii* (purple) and *I. podalirius* (teal) butterflies. The samples collected from the hybrid zone are shown in black. (B) Sampling locations of butterflies from the *Iphiclides* hybrid zone. The dashed line represents the approximate center of the hybrid zone, based on samples collected by Lafranchis et al. (2015). The circular samples resemble *I. feisthamelii*, and the triangular samples are intermediate hybrids.

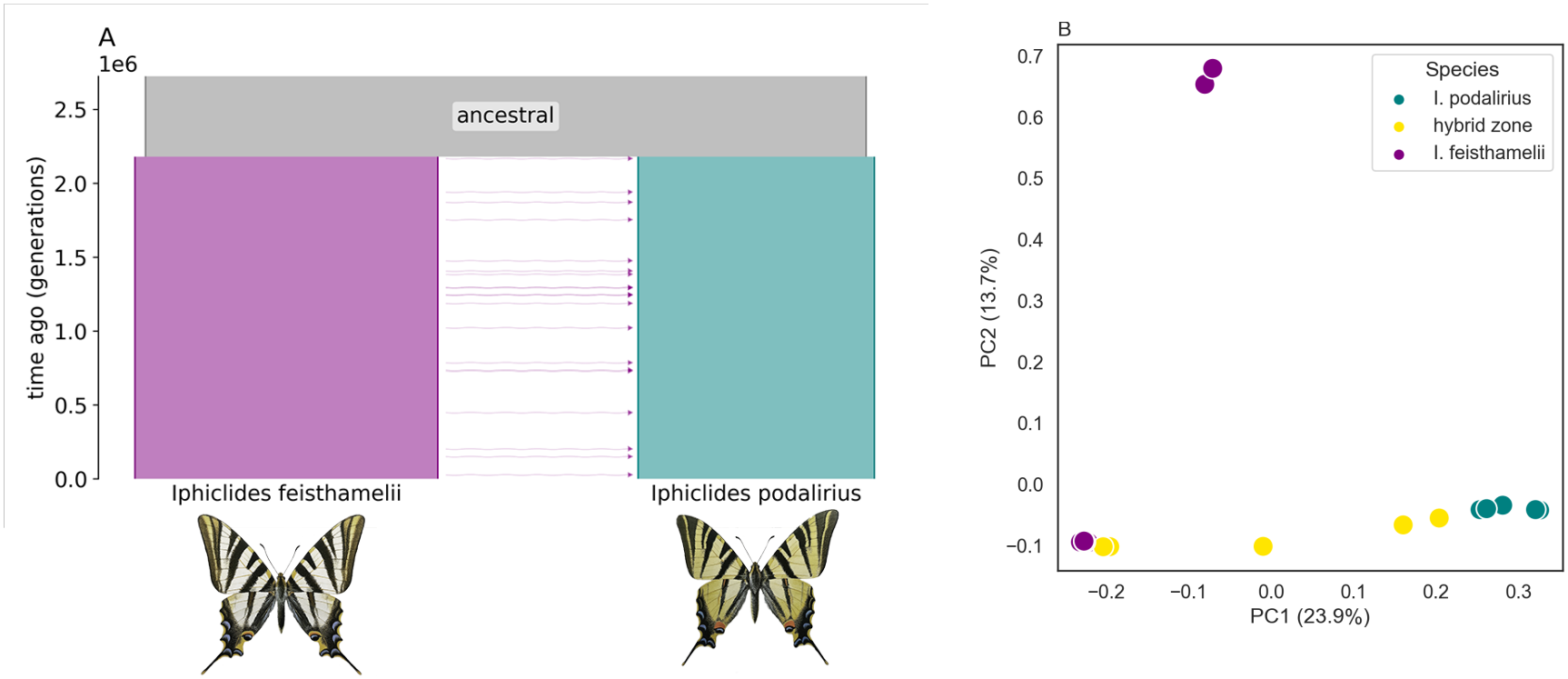

Figure 2 – (A) The background demographic history of species divergence and gene flow, the height and width of populations, is relative to the maximum likelihood estimates under the $IM_{2,\rightarrow pod}$ model (Table 2). This figure was generated using *demes* (Gower et al. 2022). (B) PCA of *Iphiclides* sampled across Europe; *I. feisthamelii* and *I. podalirius* samples are shown in purple and teal respectively. Samples from the HZ are shown in yellow. The first PC captures differences between the two taxa, and the second PC captures variation in geography, particularly the separation between North African and European *I. feisthamelii*.

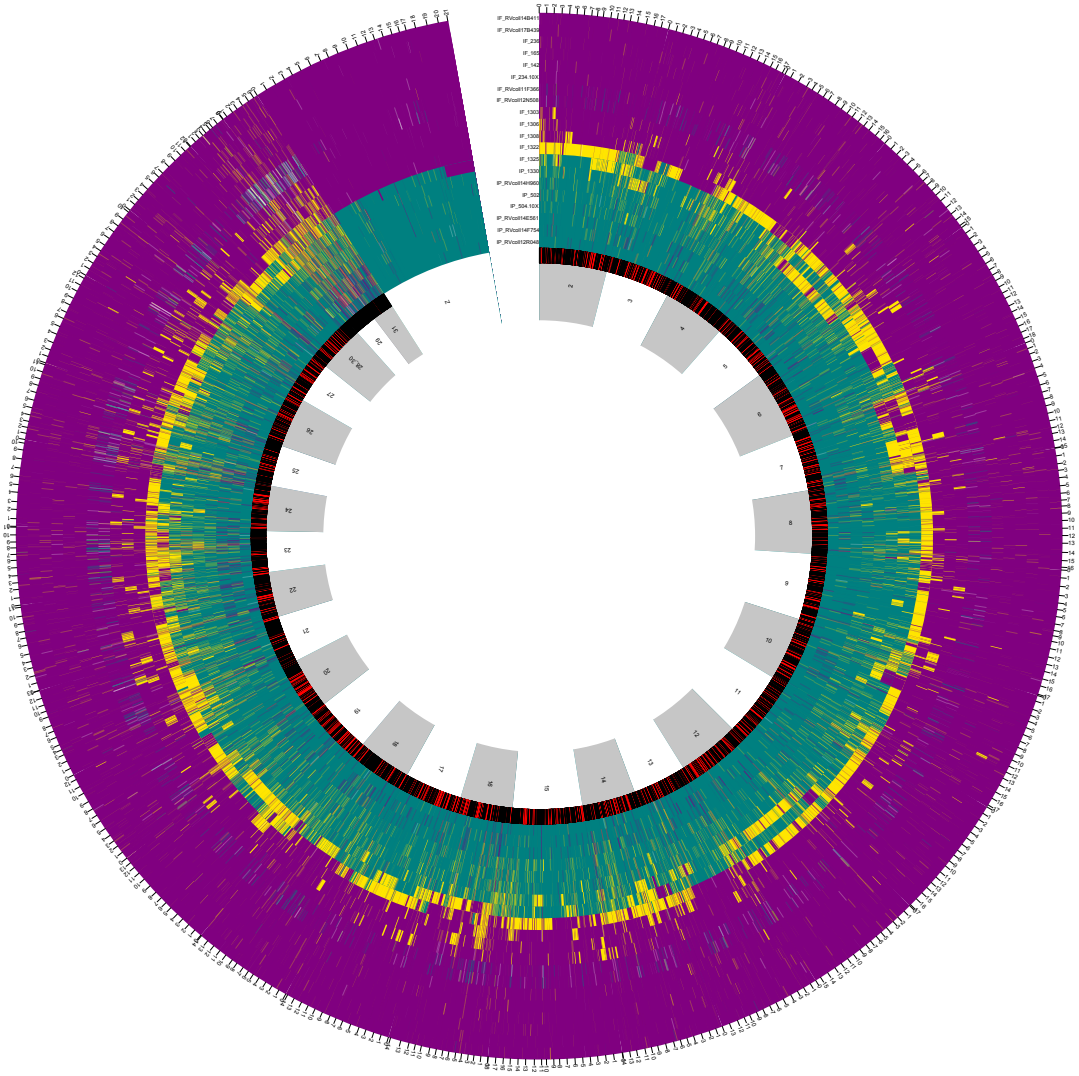

Figure 3 – Circular representation of the location of barriers identified using *gIMble* and the polarity of diagnostic markers for each sample across linkage groups. The inner ring shows the location of each chromosome in alternating grey and white. Moving outwards, the next ring indicates the location of barriers to gene flow (in red) and non-barrier regions (in black). The remaining rings show the genotype of each sample at each diagnostic marker. Teal bars are diagnostic of *I. podalirius*, purple bars are diagnostic of *I. feisthamelii*, and yellow bars are heterozygous for markers diagnostic of each species. The location of each megabase of sequence for each chromosome is indicated on the outside of the circle.

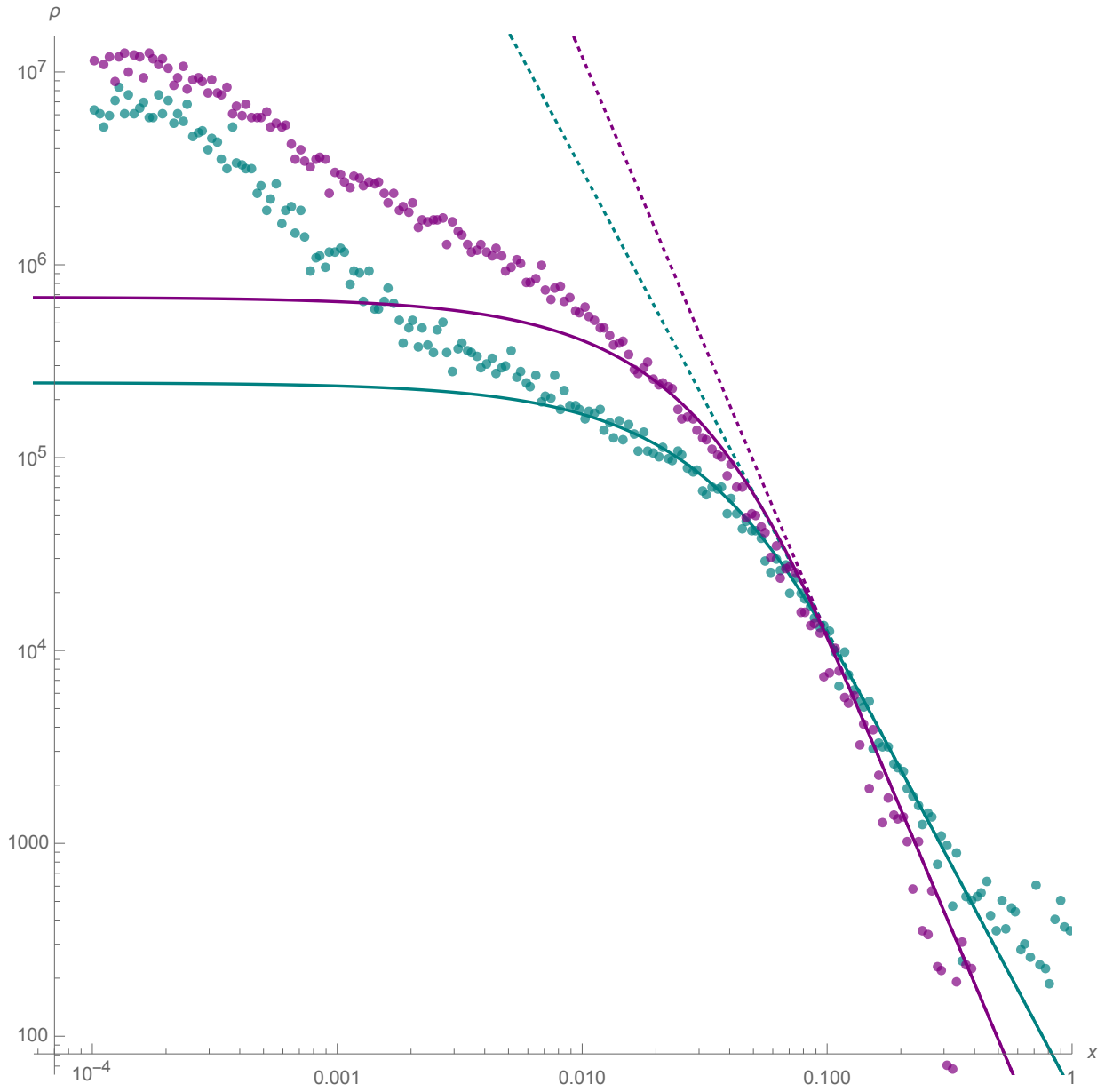

Figure 4 – Distribution of sizes of (purple points) *I. feisthamelii* tracts introgressing into *I. podalirius*, and (teal points) *I. podalirius* tracts into *I. feisthamelii*, on a log-log scale. Note, as most introgression is heterozygous, the introgressing tracts largely correspond to the yellow tracts in Figure 3. These are compared to theoretical predictions (solid lines) for exchange between two infinite demes (Appendix A-5). Introgression into *I. podalirius* is plotted for [M,S,R,T = 15, 0.11, 0.25, 150]. Introgression into *I. feisthamelii* is plotted for [M,S,R,T = 1.5, 0.0, 0.25, 275]. The difference in gradients on the right indicates stronger coupling (S/R) of the *I. podalirius* background despite more migrants M per generation. The distribution of small blocks towards the left does not match theory (see Results), making the time-since-contact estimates T lower bounds only.

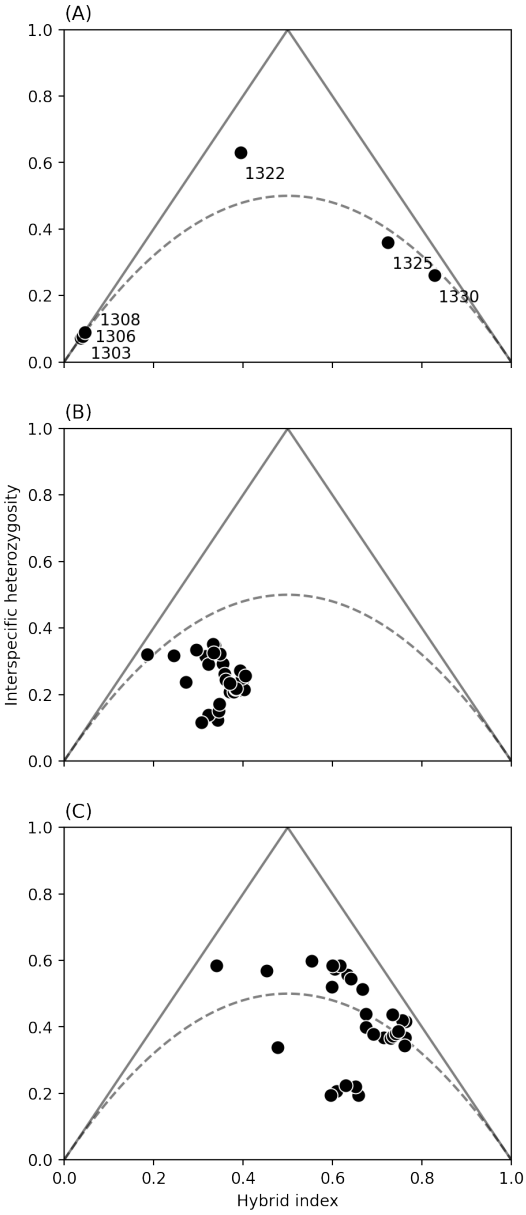

Figure 5 – Hybrid index (HI) versus interspecific heterozygosity (H) for samples collected from the *Iphiclides* hybrid zone (HZ). (A) Mean values for each HZ sample. (B) Mean values for each chromosome including all samples from the HZ. (C) Mean values for each chromosome excluding the three *I. feisthamelii*-like samples (1303, 1306 and 1308, $HI \approx 0$). The dashed line indicates the expectation under Hardy-Weinberg equilibrium.

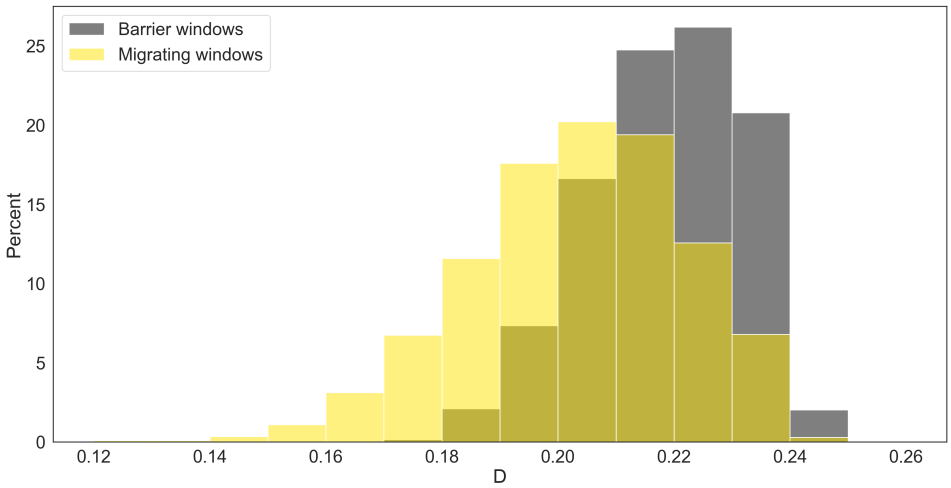

Figure 6 – The distribution of D across *gIMble* barrier windows (grey) and non-barrier windows (yellow). D calculated per window across HZ samples is greater within long-term barrier regions than in non-barrier regions (ANOVA, p < 2e-16).

Discussion

Much of the recent research on the genomics of speciation has focused either on fitting demographic models to estimate migration over timescales of N_e generations (Excoffier et al. 2013, Aeschbacher et al. 2017, Mondal et al. 2019, Fraïsse et al. 2021) or on investigating barriers to recent introgression (10s or 100s of generations) in hybrid zone and crossing studies (e.g. Buggs 2007, Teeter et al. 2008, Macholán et al. 2011, Turner & Harr 2014, Hagberg et al. 2022, Poikela et al. 2022). However, there are surprisingly few attempts to understand how barriers to gene flow on these two timescales are related. Here we have fitted explicit demographic models of speciation to infer the heterogeneity in long-term effective migration between two sister species of *Iphiclides* butterflies. We intersect this inference with signals of recent, heterogenous introgression across a hybrid zone estimated using genome polarisation.

Evidence for both long-term and recent gene flow between the scarce swallowtails

We find that the long-term demographic history of *I. podalirius* and *I. feisthamelii* is well approximated by an IM model with a low rate of unidirectional migration from *I. feisthamelii* into *I. podalirius*. Thus our analysis adds to the growing number of young species pairs for which a

signal of long-term and ongoing migration has been estimated, including great apes (Mailund et al. 2012), butterflies (Martin et al. 2013, Capblancq et al. 2019, Mackintosh et al. 2023), mollusks (Hirase et al. 2021), angiosperms (Papadopulos et al. 2011), birds (Reifová et al. 2016), *Drosophila* (Garrigan et al. 2012), and many more.

However, we note that the global effective migration rate (m_e) we infer in *Iphiclides* is considerably lower than estimates obtained for sister species pairs of *Heliconius* (Laetsch et al. 2023) and *Brenthis* (Mackintosh et al. 2023) butterflies, which are of comparable age (Fig S8). Although this low background level of genome-wide migration reduces the power to identify individual barrier regions, it still allows quantification of the variation of m_e along the genome. Furthermore, we find evidence for ongoing introgression between the two *Iphiclides* species across a HZ in the form of complex hybrids. Relating the distribution of admixture tract lengths to analytic expectations under a model of secondary contact (Baird 1995) suggests that recent gene flow into *I. podalirius* is neutral, while gene flow into *I. feisthamelii* is strongly selected against. This asymmetry is not only concordant with the inferred direction of long-term gene flow, but also consistent with genetic load arguments, which predict stronger selection against admixture tracts derived from the taxon with lower N_e (Harris & Nielsen 2016), i.e. *I. podalirius* in this case.

More generally, it is encouraging how much information about both recent and long-term introgression is contained in a small sample of individual genomes: one can locate barriers to gene flow over timescales of N_e generations by fitting models of speciation. Given that most species pairs do not have HZs, such long-term barrier inference is the only genomic information available about their speciation history. Even when HZs exist, they may not be amenable to clinal analyses due to sampling constraints. However, our analyses demonstrate that the coupling coefficient and the recent history of gene flow can be estimated from a handful of HZ individuals. It is perhaps surprising how well the admixture tract lengths we observe in *Iphiclides* fit analytic expectations under the simplest possible model of secondary contact which assumes panmixia, infinite population size and weak selection (Baird 1995). The fact that the observed length distribution deviates from this expectation for short tracts suggests that there may be additional information contained in small samples of hybrid genomes to fit more biologically realistic models of admixture. In particular, it would be interesting to fit more realistic barrier architectures that assumes a large but finite number of loci. Furthermore, classic theory on Fisher junctions for hybrid zones considers the decay of admixture tracts due to crossover events (Barton 1983). However, the internal decay of admixture tracts due to gene conversion events has so far been ignored and overlays the effects of an additional clock.

The ability to estimate the genomic distribution of block sizes (Fig 4) and relate it to theory (Baird 1995) developed long before genomic data were available is thrilling. The distribution of large blocks is expected to reach equilibrium quickly, and appears log-log linear suggesting relatively strong coupling of 0.44 at least in one direction. Neither direction reaches the 'tipping point' coupling of $S/R = 1$ (Barton 1983), where multilocus clines 'congeal', however that sharp threshold at equilibrium is misleading. Because small blocks equilibrate very slowly, it would take longer than an interglacial period for a sharp distinction in secondary contact outcome to be perceivable (Baird 1995) (Fig 2).

Long and short-term barriers to gene flow

The fact that barriers, inferred over these two time-scales of evolution, overlap shows that a significant subset of barriers is persistent and acts at very different points of the speciation continuum. This suggests that regions of the genome that maintain reproductive isolation between species in the long-term are also relevant in early species divergence. Indeed, in *Heliconius* butterflies it has been shown that the same wing pattern genes maintain species differences both across hybrid zones (Nadeau et al. 2014) and in deep time (Laetsch et al. 2023). However, in contrast the congruence in barrier landscapes across timescales we find in *Iphiclides* is not restricted to a small number of large effect genes but rather a genome-wide phenomenon, suggesting a polygenic barrier architecture.

We would argue that this internal comparison of long-term barrier landscape with short-term barriers is a more promising avenue for testing of models of speciation than comparisons/contrasts of species at different stages, which invariably differ in speciation history and barrier architecture (Johnson & Miyanishi 2008, Mérot et al. 2017, Stankowski & Ravinet 2021). Thus, an important direction for future work is to develop quantitative predictions for the temporal change in barrier landscapes. Under an allopatric null model, pairwise intrinsic incompatibilities are assumed to accumulate at random positions in the genome and genome-wide coupling is expected to increase quadratically with time. Under this model, we expect a limited amount of temporal barrier overlap which arises from the fact that barrier loci that establish early have a greater effect on long-term m_e than later barriers.

In contrast, verbal models of 'divergence hitchhiking' assume that early barriers expand locally (Wu 2001, Nosil 2008) and so predict a greater degree of temporal overlap. While theoretical models show that locally expanding barriers are possible — given sufficiently strong selection and linkage (Otto 2019) — there is so far not much empirical evidence that locally expanding 'islands

of speciation' are a common feature of speciation.

The architecture of barrier loci

We found substantial variation both in the size of barrier regions and in the proportion of barrier sequence per chromosome. Specifically, barriers to gene flow aggregate on large chromosomes and towards chromosome centres. Both are regions of the genome for which recombination rates are reduced, as we expect from barriers individually conferring small effects. This is consistent with previous research on *Heliconius* butterflies, which demonstrated that barriers to introgression are concentrated in regions of low recombination (Martin et al. 2019, Laetsch et al. 2023), and supports a polygenic barrier architecture. Such architecture was originally proposed as a null model of reproductive isolation (Barton & Charlesworth 1984) and evidence of polygenic barriers to gene flow has accumulated across a range of taxa (Szymura & Barton 1986, Macholán et al. 2007, Turner & Harr 2014, Morán & Fontdevila 2014). Thus, it appears on a broad scale that the architecture of reproductive isolation between these species is strongly linked to recombination rate heterogeneity.

It must be noted that the aforementioned low rate of background migration limits the power to detect m_e which is locally reduced but non-zero. As such, a subset of small barrier regions may simply be false positives.

We investigated whether barriers are associated with particular genomic features and failed to find enrichment of repetitive elements or coding sequence. However, one might expect this if reproductive isolation is driven by selection on genes within barrier regions. The deficit of coding sequence is likely a consequence of the strong negative correlation between gene density and chromosome length (Pearson's ρ = -0.421) on one hand and the strong positive correlation between barrier density and chromosome length on the other. These together may explain the strong negative correlation between gene and barrier density (Pearson's ρ = -0.652), which is partially driven by the numerous small chromosomes with very high numbers of genes and small numbers of barriers.

Limitations

To infer barriers to gene flow conservatively, we use the strictest possible threshold ($m_{\mathrm{e,i}} = 0$) and quantified the false positive rate (FPR) in a parametric bootstrap (see methods). Whilst this potentially excludes actual barriers with $m_{\mathrm{e,i}} > 0$ at putatively neutral flanking regions — which are the basis/input of our inference — this minimises the FPR.

Numerous factors may contribute to biases in our results (see Ravinet et al. (2017), and Laetsch et al. (2023) for *gIMble* limitations). Given that a directly estimated recombination map does not yet exist for *Iphiclides*, we cannot quantify the degree to which recombination rate heterogeneity contributes to migration rate variation nor account for the fact that uncertainty in estimates of both short and long-term barriers depends on the local recombination rate.

There may be scope to improve the power to detect weak gene flow through modelling more specific demographic scenarios. Firstly, whilst it is very likely that these species have undergone repeated rounds of separation and gene flow, we have fit a much simpler model that assumes a single continuous rate of migration. Expanding the *gIMble* framework to include isolation with initial migration (IIM) models may improve our power to detect ancient gene flow (Wilkinson-Herbots 2008, Laetsch et al. 2023). Secondly, we have modelled net gene flow as unidirectional and inferred the most likely direction by comparing models. In reality, net gene flow between these species may be bidirectional but asymmetric, a phenomenon which has been demonstrated in other systems (Yan et al. 2016, Ngeve et al. 2017, Banker et al. 2022). Whilst fitting more realistic models of gene flow is not currently feasible using this inference framework, future extensions lie on the horizon (Bisschop 2022).

The *diem* genome polarisation algorithm is designed to work off even small amounts of low quality data. There should therefore be few power limitations working with high quality genome scale data, and indeed estimates of hybrid indices, interspecific heterozygosity, and visualisation of admixture have high precision. Estimating admixture tract size distributions requires a further level of precision however: a single ancestry state error in a long block will on average halve its estimated length, error due to a miscalled variant would affect one tract, but error due to a mispolarised site could affect many. To avoid such error driven tract fragmentation, we filter polarised sites for high diagnostic index — strongly correlated with the support for correct polarity — and introduce an arbitrarily chosen scale of kernel smoothing of state along chromosomes. This has an advantage over HMM approaches in that it has no starting point chirality, but the disadvantage that it censors true small tract signal with some distribution of false negatives. This likely contributes to the departure of small tract observations from simple expectations (Fig 4).

Conclusion

We have demonstrated an association between long-term and short-term reproductive isolation in a species pair of swallowtail butterflies. Despite being several million generations old, gene flow

persists between these species and is currently concentrated at a hybrid zone. The considerable number, varying size, and location of barrier regions suggest that the genomic architecture of speciation is polygenic, or multilocus *sensu* Barton (1983). While our analysis did not account for heterogeneity in recombination rate, m_e variation both between and within chromosomes is likely shaped by recombination rate heterogeneity, a pattern previously observed in *Heliconius* butterflies (Martin et al. 2019, Laetsch et al. 2023). However, in the absence of a direct estimate of recombination heterogeneity it is impossible to know to what extent the overlap between long and short-term barriers reflects a shared/stable recombination landscape rather than shared selective targets. Our results pave the way for future quantitative analyses of the temporal evolution of the genomic landscape of species barriers. While most speciation processes are far too slow to observe directly, genomic variation clearly contains information about the interplay of forces acting at different stages of the speciation continuum.

Methods

Sampling and sequencing

In total, we generated WGS data (150 b paired-end reads) for 20 individuals (Supplementary Table 1). Field sampling was conducted in 2017 and 2018 at several locations across Southern and Central Europe (Spain, France, Romania, and Hungary). Samples were hand-netted in the field, flash-frozen from live in a liquid nitrogen dry shipper (Voyageur 12) and stored at -70 $°C$. Wings were retained for identification. DNA extractions for all individuals were performed using a Qiagen DNeasy Blood & Tissue kit. Extractions were used to prepare TruSeq Nano gel-free libraries by Edinburgh Genomics which were sequenced on a NovaSeq 6000 or HiSeq X (Supplementary Table 1). Raw reads are deposited at the ENA (accession number PRJEB76171).

QC, read mapping, variant calling, and summaries

Reads were trimmed and checked for quality using `FastQC v0.11.8` (Andrews 2016) both before and after trimming with `FastP v0.20.0` (Chen et al. 2018), using `MultiQC v1.7` (Ewels et al. 2016) to visualise the results. Trimmed reads were aligned to the *I. podalirius* reference assembly (Mackintosh et al. 2022) using `bwa-mem v0.7.17` (Li 2013). We marked duplicates using `sambamba v0.6.6` (Tarasov et al. 2015). Variants were called using `freebayes v1.3` (Garrison & Marth 2012) `(-k -w -j -T 0.02 -E -1 --report-genotype-likelihood-max)`. A principle components analysis (PCA) was performed on intergenic autosomal variants using `plink v1.9` (Purcell et al. 2007). Genetic diversity (π), mean individual heterozygosity (H) and divergence (d_{XY}) were estimated at intergenic sites using the ***gIMble*** 'gimbleprep' module (Laetsch et al. 2023).

Fitting models of speciation across the genome using the bSFS

To infer the likely speciation history of this species pair and to estimate migration rates along the genome we used the software package ***gIMble*** (Laetsch et al. 2023). This analysis was restricted to the non-HZ set of individuals. As the inference assumes that each sample contributes diploid genotypes, analysis was restricted to the autosomes to allow the inclusion of both male and female data. Demographic models fit by ***gIMble*** assume a neutral model of evolution, and so we focused analysis on intergenic sequence (39% of the assembly), i.e. genic and repeat-rich regions and contigs that were not scaffolded into chromosomes were excluded. We filtered genotype (GT) calls to a minimum depth of eight reads per sample and a maximum depth of three times the mean coverage

per sample. Additionally, GT calls were required to have a minimum PHRED quality of one and SNPs within 2 bases of indels were removed. We used the `gIMble` 'parse' module to quantify genetic diversity and divergence for the filtered subset of the data. We summarised variation in pair-blocks of 64 bases and tallied all blockwise site frequency configurations of mutations across heterospecific sample pairs (see Laetsch et al. (2023) for details).

To quantify heterogeneity in N_e we fit a grid of parameters to sliding windows of a fixed length of 28,125 (pair-blocks) with a 20% overlap along the genome. This results in a total of 14,173 windows with a minimum span of 50 kb. The grid discretised N_e from 100,000 to 1,500,000 (in increments of 100,000) for *I. feisthamelii* from 100,000 to 900,000 (in increments of 50,000) for *I. podalirius* and from 100,000 to 2,400,000 (in increments of 100,000) for the ancestral population; m_e (forwards in time from *I. feisthamelii* into *I. podalirius*) was discretised from 0 to 4e-7, in increments of 5e-9. The split time T was fixed to the global result of 2,181,320 generations (Table 2) and the mutation rate was set to $\mu = 2.9 \times 10^{-9}$ per base and generation (Keightley et al. 2014) throughout.

To adjust our barrier definition for false positives (which are expected at an appreciable rate when the background m_e is low), we ran a parametric bootstrap on the local estimates m_e: for each window, we simulated 100 replicates under the best fitting local background history using `gIMble` 'simulate' and obtained a null distribution of $\Delta_{B,0}$. Only windows for which $\Delta_{B,0}$ was greater than the largest value contained in this $\Delta_{B,0}$ distribution under the simulated null were labelled as barriers. This ensures a false positive rate < 0.01.

Estimating hybrid indices and the ancestry of hybrid zone individuals

We polarized all variable sites (across all individuals) that were not singletons (which are uninformative for genome polarisation by association in state) and estimated genome-wide hybrid indices (HIs) using the `diem` framework implemented in both R and Mathematica (Baird et al. 2023). Variants were subject to the same filtering criteria as the `gIMble` analysis. We calculated HIs per individual using sets of polarised sites with high diagnostic index (DI > -20). We used chromosome subsets of these sites to calculate HIs per chromosome. At a further reduction in scale we computed HI for each `gIMble` window and compared D across all individuals between barrier and non-barrier `gIMble` windows in an ANOVA.

Deconfounding signal from multiple barriers

diem is designed to polarise genomes with respect to one barrier at a time, however trial *diem* analysis of the full genomes dataset suggested two barriers were present in the data: the focal barrier to geneflow between 'French' *I. podalirius* and 'Iberian' *I. feisthamelii*, and a 'nuisance' barrier to geneflow between 'Iberian' and 'African' *I. feisthamelii*, presumably due to the Strait of Gibraltar. For the focal analysis, signal of the nuisance barrier was removed by censoring all sites with minority homozygous state only present in (the two) samples from north Africa. In contrast, only these sites were retained in a secondary dataset, allowing us to also polarise genomes with respect to the Gibraltar barrier. Once sites were separated in this way, *diem* was run on each barrier-specific dataset.

Kernel smoothing of genotypes along chromosomes

diem estimates the polarity of each site independently. Where a barrier exists we may expect polarisation to reveal tracts of different inheritance along chromosomes. Downstream analyses of tract size distributions requires estimates of the boundaries of these tracts. We estimate tract boundaries using kernel smoothing of the diploid state along chromosomes. Working on the physical (Mb) metric, the Laplace distribution truncated at its 95% probability boundaries provides the kernel. This is centered on each site to provide a weighted average of states 'smoothed' over the chosen physical scale. The centre-site is then estimated to have the state with greatest weight. Where variable sites are dense on the physical metric, smoothed estimates will be influenced by many flanking sites, where sparse, few. Imposing a given smoothing kernel scale assumes we are uninterested in state variation at higher frequency (involving shorter tracts). The kernel smoothing is therefore being used as a low-pass signal filter to leave larger tracts with clearer boundaries.

Untangling blocks from genotypes

When diploid genomes are polarised with respect to a single barrier, an interval of heterozygous sites within otherwise homozygous chromosome genotype suggests a single tract of minority inheritance on one of the chromosome copies. While the same pattern would result from two minority tracts, one on each chromosome copy, that 'perfectly' abut (without overlap), simple genotypes can be parsimoniously (and automatically) parsed into haplotype tracts by assuming such perfect abutment is so rare as to be negligible. Complex genotypes, where multiple minority intervals occur, would require decisions regarding tract phase (whether or not the minority tracts occupy

the same chromosome copy), if we wished to estimate tract size distributions of both minority and majority tracts. However, comparison to theory for tract size distributions (Baird 1995) requires only minority tract sizes: Then, no phasing decisions are necessary, only a decision regarding which barrier-side state is in the minority for an individual. This can be decided with little uncertainty using an estimate of the individual's HI taken over very many sites genome wide.

Acknowledgements

We would like to thank Carla and Oskar Lohse for catching the two most informative hybrid samples: individuals 1322 and 1325. This work was supported by a European Research Council starting grant (ModelGenomLand 757648) to KL which also supported DRL. SE was supported by an EastBio studentship from the British Biological Sciences Research Council (BBSRC). KL was supported by a fellowship from the Natural Environment Research Council (NERC, NE/L011522/1). RV was supported by Grant PID2019-107078GB-I00 funded by Ministerio de Ciencia e Innovación and Agencia Estatal de Investigación (MCIN/AEI/10.13039/501100011033) We thank Katy McDonald for help in the molecular lab, Alex Hayward for help with field collections and Edinburgh Genomics for generating libraries and sequence data. We are indebted to Vlad Dincă, Raluca Vodă and Leonardo Dapporto for contributing samples and to Nick Barton for helpful suggestions on the analyses of admixture tracts and Alex Mackintosh for insightful comments on an earlier version of this manuscript.

Compliance with ethical standards

Field sampling of butterflies was conducted in compliance with the School of Biological Sciences Ethics Committee at the University of Edinburgh and the ERC ethics review procedure. Permissions for field sampling were obtained from the Generalitat de Catalunya (SF/639), the Gobierno de Aragon (INAGA/500201/24/2018/0614 to Karl Wotton) and the Gobierno del Principado de Asturias (014252).

Data Availability Statement

Read data is available from the ENA at PRJEB76171. Reads for sample IP_504 were generated by a previous study (Mackintosh et al. 2022) and are available at the ENA at PRJEB51340.

APPENDIX

In the Appendix of Baird (1995, Equation A10), Barton expressed the per migrant contribution to the distribution of continuous tract length x for two infinite demes of coupling θ genomes, T rescaled generations after starting to exchange migrants, in terms of Kummer's confluent hypergeometric function ${}_1F_1\left(a;b;z\right)$ (Kummer 1837):

$$F\left(x,T\right) = Te^{-x(1+\theta)T}{}_1F_1\left(\frac{\theta-1}{\theta+1};2;-T(1+\theta)(1-x)\right) \tag{A-1}$$

This can be re-arranged such that tract length only appears as an argument of the confluent hypergeometric function, not in the initial term, by using Kummer's first transformation:

$${}_1F_1\left(a;b;z\right) = e^z{}_1F_1\left(b-a;b;-z\right) \tag{A-2}$$

Applying transformation A-2 to the confluent hypergeometric function in A-1:

$${}_1F_1\left(\frac{\theta-1}{\theta+1};2;-T(1+\theta)(1-x)\right) = e^{T(1+\theta)(1-x)}{}_1F_1\left(2-\frac{\theta-1}{\theta+1};2;T(1+\theta)(1-x)\right) \tag{A-3}$$

Multiplying both sides of A-3 by the initial term in A-1, and simplifying, tract length appears only in the confluent hypergeometric function:

$$F\left(x,T\right) = Te^{-T(1+\theta)}{}_1F_1\left(\frac{3+\theta}{1+\theta};2;T(1+\theta)(1-x)\right) \tag{A-4}$$

In this form of Barton's result the first argument of the confluent hypergeometric is the negative log-log gradient of the equilibrium distribution of tract length y from Barton (1983, Eq (5)), clarifying the link between Barton's time T and equilibrium solutions. Integrating contributions A-1 since time of contact, the distribution of tract lengths per migrant found in either infinite deme is:

$$\int_0^T F(x,t)dt = (1+\theta)^{-2}x^{-\frac{3+\theta}{1+\theta}};T \longrightarrow \infty \tag{A-5}$$

The left hand expression can be found by numerical integration for any time T, and becomes equal to the right hand equilibrium solution for progressively smaller tracts as time goes on.

References

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., Jones, J., Keller, B., Marczewski, T., Mallet, J., Martinez-Rodriguez, P., Möst, M., Mullen, S., Nichols, R., Nolte, A. W., Parisod, C., Pfennig, K., Rice, A. M., Ritchie, M. G., Seifert, B., Smadja, C. M., Stelkens, R., Szymura, J. M., Väinölä, R., Wolf, J. B. W. & Zinner, D. (2013), 'Hybridization and speciation', *Journal of Evolutionary Biology* **26**(2), 229–246.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1420-9101.2012.02599.x*

Aeschbacher, S., Selby, J. P., Willis, J. H. & Coop, G. (2017), 'Population-genomic inference of the strength and timing of selection against gene flow', *Proceedings of the National Academy of Sciences* **114**(27), 7061–7066.

Andrews, S. (2016), 'Fastqc version 0.11. 5', *A Quality Control Tool for High Throughput Sequence Data* .

Arnold, M. L., Sapir, Y. & Martin, N. H. (2008), 'Genetic exchange and the origin of adaptations: prokaryotes to primates', *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**(1505), 2813–2820.

Bachtrog, D., Mahajan, S. & Bracewell, R. (2019), 'Massive gene amplification on a recently formed drosophila y chromosome', *Nature ecology & evolution* **3**(11), 1587–1597.

Baird, S. (1995), 'A simulation study of multilocus clines', *Evolution* **49**(6), 1038–1045.

Baird, S. J., Petružela, J., Jaroň, I., Škrabánek, P. & Martínková, N. (2023), 'Genome polarisation for detecting barriers to geneflow', *Methods in Ecology and Evolution* **14**(2), 512–528.

Bank, C., Bürger, R. & Hermisson, J. (2012), 'The limits to parapatric speciation: Dobzhansky–muller incompatibilities in a continent–island model', *Genetics* **191**(3), 845–863.

Banker, S. E., Bonhomme, F. & Nachman, M. W. (2022), 'Bidirectional Introgression between *Mus musculus domesticus* and *Mus spretus*', *Genome Biology and Evolution* **14**(1). evab288.
URL: *https://doi.org/10.1093/gbe/evab288*

Barton, N. & Bengtsson, B. O. (1986), 'The barrier to genetic exchange between hybridising populations', *Heredity* **57**(3), 357–376.

Barton, N. H. (1983), 'Multilocus clines', *Evolution* pp. 454–471.

Barton, N. H. & Charlesworth, B. (1984), 'Genetic revolutions, founder effects, and speciation', *Annual review of ecology and systematics* **15**(1), 133–164.

Barton, N. H. & Gale, K. S. (1993), Genetic Analysis of Hybrid Zones, *in* 'Hybrid Zones and the Evolutionary Process', Oxford University Press.
URL: *https://doi.org/10.1093/oso/9780195069174.003.0002*

Barton, N. H. & Hewitt, G. M. (1985), 'Analysis of hybrid zones', *Annual review of Ecology and Systematics* pp. 113–148.

Bhattacharyya, T., Reifova, R., Gregorova, S., Simecek, P., Gergelits, V., Mistrik, M., Martincova, I., Pialek, J. & Forejt, J. (2014), 'X chromosome control of meiotic chromosome synapsis in mouse inter-subspecific hybrids', *PLoS genetics* **10**(2), e1004088.

Bisschop, G. (2022), 'Graph-based algorithms for laplace transformed coalescence time distributions', *bioRxiv* .

Buggs, R. (2007), 'Empirical study of hybrid zone movement', *Heredity* **99**(3), 301–312.

Bush, G. L. (1994), 'Sympatric speciation in animals: new wine in old bottles', *Trends in ecology & evolution* **9**(8), 285–288.

Campbell, C. R., Poelstra, J. & Yoder, A. D. (2018), 'What is speciation genomics? the roles of ecology, gene flow, and genomic architecture in the formation of species', *Biological Journal of the Linnean Society* **124**(4), 561–583.

Capblancq, T., Mavárez, J., Rioux, D. & Després, L. (2019), 'Speciation with gene flow: evidence from a complex of alpine butterflies (coenonympha, satyridae)', *Ecology and evolution* **9**(11), 6444–6457.

Charlesworth, B., Coyne, J. A. & Barton, N. H. (1987), 'The relative rates of evolution of sex chromosomes and autosomes', *The American Naturalist* **130**(1), 113–146.

Chen, S., Zhou, Y., Chen, Y. & Gu, J. (2018), 'fastp: an ultra-fast all-in-one fastq preprocessor', *Bioinformatics* **34**(17), i884–i890.

Coutsis, J. & Van Oorschot, H. (2011), 'Differences in the male and female genitalia between *Iphiclides podalirius* and *Iphiclides feisthamelii*, further supporting species status for the latter (lepidoptera: Papilionidae)', *Phegea* **39**(1), 12–22.

Descimon, H. & Mallet, J. (2009), 'Bad species', *Ecology of butterflies in Europe* **500**(C), 219.

Dieckmann, U. & Doebeli, M. (1999), 'On the origin of species by sympatric speciation', *Nature* **400**(6742), 354–357.

Drès, M. & Mallet, J. (2002), 'Host races in plant–feeding insects and their importance in sympatric speciation', *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **357**(1420), 471–492.
URL: *https://royalsocietypublishing.org/doi/abs/10.1098/rstb.2002.1059*

Ebdon, S., Laetsch, D. R., Dapporto, L., Hayward, A., Ritchie, M. G., Dinc, V., Vila, R. & Lohse, K. (2021), 'The pleistocene species pump past its prime: Evidence from european butterfly sister species', *Molecular Ecology* **30**(14), 3575–3589.

Ewels, P., Magnusson, M., Lundin, S. & Käller, M. (2016), 'Multiqc: summarize analysis results for multiple tools and samples in a single report', *Bioinformatics* **32**(19), 3047–3048.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C. & Foll, M. (2013), 'Robust demographic inference from genomic and snp data', *PLoS genetics* **9**(10), e1003905.

Feder, J. L., Egan, S. P. & Nosil, P. (2012), 'The genomics of speciation-with-gene-flow', *Trends in genetics* **28**(7), 342–350.

Fraïsse, C., Popovic, I., Mazoyer, C., Spataro, B., Delmotte, S., Romiguier, J., Loire, E., Simon, A., Galtier, N., Duret, L. et al. (2021), 'Dils: Demographic inferences with linked selection by using abc', *Molecular Ecology Resources* **21**(8), 2629–2644.

Garrigan, D., Kingan, S. B., Geneva, A. J., Andolfatto, P., Clark, A. G., Thornton, K. R. & Presgraves, D. C. (2012), 'Genome sequencing reveals complex speciation in the *Drosophila simulans* clade', *Genome research* **22**(8), 1499–1511.

Garrison, E. & Marth, G. (2012), 'Haplotype-based variant detection from short-read sequencing', *arXiv preprint arXiv:1207.3907* .

Gaunet, A., Dinca, V. E., Dapporto, L., Montagud, S., Voda, R., Schär, S., Badiane, A., Font, E. & Vila, R. (2019), 'Two consecutive *Wolbachia*-mediated mitochondrial introgressions obscure taxonomy in palearctic swallowtail butterflies (*Lepidoptera*, *Papilionidae*)', *Zoologica Scripta* .

Godart, J.-B. (1842), *Histoire naturelle des lépidoptères ou papillons de France*, Vol. 15, Méquignon-Marvis fils.

Gompert, Z. & Buerkle, C. (2012), 'bgc: Software for bayesian estimation of genomic clines', *Molecular Ecology Resources* **12**(6), 1168–1176.

Gompert, Z., Mandeville, E. G. & Buerkle, C. A. (2017), 'Analysis of population genomic data from hybrid zones', *Annual Review of Ecology, Evolution, and Systematics* **48**(Volume 48, 2017), 207–229.
URL: *https://www.annualreviews.org/content/journals/10.1146/annurev-ecolsys-110316-022652*

Gosset, C. & Bierne, N. (2013), 'Differential introgression from a sister species explains high f st outlier loci within a mussel species', *Journal of Evolutionary Biology* **26**(1), 14–26.

Gower, G., Ragsdale, A. P., Bisschop, G., Gutenkunst, R. N., Hartfield, M., Noskova, E., Schiffels, S., Struck, T. J., Kelleher, J. & Thornton, K. R. (2022), 'Demes: a standard format for demographic models', *Genetics* p. iyac131.
URL: *https://doi.org/10.1093/genetics/iyac131*

Grainger Hunt, W. & Selander, R. K. (1973), 'Biochemical genetics of hybridisation in european house mice', *Heredity* **31**(1), 11–33.

Hagberg, L., Celemín, E., Irisarri, I., Hawlitschek, O., Bella, J. L., Mott, T. & Pereira, R. J. (2022), 'Extensive introgression at late stages of species formation: Insights from grasshopper hybrid zones', *Molecular Ecology* **31**(8), 2384–2399.

Harris, K. & Nielsen, R. (2016), 'The Genetic Cost of Neanderthal Introgression', *Genetics* **203**(2), 881–891.
URL: *https://doi.org/10.1534/genetics.116.186890*

Harrison, R. G. et al. (1990), 'Hybrid zones: windows on evolutionary process.', *Oxford surveys in evolutionary biology* **7**, 69–128.

Hedrick, P. W. (2013), 'Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation', *Molecular ecology* **22**(18), 4606–4618.

Hewitt, G. M. (1988), 'Hybrid zones-natural laboratories for evolutionary studies', *Trends in ecology & evolution* **3**(7), 158–167.

Hirase, S., Yamasaki, Y. Y., Sekino, M., Nishisako, M., Ikeda, M., Hara, M., Merilä, J. & Kikuchi, K. (2021), 'Genomic evidence for speciation with gene flow in broadcast spawning marine invertebrates', *Molecular biology and evolution* **38**(11), 4683–4699.

Iasi, L. N. M., Chintalapati, M., Skov, L., Bossoma Mesa, A., Hajdinjak, M., Peter, B. M. & Moorjani, P. (2024), 'Neandertal ancestry through time: Insights from genomes of ancient and present-day humans', *bioRxiv* pp. 2024–05.

Johnson, E. A. & Miyanishi, K. (2008), 'Testing the assumptions of chronosequences in succession', *Ecology letters* **11**(5), 419–431.

Keightley, P. D., Ness, R. W., Halligan, D. L. & Haddrill, P. R. (2014), 'Estimation of the spontaneous mutation rate per nucleotide site in a *Drosophila melanogaster* full-sib family', *Genetics* **196**(1), 313–320.

Kingston, S. E., Parchman, T. L., Gompert, Z., Buerkle, C. & Braun, M. J. (2017), 'Heterogeneity and concordance in locus-specific differentiation and introgression between species of towhees', *Journal of evolutionary biology* **30**(3), 474–485.

Kruuk, L., Baird, S., Gale, K. & Barton, N. H. (1999), 'A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids', *Genetics* **153**(4), 1959–1971.

Kummer, E. E. (1837), 'De integralibus quibusdam definitis et seriebus infinitis.'.

Laetsch, D. R., Bisschop, G., Martin, S. H., Aeschbacher, S., Setter, D. & Lohse, K. (2023), 'Demographically explicit scans for barriers to gene flow using gimble', *PLoS Genetics* **19**(10), e1010999.

Lafranchis, T., Delmas, S. & Mazel, R. (2015), 'Le contact *Iphiclides feisthamelii-I. podalirius.* statut de ces deux taxons (lepidoptera, papilionidae)', *Revue de l'Association Roussillonnaise d'Entomologie* **24**(3), 111–132.

Lenormand, T. (2002), 'Gene flow and the limits to natural selection', *Trends in ecology & evolution* **17**(4), 183–189.

Li, H. (2013), 'Aligning sequence reads, clone sequences and assembly contigs with bwa-mem', *arXiv preprint arXiv:1303.3997* .

Macholán, M., Baird, S. J., Dufková, P., Munclinger, P., Bímová, B. V. & Piálek, J. (2011), 'Assessing multilocus introgression patterns: a case study on the mouse x chromosome in central europe', *Evolution* **65**(5), 1428–1446.

Macholán, M., Munclinger, P., Šugerková, M., Dufková, P., Bímová, B., Božíková, E., Zima, J. & Piálek, J. (2007), 'Genetic analysis of autosomal and x-linked markers across a mouse hybrid zone', *Evolution* **61**(4), 746–771.

Mackintosh, A., Laetsch, D. R., Baril, T., Ebdon, S., Jay, P., Vila, R., Hayward, A. & Lohse, K. (2022), 'The genome sequence of the scarce swallowtail, *Iphiclides podalirius*', *G3* **12**(9), jkac193.

Mackintosh, A., Vila, R., Laetsch, D. R., Hayward, A., Martin, S. H. & Lohse, K. (2023), 'Chromosome fissions and fusions act as barriers to gene flow between brenthis fritillary butterflies', *Molecular biology and evolution* **40**(3), msad043.

Mailund, T., Halager, A. E., Westergaard, M., Dutheil, J. Y., Munch, K., Andersen, L. N., Lunter, G., Prüfer, K., Scally, A., Hobolth, A. et al. (2012), 'A new isolation with migration model along complete genomes infers very different divergence processes among closely related great ape species', *PLoS genetics* **8**(12), e1003125.

Mallet, J. (1986), 'Hybrid zones of *Heliconius* butterflies in panama and the stability and movement of warning colour clines', *Heredity* **56**(2), 191–202.

Mank, J. E., Nam, K. & Ellegren, H. (2009), 'Faster-z evolution is predominantly due to genetic drift', *Molecular biology and evolution* **27**(3), 661–670.

Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., Blaxter, M., Manica, A., Mallet, J. & Jiggins, C. D. (2013), 'Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies', *Genome research* **23**(11), 1817–1828.

Martin, S. H., Davey, J. W., Salazar, C. & Jiggins, C. D. (2019), 'Recombination rate variation shapes barriers to introgression across butterfly genomes', *PLoS biology* **17**(2), e2006288.

Meier, J. I., Salazar, P. A., Kučka, M., Davies, R. W., Dréau, A., Aldás, I., Power, O. B., Nadeau, N. J., Bridle, J. R., Rolian, C., Barton, N. H., McMillan, W. O., Jiggins, C. D. & Chan, Y. F.

(2021), 'Haplotype tagging reveals parallel formation of hybrid races in two butterfly species', *Proceedings of the National Academy of Sciences* **118**(25), e2015005118.
URL: *https://www.pnas.org/doi/abs/10.1073/pnas.2015005118*

Mérot, C., Salazar, C., Merrill, R. M., Jiggins, C. D. & Joron, M. (2017), 'What shapes the continuum of reproductive isolation? lessons from *Heliconius* butterflies', *Proceedings of the Royal Society B: Biological Sciences* **284**(1856), 20170335.

Mondal, M., Bertranpetit, J. & Lao, O. (2019), 'Approximate bayesian computation with deep learning supports a third archaic introgression in asia and oceania', *Nature communications* **10**(1), 1–9.

Morán, T. & Fontdevila, A. (2014), 'Genome-wide dissection of hybrid sterility in *Drosophila* confirms a polygenic threshold architecture', *Journal of Heredity* **105**(3), 381–396.

Nadeau, N. J., Ruiz, M., Salazar, P., Counterman, B., Medina, J. A., Ortiz-Zuazaga, H., Morrison, A., McMillan, W. O., Jiggins, C. D. & Papa, R. (2014), 'Population genomics of parallel hybrid zones in the mimetic butterflies, *H. melpomene* and *H. erato*', *Genome research* **24**(8), 1316–1333.

Ngeve, M. N., Van der Stocken, T., Sierens, T., Koedam, N. & Triest, L. (2017), 'Bidirectional gene flow on a mangrove river landscape and between-catchment dispersal of *Rhizophora racemosa* (rhizophoraceae)', *Hydrobiologia* **790**(1), 93–108.

Nosil, P. (2008), 'Speciation with gene flow could be common', *Molecular Ecology* **17**(9), 2103–2106.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-294X.2008.03715.x*

Nosil, P., Funk, D. J. & Ortiz-Barrientos, D. (2009), 'Divergent selection and heterogeneous genomic divergence', *Molecular ecology* **18**(3), 375–402.

Otto, S. P. (2019), 'Evolutionary potential for genomic islands of sexual divergence on recombining sex chromosomes', *New Phytologist* **224**(3), 1241–1251.
URL: *https://nph.onlinelibrary.wiley.com/doi/abs/10.1111/nph.16083*

Papadopulos, A. S., Baker, W. J., Crayn, D., Butlin, R. K., Kynast, R. G., Hutton, I. & Savolainen, V. (2011), 'Speciation with gene flow on lord howe island', *Proceedings of the National Academy of Sciences* **108**(32), 13188–13193.

Parchman, T., Gompert, Z., Braun, M. J., Brumfield, R., McDonald, D., Uy, J. A. C., Zhang, G., Jarvis, E., Schlinger, B. & Buerkle, C. (2013), 'The genomic consequences of adaptive divergence and reproductive isolation between species of manakins', *Molecular ecology* **22**(12), 3304–3317.

Payseur, B. A. & Rieseberg, L. H. (2016), 'A genomic perspective on hybridization and speciation', *Molecular ecology* **25**(11), 2337–2360.

Poikela, N. P., Laetsch, D. R., Kankare, M., Hoikkala, A. & Lohse, K. (2022), 'Experimental introgression in *Drosophila*: asymmetric postzygotic isolation associated with chromosomal inversions and an incompatibility locus on the x chromosome', *bioRxiv* .

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., De Bakker, P. I., Daly, M. J. et al. (2007), 'Plink: a tool set for whole-genome association and population-based linkage analyses', *The American journal of human genetics* **81**(3), 559–575.

Racimo, F., Sankararaman, S., Nielsen, R. & Huerta-Sánchez, E. (2015), 'Evidence for archaic adaptive introgression in humans', *Nature Reviews Genetics* **16**(6), 359–371.

Ravinet, M., Faria, R., Butlin, R., Galindo, J., Bierne, N., Rafajlović, M., Noor, M., Mehlig, B. & Westram, A. (2017), 'Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow', *Journal of evolutionary biology* **30**(8), 1450–1477.

Reifová, R., Majerová, V., Reif, J., Ahola, M., Lindholm, A. & Procházka, P. (2016), 'Patterns of gene flow and selection across multiple species of acrocephalus warblers: footprints of parallel selection on the z chromosome', *BMC evolutionary biology* **16**(1), 1–12.

Remington, C. L. (1968), Suture-zones of hybrid interaction between recently joined biotas, *in* 'Evolutionary biology', Springer, pp. 321–428.

Rieseberg, L. H., Archer, M. A. & Wayne, R. K. (1999), 'Transgressive segregation, adaptation and speciation', *Heredity* **83**(4), 363–372.

Roux, C., Fraisse, C., Romiguier, J., Anciaux, Y., Galtier, N. & Bierne, N. (2016), 'Shedding light on the grey zone of speciation along a continuum of genomic divergence', *PLoS biology* **14**(12), e2000234.

Saarman, N. P., Opiro, R., Hyseni, C., Echodu, R., Opiyo, E. A., Dion, K., Johnson, T., Aksoy, S. & Caccone, A. (2019), 'The population genomics of multiple tsetse fly (glossina fuscipes fuscipes) admixture zones in uganda', *Molecular ecology* **28**(1), 66–85.

Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., Patterson, N. & Reich, D. (2014), 'The genomic landscape of neanderthal ancestry in present-day humans', *Nature* **507**(7492), 354–357.

Stankowski, S. & Ravinet, M. (2021), 'Defining the speciation continuum', *Evolution* **75**(6), 1256–1273.

Stankowski, S., Shipilina, D. & Westram, A. M. (2021), 'Hybrid zones', *eLS* **2**, 1–12.

Szymura, J. M. & Barton, N. H. (1986), 'Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near cracow in southern poland', *Evolution* **40**(6), 1141–1159.

Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. (2015), 'Sambamba: fast processing of ngs alignment formats', *Bioinformatics* **31**(12), 2032–2034.

Taylor, C., Touré, Y. T., Carnahan, J., Norris, D. E., Dolo, G., Traoré, S. F., Edillo, F. E. & Lanzaro, G. C. (2001), 'Gene flow among populations of the malaria vector, *Anopheles gambiae*, in mali, west africa', *Genetics* **157**(2), 743–750.

Taylor, S. A., Curry, R. L., White, T. A., Ferretti, V. & Lovette, I. (2014), 'Spatiotemporally consistent genomic signatures of reproductive isolation in a moving hybrid zone', *Evolution* **68**(11), 3066–3081.

Taylor, S. A. & Larson, E. L. (2019), 'Insights from genomes into the evolutionary importance and prevalence of hybridization in nature', *Nature ecology & evolution* **3**(2), 170–177.

Teeter, K. C., Payseur, B. A., Harris, L. W., Bakewell, M. A., Thibodeau, L. M., O'Brien, J. E., Krenz, J. G., Sans-Fuentes, M. A., Nachman, M. W. & Tucker, P. K. (2008), 'Genome-wide patterns of gene flow across a house mouse hybrid zone', *Genome research* **18**(1), 67–76.

Turelli, M. & Orr, H. A. (1995), 'The dominance theory of haldane's rule.', *Genetics* **140**(1), 389–402.

Turner, L. M. & Harr, B. (2014), 'Genome-wide mapping in a house mouse hybrid zone reveals hybrid sterility loci and Dobzhansky-Muller interactions', *Elife* **3**, e02504.

Wang, R. L., Wakeley, J. & Hey, J. (1997), 'Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives', *Genetics* **147**(3), 1091–1106.

Westram, A. M., Faria, R., Johannesson, K. & Butlin, R. (2021), 'Using replicate hybrid zones to understand the genomic basis of adaptive divergence', *Molecular ecology* **30**(15), 3797–3814.

Wiemers, M. (2003), Chromosome differentiation and the radiation of the butterfly subgenus Agrodiaetus (Lepidoptera: Lycaenidae: Polyommatus): a molecular phylogenetic approach, PhD thesis, Bonn, Univ., Diss., 2003.

Wiemers, M. & Fiedler, K. (2007), 'Does the DNA barcoding gap exist?–a case study in blue butterflies (lepidoptera: Lycaenidae)', *Frontiers in zoology* **4**(1), 1–16.

Wiemers, M. & Gottsberger, B. (2010), 'Discordant patterns of mitochondrial and nuclear differentiation in the scarce swallowtail *Iphiclides podalirius feisthamelii* (duponchel, 1832)(lepidoptera: Papilionidae)', *Entomologische Zeitschrift* **120**(3), 111–115.

Wilkinson-Herbots, H. M. (2008), 'The distribution of the coalescence time and the number of pairwise nucleotide differences in the "isolation with migration" model', *Theoretical population biology* **73**(2), 277–288.

Wu, C.-I. (2001), 'The genic view of the process of speciation', *Journal of evolutionary biology* **14**(6), 851–865.

Yan, J., Zhu, M., Liu, W., Xu, Q., Zhu, C., Li, J. & Sang, T. (2016), 'Genetic variation and bidirectional gene flow in the riparian plant *Miscanthus lutarioriparius*, across its endemic range: implications for adaptive potential', *GCB Bioenergy* **8**(4), 764–776.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/gcbb.12278*

Yang, W., Feiner, N., Laakkonen, H., Sacchi, R., Zuffi, M. A., Scali, S., While, G. M. & Uller, T. (2020), 'Spatial variation in gene flow across a hybrid zone reveals causes of reproductive isolation and asymmetric introgression in wall lizards', *Evolution* **74**(7), 1289–1300.

Supplementary materials

Supplementary figures

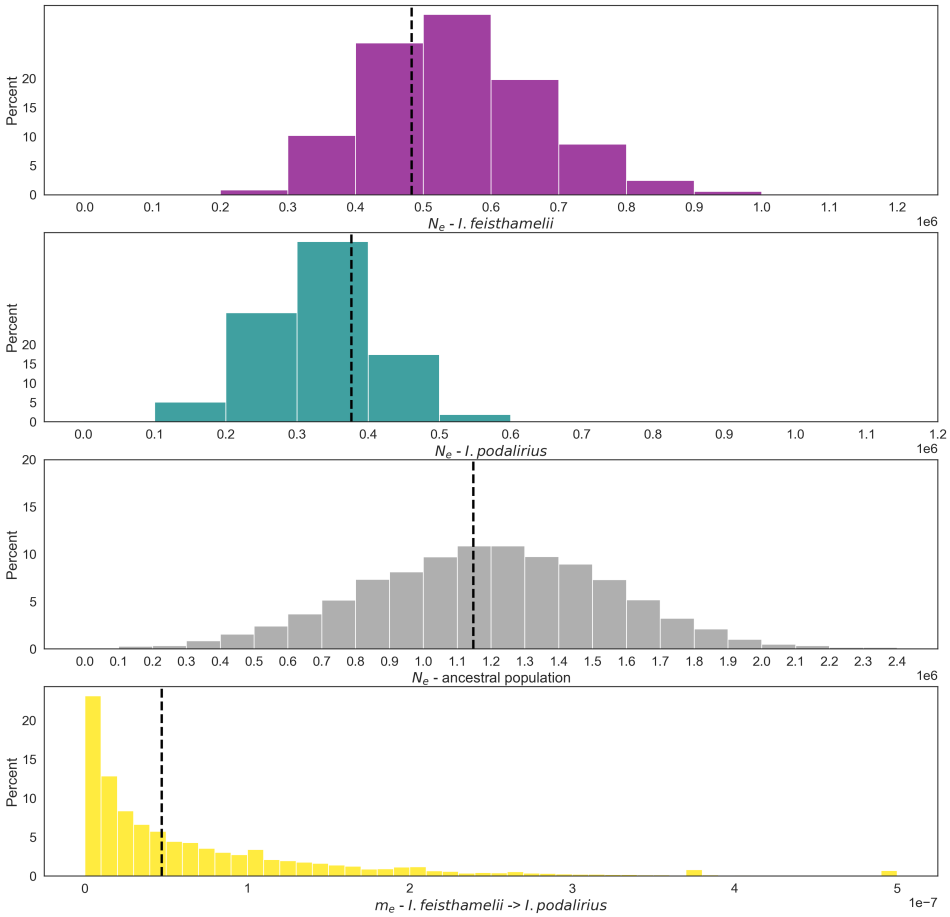

Figure S1 – The distributions of N_{fei} (purple), N_{pod} (teal), N_{anc} (grey), and $m_{e,i}$ (yellow) from *I. feisthamelii* into *I. podalirius* estimated in sliding windows using a 12x24x9x50 parameter grid. The vertical dashed lines indicate the global estimates of each parameter.

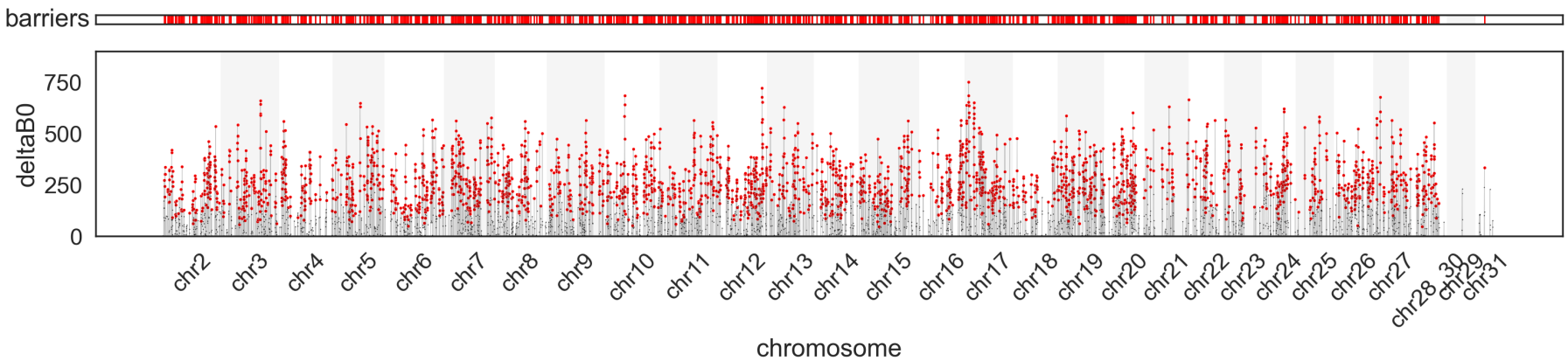

Figure S2 – Barriers to gene flow between *I. feisthamelii* and *I. podalirius*. The red bars in the top panel highlight the location of barrier windows. The bottom panel shows for each `gIMble` window, the difference in $\ln CL$ of $m_{\mathrm{e,i}} = 0$ from the background $m_{\mathrm{e}} = 1e-7$, Δ_{B0}. In windows where $\Delta_{B0} > 0$, a model of strict divergence is more likely than an IM model. We highlight these barrier windows in red.

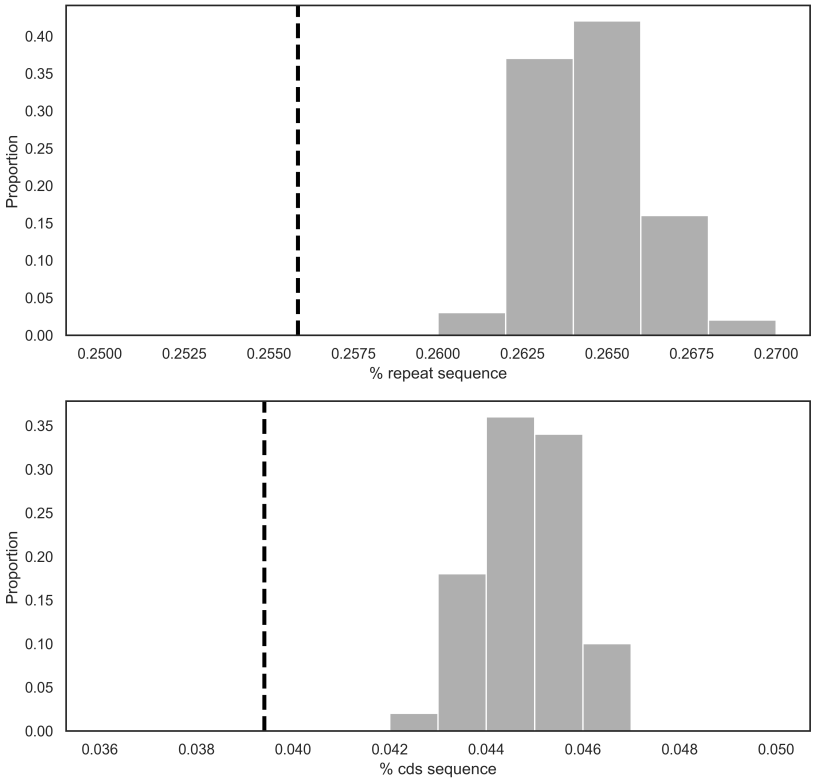

Figure S3 – The dashed lines indicate the mean density of repeats (top) and coding sequence (CDS) (bottom) in barrier regions defined via `gIMble`. Both are lower than the CDS and repeat proportion of data resampled at random using a circular resampling scheme.

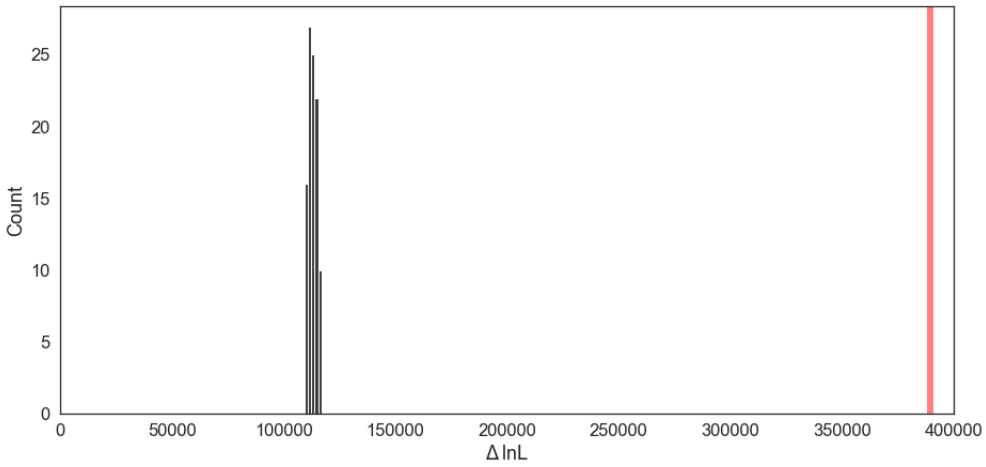

Figure S4 – The distribution of differences in log-likelihood between an isolation with migration (IM) model versus a model of strict divergence (DIV) fit to 100 replicates of data simulated under a DIV model (black) and the empirical data (red). The improvement in fit of an IM model is much greater in the empirical data than in the simulations.

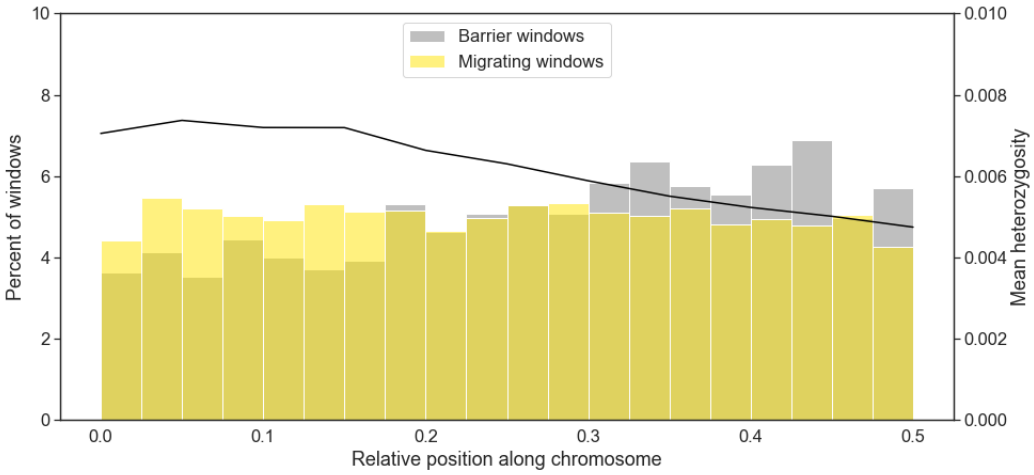

Figure S5 – The distribution of relative positions from the telomeres (at 0) to chromosome centres (at 0.5) for *gIMble* barrier windows (grey) and non-barrier windows (orange). The average heterozygosity for windows at different relative positions is shown by the black line. Barrier windows are marginally further from the telomeres than non-barrier windows (t-test, p = 0.0512).

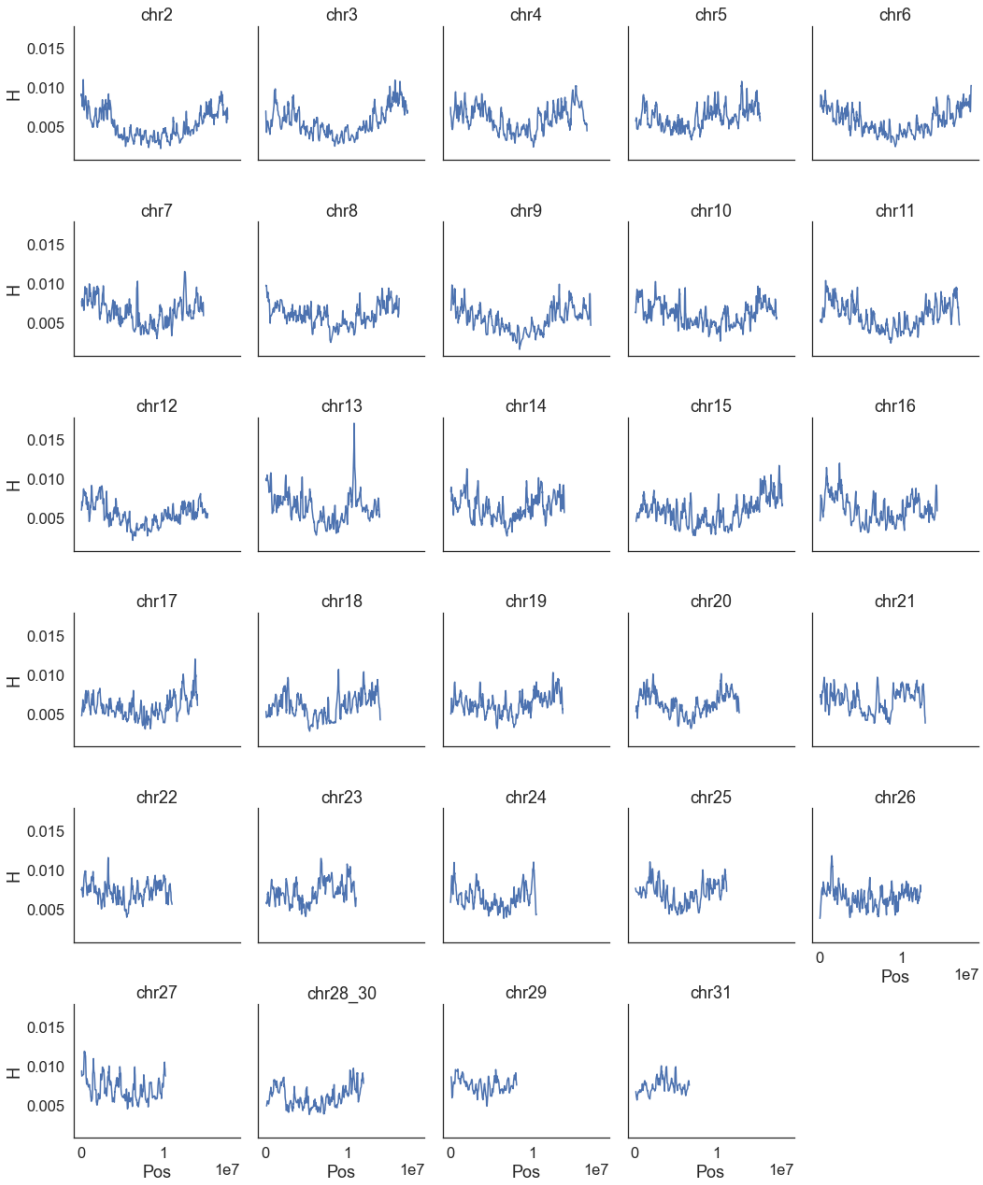

Figure S6 – Mean heterozygosity (H) in sliding windows of minimum 50kb across all chromosomes.

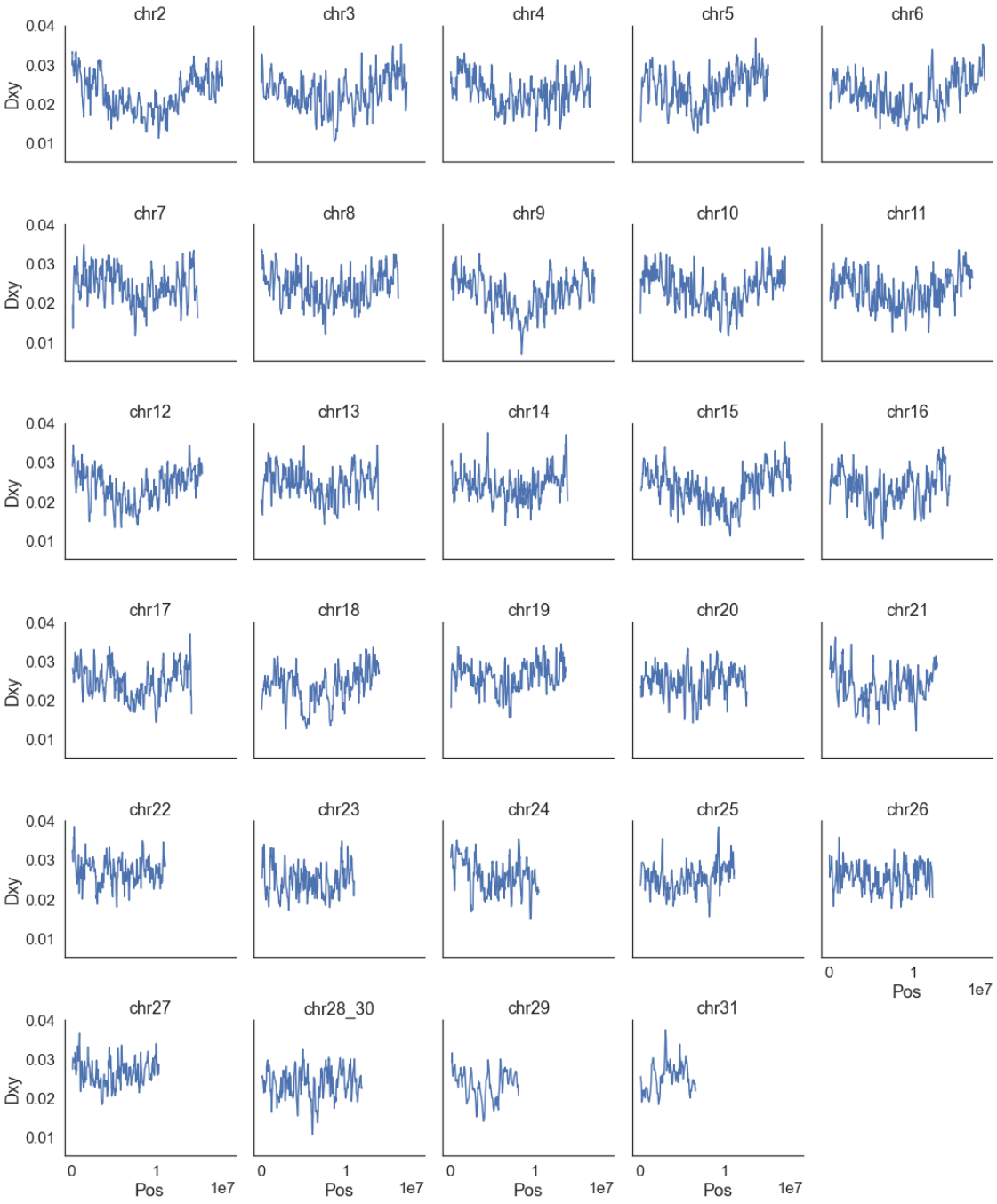

Figure S7 – Genetic divergence (d_{xy} in sliding windows of minimum 50kb across all chromosomes.

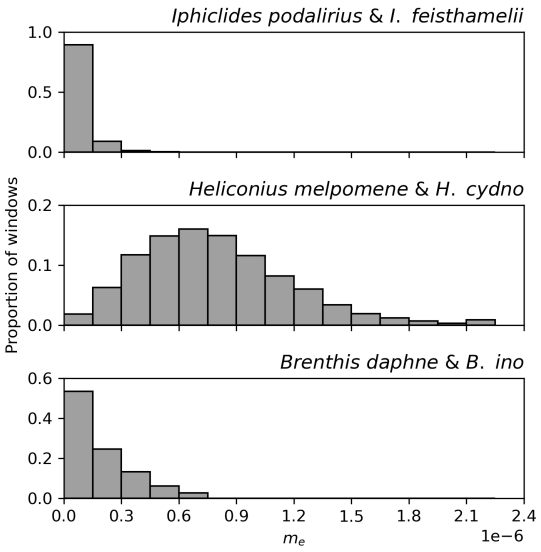

Figure S8 – The distribution of effective migration rates (m_e) between three pairs of butterfly sister species inferred by fitting IM models in windows across the genome. The rate of gene flow between *Iphiclides podalirius* and *I. feisthamelii* is lower than between *Heliconius melpomene* and *H. cydno* (Laetsch et al. 2023) and between *Brenthis daphne* and *B. ino* (Mackintosh et al. 2023).

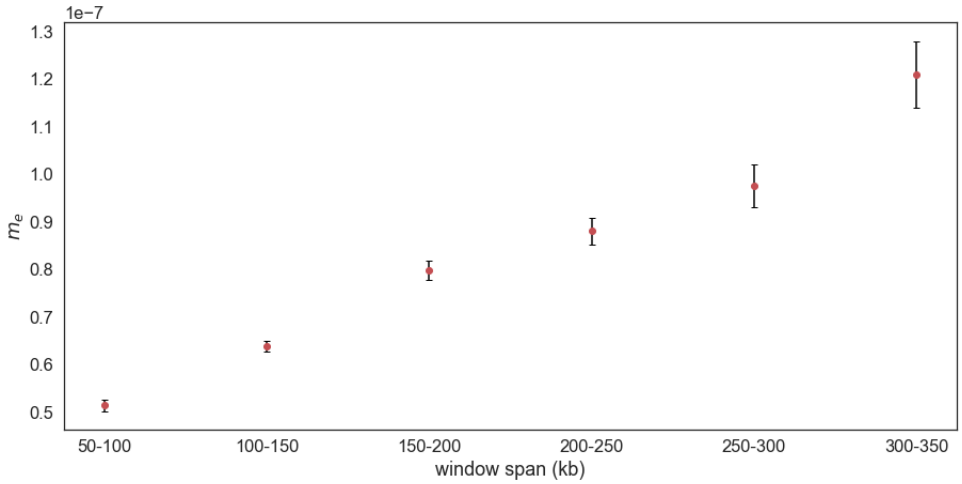

Figure S9 – Mean effective migration rates and standard errors for windows spanning different map lengths. Inferred migration rates are higher in windows spanning broader map lengths.