First chromosome scale genomes of ithomiine butterflies (Nymphalidae: Ithomiini): comparative models for mimicry genetic studies

Jérémy Gauthier, Joana Meier, Fabrice Legeai, Melanie McClure, Annabel Whibley, Anthony Bretaudeau, Hélène Boulain, Hugues Parrinello, Sam T. Mugford, Richard Durbin, et al.

First chromosome scale genomes of ithomiine butterflies (Nymphalidae: Ithomiini): comparative models for mimicry genetic studies

Jérémy Gauthier[1#], Joana Meier[2], Fabrice Legeai[3,4], Melanie McClure[5,6], Annabel Whibley[7], Anthony Bretaudeau[3,4], Hélène Boulain[8], Hugues Parrinello[9], Sam T. Mugford[10], Richard Durbin[11,12], Chenxi Zhou[11,12], Shane McCarthy[11,12], Christopher W. Wheat[13], Florence Piron-Prunier[5], Christelle Monsempes[14], Marie-Christine François[14], Paul Jay[15], Camille Noûs[16], Emma Persyn[14,17], Emmanuelle Jacquin-Joly[14], Camille Meslin[14], Nicolas Montagné[14], Claire Lemaitre[4] and Marianne Elias[5]

corresponding author: mr.gauthier.jeremy@gmail.com

1. Natural History Museum of Geneva, Geneva, Switzerland
2. Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, United Kingdom
3. BIPAA, IGEPP, INRAE, Institut Agro, Univ Rennes, Rennes, France
4. Univ Rennes, Inria, CNRS, IRISA, 35000, Rennes, France
5. Institut Systématique Évolution Biodiversité (ISYEB), Centre National de la Recherche Scientifique, MNHN, EPHE, Sorbonne Université, Université des Antilles, Paris, France
6. Laboratoire Écologie, Évolution, Interactions des Systèmes Amazoniens (LEEISA), Université de Guyane, CNRS, IFREMER, Cayenne, France
7. School of Biological Sciences, University of Auckland, Auckland, New Zealand
8. Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland
9. MGX-Montpellier GenomiX, Univ. Montpellier, CNRS, INSERM, Montpellier France
10. Dept. Crop Genetics, John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, United Kingdom
11. Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, United Kingdom
12. Tree of Life Programme, Wellcome Sanger Institute, Hinxton CB10 1SA, United Kingdom
13. Department of Zoology, Stockholm University, Stockholm, Sweden
14. Institute of Ecology and Environmental Sciences of Paris, Sorbonne Université, INRAE, CNRS, IRD, UPEC, Université de Paris, Paris, France
15. Ecologie Systématique Evolution, Bâtiment 360, CNRS, AgroParisTech, Université Paris-Saclay, 91400 Orsay, France
16. Laboratoire Cogitamus, Paris, France
17. CIRAD, UMR PVBMT, F-97410, St Pierre, Réunion, France

Running head

First chromosome scale genomes of ithomiines

Abstract

The ithomiine butterflies (Nymphalidae: Danainae) represent the largest known radiation of Müllerian mimetic butterflies. They dominate by number the mimetic butterfly communities, which include species such as the iconic neotropical *Heliconius* genus. Recent studies on the ecology and genetics of speciation in Ithomiini have suggested that sexual pheromones, colour pattern and perhaps hostplant could drive reproductive isolation. However, no reference genome was available for Ithomiini, which hinders further exploration on the genetic architecture of these candidate traits, and more generally on the genomic patterns of divergence. Here, we generated high-quality, chromosome-scale genome assemblies of two *Melinaea* species, *Melinaea marsaeus* and *Melinaea menophilus*, and a draft genome of *Ithomia salapia*. We obtained genomes with a size ranging from 396 Mb to 503 Mb across the three species and scaffold N50 of 40.5 Mb and 23.2 Mb for the two chromosome-scale assemblies. Using collinearity analyses we identified massive rearrangements between the two closely related *Melinaea* species. An annotation of transposable elements and gene content was performed as well as a specialist annotation to target chemosensory genes, which is crucial for host plant detection and mate recognition in mimetic species. A comparative genomic approach revealed independent gene expansions in ithomiines and particularly in gustatory receptor genes. These first three genomes of ithomiine mimetic butterflies constitute a valuable addition and a welcome comparison to existing biological models such as *Heliconius*, and will enable further understanding of the mechanisms of adaptation in butterflies.

Introduction

The butterfly tribe Ithomiini (Nymphalidae: Danainae), which comprises 393 species, represents the largest known radiation of Müllerianmimetic butterflies, whereby co-occurring chemically-defended species converge in wing color pattern, which acts as a warning signal learned and avoided by predators (Müller, 1879; Sherratt, 2008). Ithomiine butterflies are endemic to the neotropics, where they numerically dominate butterfly communities in lowland and mountain

forests up to 2500 m, and where they engage in mimetic interactions with many other Lepidoptera (Beccaloni, 1997).

As such, ithomiine butterflies have an important ecological relevance. It is thus no wonder that ithomiine species served as examples in Bates' (Bates, 1862) and Müller's (Muller, 1879) original descriptions of Batesian (where palatable prey mimic distasteful ones) and Müllerian mimicry, respectively. Ithomiine butterflies are also remarkable in that many species have the unusual characteristic of harbouring partially transparent or translucent wings (McClure et al., 2019b; Pinna et al., 2021).

Mimetic butterflies have long been the focus of speciation studies, as they usually feature assortative mating for wing color patterns (e.g., Jiggins, Naisbit, Coe, & Mallet, 2001), combined with selection against hybrids between forms with different color patterns (e. g., Merrill, Chia, & Nadeau, 2014), because such hybrids typically harbor intermediate, non-mimetic color patterns. The iconic genus Heliconius has been the focus of multiple such speciation studies, using both experimental (Jiggins, Naisbit, Coe, & Mallet, 2001; Merrill, Chia, & Nadeau, 2014) and genomic (Martin et al., 2013, Merrill et al., 2019, Nadeau et al., 2012) studies.

While color pattern is believed to be a strong driver of diversification of mimetic butterflies (Kozak et al., 2015), including, possibly, Ithomiini (Jiggins, Mallarino, Willmott, & Bermingham, 2006), chemosensory communication may also be involved in speciation. Selection for similarity on a mating cue among co-occurring species is likely to result in reproductive interference (Boussens & Llaurens, 2021; Estrada & Jiggins, 2008), raising the question of alternative mate recognition cues. Chemical signals such as sex pheromones have been suggested to a play a role in reproductive isolation in mimetic butterflies (González-Rojas et al., 2020; Darragh et al., 2020), particularly among co-mimetic species (Mérot, Frérot, Leppik, & Joron, 2015). In ithomiine butterflies sexual pheromones have long been studied (Schulz et al., 2004), and have been shown to diverge between closely related taxa (Mann et al., 2020; McClure et al., 2019a; Stamm, Mann, McClure, Elias, & Schulz, 2019), suggesting a possible role in reproductive isolation (McClure et al., 2019b). More broadly, butterflies are phytophagous during the larval stage, and hostplant adaptation, mediated by chemical communication, has been hypothesized to be a major driver of speciation (Ehrlich & Raven, 1964; Jousselin & Elias, 2019). In Ithomiini, where butterfly-plant interaction tends to be very specific (Willmott & Mallet, 2004), divergent selection on hostplant has been documented in a few cases (McClure & Elias, 2016). Chemosensory and associated genes

(i.e. all genes involved in chemical communication) thus represent a particularly relevant targets for the study of speciation in mimetic butterflies. In butterflies, the detection of chemical signals is mainly performed by three types of membrane receptors named Odorant Receptors (ORs), Gustatory Receptors (GRs) and Ionotropic Receptors (IRs) and two secreted proteins families, the Odorant-Binding Proteins (OBPs) and the Chemosensory Proteins (CSPs) (Pelosi et al. 2006; Robertson, 2019). The role of specific lineages of the OR gene family in the detection of volatile sex pheromones has been characterized in moths (Montagné, Wanner, & Jacquin-Joly, 2021). However, little is known of the molecular bases of pheromone detection in butterflies (Eyres et al., 2016; van Schooten et al., 2020). In Ithomiini, only one recent study addressed chemosensory genes, and found that one OR was differentially expressed between two subspecies of Melinaea marsaeus, suggesting a possible role in mate choice (Piron-Prunier et al., 2021).

In contrast to Heliconius, little is also known on the overall genomic patterns of speciation in Ithomiini. Two studies, one using microsatellites and the other relying on reduced-complexity genomic data, revealed a range of levels of genetic differentiation among subspecies in five ithomiine species (McClure et al., 2019a, Gauthier et al., 2020), calling for more in depth studies of population genetic structure and patterns of gene flow. However, currently, research on speciation in Ithomiini is hindered by the lack of reference genomes. The paucity of genomic resources for Ithomiini is surprising, given their ecological and historical importance. The closest reference genome is that of the monarch butterfly, *Danaus plexippus* (Gu et al., 2019; Zhan, Merlin, Boore, & Reppert, 2011), which belongs to the nymphalid tribe Danaini and that diverged from the Ithomiini circa 42 million years ago (Chazot et al., 2019).

Here we present the first genomes of three Ithomiini species, *Ithomia salapia* (subspecies *aquinia*), *Melinaea marsaeus* (subspecies *rileyi*) and *Melinaea menophilus* (subspecies ssp nov). *Ithomia salapia* is a typical 'clearwing' ithomiine butterfly, in that it shows transparent or translucent wings (Figure 1). Subspecies of *I. salapia* belong to large mimicry rings that include ithomiine and non ithomiine species (Beccaloni, 1997; Willmott, Robinson Willmott, Elias, & Jiggins, 2017). The genus *Ithomia* belongs to the Ithomiine 'core-group', a clade that encompasses 80% of the species of the tribe and that underwent steady diversification in the Central Andes during the Miocene before colonizing other neotropical regions (Chazot et al., 2019). A recent population genetic study in a suture zone showed that gene flow between subspecies of *I. salapia* was highly reduced, suggesting incipient speciation (Gauthier et al., 2020). The genus *Melinaea (Figure 1)* belongs to

a basal Amazonian lineage that likely experienced high extinction rates during the Miocene before diversifying at a higher pace during the last couple of million years (Chazot et al., 2019). *Melinaea* species engage in mimetic interactions with multiple other Lepidoptera, including species from the tribe Heliconiini (Beccaloni, 1997). Also, and contrasting with *I. salapia*, genetic studies based on microsatellite and coding sequences found an exceptionally low level of divergence among *Melinaea* subspecies and even species (Chazot et al., 2019; Dasmahapatra, Lamas, Simpson, & Mallet, 2010; McClure et al., 2019), which may indicate recent diversification or extensive gene flow. Another intriguing feature in the genus *Melinaea* is the high karyotypic lability, with multiple events of chromosomal fusion recorded between two closely related subspecies (Brown, Von Schoultz, & Suomalainen, 2004; McClure, Dutrillaux, Dutrillaux, Lukhtanov, & Elias, 2017).

Because the genomes of these three species are large and highly heterozygous, it has been necessary to test and combine different sequencing methods. For *M. marsaeus*, we first tested a combination of PacBio CLR, classical paired-end Illumina and 10X sequencing. Then we combined PacBio HiFi, 10x and HiC, which allowed us to obtain a genome assembled at the chromosome level. This latter strategy was applied to *M. menophilus*. The *I. salapia* genome, obtained with 10X sequencing, is more fragmented and can be considered as a draft genome. For each of the genomes we generated gene annotations using a pipeline which incorporated transcriptomic data and manually annotated the chemosensory gene families, as these families are usually badly predicted by automatic annotations.

2. Materials and Methods

2.1 Sample collection, DNA extraction, library construction and sequencing

Females of *I. salapia aquinia* were collected in Urahuasha (6°27’ S, 76°20 W, San Martin, Peru) and kept in captivity, where they were presented with potted *Witheringia solanacea* for egg-laying. Females of *M. marsaeus* rileyi and *M. menophilus* ssp nov were collected in Micaela (5°56’ S, 76°14’ W, Loreto province, Peru), and Urahuasha, respectively, and kept in captivity in Tarapoto (San Martin, Peru), where they were presented with potted *Juanulloa parasitica* on which they laid eggs. Larvae of all species were reared on their host plants until pupation, and pupae were preserved in empty plastic vials at -80 °C until extraction.

The individual for the genome *M. marsaeus* v1 was extracted following this protocol (Mugford, Wouters, Mathers, & Hogenhout, 2020) with the following modifications: frozen tissue (approx.

100 mg) was ground in a 2 ml Eppendorf containing a 2 mm steel ball bearing in a Retsch mill (Qiagen), and extracted in 500 µl CTAB buffer (2% w/v CTAB, 1.4 M NaCl, 20 mM EDTA, 100 mM Tris-HCl pH 8.0) with RNAse A (20µg/ml), incubated at room temperature for 1 hour with regular mixing. 500 µl of chloroform was added, mixed by inversion for 10 minutes, and centrifuged at 5000 g for 10 minutes. The upper phase was transferred to a new tube, an equal volume of cold isopropanol was added, and incubated at -20 degrees C for one hour. The sample was centrifuged at 5000g for 10 minutes, and the supernatant removed. The pellet was washed in 75% ethanol air-dried, and resuspended in 80 µl of TE (10 mM Tris-HCl pH 7.5, 1 mM EDTA). Library construction including adaptor ligation and size selection were performed according to the manufacturers' instructions. For the genome *M. marsaeus* v1, one lane of 250 paired-end reads were sequenced a Illumina HiSeq 4000. The PacBio library was sequenced on 10 SMRTcells of the PacBio RSII with C4-P6 chemistry. The 10x Chromium Genome Library was sequenced on one lane of the HiSeq 2500 with a 250PE-RR read metric. Sequencing was performed at the Earlham Institute (UK).

For the genomes of *M. marsaeus* v2 (ilMelMars1.1) and *M. menophilus* (ilMelMeno1.1), DNA extraction, library preparation and sequencing were performed by the Scientific Operations core at the Wellcome Sanger Institute. DNA was extracted from flash-frozen pupae of female butterflies with the Qiagen MagAttract HMW DNA kit. Pacific Biosciences (PacBio) HiFi libraries were sequenced on a PacBio SEQUEL II. 10X Genomics Chromium v2 libraries and HiC Arima v2.0 libraries were constructed according to the manufacturers' instructions and sequenced on Illumina HiSeq X instruments.

The two individuals used for the genome of *I. salapia* were extracted following a protocol adapted from (Mayjonade et al., 2016). Samples were snap frozen alive in liquid nitrogen and conserved at -80°C. DNA was extracted from the whole butterfly bodies with the exception of the head. Butterflies were ground in a frozen mortar with liquid nitrogen, 150 mg of tissue powder was mixed with 900 µl of preheated buffer and 6 µl of RNaseA. Tubes were incubated for 120 minutes at 50°C for lysis, and then at -10°C for 10 minutes, with the addition of 300 µl of Potassium acetate for the precipitation. One volume of binding buffer was added with 100 µl of Serapure beads solution. 3 washing cycles were used and DNA was resuspended in 100 µl of EB buffer. Library construction including adaptor ligation and size selection were performed according to the

manufacturers' instructions. The two 10x Chromium Genome Library libraries were sequenced on one lane of the HiSeq 2500 with a 250PE-RR read metric.

2.2 Transcriptomic data

For *M. marsaeus* and *I. salapia* transcriptomic data were generated from various tissues including (abdomen, thorax, head) and developmental stages (adult, pupae and two larval stages) (detailed in Table 1) to maximize transcript diversity. In addition, targeted tissues from pupal wing discs and antennae in *M. marsaeus* were used (Piron-Prunier et al., 2021). Tissue samples were homogenised in 600 µl of RLT buffer with TissueLyser (Qiagen, Hilden, Germany). Total RNA was then extracted according to the manufacturer's protocol (RNeasy Mini kit, Qiagen, Hilden, Germany) and eluted in 30 µl of RNase-free water. To avoid genomic contamination, RNase-free DNase treatment (Qiagen, Hilden, Germany) was performed during RNA extraction. The quality of the isolated RNA was checked on 0.8% agarose gel for the presence of 28S and 18S bands. The quality and quantity of RNA was further analyzed using Qubit 2.0 fluorometer (Invitrogen, Carlsbad, CA, USA) and RNA integrity was confirmed using an Agilent Bioanalyser 2100 (Agilent Technologies, CA, USA). Libraries were sequenced with Illumina HiSeq 2500 platform.

2.3 Genome size and heterozygosity estimation using k-mers approaches

Genome characteristics, genome size, heterozygosity, were estimated on each dataset of raw reads using k-mer spectrum distribution analysis. K-mer distribution were estimated using JELLYFISH v2.2.10 (Marçais and Kingsford 2011)and a k-mer size of 31. GENOMESCOPE2 (Ranallo-Benavidez et al. 2020) was used to estimate genome characteristics and generate plots (Supplementary Figure 1).

2.4 Genome assembly

First, for each dataset, the read quality was evaluated with FASTQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). A first attempt to assemble the *M. marsaeus* genome, called v1, was made by combining the first Illumina HiSeq, PacBio and 10X data (Table 1). This approach yielded a highly fragmented assembly, likely due to the high heterozygosity found in this species. A second approach was therefore implemented for *M. marsaeus*, and then also used for *M. menophilus*, based on 51x PacBio data, 10X Genomics

Chromium data, and Arima Hi-C data generated by the Sanger Institute Tree of Life programme (https://www.sanger.ac.uk/programme/tree-of-life/). The assembly process included the following sequence of steps: initial PacBio assembly generation with Hifiasm v0.15.1 (Cheng, Concepcion, Feng, Zhang, & Li, 2021), retained haplotig separation with purge_dups v1.2.3 (Guan et al., 2020), short-read polishing using FreeBayes v1.3.1-called variants (Garrison & Marth, 2012) from 10X Genomics Chromium reads aligned with LongRanger v2.2.2 (https://github.com/10XGenomics/longranger), and Hi-C based scaffolding with SALSA2 v2.2 (Ghurye et al., 2019) using Hi-C contact map (Supplementary Figure 2). The mitochondrial genome was assembled using MitoHifi v2 (https://github.com/marcelauliano/MitoHiFi). Finally, the assemblies were analysed and manually improved using rapid curation (Howe et al., 2021). Chromosome-scale scaffolds confirmed by the Hi-C data have been named in order of size. Genome completeness was assessed with BUSCO v5 (Manni, Berkeley, Seppey, Simão, & Zdobnov, 2021) in genome mode with the lepidoptera_odb10 ortholog dataset composed of 5,286 orthologous genes. BUSCO genes were also used to identify the Z chromosomes in both species. The putative Z chromosomes also showed reduced read coverage in both species, supporting that they are Z chromosomes of females. In *M. menophilus* a second chromosome with reduced coverage, Hi-C links to the Z chromosome and very small size (2.99 Mbp) was assigned as putative W chromosome. For *I. salapia*, all 10X libraries of the two samples were first assembled separately with Supernova v2.1.1 (Visendi, 2022) and then combined with Ragout using one genome as reference and the other one as target (Kolmogorov et al., 2018). Base accuracy (QV) was estimated using a k-mer size of 21 with Merqury (Rhie et al., 2020).

2.5 Synteny

Synteny between *M. marsaeus* and *M. menophilus* genomes was investigated using the positions of the complete non-duplicated BUSCO genes. Using a custom-made R script, we merged the BUSCO gene position files and plotted them against each other.

2.6 Gene prediction, automated and functional annotations

The transposable element annotation was realized using RepeatMasker (Tarailo-Graovac & Chen, 2009). This annotation was exported into GFF3 files and used as a mask for gene annotation. Later, repeat masking with *de novo* repeat discovery, automated curation and filtering was performed

using the EarlGrey pipeline (v1.2) (Baril, Imrie, & Hayward, 2021) with default settings in combination with the *Arthropoda* library from the Dfam database (version 3.5) (Storer, Hubley, Rosen, Wheeler, & Smit, 2021). The automated gene prediction and annotation was done using MAKER (Cantarel et al., 2008) integrating different features based on (i) the mapping of Lepidoptera proteins from LepBase (Challi, Kumar, Dasmahapatra, Jiggins, & Blaxter, 2016), (ii) the transcriptomes of each species generated by the assembly of RNA-Seq data with Trinity 2.8.4 (Haas et al., 2013) and (iii) ab initio genes predictions using Augustus (Hoff & Stanke, 2019). Reliable gene predictions were extracted according to Annotation Edit Distance (AED) ≤ 0.2 or a minimum coverage of 1000 from RNAseq data mapping after optimization using BUSCO statistics. Annotation completeness was assessed with BUSCO v5 (Manni et al., 2021) in protein mode with the lepidoptera_odb10 ortholog dataset composed of 5,286 orthologous genes. The functional annotation was performed using blastp from BLAST+ v2.5.0 (Camacho et al., 2009) to compare predicted proteins in each genome to the NCBI non-redundant database. The ten best hits below an e-value of 1e-08 without complexity masking were conserved. Interproscan (Jones et al., 2014) was used to analyze protein sequences seeking for known protein domains in the different databases available in Interproscan. Finally, we used Blast2GO (Conesa et al., 2005) to associate a protein with a GO group.

2.7 Ortholog analyses

Orthologous genes between annotated genes in each species and the seven outgroups (*Pieris napi*, *Bicyclus anynana*, *Junonia coenia*, *Melitaea cinxia*, *Heliconius erato*, *Heliconius melpomene* and *Danaus plexippus*) were identified using OrthoFinder v2.5.2 (Emms & Kelly, 2015). Single copy ortholog proteins were extracted, aligned using MAFFT v7.01775 and concatenated using AMAS (Borowiec, 2016). The species phylogeny was performed on this alignment composed of 996 orthologs for a length of 647 kb using PhyML (Anisimova & Gascuel, 2006) including a branch support estimation with 1,000 bootstrap iterations.

2.8 Manual annotation of chemosensory genes

For each of the chemosensory gene family, i.e. the odorant receptors (ORs), the gustatory receptors (GRs), the variant ionotropic receptors (IRs), the odorant-binding proteins (OBPs) and the chemosensory proteins (CSPs), amino acid sequences previously identified from the genomes of

D. plexippus, *H. melpomene*, *S. frugiperda* and *B. mori* (Briscoe et al., 2013; Gouin et al., 2017; Guo et al., 2017; Heliconius Genome Consortium, 2012; N.-Y. Liu et al., 2018; Meslin et al., 2021; Vogt, Große-Wilde, & Zhou, 2015; Zhan et al., 2011) as well as from the transcriptome of *M. marsaeus* (Piron-Prunier et al., 2021) were used as queries in a tBLASTn search against genome assemblies of the three species (e-value threshold 0.001), in order to identify scaffolds containing the genes to annotate. Query amino acid sequences were then aligned on these scaffolds with Exonerate (Slater & Birney, 2005) to identify precise intron-exon boundaries and create gene models. These models were visualized using Integrated Genomics Viewer v2.11.9 (Robinson et al., 2011), and badly predicted models were eliminated from the final sequence datasets. Nucleotide and amino acid sequences were extracted with GffRead (Pertea and Pertea 2020). To create CSP and GR trees, amino acid sequences from Ithomiini were aligned with those of the above-mentioned species (except *S. frugiperda* GRs that were not included to limit the number of sequences) with MAFFT v7 (Katoh, Rozewicki, & Yamada, 2019). Maximum-likelihood phylogenies were built using PhyML 3.0 (Guindon et al., 2010) following model selection by SMS (Lefort, Longueville, & Gascuel, 2017). Branch support was estimated via SH-like approximate likelihood-ratio test (Anisimova & Gascuel, 2006).

3. Results and Discussion

3.1 Sequencing strategy comparison

In order to obtain a high-quality reference genome for *M. marsaeu*s, we combined deeper PacBio sequencing using the new HiFi technology with low error rates, 10X sequencing and HiC data (Table 1). The use of a HiC approach, which enabled us to organise the scaffolds at the chromosome level, was particularly successful as it resulted in a final genome of 503 Mb composed of 22 scaffolds and an N50 of 40.4 Mb (Table 2). The same strategy was used for the species *M. menophilus* and yielded similar quality results with a genome of 496 Mb composed of 28 scaffolds and an N50 of 23.1 Mb (Table 2). For *I. salapia*, two 10X libraries were generated from two individuals and sequenced separately (Table 1). Largely due to the absence of HiC libraries and PacBio HiFi libraries, the genome obtained for this species is more fragmented than those of the two *Melinaea* species. The final assembly is composed of 23,973 scaffolds for a total of 395 Mb and an N50 of 1.4 Mb (Table 2). For *M. marsaeus*, the 22 scaffolds obtained could be grouped into 13 chromosomes, two sex chromosomes W and Z, the mitochondrion and six unplaced scaffolds.

For *M. menophilus*, the 28 scaffolds were grouped into 20 chromosomes, two sex chromosomes W and Z, the mitochondrion and five unplaced scaffolds. The final number of chromosomes assembled matches the number of chromosomes identified by cytogenetic techniques in *M. menophilus*, i.e. 2n = 42 (Dutrillaux et al., 2022).

3.2 Genome size and heterozygosity estimation

For each of the three genomes, the size of the final assemblies is within or slightly above the range of the size estimates from k-mer approaches on the raw reads. For *M. marsaeus* the k-mer estimates range from 330 Mb to 496 Mb (Supplementary Table 1) and the assembled genome size is 503 Mb; for *M. menophilus* the k-mer estimates range from 357 Mb to 527 Mb (Supplementary Table 1) and the assembled genome size is 496 Mb; and finally for *I. salapia*, the k-mer size estimate range is 352 Mb to 357 Mb and the assembled genome size is 395 Mb (Table 2, Supplementary Table 1). These genome sizes are at the top of the distribution of genome sizes observed in the Danainae, ranging from 249 to 455 Mb, but are below those of the largest genomes observed in the Nymphalidae, such as *Polyura nepenthes* (Nymphalidae, Charaxinae) whose genome size is estimated at 925 Mb (G. Liu et al., 2020). When comparing 10x data, almost four times more heterozygosity is observed for *M. marsaeus* than for *M. menophilus* (Supplementary Table 1). The levels of heterozygosity estimated using k-mer approaches show an heterogeneity between the different datasets but seem to show a fairly high level of heterozygosity (Supplementary Table 1). This may be related to the demographic history of the populations and, for *M. marsaeus*, to the mechanisms of divergence and hybridisation that exist in the suture zone between the Andes and the Amazon. The populations of *M. marsaeus* around Tarapoto were found to be profoundly admixed in a previous study (McClure & Elias, 2016). This high level of divergence between *M. marsaeus* populations and their hybridisation may explain the difficulty of assembly encountered during the first attempt to sequence this species.

The final assemblies show a high level of completeness, as testified by high BUSCO completeness in genome mode (Seppey, Manni, & Zdobnov, 2019). For each of the genomes, including the more fragmented genome of *I. salapia*, more than 95% of 5,286 single copy orthologs across Lepidoptera were recovered (Table 2).

In contrast to the highly colinear genomes *Heliconius* butterflies, where most species have 21 chromosomes (Seixas, Edelman, & Mallet, 2021), our closely related *Melinaea* species differ

strongly in chromosome number (14 vs 21) and show numerous massive rearrangements (Figure 2). The only two *M. marsaeus* chromosomes that fully correspond to a single *M. menophilus* chromosome, are chromosomes 7 (chr. 1 in *M. menophilus*) and the Z chromosome. The high variation in chromosome numbers in species in the genus *Melinaea* has already been observed by (Brown et al., 2004; Dutrillaux et al., 2022; McClure et al., 2017). Here we show that this variation could be the result of fusion and fission events.

3.3 Gene prediction and function annotation

Prior to the gene annotation step, an annotation of transposable and repeated elements was performed. To perform reliable gene annotation we took advantage of transcriptomic data. For *M. marsaeus*, we used assembled transcripts from a study on differential expression between two subspecies (Piron-Prunier et al., 2021), which included a reference transcriptome for that species across multiple stages (larval, pupal and imago) and transcriptomes of targeted tissues, namely pupal wing discs and antennae (Table 1). For *I. salapia*, we sequenced and assembled a reference transcriptome by sequencing transcripts from different tissues and different developmental stages (Table 1). Automated annotations combining transcriptomic data, known lepidopteran proteins and *ab initio* predictions annotated respectively 52,865 genes for *M. marsaeus*, 54,531 genes for *M. menophilus* and 32,213 for *I. salapia*. After the filtering of the reliable gene predictions, 18,670 genes were kept for *M. marsaeus*, 19,174for *M. menophilus* and 18,283 for *I. salapia*. These genes have comparable characteristics in terms of gene size, number and sizes of exons and introns (Table 3). Like the genomes, these annotations and the predicted proteins have a high completeness level identified by BUSCO in protein mode with more than 85.8% of the lepidopteran single copy orthologs recovered (Table 3).

Annotation of the repetitive elements of the genome, combining *de novo* and homology-based discovery approaches, revealed increased transposable element content with increasing genome size, with 14% total repeat content in *I. salapia* and 24% in the two *Melinaea* species (Supplementary Figure 3). The differences could be linked to sequencing strategies. The complement of different element classes differed between the species and from the repeat content described in *Danaus* species, which themselves show considerable variation within the genus (Baril & Hayward, 2022). More specifically, the ithomiine genomes all exhibit increased DNA transposon, Rolling-circle and LINE and LTR retroelement content but decreased contributions of

Penelope elements. SINE retroelements comprise nearly 3% of the genome assemblies in both Melinaea species but less than 0.2% of the *I. salapia* genome. Transposon landscape analysis supports recent transposon activity in all genomes, as indicated by the presence of several TE classifications with low genetic distance to their consensus sequences (Supplementary Figure 2). Regarding the distribution at the chromosome level, the sex chromosomes have different concentrations of repeated elements than the autosomes. The Z chromosomes present only 14% of transposable elements for both species. Conversely, the W chromosomes have a much higher concentration than the autosomes, reaching 59.72% for M. menophilus and 73.62% for M. marsaeus. However, for both the Z and W chromosomes, the composition of the different families of transposable elements is substantially similar between the sex chromosomes and with the rest of the genome (Supplementary Figure 3).

3.4 Comparison with key lepidopteran reference genomes

Orthologous genes for all annotated genes in the three focal species and seven outgroup butterfly species, including reference genomes such as *Danaus plexippus* (PRJNA564985), the species most closely related to the Ithomiini, and *Heliconius melpomene* (PRJEA71053), a species belonging to a large clade of mimetic butterflies, were identified using OrthoFinder v1.1474 (Emms & Kelly, 2015). In total, 16,736 orthology groups were identified including 93.0% of all the analysed genes from the ten species. Among them, 5,792 orthogroups are shared by all species. Larger gene numbers were observed for the *Melinaea* species. Thus, a reduced proportion of genes are shared by the ithomiines, which represent 4.4% and 3.0% of the genes for *M. marsaeus* and *M. menophilus* respectively, and 2.0% of the genes for *I. salapia* (in light orange on Figure 3). Within *Melinaea*, a large proportion of genes are associated with the *Melinaeae* genus and shared between the two species, representing 11.1% of genes for *M. marsaeus* and 10.4% of genes for *M. menophilus* (in light yellow on Figure 3). Finally, we also observed a large proportion of species-specific genes, since they reach 11.9% (including 6.2% of duplicated species-specific) for *M. marsaeus* and 14.3% (including 7.4% of duplicated species-specific) for *M. menophilus* (in green on Figure 3).

3.5 Annotation of chemosensory genes

Chemosensory cues and signals are instrumental for butterflies as they are involved in host plant detection and in mate recognition. This is especially the case in mimetic butterflies, whereby the

colour pattern may not provide an effective cue for mate recognition due to mimicry (Mérot, Frérot, Leppik, & Joron, 2015). Detection of chemosensory cues and signals by the peripheral nervous system of insects is mainly governed by transmembrane receptors located at the membrane of olfactory or gustatory neurons, responsible for signal transduction upon ligand activation. In insects, such receptors belong to three multigene families: the odorant receptors (ORs), the gustatory receptors (GRs) and the variant ionotropic receptors (IRs). Depending on insect orders, the number of genes within each family can vary from a few dozens to several hundreds (Robertson, 2019). We annotated genes belonging to these families in the three Ithomiini genomes (Table 4). The number of OR genes varied from 62 in *M. menophilus* to 70 in *I. salapia*, which is similar to the number found in any other lepidopteran genome, including the closely related species *D. plexippus* (Montagné, Wanner, & Jacquin-Joly, 2021). The same holds true for IR genes whose numbers varied from 31 in *M. marsaeus* to 36 in *I. salapia*. By contrast, we annotated an unexpectedly large number of GR genes in the three species, up to more than 200 in *M. marsaeus*. This high number of genes compared with other Nymphalidae (including *D. plexippus*) results from extensive duplications in Ithomiini that occurred in several lineages of the GR phylogeny (Figure 4). So far, such expansions of GR repertoires in Lepidoptera have been documented only in the Noctuidae family, where it has been tentatively linked to polyphagy (Gouin et al., 2017; Meslin et al., 2021). It is interesting to note that somehow similar expansions also occurred independently in Ithomiini, which are not polyphagous but rather oligophagous species (McClure & Elias, 2016; Willmott & Mallet, 2004).

Apart from transmembrane receptors, chemodetection in insects also relies on soluble proteins that can bind and transport semiochemicals in the aqueous lymph of olfactory and gustatory sensilla, so that they can reach the neurons. The genomes of Ithomiini contain 35 to 40 genes encoding odorant-binding proteins (OBPs), which is in the range of what has been observed in other Lepidoptera. On the other hand, the number of chemosensory proteins (CSPs) is exceptionally high in Ithomiini genomes, especially in both *Melinaea* species which have more than 50 CSP genes (Table 4). The phylogenetic analysis shows that all but one of the CSP lineages are highly conserved in Lepidoptera, whereas numerous gene duplications occurred in a butterfly-specific lineage (Figure 5). This expansion has been documented previously (Heliconius Genome Consortium, 2012) yet it is particularly spectacular in Ithomiini genomes, which contain up to 30 CSP genes (in *M. marsaeus*) vs. 11 in *D. plexippus* and 6 in *H. melpomene*. This further confirms

a previous observation made following the analysis of the *M. marsaeus* transcriptome (Piron-Prunier et al., 2021).

3.6 Conclusion

In this study we sequenced, de novo assembled, and annotated the genomes of three ithomiine species. We analyzed their genomic features and performed genomic content comparison and orthologous gene identification with *D. plexippus*, which belongs to the same subfamily (Danainae), and various outgroups including two *Heliconius* species (Nymphalidae: Heliconiinae), a well-studied mimetic genus that includes species that are mimetic with *Melinaea*. Manual curation of chemosensory genes in the three genomes revealed unexpected expansions of GR genes, as this has been previously observed only in polyphagous noctuids. These first genomes of ithomiine mimetic butterflies will be useful to further understand the mechanisms of adaptation and the genetic bases underpinning mimicry, and provide a welcome comparison to existing biological models of mimicry like the *Heliconius*.

Acknowledgements

ME acknowledges financial support from ANR (projects SPECREP and CLEARWING) and HFSP (RGP0014/2016). MGX acknowledges financial support from France Génomique National infrastructure, funded as part of "Investissement d'Avenir" program managed by Agence Nationale pour la Recherche (contract ANR-10-INBS-09). RD, SM and CZ acknowledge financial support from Wellcome grant WT207492, and SM and work at the Wellcome Sanger Institute were also supported by Wellcome grant WT206194. We thank members of the Wellcome Sanger Institute Scientific Operations and Tree of Life core groups for their contributions to data production and assembly for the Melinaea genomes. We thank Mario Tuanama and Ronald Mori-Pezo for help with rearing and collecting butterflies. MM acknowledges a postdoctoral fellowship by the Fonds Québécois de la Recherche sur la Nature et les Technologies (FQRNT). STM acknowledges support from the BBSRC Institute Strategy Programme (BB/P012574/1). We thank the Peruvian authorities for research permits (236-2012-AG-DGFFS-DGEFFS, 201-2013-MINAGRI-DGFFS/DGEFFS, 002-2015-SERFOR-DGGSPFFS and 373-2017-SERFOR-DGGSPFFS), the Gobierno Regional San Martín PEHCBM (permit: 124-2016-GRSM/PEHCBM-DMA/EII-ANP/JARR) and the Museo de Historia Natural and Prof. Gerardo Lamas for their

support. We thank the GenOuest BioInformatics Platform (http://www.genouest.org/), which allowed the use of a computing cluster for bioinformatic analyses.

Author Contributions

ME designed the study. MM performed sampling. JM, FL, JG, AW, HP, STM, RD, CZ, SMC, FP, PJ, CN performed the sequencing and assemblies with contributions from CWW. JG and FL performed the annotations with contributions from AB. AW performed the transposable element analyses. JM performed the synteny analyses. JG and HB performed the orthologous gene analyses. NM, EP, CMo, MCF, CMe and EJJ performed the chemosensory analyses. All authors took part in discussions concerning the analyses and result interpretations. JG and ME wrote the paper, with contributions from all authors.

Data availability

Genome assemblies are available on the NCBI under GCA_918358865.1 accession number for *M. marsaeus* and GCA_918358695.1 for *M. menophilus*. The genome of *I. salapia* is available on the BIPAA platform: https://bipaa.genouest.org/sp/ithomia_salapia/. Raw genomic data can be found under BioProjects PRJNA836751 (*M. marsaeus* v1), PRJEB48295 (*M. marsaeus* v2) and PRJEB48296 (*M. menophilus*). Raw transcriptomic data can be found under BioProjects PRJNA836751.

Code availability

The assembly and annotation pipelines including custom scripts are available in the Github repository XXXXXX (submission upon manuscript acceptance).

References

Anisimova, M., & Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic Biology*, *55*(4), 539–552. doi: 10.1080/10635150600755453

Baril, T., & Hayward, A. (2022). Migrators within migrators: exploring transposable element dynamics in the monarch butterfly, Danaus plexippus. *Mobile DNA*, *13*(1), 5. doi: 10.1186/s13100-022-00263-5

Baril, T., Imrie, R., & Hayward, A. (2021). *Earl Grey*. doi: 10.5281/zenodo.5654616

Bates, H. W. (1862). XXXII. Contributions to an Insect Fauna of the Amazon Valley. Lepidoptera:

Heliconidae. *Transactions of the Linnean Society of London*, *23*(3), 495–566. doi: 10.1111/j.1096-3642.1860.tb00146.x

Beccaloni, G. W. (1997). Vertical stratification of ithomiine butterfly (Nymphalidae: Ithomiinae) mimicry complexes: the relationship between adult flight height and larval host–plant height. *Biological Journal of the Linnean Society. Linnean Society of London*, *62*(3), 313–341. doi: 10.1006/bijl.1997.0165

Borowiec, M. L. (2016). AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ*, *4*, e1660. doi: 10.7717/peerj.1660

Boussens-Dumon, G., & Llaurens, V. (2021). 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.

Briscoe, A. D., Macias-Muñoz, A., Kozak, K. M., Walters, J. R., Yuan, F., Jamie, G. A., … Jiggins, C. D. (2013). Female Behaviour Drives Expression and Evolution of Gustatory Receptors in Butterflies. *PLoS Genetics*, *9*(7), e1003620. doi: 10.1371/journal.pgen.1003620

Brown, K. S., Jr, Von Schoultz, B., & Suomalainen, E. (2004). Chromosome evolution in Neotropical Danainae and Ithomiinae (Lepidoptera). *Hereditas*, *141*(3), 216–236. doi: 10.1111/j.1601-5223.2004.01868.x

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*, 421. doi: 10.1186/1471-2105-10-421

Cantarel, B. L., Korf, I., Robb, S. M. C., Parra, G., Ross, E., Moore, B., … Yandell, M. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research*, *18*(1), 188–196. doi: 10.1101/gr.6743907

Challi, R. J., Kumar, S., Dasmahapatra, K. K., Jiggins, C. D., & Blaxter, M. (2016). Lepbase: the Lepidopteran genome database (p. 056994). doi: 10.1101/056994

Chazot, N., Willmott, K. R., Lamas, G., Freitas, A. V. L., Piron-Prunier, F., Arias, C. F., … Elias, M. (2019). Renewed diversification following Miocene landscape turnover in a Neotropical butterfly radiation. *Global Ecology and Biogeography: A Journal of Macroecology*, (geb.12919). doi: 10.1111/geb.12919

Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods*, *18*(2), 170–175. doi: 10.1038/s41592-020-01056-5

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* , *21*(18), 3674–3676. doi: 10.1093/bioinformatics/bti610

Darragh, K., Montejo-Kovacevich, G., Kozak, K. M., Morrison, C. R., Figueiredo, C. M. E., Ready, J. S., … Jiggins, C. D. (2020). Species specificity and intraspecific variation in the chemical profiles of Heliconius butterflies across a large geographic range. Ecology and Evolution, 10(9), 3895–3918.

Dasmahapatra, K. K., Lamas, G., Simpson, F., & Mallet, J. (2010). The anatomy of a “suture zone” in Amazonian butterflies: a coalescent-based test for vicariant geographic divergence and speciation. *Molecular Ecology*, *19*(19), 4283–4301. doi: 10.1111/j.1365-294X.2010.04802.x

Dutrillaux, B., Dutrillaux, A.-M., McClure, M., Gèze, M., Elias, M., & Bed’hom, B. (2022). Improved basic cytogenetics challenges holocentricity of butterfly chromosomes (p. 2022.03.11.484012). doi: 10.1101/2022.03.11.484012

Emms, D. M., & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, *16*, 157. doi: 10.1186/s13059-015-0721-2

Ehrlich, P. R., & Raven, P. H. (1964). Butterflies and plants: A study in coevolution. Evolution; International Journal of Organic Evolution, 18(4), 586–608.

Estrada, C., & Jiggins, C. D. (2008). Interspecific sexual attraction because of convergence in

warning colouration: is there a conflict between natural and sexual selection in mimetic species? Journal of Evolutionary Biology, 21(3), 749–760.

Eyres, I., Jaquiéry, J., Sugio, A., Duvaux, L., Gharbi, K., Zhou, J.-J., … Ferrari, J. (2016). Differential gene expression according to race and host plant in the pea aphid. Molecular Ecology, 25(17), 4197–4215.

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. arXiv. Retrieved from http://arxiv.org/abs/1207.3907

Gauthier, J., de Silva, D. L., Gompert, Z., Whibley, A., Houssin, C., Le Poul, Y., … Elias, M. (2020). Contrasting genomic and phenotypic outcomes of hybridization between pairs of mimetic butterfly taxa across a suture zone. *Molecular Ecology*, *29*(7), 1328–1343. doi: 10.1111/mec.15403

Ghurye, J., Rhie, A., Walenz, B. P., Schmitt, A., Selvaraj, S., Pop, M., … Koren, S. (2019). Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLoS Computational Biology*, *15*(8), e1007273. doi: 10.1371/journal.pcbi.1007273

González-Rojas, M. F., Darragh, K., Robles, J., Linares, M., Schulz, S., McMillan, W. O., … Salazar, C. (2020). Chemical signals act as the main reproductive barrier between sister and mimetic Heliconius butterflies. Proceedings. Biological Sciences / The Royal Society, 287(1926), 20200587.

Gouin, A., Bretaudeau, A., Nam, K., Gimenez, S., Aury, J.-M., Duvic, B., … Fournier, P. (2017). Two genomes of highly polyphagous lepidopteran pests (Spodoptera frugiperda, Noctuidae) with different host-plant ranges. *Scientific Reports*, *7*(1), 11816. doi: 10.1038/s41598-017-10461-4

Guan, D., McCarthy, S. A., Wood, J., Howe, K., Wang, Y., & Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* , *36*(9), 2896–2898. doi: 10.1093/bioinformatics/btaa025

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307–321. doi: 10.1093/sysbio/syq010

Gu, L., Reilly, P. F., Lewis, J. J., Reed, R. D., Andolfatto, P., & Walters, J. R. (2019). Dichotomy of Dosage Compensation along the Neo Z Chromosome of the Monarch Butterfly. *Current Biology: CB*, *29*(23), 4071–4077.e3. doi: 10.1016/j.cub.2019.09.056

Guo, H., Cheng, T., Chen, Z., Jiang, L., Guo, Y., Liu, J., … Mita, K. (2017). Expression map of a complete set of gustatory receptor genes in chemosensory organs of Bombyx mori. *Insect Biochemistry and Molecular Biology*, *82*, 74–82. doi: 10.1016/j.ibmb.2017.02.001

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., … Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, *8*(8), 1494–1512. doi: 10.1038/nprot.2013.084

Heliconius Genome Consortium. (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, *487*(7405), 94–98. doi: 10.1038/nature11041

Hoff, K. J., & Stanke, M. (2019). Predicting Genes in Single Genomes with AUGUSTUS. *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]*, *65*(1), e57. doi: 10.1002/cpbi.57

Howe, K., Chow, W., Collins, J., Pelan, S., Pointon, D.-L., Sims, Y., … Wood, J. (2021). Significantly improving the quality of genome assemblies through curation. *GigaScience*, *10*(1). doi: 10.1093/gigascience/giaa153

Jiggins, C. D., Mallarino, R., Willmott, K. R., & Bermingham, E. (2006). The phylogenetic pattern of speciation and wing pattern change in neotropicalithomiabutterflies (Lepidoptera: Nymphalidae). Evolution; International Journal of Organic Evolution, 60(7), 1454–1466.

Jiggins, C. D., Naisbit, R. E., Coe, R. L., & Mallet, J. (2001). Reproductive isolation caused by colour pattern mimicry. Nature, 411(6835), 302–305.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., … Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* , *30*(9), 1236–1240. doi:

10.1093/bioinformatics/btu031

Jousselin, E., & Elias, M. (2019). Testing host-plant driven speciation in phytophagous insects : a phylogenetic perspective. arXiv. Retrieved from http://arxiv.org/abs/1910.09510

Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, *20*(4), 1160–1166. doi: 10.1093/bib/bbx108

Kozak, K. M., Wahlberg, N., Neild, A. F. E., Dasmahapatra, K. K., Mallet, J., & Jiggins, C. D. (2015). Multilocus species trees show the recent adaptive radiation of the mimetic heliconius butterflies. Systematic Biology, 64(3), 505–524.

Kolmogorov, M., Armstrong, J., Raney, B. J., Streeter, I., Dunn, M., Yang, F., … Pham, S. (2018). Chromosome assembly of large and complex genomes using multiple references. *Genome Research*, *28*(11), 1720–1732. doi: 10.1101/gr.236273.118

Lefort, V., Longueville, J.-E., & Gascuel, O. (2017). SMS: Smart Model Selection in PhyML. *Molecular Biology and Evolution*, *34*(9), 2422–2424. doi: 10.1093/molbev/msx149

Liu, G., Chang, Z., Chen, L., He, J., Dong, Z., Yang, J., … Li, X. (2020). Genome size variation in butterflies (Insecta, Lepidotera, Papilionoidea): a thorough phylogenetic comparison. *Systematic Entomology*, *45*(3), 571–582. doi: 10.1111/syen.12417

Liu, N.-Y., Xu, W., Dong, S.-L., Zhu, J.-Y., Xu, Y.-X., & Anderson, A. (2018). Genome-wide analysis of ionotropic receptor gene repertoire in Lepidoptera with an emphasis on its functions of Helicoverpa armigera. *Insect Biochemistry and Molecular Biology*, *99*, 37–53. doi: 10.1016/j.ibmb.2018.05.005

Llaurens, V., Joron, M., & Théry, M. (2014). Cryptic differences in colour among Müllerian mimics: how can the visual capacities of predators and prey shape the evolution of wing colours? *Journal of Evolutionary Biology*, *27*(3), 531–540. doi: 10.1111/jeb.12317

Mann, F., Szczerbowski, D., de Silva, L., McClure, M., Elias, M., & Schulz, S. (2020). 3-Acetoxy-fatty acid isoprenyl esters from androconia of the ithomiine butterfly Ithomia salapia. Beilstein Journal of Organic Chemistry, 16, 2776–2787.

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021). 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. doi: 10.1093/molbev/msab199

Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., … Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in Heliconius butterflies. Genome Research, 23(11), 1817–1828.

Mayjonade, B., Gouzy, J., Donnadieu, C., Pouilly, N., Marande, W., Callot, C., … Muños, S. (2016). Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules. *BioTechniques*, *61*(4), 203–205. doi: 10.2144/000114460

McClure, M., Dutrillaux, B., Dutrillaux, A.-M., Lukhtanov, V., & Elias, M. (2017). 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. doi: 10.1159/000487107

McClure, M., & Elias, M. (2016). Ecology, life history, and genetic differentiation in NeotropicalMelinaea(Nymphalidae: Ithomiini) butterflies from north-eastern Peru. *Zoological Journal of the Linnean Society*. doi: 10.1111/zoj.12433

McClure, M., & Elias, M. (2016). Unravelling the role of host plant expansion in the diversification of a Neotropical butterfly genus. BMC Evolutionary Biology, 16(1), 128.

McClure, M., Mahrouche, L., Houssin, C., Monllor, M., Le Poul, Y., Frérot, B., … Elias, M. (2019a). Does divergent selection predict the evolution of mate preference and reproductive isolation in the tropical butterfly genus Melinaea (Nymphalidae: Ithomiini)? *The Journal of Animal Ecology*, *88*(6), 940–952. doi: 10.1111/1365-2656.12975

McClure, M., Clerc, C., Desbois, C., Meichanetzoglou, A., Cau, M., Bastin-Héline, L., … Elias, M.

(2019b). Why has transparency evolved in aposematic butterflies? Insights from the largest radiation of aposematic butterflies, the Ithomiini. Proceedings of the Royal Society of London. Series B: Biological Sciences, 286, 20182769.

Mérot, C., Frérot, B., Leppik, E., & Joron, M. (2015). Beyond magic traits: Multimodal mating cues in Heliconius butterflies. *Evolution; International Journal of Organic Evolution*, *69*(11), 2891–2904. doi: 10.1111/evo.12789

Merrill, R. M., Chia, A., & Nadeau, N. J. (2014). Divergent warning patterns contribute to assortative mating between incipient Heliconius species. Ecology and Evolution, 4(7), 911–917.

Merrill, R. M., Rastas, P., Martin, S. H., Melo, M. C., Barker, S., Davey, J., … Jiggins, C. D. (2019). Genetic dissection of assortative mating behavior. PLoS Biology, 17(2), e2005902.

Meslin, C., Mainet, P., Montagné, N., Robin, S., Legeai, F., Bretaudeau, A., … Jacquin-Joly, E. (2021). Spodoptera littoralis genome mining brings insights on the dynamic of expansion of gustatory receptors in polyphagous noctuidae (p. 2021.12.09.471925). doi: 10.1101/2021.12.09.471925

Montagné, N., Wanner, K., & Jacquin-Joly, E. (2021). 15 - Olfactory genomics within the Lepidoptera. In G. J. Blomquist & R. G. Vogt (Eds.), *Insect Pheromone Biochemistry and Molecular Biology (Second Edition)* (pp. 469–505). London: Academic Press. doi: 10.1016/B978-0-12-819628-1.00015-8

Mugford, S., Wouters, R., Mathers, T. C., & Hogenhout, S. (2020, June 5). High quality DNA extraction from very small individual insects. Retrieved March 29, 2022, from protocols.io website: https://www.protocols.io/view/high-quality-dna-extraction-from-very-small-indivi-e6nvw9nq2gmk/v1

Muller, F. (1879). Ituna and Thyridia; a remarkable case of mimicry in butterflies. *Royal Ent. Soc. London Trans*.

Nadeau, N. J., Whibley, A., Jones, R. T., Davey, J. W., Dasmahapatra, K. K., Baxter, S. W., … Jiggins, C. D. (2012). Genomic islands of divergence in hybridizing Heliconius butterflies identified by large-scale targeted sequencing. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 367(1587), 343–353.

Pinna, C. S., Vilbert, M., Borensztajn, S., Daney de Marcillac, W., Piron-Prunier, F., Pomerantz, A., … Elias, M. (2021). Mimicry can drive convergence in structural and light transmission features of transparent wings in Lepidoptera. *eLife*, *10*. doi: 10.7554/eLife.69080

Piron-Prunier, F., Persyn, E., Legeai, F., McClure, M., Meslin, C., Robin, S., … Gauthier, J. (2021). Comparative transcriptome analysis at the onset of speciation in a mimetic butterfly, the Ithomiini *Melinaea marsaeus*. *Journal of Evolutionary Biology*. doi: 10.1111/jeb.13940

Rhie, A., Walenz, B.P., Koren, S. et al. (2020). Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biology 21, 245

Robertson, H. M. (2019). Molecular Evolution of the Major Arthropod Chemoreceptor Gene Families. *Annual Review of Entomology*, *64*, 227–242. doi: 10.1146/annurev-ento-020117-043322

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*(1), 24–26. doi: 10.1038/nbt.1754

Schulz, S., Beccaloni, G., Brown, K. S., Boppré, M., Freitas, A. V. L., Ockenfels, P., & Trigo, J. R. (2004). Semiochemicals derived from pyrrolizidine alkaloids in male ithomiine butterflies (Lepidoptera: Nymphalidae: Ithomiinae). Biochemical Systematics and Ecology, 32(8), 699–713.

Seixas, F. A., Edelman, N. B., & Mallet, J. (2021). Synteny-Based Genome Assembly for 16 Species of Heliconius Butterflies, and an Assessment of Structural Variation across the Genus. *Genome Biology and Evolution*, *13*(7). doi: 10.1093/gbe/evab069

Seppey, M., Manni, M., & Zdobnov, E. M. (2019). BUSCO: Assessing Genome Assembly and Annotation Completeness. In M. Kollmar (Ed.), *Gene Prediction: Methods and Protocols* (pp. 227–245). New York, NY: Springer New York. doi: 10.1007/978-1-4939-9173-0_14

Sherratt, T. N. (2008). The evolution of Müllerian mimicry. *Die Naturwissenschaften*, *95*(8), 681–695. doi: 10.1007/s00114-008-0403-y

Slater, G. S. C., & Birney, E. (2005). Automated generation of heuristics for biological sequence

comparison. *BMC Bioinformatics*, *6*, 31. doi: 10.1186/1471-2105-6-31
Stamm, P., Mann, F., McClure, M., Elias, M., & Schulz, S. (2019). Chemistry of the Androconial Secretion of the Ithomiine Butterfly Oleria onega. Journal of Chemical Ecology, 45(9), 768–778.
Storer, J., Hubley, R., Rosen, J., Wheeler, T. J., & Smit, A. F. (2021). The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mobile DNA*, *12*(1), 2. doi: 10.1186/s13100-020-00230-y
Tarailo-Graovac, M., & Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]*, *Chapter 4*, Unit 4.10. doi: 10.1002/0471250953.bi0410s25
Visendi, P. (2022). De Novo Assembly of Linked Reads Using Supernova 2.0. In D. Edwards (Ed.), *Plant Bioinformatics: Methods and Protocols* (pp. 233–243). New York, NY: Springer US. doi: 10.1007/978-1-0716-2067-0_12
van Schooten, B., Meléndez-Rosa, J., Van Belleghem, S. M., Jiggins, C. D., Tan, J. D., McMillan, W. O., & Papa, R. (2020). Divergence of chemosensing during the early stages of speciation. Proceedings of the National Academy of Sciences of the United States of America, 117(28), 16438–16447.
Vogt, R. G., Große-Wilde, E., & Zhou, J.-J. (2015). The Lepidoptera Odorant Binding Protein gene family: Gene gain and loss within the GOBP/PBP complex of moths and butterflies. *Insect Biochemistry and Molecular Biology*, *62*, 142–153. doi: 10.1016/j.ibmb.2015.03.003
Willmott, K. R., & Mallet, J. (2004). Correlations between adult mimicry and larval host plants in ithomiine butterflies. *Proceedings. Biological Sciences / The Royal Society*, *271 Suppl 5*, S266–S269. doi: 10.1098/rsbl.2004.0184
Willmott, K. R., Robinson Willmott, J. C., Elias, M., & Jiggins, C. D. (2017). Maintaining mimicry diversity: optimal warning colour patterns differ among microhabitats in Amazonian clearwing butterflies. *Proceedings. Biological Sciences / The Royal Society*, *284*(1855). doi: 10.1098/rspb.2017.0744
Zhan, S., Merlin, C., Boore, J. L., & Reppert, S. M. (2011). The monarch butterfly genome yields insights into long-distance migration. *Cell*, *147*(5), 1171–1185. doi: 10.1016/j.cell.2011.09.052

TABLES

Table 1. Statistics of raw read data including sequencing strategy, read length, number of reads and total sequenced bases.

Table 2. Assembly statistics and completeness evaluation.

Table 3. Annotation statistics and predicted protein completeness evaluation.

Table 4. Number of chemosensory genes annotated in different lepidopteran genomes.

FIGURE CAPTIONS

Figure 1. Wing pattern variation between subspecies of *Melinaea marsaeus*, *Melinaea menophilus* and *Ithomia salapia* (source (Mathieu Joron et al., 2006) and photograph credits: Céline Houssin).

Figure 2. Low synteny between *M. marsaeus* and *M. menophilus* despite very recent splitting time. The positions of BUSCO genes mapping uniquely to both genomes are shown in the order of the *M. marsaeus* chromosomes. The colours reflect the different *M. marsaeus* chromosomes. A fully conserved chromosome would be reflected as a single diagonal line as in *M. marsaeus* chromosome 7 which corresponds to *M. menophilus* chromosome 1. Grey lines indicate chromosome ends.

Figure 3. Phylogeny and orthologous gene numbers across 10 butterfly genomes. "Shared by some" represents orthologues shared by 8 out of the 10 species and without phylogenetic signal.

Figure 4. Maximum-likelihood phylogeny of lepidopteran GRs, built from amino acid sequences from *B. mori*, *H. melpomene*, *D. plexippus*, *I. salapia*, *M. marsaeus* and *M. menophilus*. Deep nodes highly supported by the likelihood-ratio test (aLRT>0.95) are indicated by black dots. Those that correspond to Ithomiini-specific large expansions (more than 10 genes) are shown with stars. The scale bar represents the expected number of amino acid substitutions per site.

Figure 5. Maximum-likelihood phylogeny of lepidopteran CSPs, built from amino acid sequences from *B. mori* (Bmor), *S. frugiperda* (Sfru), *H. melpomene* (Hmel), *D. plexippus* (Dple), *I. salapia* (Isal), *M. marsaeus* (Mmar) and *M. menophilus* (Mmen). Deep nodes highly supported by the likelihood-ratio test (aLRT>0.95) are indicated by black dots. The scale bar represents the expected number of amino acid substitutions per site.

SUPPLEMENTARY MATERIAL

Supplementary Table 1. Estimation of genome statistics including heterozygosity (Het), read error rate (err), genome size estimation and model fitting of the k-mer distribution (fit).

Supplementary Figure 1. k-mer profile and model fit as estimated with GenomeScopev.2.0 for each species and each sequencing library using a k-mer length of 31 bp.

Supplementary Figure 2. Hi-C contact map generated by SALSA2 for the chromosome scale genomes of M. marsaeus and M. menophilus.

Supplementary Figure 3. Transposable element content profiles of the three ithomiine species assemblies and *D. plexippus*.