Genome assembly of three Amazonian *Morpho* butterfly species reveals Z-chromosome rearrangements between closely-related species living in sympatry

Héloïse Bastide[1**], Manuela López-Villavicencio [2**], David Ogereau [1], Joanna Lledo[3], Anne-Marie Dutrillaux[2], Vincent Debat[2], Violaine Llaurens[2,*]

1 IDEEV, Bât. 680,12, route 128 91190 Gif Sur Yvette, France
2 Institut de Systématique, Evolution et Biodiversité (UMR 7205 CNRS/MNHN/SU/EPHE/UA), Muséum National d'Histoire Naturelle - CP50, 45 rue Buffon 75005 PARIS, France
3 GeT-PlaGe, Bât G2, INRAe, 24 chemin de borde rouge - Auzeville, CS 52627, 31326 CASTANET-TOLOSAN Cedex, France

***violaine.llaurens@mnhn.fr **contributed equally**

Abstract

The genomic processes enabling speciation and the coexistence of species in sympatry are still largely unknown. Here we describe the whole genome sequence of three closely-related species from the butterfly genus *Morpho* : *Morpho achilles* (Linnaeus, 1758), *M. helenor* (Cramer, 1776) and *M. deidamia* (Hübner, 1819). These large blue butterflies are emblematic species of the Amazonian rainforest. They live in sympatry in a wide range of their geographical distribution and display parallel diversification of dorsal wing colour pattern, suggesting local mimicry. By sequencing, assembling and annotating their genomes, we aim at uncovering pre-zygotic barriers preventing gene flow between these sympatric species. We found a genome size of 480 Mb for the three species and a chromosomal number ranging from $2n = 54$ for *M. deidamia* to $2n = 56$ for *M. achilles* and *M. helenor*. We also detected inversions on the sex chromosome Z that were differentially fixed between species, suggesting that chromosomal rearrangements may contribute to their reproductive isolation. The annotation of their genomes allowed us to recover in each species at least 12,000 protein-coding genes and to discover duplications of genes potentially involved in pre-zygotic isolation like genes controlling colour discrimination (*L-opsin*). Altogether, the assembly and the annotation of these three new reference genomes open new research avenues into the genomic architecture of speciation and reinforcement in sympatry, establishing *Morpho* butterflies as a new eco-evolutionary model.

Introduction

Chromosomal rearrangements are likely to play a major role in both adaptation and speciation processes [29]. Inversions, for instance, can favour the emergence of adaptive syndromes by locking together co-adapted allelic variations [26]. Chromosomal rearrangements have also been suggested to contribute to reproductive isolation between species by promoting divergent adaptation or by bringing together genetic incompatibilities [19]. Nevertheless, the role of structural variants in these evolutionary processes is still largely unknown. Recently-developed sequencing and assembly

methods now provide facilitated access to complete genomes, therefore opening the investigation of structural variation within and among species (see [43] for a review).

Here, we focus on emblematic species of the Amazonian rainforest, the blue *Morpho*. We describe the whole genomes of three closely-related *Morpho* species living in sympatry for a large range of their geographical distribution (fig. 1) : *M. helenor*, *M. achilles* and *M. deidamia* [4], thereby developing relevant resources to study the evolution of barrier to gene flow in sympatry. In Lepidoptera, specialization towards host-plant has been shown to be a major factor affecting species diversification [1]. Such ecological specialization may favour speciation and co-existence in sympatry, and may stem from the evolution of gustatory receptors enabling plant recognition by females [6].

The evolution of visual [45] and olfactory signals [55] between species may also limit gene flow between sympatric species of Lepidoptera. In the three *Morpho* species studied here, both males and females display conspicuous iridescent blue colour patterns on the dorsal side of their wings, combined with cryptic brownish colour on the ventral side [17]. Such a combination of dorso-ventral pattern, associated with a fast and erratic flight, is thought to contribute to the high escape abilities of these butterflies, promoting colour pattern convergence between sympatric species (*i.e.* escape mimicry, [50]). Parallel geographic variation of dorsal wing colour pattern has indeed been detected in the three *Morpho* species studied here, suggesting local convergence promoted by predators behaviour [36]. Given the key role of colour pattern in both sexual selection and species recognition in diurnal butterflies, such a resemblance is thought to enhance reproductive interference between sympatric species [5]. Behavioural experiments carried out in the wild revealed that males from the three mimetic *Morpho* species are indeed attracted by both intra and interspecific wing patterns [34]. Despite this heterospecific attraction of males at long distances, RAD-sequencing markers revealed a highly limited gene flow between these three sympatric species [34]. This might be due to the differences in the timing of daily activities observed between these sympatric species limiting heterospecific encountering [34]. This divergence in daily phenology may contribute to the initiation of speciation or to the reinforcement of pre-zygotic barrier to heterospecific matings.

Genetic incompatibilities may also contribute to the speciation and reinforcement processes by generating post-zygotic barriers. For instance, variation in chromosome numbers has been shown to correlate with the speciation rate in Lepidoptera [16]. Similarly, chromosomal inversions may fuel the speciation process: by capturing genetic variations, inversions may lead to increased genetic divergence between species. Such divergence may lead to maladaption in hybrids and further limit gene flow between species living in sympatry.

Here, we investigate the structural variations as well as variations in genes potentially contributing to pre-zygotic isolation in the genomes of three sympatric species of *Morpho* butterflies to shed light on the genomic processes involved in sympatric speciation and reinforcement.

We thus study the karyotypes of these three mimetic species and use PacBio-Hifi sequencing of fresh samples to generate complete *de novo* genome assemblies, aiming at detecting chromosomal rearrangements. We also provide their mitogenomes, study their TE contents and annotate the whole genomes. These new genomic resources will open new research avenues into the understanding of adaptive processes, such as convergence evolution of colour pattern or divergence in visual systems, as well as speciation and co-existence of sister-species in sympatry, establishing *Morpho* butterflies as a new eco-evolutionary model.

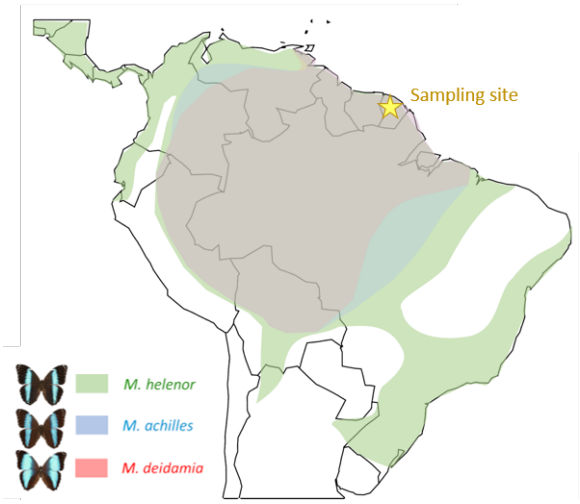

Figure 1. Geographical distribution of the three neotropical species *M. helenor* (green areas), *M. achilles* (blue areas) and *M. deidamia* (red areas). *M. helenor* has the widest distribution, from central America to Southern Brazil, while *M. achilles* and *M. deidamia* are restricted to the Amazonian basin. The three species are in sympatry throughout the Amazonian rainforest, including French Guiana (marked with the yellow star) where the samples studied here were collected.

Materials and Methods

Butterfly sampling

Males from the species *M. helenor*, *M. achilles* and *M. deidamia* were caught with a handnet at the Patawa waterfall, located in the Kaw mountain area of French Guiana (GPS location: 4.54322; 52.15832). In these species, males typically patrol in river beds and are easy to catch, while females are much rarely encountered. We therefore focused on males only. Because in butterflies sex is controlled by a ZW sex chromosome system (females being the heterogametic sex), we were thus able to access the Z sex chromosome but not the W chromosome.

Karyotype study

Cytogenetic techniques were applied to two to four wild caught males per species that were collected at the above-mentioned location in 2019. Their testicles were dissected and processed shortly after capture following the protocol described in [41]. The obtained cell suspension was conserved in fixative at about 4°C. The cell spreading and staining were then performed as described in [41].

DNA extractions and genome sequencing

Live butterflies captured in 2021 at the same site in French Guiana were killed in the lab and their body immediately placed in liquid nitrogen. The DNA extraction was carried out the following day using the Qiagen Genomic-tip 100/G kit and following supplier instructions. The extracted DNA of a single male from each species was used (see supp. fig. 1 for pictures of the wings of the sequenced specimens). Library preparation and sequencing were performed at GeT-PlaGe core facility (INRAe Toulouse) according to the manufacturer's instructions "Procedure and Checklist Preparing HiFi SMRTbell Libraries using SMRTbell Express Template Prep Kit 2.0".

At each step, DNA was quantified using the Qubit dsDNA HS Assay Kit (Life Technologies). DNA purity was tested using the nanodrop (Thermofisher) and size distribution and degradation assessed using the Femto pulse Genomic DNA 165 kb Kit (Agilent). Purification steps were performed using AMPure PB beads (PacBio). 15μg of DNA was purified then sheared at 15kb (speed 31 and 32) with the Megaruptor3 system (Diagenode). Using SMRTbell Express Template prep kit 2.0, a Single strand overhangs removal, a DNA and END damage repair step were performed on 10μg of sample. Blunt hairpin adapters were then ligated to the library, which was treated with an exonuclease cocktail to digest unligated DNA fragments. A size selection step using a 10kb cutoff was performed on the BluePippin Size Selection system (Sage Science) with the "0.75 percent DF Marker S1 6-10 kb vs3 Improved Recovery" protocol. Using Binding kit 2.2 and sequencing kit 2.0, the primer V5 annealed and polymerase 2.2 bounded library was sequenced by diffusion loading onto 1 SMRTcells per sample on SequelII instrument at 80 pM with a 2 hours pre-extension and a 30 hours movie.

K-mer analysis, genome size and heterozygosity estimation

We used Jellyfish (v.2.3.0) [40] to perform a k-mer analysis on each of the PacBioHiFi dataset with a k-mer size of 21. For each HiFi read dataset k-mers were counted and aggregated (jellyfish count option) and histograms were generated using the histo command. The resulting histograms allowed the estimation of genome length and heterozygosity with GenomeScope version 2.0 [52] using the web application.

Nuclear and mitochondrial genome assembly

For the assembly of the nuclear genomes, we compared three long-read assembly tools: IPA-Improved Phased Assembler (v1.0.3-0) (https://github.com/PacificBiosciences/pbipa), Flye (v2.9) [30] and Hifiasm (v0.16.1 with the option -l3 to purge all types of haplotigs in the most aggressive way) [12]. For each assembler, we estimated basic assembly statistics such as scaffold count and N50 using the "stats.sh" program from the BBMap v38.93 package [8]. The completeness of the assembly was assessed using BUSCO v5.2.2 with the *lepidoptera_odb*10 database [39]. We retained the Hifiasm assembly because it had the highest BUSCO score, the highest contiguity (N50) and longest contig. Despite the high level of purge performed by Hifiasm, two species (*M. helenor* and *M. achilles* respectively) retained a high level of duplicates in the BUSCO score. To remove false haplotypic duplications in these two species, we used Purge_dups v1.2.5 setting the cutoffs manually [25]. The completeness of the purged genomes was then reassessed using BUSCO.

The mitochondrial genome of each species was assembled and circularized using Rebaler (https://github.com/rrwick/Rebaler) directly from the PacBio Hifi reads and using the mitochondrial genome of the closely related species *Pararge aegeria* as a reference.

Annotation of repetitive regions

The annotation of repetitive regions in the three species was performed following two main steps. First, we used RepeatModeler v2.0.2a [21] with the option -s (slow search) and -a (to get a .align output file) to create *de novo* libraries of repetitive elements for each species. The library was then used to hardmask the corresponding genome assembly using RepeatMasker 4.1.2.p1 [21] . A summary of the repeated elements was generated with the script 'buildSummary.pl' included in RepeatMasker.

Genome annotation

Each of the three genomes was independently annotated using Maker v2.31.10 [10], following the protocol given in [44]. In a nutshell, Maker is usually run several times successively and use the gene models generated in one round to train *ab initio* gene-predictors and improve the initial gene models in the next round (see below). We used the above-mentioned hardmasked genomes and carried out their annotation using the proteomes of three closely-related species, namely *Pararge aegeria* [18], *Maniola hyperantus* [42] and *Bicyclus anynana* [47]. For each species, the output files were merged into a gff3 file that was then used to generate the necessary files to train SNAP (version 2006-07-28), an *ab initio* gene finding program [31]. A second run of Maker with the above-mentioned gff3 file and the .hmm file provided by SNAP resulted in a second gff3 file that was used to train SNAP a second time. A third round of Maker with the second gff3 and .hmm files was followed by the training of Augustus (3.3.3), another gene prediction tool [33], with the third gff3 file. A final round of Maker with the third gff3 file and the files generated by Augustus led to the fourth and last gff3 file, containing all the genome features for each species.

Protein-Protein BLAST 2.9.0+ (-evalue 1e-6 -max_hsps 1 -max_target_seqs 1) was then used to assess putative protein functions in each *Morpho* species by comparing the protein sequences given by Maker to the protein sequences from the annotated genomes of *Maniola jurtina* [18], *P. aegeria* [18] and *B. anynana* [47]. To specifically compare the exon sequences of the opsins detected in the *Morpho* genomes to the opsins described in other Lepidoptera, we retrieved the coding sequences of opsins from NCBI and used the software Mega v.11 [58] to build a maximum likelihood tree and compute the associated bootstrap values.

Synteny and rearrangement detection

To assess variations in chromosome-scale synteny, we compared the assemblies of each *Morpho* to the assembly of *M. jurtina*, the closest relative of *Morpho* for which a high quality chromosome-level assembly (based in N50 values and Busco score, accession ID GCF_905333055.1) is available [18]. We used MUMmer 3.23 [32] to align the masked assembled genomes of *M. helenor*, *M. achilles* and *M. deidamia* to the *M. jurtina* genome. The output produced by MUMmer is an ASCII delta file that was then filtered and parsed using the utility programs delta-filter and show-coords from MUMmer. We removed short scaffolds, short alignments and low identity alignments with the R script proposed in [11]. Synteny was visualized in R with the packages circlize v 0.4.12 [24] and Paletteer (https://github.com/EmilHvitfeldt/paletteer) using the Rscript from [49].

In order to detect potential genome rearrangements between *Morpho* and closely-related species, we estimated the whole-genome collinearity between the *Morpho* assemblies and five closely-related Nymphalidae species whose genomes exhibit a good-quality assemblies in the NCBI genome database: *M. jurtina* (GCA_905333055.1), *P. aegeria* (GCA_905333055.1), *Erebia ligea* (GCA_923060345.2), *Melanargia galathea* (GCA_920104075.1) and *Lasiommata megera* (GCA_928268935.1) using D-GENIES [9]. Paired alignments between a *Morpho* species and one Nymphalidae species were performed using the minimap2 aligner [35] in D-GENIES, treating each *Morpho* species genome as the query and the Nymphalidae species genome as the target reference. We also used D-GENIES to pair-compare the genomes of the three *Morpho* species. As D-GENIES revealed differences between *Morpho* species in the scaffold corresponding to the Z chromosome (see results), we used SyRI [23] to study in details the rearrangements in the sequences of this scaffold between the three species. We generated paired alignments of the Z scaffold with minimap2 and ran SyRI with the option -c on .sam files. SyRI requires that the two compared genomes represent the

same strand and in the case of *M. achilles*, the orientation of the sequence produced by HiFiasm was the complementary to the sequences of *M. helenor* and *M. deidamia*. We then reverse-complemented this sequence in order to make the alignments. We plotted the genomic structures predicted by SyRI with plotsr [22].

Results

Comparing karyotypes between species

First, we characterized the caryotypes of the three studied species (see sup. fig. 2 to visualize the chromosomes). In *M. helenor*, the detected number of diploid chromosomes ranged from 54 to 56 in the different replicates of mitoses, with a discreet mode at $2n = 56$. This variation is probably due to technical difficulties. The presence of $n = 28$ bivalents in metaphase confirmed the diploid number of $2n = 56$ chromosomes. In *M. achilles*, four specimens had the same modal chromosome counts: mitoses: $2n = 56$ chromosomes; pachynema: $n = 28$ bivalents; Metaphases I: $n = 28$ bivalents; Metaphases II: $n = 28$ chromosomes with 2 chromatids. Surprisingly, the karyotype of the last male was quite different, with a modal number of 84 mitotic chromosomes. Interestingly, there was the same number ($n = 28$) of elements as above at the pachynema stage, indicating that they were trivalents. They were thicker than bivalents and a more careful analysis showed the recurrent asynapsis of one of the 3 chromosomes (sup. fig. 3). No "normal" metaphase I or II was observed. It was concluded that this specimen was triploid with $3n = 84$, and probably sterile. In *M. deidamia*, the diploid chromosome number had a discreet mode of $2n = 54$, suggesting a slightly smaller number of chromosome pairs ($n = 27$) in this more distantly-related species.

Nuclear and mitochondrial genome assembly

Genome size and scaffold number

GenomeScope analyses suggested relatively similar genome sizes (between 470 and 489 Mb) and very high levels of heterozygosity for the three species (table 1). In all of them, the N50 and scaffold sizes were generally larger in the assemblies produced by Hifiasm than in IPA and Flye assemblies (see sup. table 1). The BUSCO scores revealed a very high percentage of repeated sequences, especially in the assemblies produced by IPA and Flye. The use of purge_dups significantly reduced the number of duplicates, the estimated size of the genome and the number of final scaffolds (see sup. fig. 4 and sup. table 1). Hifiasm and the post treatment with Purge_dups v1.2.5 gave an assembly of 143 scaffolds for *M. helenor* (size of the longest scaffold: 42411663bp), of 32 scaffolds for *M. achilles* (size of the longest scaffold: 24854087bp) and of 58 scaffolds for *M. deidamia* (size of the longest scaffold: 22518629bp) (sup. table 1). The Rebaler pipeline identified a circular mitochondrial genome of 15,336 bp for the species *M. helenor*, 15,340 bp for *M. achilles* and 15,196 bp for *M. deidamia*.

Table 1. Haploid size and heterozygosity of the three *Morpho* species estimated by GenomeScope

Species	Haploid size	Heterozygosity(%)
M. helenor	355,313,687 bp	3.35
M. achilles	363,683,954 bp	2.78
M. deidamia	380,442,226 bp	1.68

Annotation of repetitive region

In each out of the three species of *Morpho*, we annotated around 50% of the genome as repeated elements (supplementary figure 5). In *M. helenor*, 241,166,073 bp (51.28% of the genome) corresponded to repeated elements, 261,488,514 bp (54.65% of the genome) in *M. achilles* and 255,779,512bp (52.75% of the genome) in *M. deidamia*. The repetitive elements categories are shown in supplementary figure 5. For the three species, long interspersed nuclear elements (LINE's) accounted for the largest percentage (between 13.53% and 17.22%) of the repeated elements in the genomes.

Genome annotation

We recovered respectively 12,651, 12,978 and 12,093 protein-coding genes in the genomes of *M. helenor*, *M. achilles* and *M. deidamia*. These values are comparable to what was found in *Maniola hyperantus* (13,005 protein-coding genes) and *P. aegeria* (13,515 protein-coding genes), but were lower than in *M. jurtina* (13,777 protein-coding genes) and *B. anynana* (14,413 protein-coding genes). In order to assess if the annotations were complete, we estimated in each species the percentage of proteins with a Pfam domain as this value has been found to vary between 57% and 75% in eukaryotes [60]. This value ranged from 65,50% in *M. achilles* to 71,32% in *M. helenor* with an intermediate value of 70,42% in *M. deidamia*, thus showing that the annotations were of good quality. We were thus able to further investigate gene families that could be involved in pre-zygotic isolation through duplication or loss events. This includes genes having a role in vision (*L-opsin*) but also chemosensory genes such as odorant and gustatory receptors that reflect the degree of species specialization.

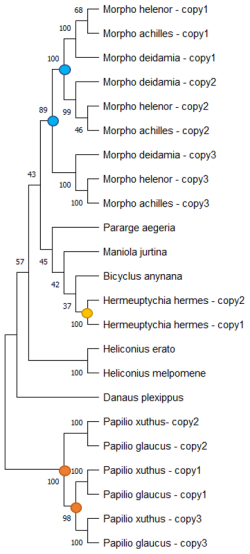

Figure 2. Maximum Likelihood tree of L-opsin exon sequences detected in the genomes of *M. helenor*, *M. achilles* and *M. deidamia* and other butterflies species, with bootstrap values. The colored dots indicate the putative locations of the duplication events on the tree: the putative origin of duplications of the L-opsin observed within the genus *Morpho* appear in blue, while the duplications that occured in the *Hermeuptychia hermes* clade and in the *Papilio* clade appear in yellow and orange respectively

Duplications in opsins genes

Vision in butterflies notably relies on opsins, for which three major types of molecules have been described depending on their wavelength of peak absorbance: in the ultraviolet (UV, 300-400 nm), blue (B, 400-500 nm) and long wavelength (LW, 500-600 nm) part of the visible spectrum. Opsins are respectively encoded by UV, B and LW opsin genes. We investigated the number of copies for each of the opsin gene in the three *Morpho* species. We consistently found one copy of the UV opsin gene and two copies of the B opsin genes in the three *Morpho* species. Duplications of the L-opsin were observed in *M. achilles*, *M. deidamia* and *M. helenor*. In the other reference genomes *M. jurtina*, *B. anynana* and *P. aegeria*, a single copy of the UV opsin gene, the B opsin gene and the LW opsin gene were found. By comparing the L-opsin sequences using a maximum likelihood tree based on the exon sequences, (fig. 2) we showed that the duplications observed in *Morpho* butterflies probably occurred independently from previously described duplications that happened in other clades of Lepidoptera. The phylogenetic relationships between the copies in the three species reveal that the duplications observed in the three *Morpho* species probably occurred before their speciation (fig. 2).

Odorant and gustatory receptors

In order to estimate the number of *Or* and *Gr* genes in the three *Morpho* species, we used the species *Spodoptera littoralis* as a reference. In this moth species, 60 *Or* and 16 *Gr* genes were curated [59]. We independently blasted the proteic sequences predicted by Maker for the three *Morpho* species as well as the annotated sequences of *M. jurtina*, *B. anynana* and *P. aegeria* to the proteic sequences of *Spodoptera littoralis*. Interestingly, we recovered only 31 *Or* genes in *M. helenor*, 32 in *M. achilles* and 36 in *M. deidamia*, while we found 14 *Gr* genes in *M. helenor* and 16 in *M. achilles* and *M. deidamia*. Our three reference species showed a much higher number of *Or* and *Gr* genes with respectively 61 *Or* and 28 *Gr* in *M. jurtina*, 60 *Or* and 35 *Gr* in *B. anynana* and 50 *Or* and 20 *Gr* in *P. aegeria*. The drastic reduction of chemosensory receptors, particularly in the number of *Or* genes in the three *Morpho* species is the sign of a high specialization of individuals to their biochemical environment.

Synteny and rearrangement detection

Conserved synteny with other Lepidoptera species

We found a high concordance between the $n = 29$ chromosomes of *M. jurtina* and the scaffolds of the three *Morpho* species (fig. 3). The MUMmer alignment and the post alignment treatment to remove short scaffolds and low identity alignments reduced the assembly to 27 scaffolds containing 97% of the total genome for *M. helenor* (removing 117 short scaffolds from the original assembly), 29 scaffolds (98% of the genome) for *M. achilles* (3 scaffolds removed) and 27 for *M. deidamia* (31 scaffolds removed). In *M. helenor*, the Hifiasm assembly assigned a single scaffold (ptg000028l) to two different chromosomes from the *M. jurtina* assembly (chromosomes 2 and 6, NC_060030.1 and NC_060034.1) (fig. 3). Similarly, chromosomes NC_060053.1 and NC_060056.1 of *M. jurtina* were not assigned into single scaffolds in *M. helenor* but were distributed into several other scaffolds. In *M. deidamia*, the Hifiasm assembly showed a single scaffold ptg000028l containing chromosomes NC_060051.1 and NC_060052.1 from *M. jurtina*. For the three *Morpho* species, we were able to identify a single scaffold corresponding to the chromosome Z (NC_060058.1) in *M. jurtina* (scaffold ptg000030l in *M. helenor*, scaffold ptg000024l in *M. achilles* and scaffold ptg000019l in *M. deidamia*).

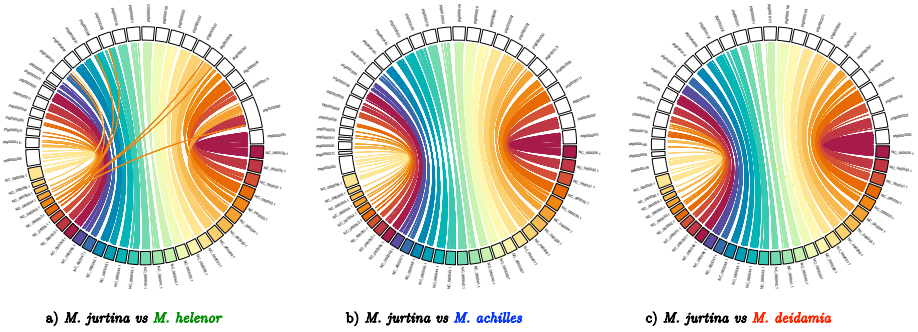

Figure 3. Synteny between the chromosome-assembled genome of *Maniola jurtina* (colored chromosomes) and the genome assemblies of the species *Morpho helenor* (a),*M. achilles* (b) and *M. deidamia* (c). Equivalent chromosomes/scaffolds are linked by same color ribbons

We also found a a high level of colinearity between the genomes of the three *Morpho* species and the five Nymphalidae species used for comparisons. The alignment between *M. jurtina* and the three *Morpho* species (fig. 3) was very similar to the alignments obtained for the other Nymphalidae (sup. fig. 6) and confirmed that the assembly of the genome of *M. helenor* by hifiasm might have merge together two chromsomes: the single scaffold ptg000028l was scattered into two chromosomes in the other Nymphalidae. Although collinearity was generally high, we detected some putative inversions located in regions that varied among pairs for the three *Morpho* species in comparison with the Nymphalidae (see sup. fig. 6). Interestingly, the scaffold corresponding to the chromosome Z was the only one consistently showing inversions in the pairwise genome-wide alignments (see sup. fig. 6).

Inversions in the Z-chromosome between the three sympatric *Morpho* species

The dot-plots from the paired comparisons between the three *Morpho* using D-GENIES showed a very high similarity between genomes (see sup. fig. 7). The only scaffold that differed between species was the one corresponding to the Z chromosome. SyRI identified one inversion of 1.6 Mb between *M. helenor* and *M. deidamia*, five inversions (comprising one of more than 1.8 Mb) between *M. helenor* and *M. achilles* and two between *M. deidamia* and *M. achilles* with one of 1.6 Mb (fig. 4). Interestingly, the inversion found in *M. deidamia* when compared to *M. achilles* or *M. helenor* has the same size and is located in exactly the same position of the chromosome (from bp 1567583 to 3192401), suggesting that this inversion is ancestral to the speciation of *M. achilles* and *M. helenor*. In the case of *M. achilles vs. M. helenor* two inversions were found flanking the site of the putative ancient inversion and a bigger inversion was found at the end of the chromosome (fig. 4).

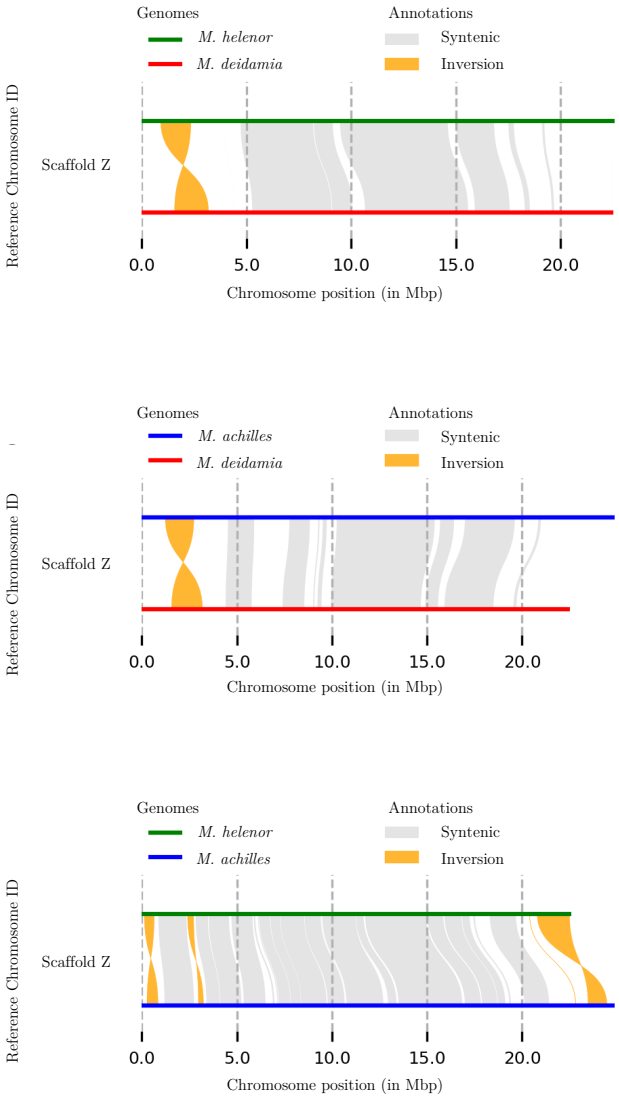

Figure 4. Rearrangement (SyRI) plot of the paired comparisons for the Z scaffold between the three *Morpho* species. Upper figure: *M. helenor* and *M. deidamia*; middle: *M. deidamia* and *M. achilles*; lower: *M. helenor* and *M. deidamia*. SyRI results were plotted using plotsr

Discussion

Assembly of heterozygous Lepidoptera genomes with a high proportion of repeated elements

We generated *de novo*, reference-quality genome assemblies for three emblematic species of Amazonian butterflies: *M. helenor*, *M. achilles* and *M. deidamia*. Our results indicate genome sizes comprised between 470 Mb and 489 Mb, similarly to most of the closely-related Nymphalidae species sequenced so far, *e.g. B. anynana* (475 Mb), *P. aegeria* (479 Mb) or *M. jurtina* (429 Mb). The final number of scaffolds within each of the three species ranged from 27 to 29, close to the number of chromosome pairs observed in our cytogenetics study. The numbers of chromosomes found in those French

Guiana samples (*i.e.* in the subspecies *M. helenor helenor* and *M. achilles achilles*) is consistent with those found in other subspecies of both species in previous studies [7]. The available sequenced species of Nymphalidae that are closely-related to the genus *Morpho* also generally show 29 pairs of chromosomes (28 autosomes, plus Z and W sex chromosomes), which is close to the chromosomal numbers observed in the three *Morpho* species studied here. The mapping between the assemblies of *Morpho* species to the chromosome-level assembly of *Maniola jurtina* and the post-treatment to eliminate small scaffolds allowed us to identify between 27 and 29 scaffolds in *Morpho* that were homologous to *Maniola jurtina* chromosomes, including the scaffold corresponding to the *Z* chromosome. This suggests a high conservation of chromosomal synteny among closely-related Nymphalidae species, which is consistent with the high level of syntheny observed throughout the whole Lepidoptera clade [14]. In the three species, genome heterozygosity was very high (from 1.68% in *M. deidamia* to 3.35% in *Morpho helenor*) and heterozygosity presents a major challenge in *de novo* assembly of diploid genomes. Indeed, levels of heterozygosity of 1% or above are considered "moderate to high" and most assemblers struggle when two divergent haplotypes are sequenced together, as heterozygosity may impair the distinction of different alleles at the same locus from paralogs at different loci [3]. Then, final assemblies of heterozygous genomes are expected to be of poor-quality, highly fragmented and containing redundant contigs [51]. Hifiasm generated the most completely haplotype-resolved assemblies, nevertheless the level of heterozygosity clearly impacted the quality of the assemblies and a post treatment to remove duplicated sequences was necessary for the two most heterozygous genomes (*M. helenor* and *M. achilles*), showing the difficulty that heterozygosity still imposes to long-read heterozygosity-aware assemblers. Such a high heterozygosity has been observed in other genomes of Lepidoptera [47] and can be a signature of high effective population sizes. The wide Amazonian distribution of these species, and their flight activity could contribute to such high level of genetic diversity within population, because elevated dispersal contribute to increase gene flow within each species throughout their geographic range. Our results also showed that around 50% of the genomes of the sequenced *Morpho* was composed of repeated elements, a very high proportion as compared to other genomes of Lepidoptera. In lepidoptera, TE content has been found to be correlated with genome size [57], but in the case of the three *Morpho* species studied here, the repeat content is higher than for other species with similar genome sizes such as the *Bombyx mori* moth, with a genome size estimated at 530 Mb and a TE content of 35% [48] or the more closely-related species *Bicyclus anynana* with a genome size of 475 Mb and a repeat content of 26% [47].

Structural variations between genomes of sympatric species

The karyotype and assembly analyses suggest some differences in chromosome number between the three sympatric *Morpho* species studied here, particularly between *M. deidamia* (27 chromosome pairs) and *M. achilles* (28 or 29 chromosome pairs). Differences in chromosome numbers and other chromosomal rearrangements may strongly affect reproductive barriers. Two groups of models have been proposed to explain how chromosomal rearrangements prevent gene-flow and contribute to species maintenance and speciation. First, hybrid-sterility models suggest reduced fertility or viability in individuals heterozygous for chromosomal rearrangements. These models are considered to be inconsistent and difficult to evaluate [19]. More recently, suppressed-recombination models propose that chromosomal rearrangements permit speciation in sympatry because they reduce recombination between chromosomes carrying different rearrangements [19]. Indeed, in Lepidoptera, differences in chromosome number are proposed to be an important mechanism leading to species diversification in *Agrodiaetus*, *Erebia* and *Lysandra* butterflies ([37, 38, 56]).

Besides differences in chromosome numbers, we systematically found inversions in the Z chromosome when comparing the genomes of *Morpho* to the other Nymphalidae and between the three different *Morpho* species. Inversions are also a type of chromosomal rearrangement known to occur throughout evolution and are considered an important mechanism for speciation particularly for species living in sympatry ([19, 29]). Empirically and theoretically, it has been suggested that inversions may have contributed to speciation in sympatry in different groups of animals. In two ascidians species of the genus *Ciona* and in insects like *Drosophila* inversions may promote speciation by reduction of the fitness or by causing sterility of heterozygotes. In the *Anopheles gambiae* species complex, inversions may allow for ecotypic differentiation and niche partitioning leading to different sympatric and genetically isolated populations ([13, 46, 54]). In groups like paserine birds where sexual differentiation is controlled by a ZW sex chromosome system (females being the heterogametic sex), inversions in the Z chromosome in particular seem to explain speciation in sympatry between close species. Data show that across the Passeriformes, the Z chromosome has accumulated more inversions than any other autosome and that the inversion fixation rate on the Z chromosome is 1.4 times greater than the average autosome. Interestingly, inversions on the Z chromosome are significantly more common in sympatric than in allopatric closely related clades ([27, 28]).

In Lepidoptera, the role of inversions in speciation in sympatry has been studied in the species *Heliconius melpomene* and *H. cydno*, two sympatric species that can hybridize (although rarely) in the wild. The analyses of the genomic differences between the two species showed some small inversions (less than 50 kb) and there was no evidence for a reduction of recombination in hybrids, suggesting that in this case, inversions were not involved in the maintenance of the species barriers and other processes as strong mate preference could prevent hybridisation in the wild [15]. In the *Morpho* studied here however, we found inversions between *Morpho* Z chromosomes that were longer than 1.5 Mb. Models suggest that to be associated with adaptive traits or species barriers, inversions should typically be megabases long in order to be fixed in populations [15]. The position of the inversion in the Z scaffold when comparing *M. helenor* or *M. achilles* to *M. deidamia* is at the exact same place in *M. deidamia*'s genome, suggesting that this specific inversion likely occurred before the speciation between *M. achilles* and *M. helenor*. When comparing *M. helenor* to *M. achilles*, we found two different smaller inversions that are not found in *M. deidamia* and that are close to the putative ancestral inversion region, suggesting that these two smaller inversions could have appeared after the speciation between *M. achilles* and *M. helenor*. At the moment, we do not know what is the frequency of the inversions in the different *Morpho* populations or whether they are fixed. Further population analyses are needed to answer this question and to enlighten what evolutionary forces could be acting to maintain them. The copy number variation detected in genes involved in colour perception (*i.e. L-opsin*) may also play a significant role in reproductive isolation in these sympatric species. For instance, the three copies of LW opsins found in the Papilio genus (fig.) have been found to also show subfunctionalization and neofunctionalization [2]. The duplication followed by genetic divergence observed in these three mimetic *Morpho* species may improve their visual discrimination capacities, and facilitate species recognition, therefore reinforcing barrier to gene flow in sympatry. Genes potentially involved in colour pattern variations (*e.g.* $bric-a-brac$ or *bab*) may also play a role in prezygotic isolation but they were not thoroughly investigated here as their functional evolution involves changes in regulatory sequences rather than events of duplication or gene loss [53]. Interestingly, a blast of the putative proteic sequences of each *Morpho* species against those of *M. jurtina* allowed us to uncover different copy numbers of the gene $bric-a-brac$, which play a significant role in differences of UV

iridescence between males of two incipient species of sulphur butterflies [20]. The copy responsible for the presence/absence of UV iridescence is located on the Z chromosome and in the three *Morpho* species, we found one or more copies of $bric-a-brac$ on the scaffolds that correspond to the Z chromosome: *M. deidamia* had one copy of $bric-a-brac$, while *M. helenor* and *M. achilles* displayed two copies of this gene. It seems however that the second copy in *M. helenor* and *M. achilles* correspond to truncated copies of $bric-a-brac$. While this is certainly the sign of an ancient duplication followed by a pseudogenization event, this could lead to further investigations of putative functions of the truncated copies. It is worth noting that variations in the number of *bab* copies was also observed in the three reference genomes used for the blast: *M. jurtina* had two copies on the Z chromosome (including a truncated copy), *B. anynana* had only one and *P. aegeria* had none.

Altogether, the assembly and annotation of these three mimetic species of *Morpho* butterflies reveal differences in chromosome numbers, the presence of several Mb-long inversions in the Z chromosome, as well as copy number variation and genetic divergence among copies of genes that may play a significant role in reproductive isolation. Our study thus open new avenues into the investigation of the ecological and genomic factors involved in sympatric speciation and its reinforcement.

Supporting Information

Genome assemblies were uploaded at the ENA web site (https://www.ebi.ac.uk/ena/browser/home) under the project number PRJEB56642. Assembled genomes will be available upon acceptance of the submitted manuscript with accession numbers ERZ14213098 for *Morpho helenor*, ERZ14213099 for *M. achilles* and ERZ14213100 for *M. deidamia*. Please see Supplementary material file for supplementary figures and tables.

Acknowledgments

The authors would like to thank Patrick Blandin for his continuous support on our *Morpho* studies. They also thank Mélanie McClure, Mathieu Chouteau, Camille Le Roy and Ombeline Sculfort for their help during field work in French Guiana. We thank Elise Gay, Romuald Laso-Jadart, Pierre Lesturgie, Christelle Fraisse, Clement Gilbert and Quentin Rougemont for help with some scripts and for discussions of previous versions of the manuscript.

References

1. R. Allio, B. Nabholz, S. Wanke, G. Chomicki, O. A. Pérez-Escobar, A. M. Cotton, A.-L. Clamens, G. J. Kergoat, F. A. Sperling, and F. L. Condamine. Genome-wide macroevolutionary signatures of key innovations in butterflies colonizing new host plants. *Nature communications*, 12(1):1–15, 2021.

2. K. Arikawa. Spectral organization of the eye of a butterfly, papilio. *Journal of comparative physiology. A, Neuroethology, sensory, neural, and behavioral physiology*, 189(11):791—800, November 2003.

3. K. C. Asalone, K. M. Ryan, M. Yamadi, A. L. Cohen, W. G. Farmer, D. J. George, C. Joppert, K. Kim, M. F. Mughal, R. Said, M. Toksoz-Exley, E. Bisk, and J. R. Bracht. Regional sequence expansion or collapse in heterozygous genome assemblies. *PLOS COMPUTATIONAL BIOLOGY*, 16(7), JUL 2020.

4. P. Blandin and B. Purser. Evolution and diversification of neotropical butterflies: Insights from the biogeography and phylogeny of the genus morpho fabricius, 1807 (nymphalidae: Morphinae), with a review of the geodynamics of south aamerica. *Tropical Lepidoptera Research*, pages 62–85, 2013.

5. G. Boussens-Dumon and V. Llaurens. Sex, competition and mimicry: an eco-evolutionary model reveals unexpected impacts of ecological interactions on the evolution of phenotypes in sympatry. *Oikos*, 130(11):2028–2039, 2021.

6. A. D. Briscoe, A. Macias-Munoz, K. M. Kozak, J. R. Walters, F. Yuan, G. A. Jamie, S. H. Martin, K. K. Dasmahapatra, L. C. Ferguson, J. Mallet, et al. Female behaviour drives expression and evolution of gustatory receptors in butterflies. *PLoS genetics*, 9(7):e1003620, 2013.

7. K. S. Brown, A. V. Freitas, B. Von Schoultz, A. O. Saura, and A. Saura. Chromosomal evolution of south american frugivorous butterflies in the satyroid clade (nymphalidae: Charaxinae, morphinae and satyrinae). *Biological Journal of the Linnean Society*, 92(3):467–481, 2007.

8. B. Bushnell. Bbmap: A fast, accurate, splice-aware aligner.[www document]. *URL https://www. osti. gov/se rvlets/purl/1241166*, 2014.

9. F. Cabanettes and C. Klopp. D-genies: dot plot large genomes in an interactive, efficient and simple way. *PeerJ*, 6:e4958, 2018.

10. B. L. Cantarel, I. Korf, S. M. Robb, G. Parra, E. Ross, B. Moore, C. Holt, A. Sánchez Alvarado, and M. Yandell. Maker: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research*, 18(1):188–196, 2008.

11. M. d. l. P. Celorio-Mancera, P. Rastas, R. A. Steward, S. Nylin, and C. W. Wheat. Chromosome level assembly of the comma butterfly (polygonia c-album). *Genome biology and evolution*, 13(5):evab054, 2021.

12. H. Cheng, G. T. Concepcion, X. Feng, H. Zhang, and H. Li. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *NATURE METHODS*, 18(2):170+, FEB 2021.

13. M. Coluzzi, A. Sabatini, A. della Torre, M. Di Deco, and V. Petrarca. A polytene chromosome analysis of the anopheles gambiae species complex. *SCIENCE*, 298(5597):1415–1418, NOV 15 2002.

14. E. d'Alencon, H. Sezutsu, F. Legeai, E. Permal, S. Bernard-Samain, S. Gimenez, C. Gagneur, F. Cousserans, M. Shimomura, A. Brun-Barale, et al. Extensive synteny conservation of holocentric chromosomes in lepidoptera despite high rates of local genome rearrangements. *Proceedings of the National Academy of Sciences*, 107(17):7680–7685, 2010.

15. J. W. Davey, S. L. Barker, P. M. Rastas, A. Pinharanda, S. H. Martin, R. Durbin, W. O. McMillan, R. M. Merrill, and C. D. Jiggins. No evidence for maintenance of a sympatric heliconius species barrier by chromosomal inversions. *EVOLUTION LETTERS*, 1(3):138–154, AUG 2017.

16. J. M. de Vos, H. Augustijnen, L. Bätscher, and K. Lucek. Speciation through chromosomal fusion and fission in lepidoptera. *Philosophical Transactions of the Royal Society B*, 375(1806):20190539, 2020.

17. V. Debat, S. Berthier, P. Blandin, N. Chazot, M. Elias, D. Gomez, and V. Llaurens. Why are morpho blue? In *Biodiversity and evolution*, pages 139–174. Elsevier, 2018.

18. E. A. Ellis, C. G. Storer, and A. Y. Kawahara. *De novo* genome assemblies of butterflies. *GigaScience*, 10(6), 06 2021. giab041.

19. R. Faria and A. Navarro. Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends in ecology & evolution*, 25(11):660–669, 2010.

20. V. Ficarrotta, J. J. Hanly, L. S. Loh, C. M. Francescutti, A. Ren, K. Tunström, C. W. Wheat, A. H. Porter, B. A. Counterman, and A. Martin. A genetic switch for male uv iridescence in an incipient species pair of sulphur butterflies. *Proceedings of the National Academy of Sciences*, 119(3):e2109255118, 2022.

21. J. M. Flynn, R. Hubley, C. Goubert, J. Rosen, A. G. Clark, C. Feschotte, and A. F. Smit. Repeatmodeler2 for automated genomic discovery of transposable element families. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 117(17):9451–9457, APR 28 2020.

22. M. Goel and K. Schneeberger. plotsr: visualizing structural similarities and rearrangements between multiple genomes. *Bioinformatics*, 38(10):2922–2926, 2022.

23. M. Goel, H. Sun, W.-B. Jiao, and K. Schneeberger. Syri: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome biology*, 20(1):1–13, 2019.

24. Z. Gu, L. Gu, R. Eils, M. Schlesner, and B. Brors. Circlize implements and enhances circular visualization in r. *Bioinformatics*, 30(19):2811–2812, 2014.

25. D. Guan, S. A. McCarthy, J. Wood, K. Howe, Y. Wang, and R. Durbin. Identifying and removing haplotypic duplication in primary genome assemblies. *BIOINFORMATICS*, 36(9):2896–2898, MAY 1 2020.

26. A. A. Hoffmann, C. M. Sgrò, and A. R. Weeks. Chromosomal inversion polymorphisms and adaptation. *Trends in Ecology & Evolution*, 19(9):482–488, 2004.

27. D. M. Hooper, S. C. Griffith, and T. D. Price. Sex chromosome inversions enforce reproductive isolation across an avian hybrid zone. *MOLECULAR ECOLOGY*, 28(6, SI):1246–1262, MAR 2019.

28. D. M. Hooper and T. D. Price. Chromosomal inversion differences correlate with range overlap in passerine birds. *NATURE ECOLOGY & EVOLUTION*, 1(10):1526–1534, OCT 2017.

29. M. Kirkpatrick and N. Barton. Chromosome inversions, local adaptation and speciation. *Genetics*, 173(1):419–434, 2006.

30. M. Kolmogorov, J. Yuan, Y. Lin, and P. A. Pevzner. Assembly of long, error-prone reads using repeat graphs. *NATURE BIOTECHNOLOGY*, 37(5):540+, MAY 2019.

31. I. Korf. Gene finding in novel genomes. *BMC bioinformatics*, 5(1):1–9, 2004.

32. S. Kurtz, A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway, C. Antonescu, and S. L. Salzberg. Versatile and open software for comparing large genomes. *Genome biology*, 5(2):1–9, 2004.

33. S. König, L. W. Romoth, L. Gerischer, and M. Stanke. Simultaneous gene finding in multiple genomes. *Bioinformatics*, 32(22):3388–3395, 07 2016.

34. C. Le Roy, C. Roux, E. Authier, H. Parrinello, H. Bastide, V. Debat, and V. Llaurens. Convergent morphology and divergent phenology promote the coexistence of morpho butterfly species. *Nature communications*, 12(1):1–9, 2021.

35. H. Li. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18):3094–3100, 2018.

36. V. Llaurens, Y. Le Poul, A. Puissant, P. Blandin, and V. Debat. Convergence in sympatry: Evolution of blue-banded wing pattern in morpho butterflies. *Journal of Evolutionary Biology*, 34(2):284–295, 2021.

37. K. Lucek. Evolutionary mechanisms of varying chromosome numbers in the radiation of erebia butterflies. *GENES*, 9(3), MAR 2018.

38. V. Lukhtanov, N. Kandul, J. Plotkin, A. Dantchenko, D. Haig, and N. Pierce. Reinforcement of pre-zygotic isolation and karyotype evolution in agrodiaetus butterflies. *NATURE*, 436(7049):385–389, JUL 21 2005.

39. M. Manni, M. R. Berkeley, M. Seppey, F. A. Simao, and E. M. Zdobnov. Busco update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *MOLECULAR BIOLOGY AND EVOLUTION*, 38(10):4647–4654, OCT 2021.

40. G. Marcais and C. Kingsford. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *BIOINFORMATICS*, 27(6):764–770, MAR 15 2011.

41. M. McClure, B. Dutrillaux, A.-M. Dutrillaux, V. Lukhtanov, and M. Elias. Heterozygosity and chain multivalents during meiosis illustrate ongoing evolution as a result of multiple holokinetic chromosome fusions in the genus melinaea (lepidoptera, nymphalidae). *Cytogenetic and Genome Research*, 153(4):213–222, 2017.

42. D. Mead, I. Saccheri, C. J. Yung, K. Lohse, C. Lohse, P. Ashmole, M. Smith, C. Corton, K. Oliver, J. Skelton, et al. The genome sequence of the ringlet, *Aphantopus hyperantus* linnaeus 1758. *Wellcome Open Research*, 6(165):165, 2021.

43. C. Mérot, R. A. Oomen, A. Tigano, and M. Wellenreuther. A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends in Ecology & Evolution*, 35(7):561–572, 2020.

44. H. Muller, D. Ogereau, J.-L. Da Lage, C. Capdevielle, N. Pollet, T. Fortuna, R. Jeannette, L. Kaiser, and C. Gilbert. Draft nuclear genome and complete mitogenome of the Mediterranean corn borer, *Sesamia nonagrioides*, a major pest of maize. *G3 Genes—Genomes—Genetics*, 11(7), 05 2021. jkab155.

45. R. E. Naisbit, C. D. Jiggins, and J. Mallet. Disruptive sexual selection against hybrids contributes to speciation between heliconius cydno and heliconius melpomene. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1478):1849–1854, 2001.

46. M. Noor, K. Grams, L. Bertucci, and J. Reiland. Chromosomal inversions and the reproductive isolation of species. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 98(21):12084–12088, OCT 9 2001.

47. R. W. Nowell, B. Elsworth, V. Oostra, B. J. Zwaan, C. W. Wheat, M. Saastamoinen, I. J. Saccheri, A. E. van't Hof, B. R. Wasik, H. Connahs, M. L. Aslam, S. Kumar, R. J. Challis, A. Monteiro, P. M. Brakefield, and M. Blaxter. A high-coverage draft genome of the mycalesine butterfly *Bicyclus anynana*. *GigaScience*, 6(7), 05 2017. gix035.

48. M. Osanai-Futahashi, Y. Suetsugu, K. Mita, and H. Fujiwara. Genome-wide screening and characterization of transposable elements and their distribution analysis in the silkworm, bombyx mori. *INSECT BIOCHEMISTRY AND MOLECULAR BIOLOGY*, 38(12):1046–1057, DEC 2008.

49. J. B. Pettengill, A. W. Pightling, J. D. Baugher, H. Rand, and E. Strain. Real-time pathogen detection in the era of whole-genome sequencing and big data: comparison of k-mer and site-based methods for inferring the genetic distances among tens of thousands of salmonella samples. *PLoS One*, 11(11):e0166162, 2016.

50. C. Pinheiro, A. Freitas, V. Campos, P. DeVries, and C. Penz. Both palatable and unpalatable butterflies use bright colors to signal difficulty of capture to predators. *Neotropical Entomology*, 45(2):107–113, 2016.

51. L. P. Pryszcz and T. Gabaldon. Redundans: an assembly pipeline for highly heterozygous genomes. *NUCLEIC ACIDS RESEARCH*, 44(12), JUL 8 2016.

52. T. R. Ranallo-Benavidez, K. S. Jaron, and M. C. Schatz. Genomescope 2.0 and smudgeplot for reference-free profiling of polyploid genomes. *NATURE COMMUNICATIONS*, 11(1), MAR 18 2020.

53. M. Rebeiz and T. M. Williams. Using drosophila pigmentation traits to study the mechanisms of *cis*-regulatory evolution. *Current Opinion in Insect Science*, 19:1–7, 2017.

54. Y. Satou, A. Sato, H. Yasuo, Y. Mihirogi, J. Bishop, M. Fujie, M. Kawamitsu, K. Hisata, and N. Satoh. Chromosomal inversion polymorphisms in two sympatric ascidian lineages. *GENOME BIOLOGY AND EVOLUTION*, 13(6), JUN 2021.

55. C. Smadja and R. Butlin. On the scent of speciation: the chemosensory system and its role in premating isolation. *Heredity*, 102(1):77–97, 2009.

56. G. Talavera, V. A. Lukhtanov, L. Rieppel, N. E. Pierce, and R. Vila. In the shadow of phylogenetic uncertainty: The recent diversification of lysandra butterflies through chromosomal change. *MOLECULAR PHYLOGENETICS AND EVOLUTION*, 69(3):469–478, DEC 2013.

57. V. Talla, A. Suh, F. Kalsoom, V. Dinca, R. Vila, M. Friberg, C. Wiklund, and N. Backstrom. Rapid increase in genome size as a consequence of transposable element hyperactivity in wood-white (leptidea) butterflies. *GENOME BIOLOGY AND EVOLUTION*, 9(10):2491–2505, OCT 2017.

58. K. Tamura, G. Stecher, and S. Kumar. Mega11: molecular evolutionary genetics analysis version 11. *Molecular biology and evolution*, 38(7):3022–3027, 2021.

59. W. B. Walker, A. Roy, P. Anderson, F. Schlyter, B. S. Hansson, and M. C. Larsson. Transcriptome analysis of gene families involved in chemosensory function in spodoptera littoralis (lepidoptera: Noctuidae). *BMC genomics*, 20(1):1–20, 2019.

60. M. Yandell and D. Ence. A beginner's guide to eukaryotic genome annotation. *Nature reviews. Genetics*, 13(5):329—342, April 2012.