



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## Chromosome fissions and fusions act as barriers to gene flow between *Brenthis fritillaria* butterflies

**Citation for published version:**

Mackintosh, A, Vila, R, Laetsch, DR, Hayward, A, Martin, SH, Lohse, K & Rogers, R (ed.) 2023, 'Chromosome fissions and fusions act as barriers to gene flow between *Brenthis fritillaria* butterflies', *Molecular Biology and Evolution*, vol. 40, no. 3, msad043. <https://doi.org/10.1093/molbev/msad043>

**Digital Object Identifier (DOI):**

[10.1093/molbev/msad043](https://doi.org/10.1093/molbev/msad043)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

*Molecular Biology and Evolution*

**General rights**


Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Chromosome Fissions and Fusions Act as Barriers to Gene Flow between *Brenthis* Fritillary Butterflies

Alexander Mackintosh,<sup>\*,1</sup> Roger Vila ,<sup>2</sup> Dominik R. Laetsch,<sup>1</sup> Alex Hayward ,<sup>3</sup> Simon H. Martin ,<sup>†,1</sup> and Konrad Lohse<sup>†,1</sup>

<sup>1</sup>Institute of Ecology and Evolution, University of Edinburgh, Edinburgh EH9 3FL, United Kingdom

<sup>2</sup>Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Passeig Marítim de la Barceloneta 37, ESP-08003 Barcelona, Spain

<sup>3</sup>Centre for Ecology and Conservation, University of Exeter, Penryn Campus, Cornwall TR10 9FE, United Kingdom

<sup>†</sup>These authors contributed equally to this work.

\*Corresponding author: E-mail: a.j.f.mackintosh@sms.ed.ac.uk.

Associate editor: Rebekah Rogers

## Abstract

Chromosome rearrangements are thought to promote reproductive isolation between incipient species. However, it is unclear how often, and under what conditions, fission and fusion rearrangements act as barriers to gene flow. Here we investigate speciation between two largely sympatric fritillary butterflies, *Brenthis daphne* and *Brenthis ino*. We use a composite likelihood approach to infer the demographic history of these species from whole-genome sequence data. We then compare chromosome-level genome assemblies of individuals from each species and identify a total of nine chromosome fissions and fusions. Finally, we fit a demographic model where effective population sizes and effective migration rate vary across the genome, allowing us to quantify the effects of chromosome rearrangements on reproductive isolation. We show that chromosomes involved in rearrangements experienced less effective migration since the onset of species divergence and that genomic regions near rearrangement points have a further reduction in effective migration rate. Our results suggest that the evolution of multiple rearrangements in the *B. daphne* and *B. ino* populations, including alternative fusions of the same chromosomes, have resulted in a reduction in gene flow. Although fission and fusion of chromosomes are unlikely to be the only processes that have led to speciation between these butterflies, this study shows that these rearrangements can directly promote reproductive isolation and may be involved in speciation when karyotypes evolve quickly.

**Key words:** chromosome evolution, speciation, population genomics, chromosome rearrangements, demographic inference.

## Introduction

### Chromosomal Speciation

The process of speciation, where groups of individuals become reproductively isolated from one another, is driven by evolutionary forces that prevent gene flow. Many closely related species show differences in karyotype and there has been much discussion about the role of chromosome rearrangements (e.g. inversions, translocations, fissions, and fusions) in preventing gene flow and promoting speciation. Early work on *Drosophila* demonstrated that inversions suppress recombination (Sturtevant 1921; Dobzhansky and Epling 1948). More recently, both theoretical models (Noor et al. 2001; Navarro and Barton 2003; Kirkpatrick and Barton 2006) and examples in a variety of organisms (Wellenreuther and Bernatchez 2018) have shown that inversions can facilitate local adaptation, promote the evolution of genetic incompatibilities and act as barriers between recently diverged species. It is less clear, however, whether fission and fusion rearrangements have a similarly important role in

speciation (Rieseberg 2001). These rearrangements do not typically confer the same change in recombination as inversions do, yet there is evidence for increased speciation rates in groups where fissions and fusions happen more often (Bush et al. 1977; Leaché et al. 2016; de Vos et al. 2020). Fissions and fusions could act as barriers to gene flow if hybrid individuals that are heterozygous for a rearrangement suffer from underdominance (heterozygote disadvantage). This will happen when karyotypic heterozygosity generates multivalents at meiosis, which are prone to unbalanced segregation. Although there is indeed evidence for fissions and fusions causing underdominance through aneuploidy (Dutrillaux and Rumpler 1977; Castiglia and Capanna 2000; Lukhtanov et al. 2018), models of chromosomal speciation that assume underdominance are paradoxical; for hybrids to suffer from underdominance, the rearrangement must be at high frequency in one population, but how does a rearrangement rise to high frequency if it causes underdominance? Proposed solutions to this paradox include fixation by meiotic drive (White 1968), strong drift in a founder population

© The Author(s) 2023. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Open Access

(Templeton 1981; but see Barton and Charlesworth 1984), and complex rearrangements that evolve in a stepwise manner, where each step has a small fitness effect (White 1978b; Baker and Bickham 1986). This limits the conditions under which underdominant chromosomal speciation can happen, and it is therefore perhaps unsurprising that there are few convincing empirical examples (see Basset et al. 2006; Yannic et al. 2009; Yoshida et al. 2023).

Not all models of chromosomal speciation require underdominance. For example, fusions could affect gene flow by bringing preexisting barrier loci onto the same chromosome. Guerrero and Kirkpatrick (2014) showed that for two polymorphic loci maintained by selection–migration balance, a fusion will rise in frequency if it brings two locally adapted alleles into strong linkage disequilibrium (LD). This process has the potential to strengthen the combined effect of barrier loci by reducing recombination between them, thus promoting reproductive isolation. Although Guerrero and Kirkpatrick (2014) do not include underdominance in their model, the process they describe is not mutually exclusive with underdominant chromosomal speciation, and may offer an additional way for fusions to evolve in spite of underdominance.

Fission and fusion rearrangements can also influence the accumulation of reproductive isolation when a barrier to gene flow is highly polygenic. Given such a barrier, the probability that a neutral allele migrates is partly determined by whether it can recombine away from the foreign deleterious alleles that it was introgressed with (Aeschbacher et al. 2017). Fissions and fusions can alter the per-base recombination rates of chromosomes by changing their length and they can therefore influence effective migration. Recently, Martin et al. (2019) showed that recombination rate was the main determinant of the amount of introgression between species of *Heliconius* butterflies, with long fused chromosomes having less introgression than short non-fused ones. These fusions cannot be barriers themselves because they are shared among the species. Instead, because of their length, the fused chromosomes have a low per-base crossover rate (Davey et al. 2017), which reduces effective migration when barrier loci are common. Although the fusions in these *Heliconius* butterflies are shared, similar logic applies to a fusion that generates a long chromosome in just one population.

Importantly, a chromosome rearrangement may arise and fix long after a particular species split and so have no role in speciation. Alternatively, if rearrangements are present during the early stages of speciation, they may not have any effect on gene flow. This would be the case if underdominance was weak enough for a rearrangement to be effectively neutral. Moreover, even if rearrangements do have underdominant or recombination modifying effects, there may be barriers of very large effect which have played a much greater role in speciation. It is therefore important to quantify the effect of fission and fusion rearrangements on gene flow, rather than assuming that these conspicuous changes in the genome must play an important role in the speciation process.

## Chromosome Evolution in Butterflies

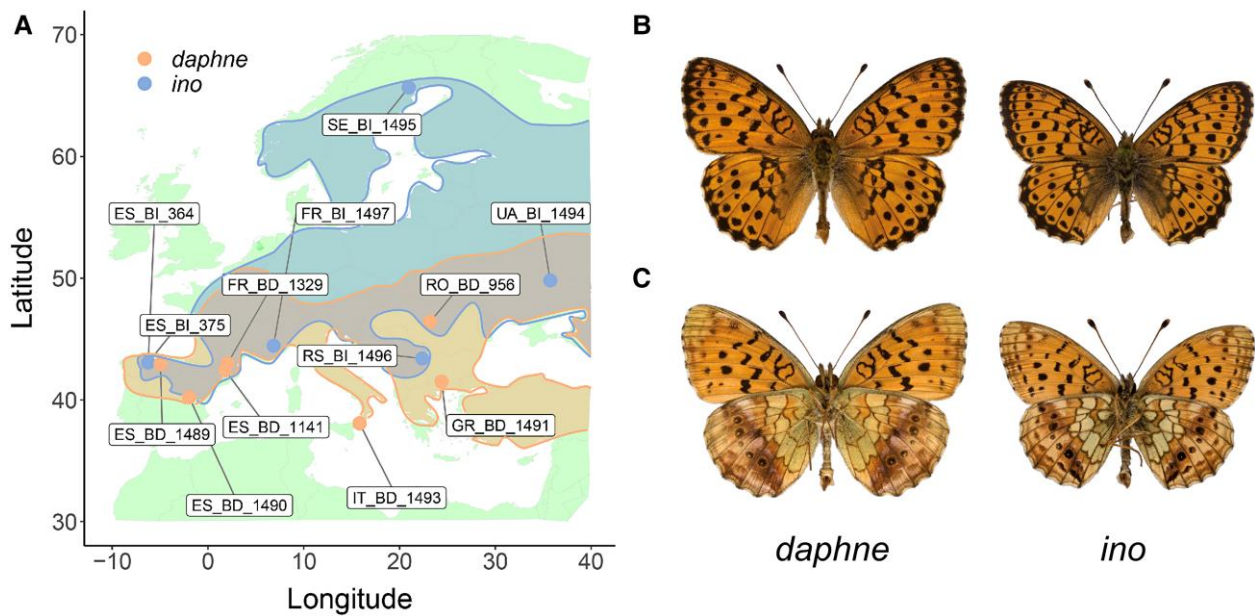
Most Lepidoptera (moths and butterflies) have similar karyotypes, consisting of around 30 pairs of autosomes and ZW sex chromosomes (de Vos et al. 2020). However, there are notable exceptions. For example, *Pieris* butterflies have a reduced karyotype where chromosomes have undergone substantial reorganization via inter-chromosomal rearrangements (Hill et al. 2019). There are also taxa with highly variable chromosome counts, such as the butterfly genera *Erebia*, *Lysandra*, *Polyommatus*, and *Leptidea*. In each of these genera, it has been suggested that rearrangements have facilitated speciation (Lukhtanov et al. 2005, 2011; Talavera et al. 2013; Augustijnen et al. 2023), although the extent to which rearrangements have affected reproductive isolation remains unclear.

Another group of butterflies in which karyotypes vary is the genus *Brenthis* (Nymphalidae) which consists of four species. Although 34 chromosome pairs have been observed in *Brenthis hecate* spermatocytes (de Lesse 1961; Saitoh and Lukhtanov 1988), *B. daphne* and *B. ino* are reported to have only 12–14 pairs of chromosomes (Federley 1938; Maeki and Makino 1953; de Lesse 1960; Saitoh 1986, 1987; Saitoh et al. 1989; Saitoh 1991). We recently assembled a *B. ino* reference genome (Mackintosh et al. 2022) with 14 pairs of chromosomes. We found that the genome was highly rearranged compared with the ancestral nymphalid karyotype and that a male individual was heterozygous for a Z-autosome chromosome fusion. These results are consistent with rapid, and likely still ongoing, chromosome evolution in the genus *Brenthis*.

The sister species *B. daphne* and *B. ino* are largely sympatric (fig. 1), have differences in larval host plant preference, and are estimated to have split approximately 3 Mya (Ebdon et al. 2021). Interspecific mating experiments have shown that female *B. daphne* and male *B. ino* can produce fertile offspring, suggesting that reproductive isolation between these species is incomplete (Kitahara 2008, 2012). Additionally, putative F1 hybrids have been observed in Japan (Kitahara 2012). Similar chromosome numbers have been observed for males of either species, 12–13 for *B. daphne* and 13–14 for *B. ino*, suggesting some intraspecific variation in karyotype, but no large differences between species. However, chromosome numbers will be unchanged by reciprocal translocations or an equal number of chromosome fission and fusion events. Such “cryptic” rearrangements are best identified by comparing genome assemblies. If *B. daphne* and *B. ino* possess cryptic inter-chromosomal rearrangements, then their recent divergence and potential for ongoing gene flow makes them a useful model for investigating the effects of rearrangements on reproductive isolation.

## Overview

Here we show that the genomes of *B. daphne* and *B. ino* differ by multiple fission and fusion rearrangements. More specifically, almost half of the chromosomes are involved in rearrangements, whereas the rest are syntenic. We



**FIG. 1.** (A) Sampling locations of *Brenthis daphne* (points in orange and IDs containing BD) and *B. ino* individuals (points in blue and IDs containing BI) across Europe. Approximate distributions are also shown using the same color scheme. (B) Uppersides of male *B. daphne* and male *B. ino*. (C) Undersides of male *B. daphne* and male *B. ino*.

estimate the demographic history of these species as well as genome-wide variation in effective migration rate ( $m_e$ ). By intersecting estimates of  $m_e$  with chromosome rearrangements, we test whether fissions and fusions have acted as barriers to gene flow. We consider the following scenarios:

- **No effect:** Fission and fusion rearrangements are selectively neutral and have had no effect on the effective rate of gene flow, either directly or indirectly.
- **Underdominance:** Fissions and fusions produce direct, localized barriers to gene flow because early generation hybrids and backcrosses with heterokaryotypes suffer reduced fitness. This would result in decreased post-divergence gene flow on rearranged chromosomes. Assuming that heterokaryotypes still undergo recombination, the reduction in gene flow would be strongest for loci that are closely linked to rearrangement points.
- **Fused barriers:** Fusions are not barriers to gene flow themselves, but have brought individual barrier alleles of large effect into linkage, thus strengthening the barrier effect of these loci. If most fusions put large effect loci into linkage, then this would cause a reduction in gene flow on rearranged chromosomes and the effect would be strongest close to fusion points. This scenario makes no predictions about the effect of chromosome fissions on gene flow.
- **Polygenic barriers:** In the presence of polygenic barriers, fissions and fusions affect gene flow by modifying chromosome lengths and therefore recombination rates. This scenario predicts a negative correlation between gene flow and chromosome length.

## Results

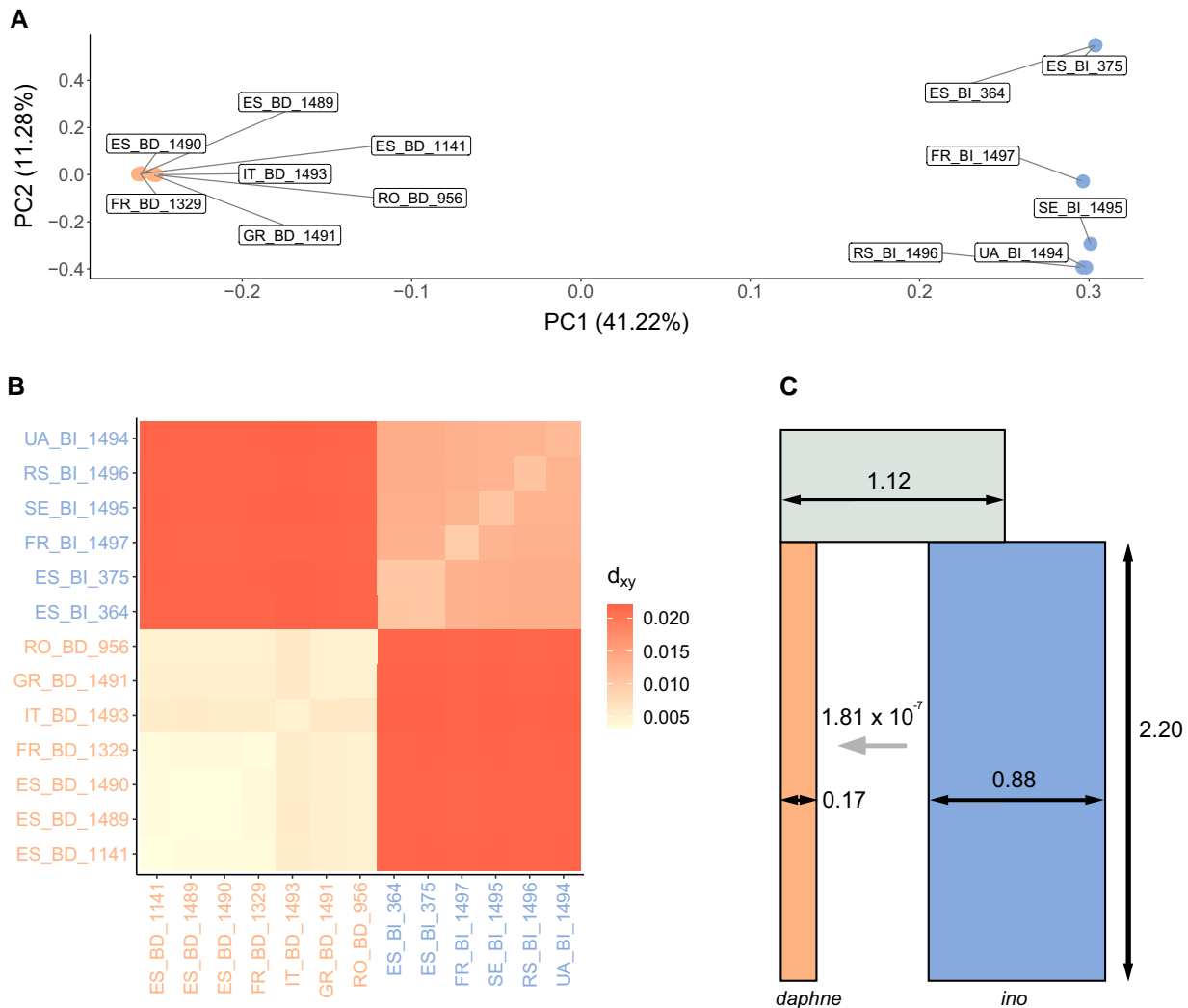
### Diversity and Divergence

Using our previously published *B. ino* genome assembly (Mackintosh et al. 2022) as a reference, we analyzed whole-genome sequence data for seven *B. daphne* and six *B. ino* individuals (fig. 1; supplementary table S1, Supplementary Material online). We restricted our analyses to intergenic regions of the genome, as these typically evolve under less selective constraint than genic regions. Consistent with a previous analysis of transcriptomic data (Ebdon et al. 2021), we find that per-site heterozygosity is greater in *B. ino* (0.0111) than in *B. daphne* (0.0043) and that interspecific divergence is considerable ( $d_{xy} = 0.0228$ ,  $F_{ST} = 0.4976$ ). We also find evidence of population structure within each species (fig. 2A and B). For example, pairwise  $F_{ST}$  is  $\sim 0.1$  for *B. daphne* individuals sampled in different glacial refugia (Iberia, Italy, or the Balkans) and there are similar levels of differentiation between *B. ino* individuals sampled from Iberia and elsewhere in Europe. Although this shows that European *B. daphne* and *B. ino* are not panmictic populations, this should only have a small effect on our analyses of long-term divergence and gene flow between the two species (see below).

### Demographic History

We use gIMble (Laetsch et al. 2022), a recent implementation of a blockwise likelihood calculation (Lohse et al. 2016), to infer the demographic history of speciation between *B. daphne* and *B. ino*. gIMble calculates the blockwise site frequency spectrum (bSFS) of all possible interspecific pairwise comparisons, that is, sampling a single diploid genome from each species and tallying mutations in short blocks of sequence (see Materials and





**Fig. 2.** Diversity and divergence between *B. daphne* and *B. ino*. (A) A PCA of individuals sampled across Europe, with PC1 capturing interspecific variation. Orange points (labelled with IDs containing BD) are *B. daphne* individuals and blue points (labelled with IDs containing BI) are *B. ino* individuals. The same color scheme is used in subplots (B) and (C). (B) A heatmap showing  $d_{xy}$  between pairs of individuals with the diagonal showing heterozygosity within individuals. (C) The best fitting demographic model, with parameter values inferred from the genome-wide bSFS. The  $N_e$  (indicated by horizontal black arrows) and split time (vertical black arrow) parameter estimates are in units of  $10^6$  individuals and years, respectively. The horizontal grey arrow indicates the direction of gene flow, from *B. ino* to *B. daphne*, forwards in time.

Methods). We fit three demographic models to the bSFS: strict divergence (*DIV*) and two scenarios of isolation with migration ( $IM_{\rightarrow Bda}$  and  $IM_{\rightarrow Bin}$ ). The *DIV* model has three  $N_e$  parameters (*B. daphne*, *B. ino*, ancestral) and a split time parameter. The *IM* models have an additional parameter, that is, they assume a constant rate of effective migration ( $m_e$ ) either from *B. ino* into *B. daphne* forwards in time ( $IM_{\rightarrow Bda}$ ) (fig. 2C) or in the opposite direction ( $IM_{\rightarrow Bin}$ ). By optimizing the parameters under each model, we found that the  $IM_{\rightarrow Bda}$  model fits best (table 1; fig. 2C). The *DIV* and  $IM_{\rightarrow Bin}$  models converged to the same parameter values and composite likelihood (table 1), that is, the maximum composite likelihood (MCL) estimate of  $m_e$  under the  $IM_{\rightarrow Bin}$  model is 0. By contrast, the MCL estimate of  $m_e$  from *B. ino* to *B. daphne* under the best fitting ( $IM_{\rightarrow Bda}$ ) model is  $1.811 \times 10^{-7}$ , which is equivalent to

0.124 effective migrants per generation. As a result of this migration, the  $IM_{\rightarrow Bda}$  model also has an older split time ( $\approx 2.2$  MY) than the *DIV*/ $IM_{\rightarrow Bin}$  model ( $\approx 1.2$  MY) (table 1).

Given the nesting of models, an *IM* model has to fit the data equally well or better than a *DIV* model because it includes an additional parameter,  $m_e$ . To test whether the  $IM_{\rightarrow Bda}$  model fits significantly better than *DIV* (see Laetsch et al. 2022), we simulated parametric bootstrap replicates for the MCL estimates under the *DIV* history and optimized both the *DIV* and  $IM_{\rightarrow Bda}$  models. The improvement in fit ( $\Delta \ln CL$ ) between *DIV* and  $IM_{\rightarrow Bda}$  models for parametric bootstrap replicates was far below what we observe in the data (supplementary fig. S1, Supplementary Material online). An *IM* demographic history, with migration from *B. ino* to *B. daphne*, is therefore well supported.

**Table 1.** Maximum Composite Likelihood Parameters for Three Demographic Models. The  $N_e$  and split time parameter estimates are in units of  $10^6$  individuals and years, respectively. The  $IM_{\rightarrow Bda}$  model has the highest lnCL.

Model	$N_e$ <i>daphne</i>	$N_e$ <i>ino</i>	$N_e$ ancestral	$m_e$	Split Time	lnCL
DIV	0.252	0.683	1.433	–	1.183	–234,678,837
$IM_{\rightarrow Bda}$	0.171	0.880	1.116	$1.811 \times 10^{-7}$	2.202	–233,968,576
$IM_{\rightarrow Bin}$	0.252	0.683	1.433	0.000	1.183	–234,678,837

### Synteny

To compare synteny between *B. daphne* and *B. ino*, we generated a chromosome-level assembly for a female *B. daphne* individual, collected in Catalunya, Spain. The assembly is 419.1 Mb in length, with a scaffold N50 of 30.6 Mb and a contig N50 of 13.4 Mb. The *B. daphne* assembly is scaffolded into 13 chromosome-level sequences (hereafter simply referred to as chromosomes) corresponding to 12 autosomes and the Z sex chromosome (fig. 3 and supplementary fig. S2, Supplementary Material online). We failed to scaffold the W chromosome which is likely contained within the remaining 35 contigs that total 5.3 Mb.

A pairwise alignment between the *B. daphne* and *B. ino* assemblies shows that only eight chromosomes have one-to-one homology, with the others showing more complex relationships (fig. 3). For example, *B. daphne* chromosome 1 is homologous to parts of *B. ino* chromosomes 1, 3, and 8 (fig. 3). Altogether, we find that five *B. daphne* chromosomes and six *B. ino* chromosomes are involved in a total of nine inter-chromosomal rearrangements. Hereafter we refer to these chromosomes as rearranged. Additionally, we define rearrangement points as chromosome ends involved in fissions fusions or sites where alignments on either side connect different *B. daphne* and *B. ino* chromosomes. All nine rearrangements points are supported by both HiC data and contig sequences.

From a single pairwise comparison, it is not possible to tell whether a genome possesses a rearrangement in the ancestral or derived state. Therefore, to polarize these rearrangements, we analyzed the assemblies alongside a publicly available genome assembly of *Fabriciana adippe* (see Materials and Methods). We infer a maximally parsimonious history of rearrangements where the common ancestor of *B. daphne* and *B. ino* had 16 chromosomes, with two fissions and five fusions in the *B. daphne* lineage and two fusions in the *B. ino* lineage. This inferred rearrangement history involves two small ancestral chromosomes (approximately 6.6 and 8.4 Mb), which fused independently to different chromosomes in either species (fig. 3).

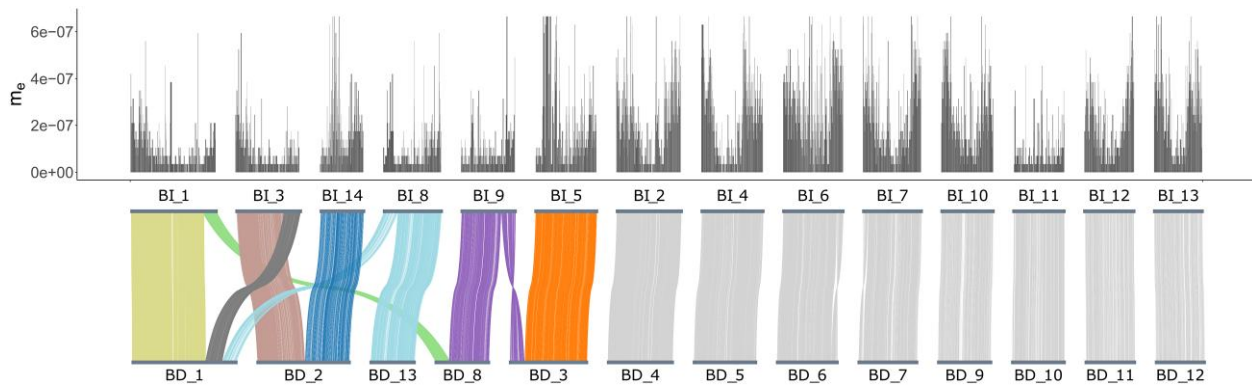
### Variation in $m_e$ Across the Genome

To investigate the effect of rearrangements on reproductive isolation, we followed the approach of Laetsch et al. (2022) by inferring effective population sizes ( $N_e$ ) and the effective migration rate ( $m_e$ ) in windows along the genome. We assume that the species split time is fixed to the MCL estimate under the  $IM_{\rightarrow Bda}$  model (table 1). We used simulations to confirm that, given plausible (but conservative) assumptions about recombination, demographic

parameters could be inferred for windows containing 30,000 consecutive sequence blocks (supplementary note 1 and fig. S3, Supplementary Material online). To infer parameters for the real data, we set up a grid of 67,500 possible parameter value combinations: 15 *B. daphne*  $N_e$  values (20,000–720,000), 15 *B. ino*  $N_e$  values (50,000–2,850,000), 15 ancestral  $N_e$  values (50,000–2,010,000), and 20  $m_e$  values ( $0$ – $6.65 \times 10^{-7}$ ). We identified the best fitting parameter combination for each window (30,000 consecutive blocks, median length = 122 kb). Estimates of local  $m_e$  have a long tailed distribution with a peak at  $3.5 \times 10^{-8}$  (fig. 4). Consistent with the genome-wide estimate, the mean  $m_e$  across windows is  $1.845 \times 10^{-7}$ . We find that  $m_e$  is lower on rearranged chromosomes compared with non-rearranged chromosomes (mean  $m_e = 1.281 \times 10^{-7}$  vs.  $2.292 \times 10^{-7}$  respectively; figs. 3 and 4; one-tailed permutation test  $p < 0.005$ ). This suggests that inter-chromosomal rearrangements are associated with reduced gene flow.

An alternative approach to estimating  $m_e$  for each window is to identify “barrier windows” where there is statistical support for a reduction in gene flow (compared with the background  $m_e$ ). Following Laetsch et al. (2022), we defined barrier windows as those where  $m_e = 0$  has a greater lnCL than  $m_e = 1.75 \times 10^{-7}$  (the grid value nearest to the genome-wide  $m_e$  estimate). Under this criterion, 23.08% of windows are barriers and these are distributed across all 14 *B. ino* chromosomes. However, the number of barrier windows is not equal among *B. ino* chromosomes, for example, 48.11% and 4.22% of windows are barriers on chromosome 3 and chromosome 10, respectively. Windows on rearranged chromosomes are twice as often classified as barriers than windows on non-rearranged chromosomes (32.91% vs. 15.27%; one-tailed permutation test  $p < 0.01$ ). The window with the greatest barrier support ( $\Delta$  lnCL) is located on *B. ino* chromosome 8, with the start of this window being less than 200 kb from a rearrangement point. This alternative, but not independent, estimation of  $m_e$  variation provides further evidence for an association between fission and fusion rearrangements and a reduction in gene flow.

Under the best fitting demographic model (fig. 2C) *B. daphne* receives gene flow from *B. ino*. As a result, low recombination regions in the *B. daphne* genome are expected to have reduced  $m_e$  under the *polygenic barriers* scenario (see Introduction). With this in mind, it is therefore possible that the reduced  $m_e$  for rearranged chromosomes is not the result of a direct barrier effect, but instead an indirect consequence of rearrangements producing large *B. daphne* chromosomes with low recombination rates (e.g. *B. daphne* chromosomes 1, 2, and 3; see fig. 3). To test this possibility, we assigned each genomic window to a *B. daphne*



**Fig. 3.** A whole-genome alignment between *B. daphne* and *B. ino*, with effective migration ( $m_e$ ) estimates for windows along the *B. ino* genome plotted above. Alignments between non-rearranged chromosomes are colored in gray. Alignments between rearranged chromosomes are colored by the inferred chromosomes of the common ancestor of *B. daphne* and *B. ino*. The Z chromosome is labeled as BD\_10 in the *B. daphne* genome and BI\_11 in the *B. ino* genome.

chromosome using a whole-genome alignment (fig. 3) and calculated the mean  $m_e$  of each *B. daphne* chromosome. There is no significant linear relationship between *B. daphne* chromosome length and mean  $m_e$  (Spearman's  $\rho_{df=11} = -0.0769$ ,  $p = 0.8065$ ; fig. 4). Although the largest chromosomes, which happen to be rearranged, do indeed have relatively low  $m_e$ , short rearranged chromosomes also have low  $m_e$ . Additionally, the Z chromosome (*B. daphne* chromosome 10, *B. ino* chromosome 11), which is not rearranged and is short, has low mean  $m_e$ . Chromosome size alone is therefore unlikely to explain the association between chromosome rearrangements and reduced  $m_e$ .

If fission and fusion rearrangements act as direct barriers to gene flow, such as in the *fused barriers* and *underdominance* scenarios, then we would expect loci that are closely linked to rearrangement points to have the greatest reduction in  $m_e$ . This is because loci that are on the same chromosome but are less closely linked will be more likely to recombine away following introgression. Selection against foreign rearrangements will therefore only have a weak effect on loosely linked loci. We indeed find that genomic windows which are located within 1 Mb of a rearrangement point have a lower  $m_e$  (mean =  $5.618 \times 10^{-8}$ ) than those located elsewhere on rearranged chromosomes (mean =  $1.328 \times 10^{-7}$ ) (fig. 4; one-tailed permutation test  $p < 0.0005$ ). All 76 of these windows have estimated  $m_e$  values (between 0 and  $1.75 \times 10^{-7}$ ; fig. 4) that are below the genome-wide estimate ( $1.811 \times 10^{-7}$ ). Additionally, 59.21% of them are classified as barrier windows. The signal of reduced  $m_e$  at closely linked sites provides support for rearrangements having acted as barriers to gene flow.

## Discussion

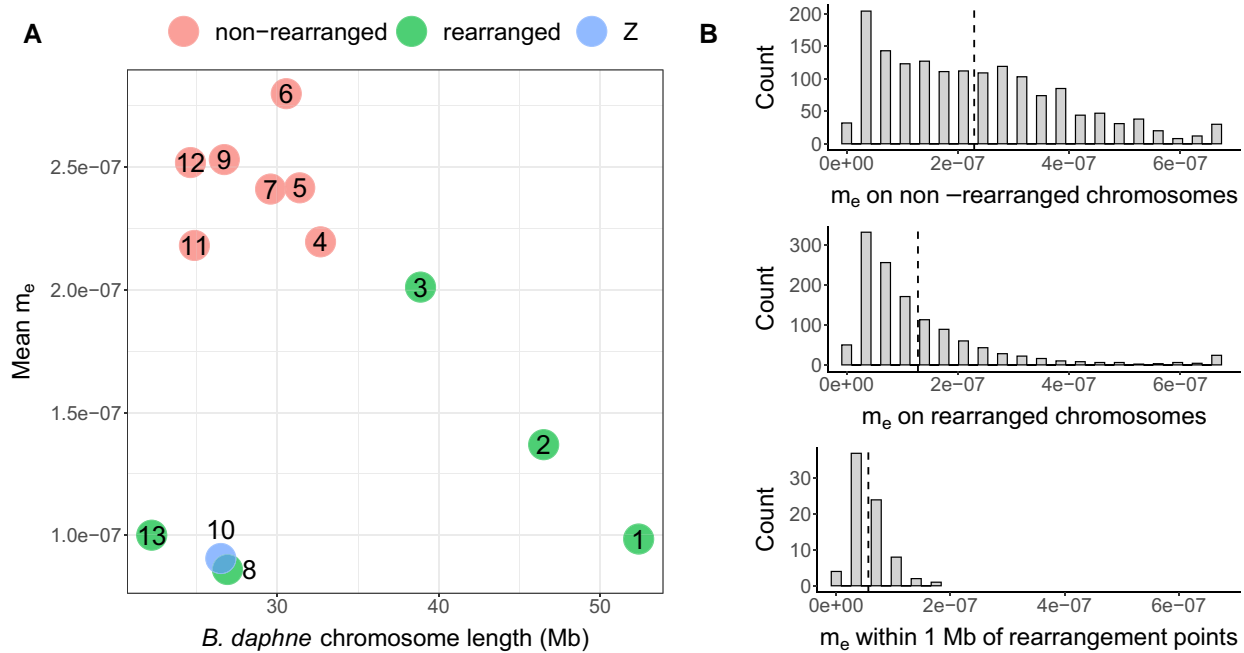
### The Effect of Fission and Fusion Rearrangements on Gene Flow

We have shown that the fritillary butterflies *Brenthis daphne* and *B. ino* possess different karyotypes due to multiple fission and fusion rearrangements, and that these

rearrangements are associated with reduced  $m_e$ . We can therefore reject the *no effect* scenario where rearrangements are only coincidental with speciation.

We considered the possibility that the association between rearrangements and low  $m_e$  could be solely driven by the modification of chromosome lengths, and therefore recombination rate, in the presence of polygenic barriers. Indeed fusions in the *B. daphne* population have generated large (up to 52 Mb) chromosomes with presumably low recombination rates and low  $m_e$ . However, the fact that small chromosomes that are involved in fissions and fusions have reduced  $m_e$  (fig. 4) is not well explained by the *polygenic barriers* scenario where rearrangements only modify the size of chromosomes. We do expect recombination rate to play some role in determining variation in  $m_e$  across the genome (see below). However, given the small number of chromosomes in the focal *Brenthis* pair, the relationship between chromosome length and  $m_e$  variation remains difficult to quantify precisely. Nevertheless, our results—especially the finding of reduced  $m_e$  around rearrangement points—are better explained by localized natural selection against introgression around rearrangements. In other words, rearrangements have acted as barriers to gene flow.

The association between rearrangements and  $m_e$  that we find is consistent with two scenarios, *underdominance* and *fused barriers*. Under the *underdominance* scenario we would expect rearranged chromosomes to have lower  $m_e$  and we would also expect  $m_e$  to be further reduced near rearrangement points. We find both of these patterns in our data (fig. 4). The expectations under the *fused barriers* scenario are more variable. If the number of initial barrier loci is small, and fusions that put two or more barrier loci into strong LD rise in frequency due to natural selection (Guerrero and Kirkpatrick 2014), then we would indeed expect lower  $m_e$  on rearranged chromosomes as well as particularly low  $m_e$  around fusion points. However, if there were enough initial barrier loci so that some were in strong LD by chance alone, then the  $m_e$  of barrier loci brought together by a fusion would be unremarkable. We find that all



**Fig. 4.** Differences in effective migration ( $m_e$ ) between rearranged and non-rearranged chromosomes. (A) Mean  $m_e$  for each *B. daphne* chromosome plotted against its length. Points are colored green if the chromosome is rearranged (chromosomes 1, 2, 3, 8 and 13) and red if not. The Z chromosome (chromosome 10), which is not rearranged, is colored blue. (B) The distribution of window-wise  $m_e$  estimates across non-rearranged chromosomes (top), rearranged chromosomes (middle), and within regions near rearrangement points (bottom). For each plot, the mean is plotted as a dashed vertical line.

rearranged chromosomes have reduced  $m_e$  when compared with other autosomes (fig. 4), which can only be explained by the *fused barriers* scenario if fusions always put barrier loci into strong LD, with their combined effects being greater than barrier loci on non-rearranged chromosomes. One way to discern between the *fused barriers* and *underdominance* scenarios would be to compare  $m_e$  around fission points, as it is only expected to be reduced in the *underdominance* scenario. However, the two fission events in the *B. daphne* lineage are both followed by fusions, making this test inappropriate. So while the *fused barriers* scenario requires a particular number and distribution of initial barrier loci, it is still consistent with our results. Note, again, that the *fused barriers* and *underdominance* scenarios are not mutually exclusive, and both processes could have contributed to fissions and fusions acting as barriers to gene flow between *B. daphne* and *B. ino*.

### The Underdominance Paradox

Earlier we noted that chromosomal speciation models involving underdominance are often paradoxical (see Introduction). So, how could rearrangements rise to high frequency in the *B. daphne* and *B. ino* populations if heterokaryotypes are selected against? The *fused barriers* scenario is one way in which underdominance could be overcome within a population because this scenario involves natural selection favoring the fusions to enhance local adaptation. Although it can only explain the evolution of fission rearrangements if they were translocations instead. Another

solution is that the fitness consequences of heterozygosity for a single fission/fusion are effectively neutral. This is more likely to be the case when chromosomes are holocentric (Lucek et al. 2022), as they are in butterflies (although see Dutrillaux et al. 2022). A single rearrangement could therefore fix in a population and, over time, karyotypes could evolve in a stepwise process. By contrast, heterozygosity for multiple fissions/fusions could have a larger fitness cost due to the difficulty of properly segregating multiple, potentially complex, multivalents (Dutrillaux and Rumpler 1977; Castiglia and Capanna 2000). If *B. daphne* and *B. ino* evolved multiple rearrangements through a stepwise process during a period of allopatry, then rearrangements could act as barriers once the populations came back into contact. This scenario, which has similarities with the stepwise accumulation of Bateson–Dobzhansky–Muller incompatibilities (Dobzhansky 1934), has been previously described by White (1978a), and is known as the chain model (Rieseberg 2001). Although the rearrangements between *B. daphne* and *B. ino* are numerous and complex (fig. 3), consistent with the chain model, we have not tested whether these populations underwent a period of allopatry followed by secondary contact. There may be enough information in the two-diploid bSFS to fit such a model, but no exact likelihood implementation exists yet (although see Beeravolu et al. 2018; Bisschop 2022) and so we have had to assume a simpler IM model in order to investigate variation in  $m_e$  across the genome. Importantly, if the chain model does apply here, it has only generated partial barriers to gene flow and has not



resulted in complete reproductive isolation. If hybrids with heterokaryotypes were sterile, then gene flow would cease across the entire genome. We instead find that gene flow is reduced on rearranged chromosomes, which means that heterokaryotype hybrids must have been able to backcross.

### Variation Among Rearrangements

In our analysis, we grouped chromosomes into two categories, rearranged and non-rearranged. Although this simplification is convenient, it ignores potentially important variation among rearrangements. For example, rearrangements could vary in their effect on meiosis. Although most rearrangements will result in multivalents, particularly complex multivalent chains could cause recombination suppression if crossover formation is physically constrained (Borodin et al. 2019). Rearrangements are also likely to vary in terms of their time of origin, with some arising around the split time of *Brenthis daphne* and *B. ino* ( $\approx 2.2$  My), affecting gene flow during the early stages of speciation. Others may have arisen much more recently, and so have made a relatively small addition to existing reproductive isolation. It is also possible that some of the rearrangements we have identified are still polymorphic within species (i.e., not fixed between species). Interestingly, a polymorphic rearrangement could act as a barrier to gene flow within a species. An analysis of intraspecific gene flow (supplementary note 2 and table S2, Supplementary Material online) suggests that the rearrangements we have identified only reduce gene flow between species, rather than between different refugial populations of the same species (supplementary fig. S4, Supplementary Material online). Nonetheless, it is likely that at least some rearrangements are polymorphic given variation in chromosome number within both *B. daphne* (de Lesse 1960; Saitoh 1986) and *B. ino* (Federley 1938; Maeki and Makino 1953; Saitoh 1991). We cannot yet infer an evolutionary history for each rearrangement that is detailed enough to capture its time of origin and frequency over time. However, such detailed reconstructions may become a realistic goal as the quality of data and inference methods improve.

### Other Determinants of $m_e$ Variation

We have focused on whether chromosome rearrangements, the most conspicuous genomic difference between these species, have acted as barriers to gene flow. Yet variation in  $m_e$  across the genome cannot be explained by rearrangements alone. Firstly, the centers of non-rearranged chromosomes clearly have lower  $m_e$  estimates than regions near chromosome ends (fig. 3). This can be explained by variation in recombination rate, with crossovers concentrated towards telomeres (Haenel et al. 2018), as neutral alleles are more likely to introgress if they can quickly recombine away from the barrier loci they are linked to. The fact that chromosome centers consistently have lower  $m_e$  suggests that there are other barriers to gene flow distributed across the genome, not only rearrangement points. Secondly, the Z chromosome has a considerably lower mean  $m_e$  than all other non-rearranged

chromosomes (fig. 4), which cannot be because of rearrangements or low recombination (the Z recombines more frequently than autosomes in Lepidoptera due to achiasmatic meiosis in females with ZW sex chromosomes; Maeda 1939; Turner and Sheppard 1975). Instead, low  $m_e$  on the Z may be a result of recessive barrier loci being exposed to selection in females (Turelli and Orr 1995). Additionally, if the Z evolves faster than autosomal chromosomes (Mongue et al. 2021), then barrier loci, both recessive and dominant, may accumulate faster. Reduced gene flow on the *Brenthis* Z chromosome mirrors findings in other butterfly systems (Rosser et al. 2022; Xiong et al. 2022), as well as in birds (Irwin 2018; Ottenburghs 2022), suggesting that Z chromosomes often accumulate reproductive isolation at a faster rate than autosomes. Given that there are likely many barriers to gene flow between *B. daphne* and *B. ino*, especially on the Z, it may be inaccurate to describe the history of these species as “chromosomal speciation.” Instead, fission and fusion rearrangements are likely one of several processes that have promoted reproductive isolation.

### Outlook

The particular process we have investigated here, where fissions and fusions act as barriers to gene flow, likely modulates speciation more strongly in certain groups of organisms than in others. For example, the majority of butterfly species have very slow karyotypic evolution and thus speciation will have happened through the accumulation of other genetic barriers. Nevertheless, radiations of butterflies where karyotypes evolve quickly (e.g., the genera *Erebia*, *Lysandra*, and *Polyommatus*) may be partly explained by fissions and fusions acting as barriers to gene flow. This could also be true for other radiations in which karyotypes vary, such as Rock-wallabies (Potter et al. 2017), Morabine grasshoppers (White et al. 1964; Kawakami et al. 2011), and *Carex* sedges (Márquez-Corro et al. 2021). Evidence for fissions and fusions promoting speciation has often been macro-evolutionary, where analyses of large phylogenetic trees have shown an association between rearrangement and diversification rates. By contrast, focusing on a single pair of species, we have shown that fissions and fusions can act as barriers to gene flow and that their effect can be quantified from genomic data.

## Materials and Methods

### Sampling

Butterflies were collected by hand netting. Individuals collected by KL were flash frozen in a liquid nitrogen dry shipper (supplementary table S1, Supplementary Material online); those collected by RV and collaborators were dried and, after some days, stored in ethanol at  $-20^\circ\text{C}$  (supplementary table S1, Supplementary Material online).

### Sequencing

Previously published data—the *B. ino* genome assembly and whole-genome sequencing (WGS) data from three

individuals (NCBI accessions: GCA\_921882275.1; ERX7241006; ERX7249694; ERX7250096)—were used in this study (supplementary table S1, Supplementary Material online). The sequencing process for generating these data are described in Mackintosh et al. (2022). Additional sequence data—Pacbio long reads, HiC data, and WGS data for ten individuals—were generated for this study (supplementary table S1, Supplementary Material online).

A high molecular weight (HMW) DNA extraction was performed for *B. daphne* individual ES\_BD\_1141 (supplementary table S1, Supplementary Material online), using a salting-out protocol (see Mackintosh et al. 2022 for details). A SMRTbell sequencing library was generated from the HMW extraction by the Exeter Sequencing Service. This was sequenced on three SMRT cells on a Sequel I instrument to generate 20.4 Gb of Pacbio continuous long read (CLR) data.

A second *B. daphne* individual (FR\_BD\_1329; supplementary table S1, Supplementary Material online) was used for chromatin conformation capture (HiC) sequencing. The HiC reaction was done using an Arima-HiC kit, following the manufacturer's instructions for flash frozen animal tissue. The Illumina TruSeq library was sequenced on an Illumina NovaSeq 6000 at Edinburgh Genomics, generating 9.9 Gb of paired-end reads.

DNA extractions were performed for nine individuals using a Qiagen DNeasy Blood & Tissue kit, following the manufacturers' instructions. TruSeq Nano gel free libraries were prepared from these extractions as well as the HMW extraction of individual ES\_BD\_1141. All ten libraries were sequenced on a NovaSeq 6000 at Edinburgh Genomics, generating between 10.1 and 40.0 Gb of paired-end reads for each sample.

### Genome Assembly

A *B. daphne* genome sequence was assembled from the Pacbio long reads (ES\_BD\_1141), HiC data (FR\_BD\_1329), and WGS data (ES\_BD\_1141) using the same pipeline described in Mackintosh et al. (2022) (Durand et al. 2016; Laetsch and Blaxter 2017; Robinson et al. 2018; Guan et al. 2020; Aury and Istace 2021; Hu 2021), with one modification; YaHS (Zhou et al. 2022) was used to scaffold the contig assembly into chromosomes rather than 3d-dna (Dudchenko et al. 2017).

### Synteny Analysis

To identify rearrangements, the *B. daphne* and *B. ino* assemblies were aligned with minimap2 v2.17 (Li 2018) using the option `-x asm10`. Alignments longer than 50 kb and with a mapping quality of 60 (2563 in total with a mean length of 132 kb) were visualized with minimap2synteny.py. This script (see Data availability) plots the chromosomes of each genome with ribbons connecting regions that align to each other (fig. 3). Fission and fusion

rearrangements were identified from the plot and breakpoints were defined using the paf file generated by minimap2.

To polarize rearrangements and infer ancestral chromosomes, the *Brenthis* assemblies were analyzed alongside a *Fabriciana adippe* genome assembly (NCBI accession: GCA\_905404265.1; Lohse et al. 2022). Single-copy orthologs were identified in each genome using BUSCO v5.3.2 (Simão et al. 2015) with the lepidoptera\_odb10 dataset. Complete and Fragmented BUSCO genes were analyzed with syngaph (<https://github.com/DRL/syngaph>). In brief, syngaph identifies sets of markers, in this case BUSCO genes, that are found on the same chromosome in all three assemblies. Which sets of markers are found together on extant chromosomes is also recorded. Then, given a phylogenetic tree, parsimony is used to estimate the marker content of ancestral chromosomes and the inter-chromosomal rearrangements on each branch.

### Variant Calling and Filtering

Raw WGS reads were adapter and quality trimmed with fastp v0.2.1 (Chen et al. 2018) and aligned to the *B. ino* assembly (GCA\_921882275.1) with bwa-mem v0.7.17 (Li 2013). Duplicates were marked using sambamba v0.6.6 (Tarasov et al. 2015). Variants were called with freebayes v1.3.2-dirty (Garrison and Marth 2012), using the following options: `--limit-coverage 250 --use-best-n-alleles 8 --no-population-priors --use-mapping-quality --ploidy 2 --haplotype-length -1`. This generated a VCF file containing unfiltered SNP and indel calls. Note that the `--limit-coverage 250` and `--use-best-n-alleles 8` options are for computational efficiency only and should not affect whether variants are called at a given site.

Variant calls were filtered using gIMble preprocess (Laetsch et al. 2022), with the following options: `--snp-gap 2 --min_qual 10 --min_depth 8 --max_depth 3`, where `--max_depth` is in units of mean coverage. This generated a VCF of filtered SNPs, where SNPs were not within two bases of an indel and QUAL scores of SNPs were  $\geq 10$ . Individual genotypes were set to missing if read depth was below the minimum depth or above the maximum depth. Sites with multiallelic SNPs were retained if they satisfied all other filtering criteria.

Callable sites for each individual were identified with mosdepth v0.3.2 (Pedersen and Quinlan 2017), called through gIMble preprocess. To restrict downstream analyses to intergenic regions of the genome, the callable sites bed file was stripped of sites belonging to genic and/or repeat regions.

### Summaries of Diversity and Divergence

Variants in intergenic regions of autosomal chromosomes, where all individuals had a genotype, were used to generate a PCA with plink v1.90b6.18 (Purcell et al. 2007).

Genome-wide averages of  $d_{xy}$  and  $F_{ST}$  were calculated from the same set of variants using VCF\_stats.py. The denominator for  $d_{xy}$  was the total number of autosomal intergenic sites that were callable across all individuals (123 Mb out of a possible 150 Mb).

### Demographic Modeling with gIMble

To fit a genome-wide demographic model, autosomal variants were analyzed with gIMble. Blocks of 64 bases, with a max span of 128 bases, were generated for all interspecific pairwise comparisons. A bSFS with a  $k_{max}$  values of 2 was tallied from these blocks. The bSFS contained mutation counts for 81,104,834 interspecific blocks, each of length 64 bases, distributed over 139 Mb of intergenic sequence. Three models (*DIV*, *IM*<sub>→*Bda*</sub>, *IM*<sub>→*Bin*</sub>) were fit to the genome-wide bSFS and the model with the highest lnCL (*IM*<sub>→*Bda*</sub>) was used for downstream analysis. Absolute parameter estimates were calculated by assuming the *de novo* mutation rate estimate for *Heliconius melpomene* ( $2.9 \times 10^{-9}$  mutations per site per generation; Keightley et al. 2015) and a generation time of one year.

Parametric bootstrap simulations were performed with msprime v1.0.2 (Baumdicker et al. 2021), called through gIMble simulate. The simulations were parameterized with the maximum composite likelihood (MCL) *DIV* values, that is, the best fitting history without gene flow, and a per-base recombination rate of  $8.5 \times 10^{-9}$  (equivalent to a single crossover per male meiosis for 14 chromosome pairs). A total of 100 replicates were performed. Each simulated bSFS was optimized under the *DIV* and *IM*<sub>→*Bda*</sub> models.

To estimate variation in  $m_e$  and  $N_e$  across the genome, genomic windows containing 30,000 consecutive blocks were defined. Next, likelihood calculations were generated for a grid of 67,500 parameter combinations using gIMble makegrid. The lnCL of each windowed bSFS was then calculated for every grid-point. The MCL grid-point was recorded for each window. Additionally, MCLs were recorded for each window conditioning on each  $m_e$  value, for example, the MCL of a window considering all grid-points where  $m_e = 0$ .

Variation in  $m_e$  across the Z chromosome was estimated as above, with the following modification: only male individuals (two *B. daphne*, three *B. ino*, [supplementary table S1, Supplementary Material](#) online) were analyzed (since females only have a single copy of the Z). Given the smaller number of interspecific comparisons (6 vs. 42 for the autosomal analysis), we reduced the number of blocks per window accordingly (4,286 consecutive blocks rather than 30,000) to achieve windows of a comparable span.

Demographic models and variation in  $m_e$  were also estimated at the intraspecific level ([supplementary note 2, Supplementary Material](#) online). Individuals within each species were grouped as Iberian if collected in Spain, and Balkan if collected in Serbia, Greece, Romania, or Ukraine. Note that we use the terms Iberian and Balkan to refer to the likely glacial refugia in which populations

expanded from. Genome-wide demographic models were fit to the Iberian–Balkan bSFS for each species. For the *B. daphne* analysis, where the genome-wide model suggested post-divergence gene flow, windows of 4,286 blocks were defined and a grid of 10,000 parameter values was calculated. Windows were then run across the grid (as described above) to obtain  $m_e$  estimates for each window.

### Statistical Analysis

Permutations were used to test whether differences in  $m_e$  between chromosomes were statistically significant. First, a label-switching operation was performed to randomize whether a *B. ino* chromosome was defined as rearranged or non-rearranged, with the rearranged group always consisting of six chromosomes. For each permutation, the differences in mean  $m_e$  and barrier window frequency between the randomly defined groups were calculated and used to build null distributions. The observed differences in mean  $m_e$  and frequency of barrier windows between rearranged and non-rearranged chromosomes were then compared with these distributions to calculate *P*-values.

A second permutation test was used to approximate a null distribution for the difference in mean  $m_e$  between windows within 1 Mb of rearrangement points, and windows that are elsewhere on rearranged chromosomes. For each permutation, nine points were randomly chosen from rearranged chromosomes and adjacent windows around these points were sampled. The number of adjacent windows sampled for each point was matched to a number of adjacent windows within 1 Mb of a rearrangement point in the real data. Permutations where any window was sampled multiple times were discarded. To avoid under-sampling windows near the ends of chromosomes, adjacent windows were allowed to roll over on to the next rearranged chromosome. The difference in mean  $m_e$  between windows adjacent to randomly sampled points and all other windows on rearranged chromosomes, was calculated for each permutation. A total of 100,000 permutations were done to approximate the null distribution. The difference in mean  $m_e$  between windows within 1 Mb of rearrangement points and windows that are elsewhere on rearranged chromosomes, was compared with the null distribution to calculate a *P*-value.

Spearman's  $\rho$  was calculated for chromosome length and mean  $m_e$ . All analyses were performed in R (R Core Team 2021).

### Supplementary Material

[Supplementary data](#) are available at *Molecular Biology and Evolution* online.

### Acknowledgments

We would like to thank Marian Thompson and Robert Foster (both Edinburgh Genomics) for preparing Pacbio and HiC sequencing libraries and Katy MacDonald for



help in the molecular lab. We also thank Maria Jesus Cañal Villanueva and Luis Valledor (Universidad de Orviedo) for help with fieldwork logistics, as well as Vlad Dincă, Raluca Vodă, and Sabina Vila for contributing samples. We would like to thank Staffan Bensch and Deborah Charlesworth for insightful comments on an earlier version of the manuscript and Sam Ebdon for helping to improve figure 1. A.M. is supported by an E4 PhD studentship from the Natural Environment Research Council (NE/S007407/1). K.L. is supported by a fellowship from the Natural Environment Research Council (NERC, NE/L011522/1). R.V. is supported by Grant PID2019-107078GB-I00 funded by Ministerio de Ciencia e Innovación and Agencia Estatal de Investigación (MCIN/AEI/10.13039/501100011033). S.H.M. is supported by a Royal Society University Research Fellowship (URF/R1/180682). This work was supported by a European Research Council starting grant (ModelGenomLand 757648) to K.L. and a David Phillips Fellowship (BB/N020146/1) by the Biotechnology and Biological Sciences Research Council (BBSRC) to A.H.

## Data availability

Raw sequencing reads and the *Brenthis daphne* genome assembly are available at the European Nucleotide Archive under project accession PRJEB56310. The scripts VCF\_stats.py and minimap2synteny.py, as well as the R code for carrying out permutation tests are available at [https://github.com/A-J-F-Mackintosh/Mackintosh\\_et\\_al\\_2022\\_Binodaphne](https://github.com/A-J-F-Mackintosh/Mackintosh_et_al_2022_Binodaphne).

**Conflicts of interest.** The authors declare no conflicts of interest.

## References

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

Augustijn H, Baetscher L, Cesanek M, Chkhartishvili T, Dincă V, Iankoshvili G, Ogawa K, Vila R, Klopstein S, de Vos J, et al. 2023. A macroevolutionary role for chromosomal fusion and fission in *Erebia* butterflies. bioRxiv. Unpublished. doi:10.1101/2023.01.16.524200.

Aury JM, Istace B. 2021. Hapo-G, haplotype-aware polishing of genome assemblies with accurate reads. *NAR Genom Bioinform*. **3**:lqab034.

Baker RJ, Bickham JW. 1986. Speciation by monobrachial centric fusions. *Proc Natl Acad Sci*. **83**:8245–8248.

Barton NH, Charlesworth B. 1984. Genetic revolutions, founder effects, and speciation. *Annu Rev Ecol Syst*. **15**:133–164.

Basset P, Yannic G, Brünner H, Hausser J. 2006. Restricted gene flow at specific parts of the shrew genome in chromosomal hybrid zones. *Evolution*. **60**:1718–1730.

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

Beravolu CR, Hickerson MJ, Frantz LA, Lohse K. 2018. ABLE: block-wise site frequency spectra for inferring complex population histories and recombination. *Genom Biol*. **19**:1–16.

Bisschop G. 2022. Graph-based algorithms for Laplace transformed coalescence time distributions. *PLOS Comput Biol*. **18**:1–13.

Borodin PM, Fedyk S, Chetnicki W, Torgasheva AA, Pavlova SV, Searle JB. 2019. Shrews, chromosomes and speciation.

Bush GL, Case S, Wilson A, Patton JL. 1977. Rapid speciation and chromosomal evolution in mammals. *Proc Natl Acad Sci*. **74**:3942–3946.

Castiglia R, Capanna E. 2000. Contact zone between chromosomal races of *Mus musculus domesticus*. 2. Fertility and segregation in laboratory-reared and wild mice heterozygous for multiple robertsonian rearrangements. *Heredity*. **85**:147–156.

Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. **34**:i884–i890.

Davey JW, Barker SL, Rastas PM, Pinharanda A, Martin SH, Durbin R, McMillan WO, Merrill RM, Jiggins CD. 2017. No evidence for maintenance of a sympatric *Heliconius* species barrier by chromosomal inversions. *Evol Lett*. **1**:138–154.

de Lesse H. 1960. Spéciation et variation chromosomique chez les lépidoptères rhopalocères. *Ann Soc Nat, Zool*. **12**(2):1–223.

de Lesse H. 1961. Signification supraspécifique des formules chromosomiques chez les lépidoptères. *Bull Soc Entomol Franc*. **66**:71–83.

de Vos JM, Augustijn H, Båtscher L, Lucek K. 2020. Speciation through chromosomal fusion and fission in Lepidoptera. *Philos Trans R Soc Lond B, Biol Sci*. **375**:20190539.

Dobzhansky T. 1934. Studies on hybrid sterility. *Z Zellforsch Mikrosk Anat*. **21**:169–223.

Dobzhansky T, Epling C. 1948. The suppression of crossing over in inversion heterozygotes of *Drosophila pseudoobscura*. *Proc Natl Acad Sci*. **34**:137–141.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, et al. 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*. **356**:92–95.

Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst*. **3**:95–98.

Dutrillaux B, Dutrillaux AM, McClure M, Gèze M, Elias M, Bed'Hom B. 2022. Improved basic cytogenetics challenges holocentricity of butterfly chromosomes. bioRxiv. Unpublished. doi:10.1101/2022.03.11.484012.

Dutrillaux B, Rumpel Y. 1977. Chromosomal evolution in Malagasy lemurs. *Cytogenet Genome Res*. **18**:197–211.

Ebdon S, Laetsch DR, Dapporto L, Hayward A, Ritchie MG, Dincă V, Vila R, Lohse K. 2021. The Pleistocene species pump past its prime: evidence from European butterfly sister species. *Mol Ecol*. **30**:3575–3589.

Federley H. 1938. Chromosomenzahlen Finnländischer Lepidopteren. *Hereditas*. **24**:397–464.

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. ArXiv e-prints. Unpublished.

Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*. **36**:2896–2898.

Guerrero RF, Kirkpatrick M. 2014. Local adaptation and the evolution of chromosome fusions. *Evolution*. **68**:2747–2756.

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

Hill J, Rastas P, Hornett EA, Neethiraj R, Clark N, Morehouse N, Cols JC, Dirksen H, Meslin C, et al. 2019. Unprecedented reorganization of holocentric chromosomes provides insights into the enigma of lepidopteran chromosome evolution. *Sci Adv*. **5**:eaau3648.

Hu J. 2021. Nextdenovo v2.4.0. Available from: <https://github.com/Nextomics/NextDenovo>.

Irwin DE. 2018. Sex chromosomes and speciation in birds and other ZW systems. *Mol Ecol*. **27**:3831–3851.

Kawakami T, Butlin RK, Cooper SJ. 2011. Chromosomal speciation revisited: modes of diversification in Australian morabine grasshoppers (*Vandiemenella*, *viatica* species group). *Insects*. **2**:49–61.



- Keightley PD, Pinharanda A, Ness RW, Simpson F, Dasmahapatra KK, Mallet J, Davey JW, Jiggins CD. 2015. Estimation of the spontaneous mutation rate in *Heliconius melpomene*. *Mol Biol Evol.* **32**:239–243.
- Kirkpatrick M, Barton N. 2006. Chromosome inversions, local adaptation and speciation. *Genetics* **173**:419–434.
- Kitahara H. 2008. Interspecific hybrid between *Brenthis daphne* and *B. ino* (Lepidoptera, Nymphalidae). *Lepidopteral Soc Jpn.* **59**:144–148.
- Kitahara H. 2012. Artificial interspecific and natural hybrids between sympatric *Brenthis daphne* and *B. ino* (Lepidoptera, Nymphalidae) in Nagano Prefecture, Japan. *Lepidopteral Soc Jpn.* **63**:142–150.
- Laetsch DR, Bisschop G, Martin SH, Aeschbacher S, Setter D, Lohse K. 2022. Demographically explicit scans for barriers to gene flow using gIMble. bioRxiv. Unpublished. 10.1101/2022.10.27.514110.
- Laetsch D, Blaxter M. 2017. Blobtools: interrogation of genome assemblies. *F1000Research* **6**:1287.
- Leaché AD, Banbury BL, Linkem CW, de Oca ANM. 2016. Phylogenomics of a rapid radiation: is chromosomal evolution linked to increased diversification in north american spiny lizards (Genus *Sceloporus*)? *BMC Evol Biol.* **16**:1–16.
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**:3094–3100.
- Lohse K, Chmelik M, Martin SH, Barton NH. 2016. Efficient strategies for calculating blockwise likelihoods under the coalescent. *Genetics* **202**:775–786.
- Lohse K, Vila R, Hayward A, Laetsch DR, Wahlberg N, Darwin Tree of Life Consortium, et al. 2022. The genome sequence of the high brown fritillary, *Fabriciana adippe* (Dennis & Schiffermüller, 1775). *Wellcome Open Res.* **7**:298.
- Lucek K, Augustijnen H, Escudero M. 2022. A holocentric twist to chromosomal speciation? *Trends Ecol Evol.* **37**(8):655–662.
- Lukhtanov VA, Dincă V, Friberg M, Šichová J, Olofsson M, Vila R, Marec F, Wiklund C. 2018. Versatility of multivalent orientation, inverted meiosis, and rescued fitness in holocentric chromosomal hybrids. *Proc Natl Acad Sci.* **115**:E9610–E9619.
- Lukhtanov VA, Dincă V, Talavera G, Vila R. 2011. Unprecedented within-species chromosome number cline in the Wood White butterfly *Leptidea sinapis* and its significance for karyotype evolution and speciation. *BMC Evol Biol.* **11**:1–11.
- Lukhtanov VA, Kandul NP, Plotkin JB, Dantchenko AV, Haig D, Pierce NE. 2005. Reinforcement of pre-zygotic isolation and karyotype evolution in *Agrodiaetus* butterflies. *Nature* **436**:385–389.
- Mackintosh A, Laetsch DR, Baril T, Foster RG, Dincă V, Vila R, Hayward A, Lohse K. 2022. The genome sequence of the lesser marbled fritillary, *Brenthis ino*, and evidence for a segregating neo-Z chromosome. *G3* **12**:jkac069.
- Maeda T. 1939. Chiasma studies in the silk worm, *Bombyx mori* L. *Jpn J Genet.* **15**:118–127.
- Maeki K, Makino S. 1953. Chromosome numbers of some Japanese Rhopalocera. *Lepid News.* **7**:36–38.
- Márquez-Corro JI, Martín-Bravo S, Jiménez-Mejías P, Hipp AL, Spalink D, Naczi RF, Roalson EH, Luceño M, Escudero M. 2021. Macroevolutionary insights into sedges (*Carex*: Cyperaceae): the effects of rapid chromosome number evolution on lineage diversification. *J Syst Evol.* **59**:776–790.
- Martin SH, Davey JW, Salazar C, Jiggins CD. 2019. Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLoS Biol.* **17**:e2006288.
- Mongue AJ, Hansen ME, Walters JR. 2021. Support for faster and more adaptive Z chromosome evolution in two divergent lepidopteran lineages. *Evolution* **76**:332–345.
- Navarro A, Barton NH. 2003. Accumulating postzygotic isolation genes in parapatry: A new twist on chromosomal speciation. *Evolution* **57**:447–459.
- Noor MA, Grams KL, Bertucci LA, Reiland J. 2001. Chromosomal inversions and the reproductive isolation of species. *Proc Natl Acad Sci.* **98**:12084–12088.
- Ottenburghs J. 2022. Avian introgression patterns are consistent with Haldane's rule. *J Hered.* **113**:363–370.
- Pedersen BS, Quinlan AR. 2017. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **34**:867–868.
- Potter S, Bragg JG, Blom MP, Deakin JE, Kirkpatrick M, Eldridge MD, Moritz C. 2017. Chromosomal speciation in the genomics era: disentangling phylogenetic evolution of rock-wallabies. *Front Genet.* **8**:10.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Human Genet.* **81**:559–575.
- R Core Team. 2021. R: A language and environment for statistical computing.
- Rieseberg LH. 2001. Chromosomal rearrangements and speciation. *Trends Ecol Evol.* **16**:351–358.
- Robinson JT, Turner D, Durand NC, Thorvaldsdóttir H, Mesirov JP, Aiden EL. 2018. Juicebox.js provides a cloud-based visualization system for Hi-C data. *Cell Syst.* **6**:256–258.e1.
- Rosser N, Edelman NB, Queste LM, Nelson M, Seixas F, Dasmahapatra KK, Mallet J. 2022. Complex basis of hybrid female sterility and Haldane's rule in *Heliconius* butterflies: Z-linkage and epistasis. *Mol Ecol.* **31**:959–977.
- Saitoh K. 1986. On the haploid chromosome number of *Brenthis daphne iwatensis* (Lepidoptera Nymphalidae) from hamakoshimizu, hokkaido. *Tyo To Ga.* **37**:101–102.
- Saitoh K. 1987. A note on the haploid chromosome number of *Brenthis ino* (Rottemburg, 1775) from Finland (Lepidoptera, Nymphalidae). *Nota Lepidopteral.* **10**:131–132.
- Saitoh K. 1991. Chromosome number of *Brenthis ino* (Rottemburg, 1775) from Sweden (Lepidoptera, Nymphalidae). *Nota Lepidopteral.* **14**:241–243.
- Saitoh K, Abe A, Kumagai Y, Hiroshi O. 1989. Chromosomes of the fritillaries of the genus *Brenthis* (Lepidoptera, Nymphalidae) from Japan II. A chromosome survey in males of *Brenthis ino matsuensis* (Kono, 1931). *Tyo To Ga.* **40**:253–257.
- Saitoh K, Lukhtanov V. 1988. Some chromosomal aspects of *Brenthis hecate* [Denis & Schiffermüller], 1775 from South Altai, USSR (Lepidoptera, Nymphalidae). *Nota Lepidopteral.* **11**:234–236.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**:3210–3212.
- Sturtevant AH. 1921. A case of rearrangement of genes in *Drosophila*. *Proc Natl Acad Sci U S A.* **7**:235–237.
- Talavera G, Lukhtanov VA, Rieppel L, Pierce NE, Vila R. 2013. In the shadow of phylogenetic uncertainty: The recent diversification of *Lysandra* butterflies through chromosomal change. *Mol Phylogenet Evol.* **69**:469–478.
- Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. 2015. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**:2032–2034.
- Templeton AR. 1981. Mechanisms of speciation—a population genetic approach. *Annu Rev Ecol Syst.* **12**:23–48.
- Turelli M, Orr HA. 1995. The dominance theory of Haldane's rule. *Genetics* **140**:389–402.
- Turner J, Sheppard P. 1975. Absence of crossing-over in female butterflies (*Heliconius*). *Heredity* **34**:265–269.
- Wellenreuther M, Bernatchez L. 2018. Eco-evolutionary genomics of chromosomal inversions. *Trends Ecol Evol.* **33**:427–440.
- White MJD. 1968. Models of speciation. *Science* **159**:1065–1070.
- White M. 1978a. Chain processes in chromosomal speciation. *Syst Zool.* **27**:285–298.
- White MJD. 1978b. *Modes of speciation*. San Francisco: W. H. Freeman.

- White M, Carson H, Cheney J. 1964. Chromosomal races in the Australian grasshopper *Moraba viatica* in a zone of geographic overlap. *Evolution* **18**:417–429.
- Xiong T, Li X, Yago M, Mallet J. 2022. Admixture of evolutionary rates across a butterfly hybrid zone. *Elife* **11**:e78135.
- Yannic G, Basset P, Hausser J. 2009. Chromosomal rearrangements and gene flow over time in an inter-specific hybrid zone of the *Sorex araneus* group. *Heredity* **102**:616–625.
- Yoshida K, Rödelberger C, Röseler W, Riebesell M, Sun S, Kikuchi T, Sommer RJ. 2023. Chromosome fusions repatterned recombination rate and facilitated reproductive isolation during *Pristionchus* nematode speciation. *Nat Ecol Evol.* 1–16.
- Zhou C, McCarthy SA, Durbin R. 2022. YaHS: yet another Hi-C scaffolding tool. bioRxiv. Unpublished. <https://doi.org/10.1101/2022.06.09.495093>.