The *Hypolimnas misippus* genome supports a common origin of the W chromosome in Lepidoptera

Anna Orteu[1,2*], Shane A. McCarthy[2,3], Emily A. Hornett[4,5], Mathew Gemmel[4], Louise A. Reynolds[4], Ian A. Warren[1], Ian J. Gordon[6], Gregory D. D. Hurst[4], Richard Durbin[3], Simon H. Martin[7], Chris D. Jiggins[1]

1. Department of Zoology, University of Cambridge, Cambridge, United Kingdom
2. Tree of Life Programme, Wellcome Sanger Institute, Hinxton, United Kingdom
3. Department of Genetics, University of Cambridge, Cambridge, United Kingdom
4. IVES, University of Liverpool, Liverpool, United Kingdom
5. Vector Biology, Liverpool School of Tropical Medicine, Liverpool, United Kingdom
6. Centre of Excellence in Biodiversity and Natural Resource Management, University of Rwanda, Huye Campus, Huye, Rwanda
7. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom

* Author for Correspondence: Anna Orteu, Tree of Life Programme, Wellcome Sanger Institute, Hinxton, United Kingdom, ao20@sanger.ac.uk

Abstract

Moths and butterflies (Lepidoptera) have a heterogametic sex chromosome system with females carrying ZW chromosomes and males ZZ. The lack of W chromosomes in early diverging lepidopteran lineages has led to the suggestion of an ancestral Z0 system in this clade and a B chromosome origin of the W. This contrasts with the canonical model of W chromosome evolution in which the W would have originated from the same homologous autosomal pair as the Z chromosome. Despite the distinct models proposed, the rapid evolution of the W chromosome has hindered the elucidation of its origin. Here, we present high-quality, chromosome-level genome assemblies of two *Hypolimnas* species (*Hypolimnas bolina* and *Hypolimnas misippus)* and use the *H. misippus* assembly to explore the evolution of W chromosomes in butterflies and moths. We show that in *H. misippus* the W chromosome has higher similarity to the Z chromosome than any other chromosome, which could suggest a possible origin from the same homologous autosome pair as the Z chromosome. However, using genome assemblies of closely related species (ditrysian lineages) containing assembled W chromosomes, we present contrasting evidence suggesting that the W chromosome might have evolved from a B chromosome instead. Crucially, by using a synteny analysis to infer homology, we show that W chromosomes are likely to share a common evolutionary origin in Lepidoptera. This study highlights the difficulty of studying the evolution of W chromosomes and contributes to better understanding its evolutionary origins.

Key words

B chromosome, sex chromosome evolution, synteny, genome assembly, neo-Z chromosome

Significance

Butterflies and moths have a sex determination system in which females carry two different sex chromosomes, Z and W, while males carry two copies of the Z. The evolutionary origin of the W chromosome has been elusive, with many possible scenarios being suggested, such as the independent evolution of W chromosomes in many butterfly and moth species. Here, we present genome assemblies of two *Hypolimnas* butterfly species and use one of them to shed light on the evolution of the W chromosome. We show that W chromosomes across butterflies and moths are very similar which suggests a shared common origin.

Introduction

Sex chromosomes are highly variable in eukaryotes and have evolved independently multiple times (Bachtrog et al. 2014, 2011; Beukeboom & Perrin 2014). In animals there are multiple types of chromosomal sex determination systems but two are predominant: male heterogamety as seen in mammals where males are XY and females XX; and female heterogamety as seen in birds where females are ZW and males ZZ (Beukeboom & Perrin 2014). These heteromorphic sex chromosomes (XY and ZW) can potentially originate from different processes. One possibility is that they initially arise from a pair of autosomes that evolve genetic sex determination and, through a process of reduced recombination, sex specific mutations accumulate (Wright et al. 2016). This reduction in recombination can also lead to gene depletion and the accumulation of repeat and transposable elements (TEs), which are common characteristics of sex specific (Y or W) chromosomes (Bachtrog et al. 2011; Wright et al. 2016). However, discerning between cause and consequence can be difficult, as repeats and TEs could be enhancing the reduction in recombination rather than resulting from it. Additionally, other autosomes may fuse to the sex chromosomes and become differentiated. Alternatively, the W/Y chromosomes can originate from the recruitment of a B

chromosome, which are dispensable chromosomes that are found variably in populations and species (Yoshida et al. 2011).

Moths and butterflies commonly have a ZW sex chromosome system in which females are heterogametic and show a lack of recombination (Turner & Sheppard 1975). Whilst the Z chromosome has been shown to be highly conserved across the Lepidoptera (Fraïsse et al. 2017), the origin and evolution of the W chromosome remains unclear and several putative origins have been hypothesised. First, some recent evidence has been shown to suggest that the W chromosome originated from the same homologous chromosome pair as the Z (Dai et al. 2022)(Figure 1A). However, the possible absence of W chromosomes in early-diverging ditrysian lineages (a lepidopteran clade containing all butterflies and most moths) and the deep conservation of the Z chromosome have been suggested to be evidence of an ancestral Z0 sex determination system for Lepidoptera, and an origin for the W chromosome that is independent of the Z chromosome (Dalíková et al. 2017; Fraïsse et al. 2017; Hejníčková et al. 2019; Lukhtanov 2000; Sahara et al. 2012). Two main alternative hypotheses have been proposed for the origin of the W chromosome (that is independent of the Z chromosome): evolution from a B chromosome (Dalíková et al. 2017; Lewis et al. 2021; Lukhtanov 2000) (Figure 1B), or evolution from the homologous pair of an autosome that fused to the Z chromosome (Sahara et al. 2012)(Figure 1C). Furthermore, and irrespective of the specific origin, the number of evolutionary events leading to the formation of the W chromosome has also been debated. That is whether all W chromosomes in the Lepidoptera share a common evolutionary origin of if their inception was the result of multiple independent events, with the most recent evidence supporting the latter (Dai et al. 2022; Lewis et al. 2021)(Figure 1D-E).

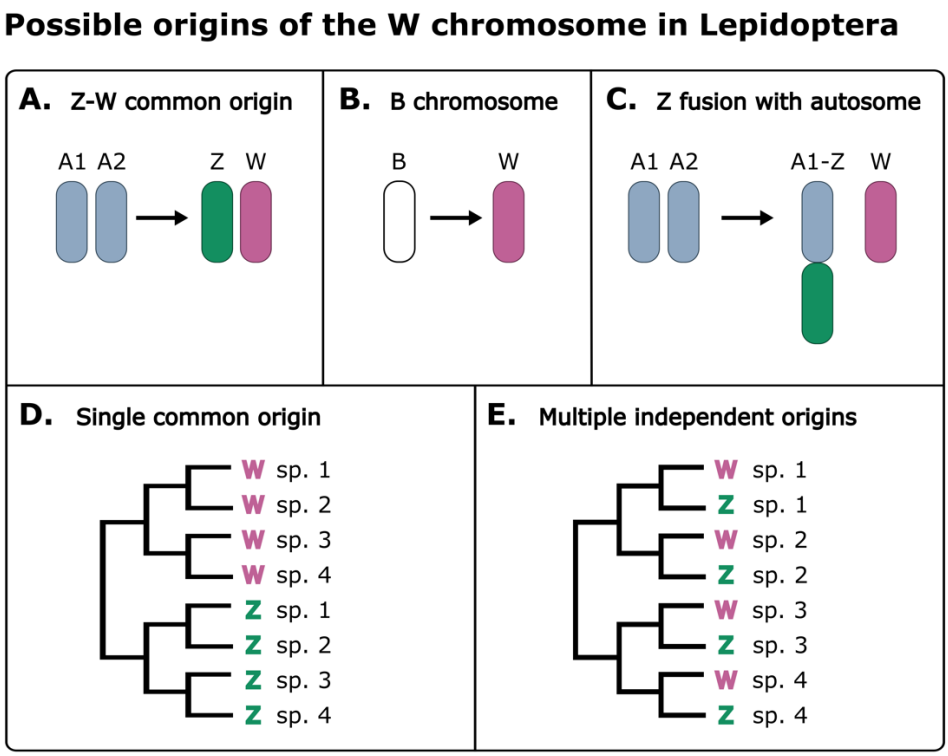


Fig. 1. Possible origins of the W chromosome in Lepidoptera. **A.** One hypothesis on the origin of W chromosomes is that Z and W evolved from the same autosomal pair. **B.** Another hypothesis is that the W chromosome evolved from a B chromosome, which are non-essential chromosomes that vary in number in populations and species. **C.** Finally, the inception of W chromosomes could have involved the fusion of the Z chromosome to an autosome and subsequent formation of the W from the autosomal pair. **D.** In any of these cases, if the W chromosome has originated only once in Lepidoptera, W chromosomes of all species would be more similar to each other than to the Z chromosomes. **E.** Whereas if they evolved multiple times from the same autosomal pair as the Z, W chromosomes would be more similar to the Z of the same species than to other Ws. While, multiple B chromosome recruitments would result in a lack of similarity between Z and W chromosomes as well as among Ws.

Despite interest in understanding the evolution of the W chromosome in the Lepidoptera, the absence of high-quality reference genomes containing assembled W chromosomes has limited its study. The elevated repeat and TE content have made its assembly challenging until recently. Long read sequencing technologies and the decrease in sequencing price have enabled the production of high-quality Lepidopteran genome assemblies containing W chromosomes, which makes it increasingly possible to elucidate the enigmatic origin of the Lepidoptera W chromosome.

Hypolimas butterflies, commonly known as eggflies, are a phenotypically diverse genus that has served as a model for the study of ecology and evolutionary biology. Many *Hypolimnas* species are mimics of toxic species, which has shaped the diversification of wing colour pattern in the genus (Vane-Wright et al. 1977). Historically, two species, *Hypolimnas bolina* and *Hypolimnas misippus*, have received most of the attention. *Hypolimnas bolina* and *H. misippus* diverged 8 million years ago and share many similarities, both have a pantropical distribution and the females of both species are polymorphic Batesian mimics of toxic models (Marsh et al. 1977; Sahoo et al. 2018; Smith 1976). In contrast, males are monomorphic and have retained what is likely to be the ancestral phenotype of white-spotted black wings. Nonetheless, the two species also differ in many aspects. *Hypolimnas misippus* females are mimics of the four morphs of the African Queen, *Danaus chrysippus*, while *H. bolina* females are mimics of several *Euploea* species. *Hypolimnas bolina* has also received special interest for its coevolution with the endosymbiont *Wolbachia* (Charlat et al. 2009; Dyson et al. 2002). In *H. bolina*, *Wolbachia* has a male-killing phenotype that promotes spread of the endosymbiont through females but has counter-evolved a suppressor locus that rescues male butterflies (Hornett et al. 2006).

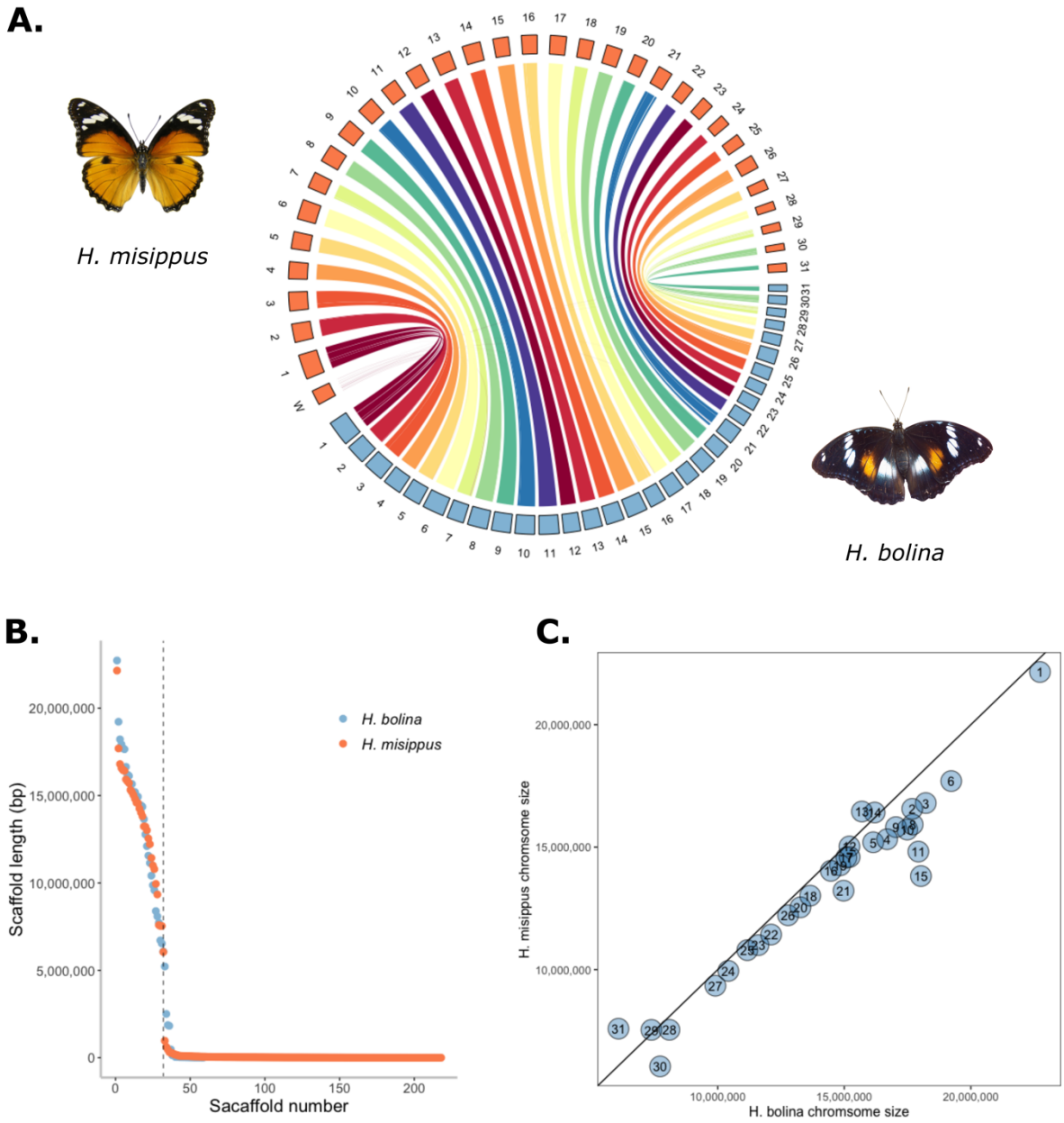
124 The diversity in phenotype, precision of mimicry and intricate coevolution, make *Hypolimnas*
125 a remarkable genus for evolutionary biology studies. However, there are few genomic
126 resources to date. Here, we present chromosome-level assemblies for *H. misippus*
127 (HypMis_v2) and *H. bolina* (HypBol_v1) and use our high-quality *H. misippus* assembly
128 containing the Z and W chromosomes to explore the evolution of sex chromosomes in
129 Lepidoptera. First, we compare synteny and TE content between the two *Hypolimnas*
130 assemblies. Next, we evaluate the annotation completeness and gene content by comparison
131 to the closely related painted lady butterfly, *Vanessa cardui*. Then, we investigate the
132 hypothesis of a B chromosome origin of the W chromosome by comparing homology
133 between the W chromosome, autosomes, and Z chromosome within *H. misippus*. Finally, we
134 examine the origin of the W and Z chromosomes across the Lepidoptera by analysing synteny
135 between *Hypolimnas misippus* and a diverse set of 10 Lepidoptera species in different
136 ditrysian families.

137 **Results**

138 **Genome assemblies and synteny between *HypMis_v2* and *HypBol_v1***

139 The size of the final assemblies were 438.07 MB for HypMis_v2 and 444.68 MB for
140 HypBol_v1, assembled into 218 and 59 scaffolds respectively. HypMis_v2 was sequenced
141 using a trio-binning strategy, which, by using a combination of short and long-read
142 sequencing of the parents and offspring of a cross, allows for the assembly of two parental
143 haplotypes (Yen et al. 2020). We assembled both parental haplotypes to a quasi-chromosome
144 level and then scaffolded haplotype 2, which corresponds to the maternal haplotype, using
145 Hi-C data. Hereafter, all mentions of the *H. misippus* assembly, HypMis_v2, refer to the Hi-C
146 scaffolded haplotype 2 assembly.

147 HypMis_v2 was assembled into 32 chromosome-level scaffolds (>6 MB; 30 autosomes and
148 the Z and W chromosomes) and 186 unplaced scaffolds smaller than 1MB (Figure 2B) and
149 has an N50 of 14.6 MB; while HypBol_v1 was assembled into 59 scaffolds all placed onto 31
150 chromosomes, with an N50 of 15.2 MB. Thus, both species have retained the ancestral
151 karyotype of 31 chromosomes. This karyotype is present in other Nymphalinae species such
152 as the painted lady *Vanessa cardui*, which has a comparable genome size (424.8 Mb) to the
153 two *Hypolimnas* species. In general, all chromosomes were slightly larger in *HypBol_v1*
154 compared to *HypMis_v2* with only 4 exceptions (chromosomes 13, 14, 28 and 31; Figure
155 2C). When aligning the two *Hypolimnas* assemblies, there were no fusions or fissions among
156 chromosomes but with multiple large rearrangements (Figure 2A; Supplementary Figure 1).
157 The only caveat to these conclusions is that *HypMis_v2* was used to scaffold *HypBol_v1*
158 before using the linkage map, which might have altered the *H. bolina* chromosomal structure.
159 Nonetheless the linkage map applied after the Ragout scaffolding should correct for any
160 large-scale errors in scaffolding. Finally, 12 large inversions on multiple chromosomes are
161 apparent in the alignment of the two assemblies (Supplementary Figure 1).



162

163 **Figure 2.** Chromosome level assemblies for *Hypolimnas misippus* and *H. bolina*. **A.**

164 Chromosomal synteny is conserved in both species. **B.** Chromosome level scaffolds have

165 been assembled for both species. **C.** In general, *H. bolina* chromosomes are larger.

166 Chromosome 1 refers to the Z chromosome.

167

168 **Transposable elements and repeat content**

169 In total, the two *Hypolimnas* genomes contain similar levels of repeats, 43.86% for
170 HypBol_v1 and 42.34% for *H. misippus*, which are slightly higher than in the painted lady
171 *Vanessa cardui* (37.27%) of comparable genome size (424.8 Mb). The composition of
172 repeats is similar in the two *Hypolimnas* genomes, while they differ substantially from the
173 painted lady (Figure 3). The contribution of rolling-circles (also known as *Helitrons*) and
174 DNA-transposons is more than double in *Hypolimnas,* while retroelements are >1.5 fold
175 higher in the painted lady (Figure 3). This suggests a shift in TE activity, with *Helitrons* and
176 DNA-transposons playing a more important role in *Hypolimnas* species. In both *Hypolimnas*,
177 there has been both a recent and a more ancient expansion of *Helitron* family transposable
178 elements (Figure 3B). Finally, the percentage of unclassified repeats and GC content is
179 broadly the same in the three species.

180

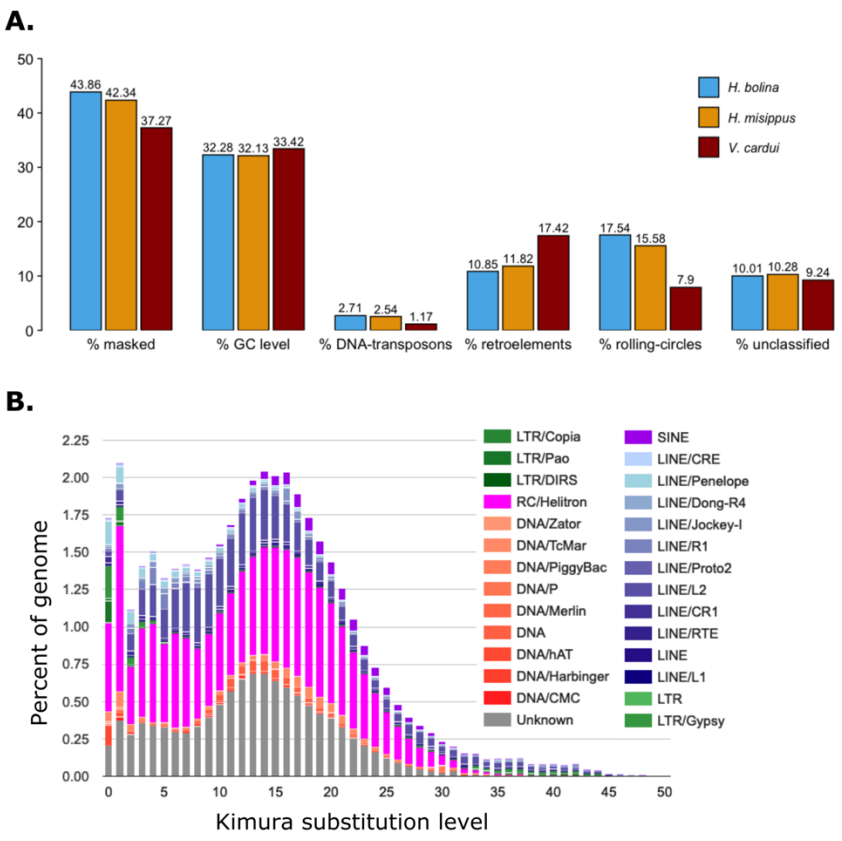


Fig. 3. A. Repeat content of the *Hypolimnas bolina*, *Hypolimnas misippus* and *Vanessa cardui* assemblies. **B**. Repeat landscape of the *H. misippus* assembly. Helitrons and LTR (Long Terminal Repeat) retrotransposons have undergone a recent expansion, as seen in the higher percentage of the genome covered by *Helitrons*.

Gene content and completeness of the annotation

In total 19,721 genes were annotated in HypBol_v1 and 21,784 coding mRNAs including all isoforms of the same gene; while for HypMis_v2 there were 20,293 genes and 22,468 coding mRNAs. These numbers are higher than other Nymphalidae species such as the painted lady butterfly (*Vanessa cardui)*, whose latest annotation (v2.1) includes 13,223 genes and 19,836 mRNAs, with the number of genes being considerably larger in the two *Hypolimnas* species (Supplementary Figure 2). Analysis with BUSCO using the *insecta_odb10* benchmarking set

193 showed that the completeness of the genome and annotation were 98.7% and 98.6% for
194 HypMis_v2, and 94% and 92.4% for HypBol_v1 (Table 1). These scores are comparable to
195 other published Nymphalidae assemblies and annotations such as that of *Danaus chrysippus*
196 (Singh et al. 2022), with those for *HypBol_v1* slightly lower possibly due to errors introduced
197 by Nanopore sequencing, which has a high per base error rate.

	Species	**Complete**	**Single**	**Dupl.**	**Frag.**	**Miss.**
Genomes	*Hypolimnas misippus* haplotype 1 (parental)	98.7	98.5	0.2	0.4	0.9
	Hypolimnas misippus haplotype 2 (maternal)	98.7	98.5	0.2	0.4	0.9
	Hypolimnas bolina	94.0	93.6	0.4	1.8	4.2
Annotations	*Hypolimnas misippus*	98.6	89.8	8.8	0.6	0.8
	Hypolimnas bolina	92.4	84.1	8.3	2.9	4.7

198 **Table 1.** BUSCO scores for the genome assemblies and gene annotations of *H. misippus* and
199 *HypBol_v1* calculated using the Insecta_odb10 (n=1367) set of genes.

200 Despite having a larger number of genes and mRNAs, the two *Hypolimnas* annotations show
201 a smaller total number of exons and introns than *V. cardui* (Supplementary Figure 2). This is
202 because the two *Hypolimnas* annotations have more single exon genes and their mRNAs
203 have, on average, a smaller number of exons, which might be a difference produced by the
204 distinct annotation pipelines. The total length of mRNAs and genes in the two *Hypolimnas* is
205 shorter then in *V. cardui*, which results in a smaller percentage of the genome covered by
206 them. However, this trend is different for exons, which are on average the same length in the
207 three species, have a comparable total length and cover a similar percentage of the genome.

208 Thus, the longer total length of mRNAs and genes in *V. cardui* is due to a total longer length
209 of introns, due to a higher number of introns per mRNA and an average longer length.

210 **Mixed evidence on the origin of W chromosomes from the same autosome pair as the Z**

211 Using our HypMis_v2 assembly that contains a W chromosome, we set to explore the
212 evolutionary origins of the W chromosome in Lepidoptera. First, we wanted to clarify
213 whether the W chromosome had originated from a B chromosome or form the same
214 autosomal pair as the Z (Figure 1A-B). To specifically test the hypothesis that the Z and W
215 have a common autosomal origin, we used a synteny analysis to infer sequence homology
216 between the *H. misippus* W chromosome and the remaining 31 chromosomes. We would
217 expect that if the W and the Z share a common autosomal origin, they would share more
218 similarities between them than with any other chromosome. We found that the Z
219 chromosome shows a higher degree of similarity to the W chromosome than any autosome,
220 defined as the percentage of the W that is covered by non-overlapping matches to the target
221 chromosome (Figure 4A). However, the W chromosome is the chromosome with the highest
222 percentage of repeat regions (71% in *H. misippus*), which could be interfering with the
223 analysis. Furthermore, the Z chromosome is the longest chromosome in *H. misippus* and has
224 the highest absolute length of repeats (Supplementary Figure 3 and 4). To take this into
225 account, we filtered out of the synteny analysis all the repeat regions in the *H. misippus* W
226 chromosome. With this correction, the Z chromosome was the second most similar
227 chromosome to the W after chromosome 2.

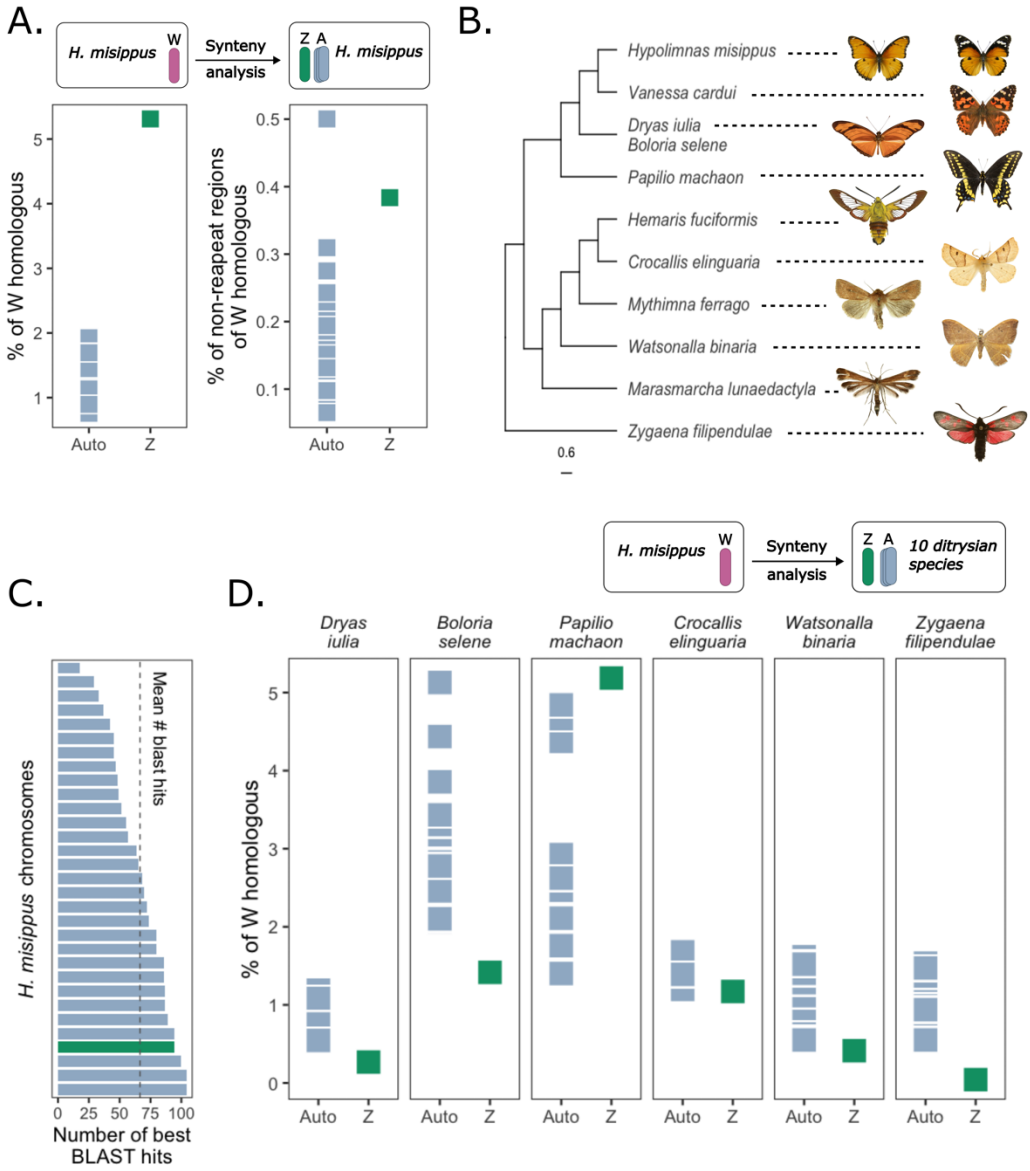


Fig. 4. Genomic synteny of the *H. misippus* W chromosome provides mixed evidence on its origin. **A.** Synteny analysis of the *H. misippus* W chromosome and the autosomes and Z chromosome shows that the W and Z chromosome have the highest similarity. The analysis of non-repeat regions shows that the Z chromosome is the second most similar chromosome to the W after chromosome 19 (right). A schematic of the method is shown on top. **B.** Phylogenetic tree modified from Kawahara, et al 2019 (Kawahara et al. 2019) including the species used in the synteny analysis. **C.** Number of best BLAST hits of the proteins found in the *H. misippus* W chromosome to the other chromosomes. Numbers summarised by

237 chromosome with the Z shown in green. **D.** Synteny analysis of the *H. misippus* W
238 chromosome compared to several Lepidoptera genome assemblies shows homology between
239 W chromosomes across many species and suggests a common origin. A schematic of the
240 method is shown on top.

241

242 To further explore the possible homology between the W and Z chromosomes, we extracted
243 the predicted sequences of the proteins in the *H. misippus* W chromosome and searched for
244 homology in the remaining chromosomes using BLASTp. Contrarily to the synteny results,
245 we found that the Z chromosome placed 4th in the number of best BLAST protein hits with 94
246 proteins mapping to it, while 104 mapped to chromosome 2, 3 and 14 (Figure 4C). The
247 average number of best BLAST hits was 66. These results suggests that the W chromosome
248 did not evolve from the same homologous autosomal pair as the Z. However, as a result of
249 the arrest of recombination and subsequent accumulation of mutations, sex specific
250 chromosomes as the W are often gene poor and highly degenerated compared to their
251 autosomal counterparts (Wright et al. 2016). Additionally, lepidopteran W chromosomes
252 have been shown to evolve rapidly (Yoshido et al. 2013; Vítková et al. 2007). Thus, the lack
253 of homology between proteins of the W and other chromosomes is expected given the rapid
254 turnover seen in the W chromosome and is in line with results shown in other Lepidoptera
255 species (Yang et al. 2020; Singh et al. 2022; Lewis et al. 2021). Interestingly, chromosome 2
256 showed the highest sequence similarity in non-repeat regions and the highest number of best
257 BLAST hits (Figure 5C).

258 In view of the differences between these results and previously published homology tests in
259 other species, we decided to expand the analysis to include other Lepidoptera species. We
260 performed synteny analysis between the W and the other 31 chromosomes in 6 diverse taxa:

261 *Dryas iulia, Boloria selene, Papilio machaon, Crocallis elinguaria, Watsonalla binaria,* and
262 *Zygaena filipendulae*. These species were chosen as they have a publicly available high-
263 quality assembly containing a W chromosome, and cover distinct ditrysian lineages
264 (including 4 families: *Zyganoidea, Papilonoidea, Drepanoidea* and *Geometroidea*). No
265 homology was observed between the W and Z in any species except for *P. machaon*, which
266 contrasts with the results in *H. misippus* (Figure 4D). However, these analyses were
267 performed using all regions as no repeat annotation was available for these species, which
268 could be interfering with the results.

269 **Synteny analysis of the *H. misippus* W chromosome reveals a possible single common**
270 **origin of the W chromosome in ditrysian taxa**

271 The second question we wanted to clarify was whether the W has a single origin in the
272 Lepidoptera or evolved independently multiple times (Figure 1D-E). To explore the evolution
273 of W chromosomes within ditrysian lineages, we performed a synteny analysis of the *H.*
274 *misippus* W chromosome compared to a diverse set of 10 Lepidoptera species, including the
275 6 species from the previous analysis plus *Vanessa cardui, Hemaris fuciformis, Mythimna*
276 *farrago* and *Marasmarcha lunaedactyla*. These species cover the breadth of ditrysian
277 lineages, including 8 distinct families: *Zyganoidea, Papilonoidea, Pterophonoidea,*
278 *Pyraloidea, Drepanoidea, Noctuoidea, Geometroidea* and *Bombycoidea*. We performed the
279 comparisons including all the chromosomes of the target species (including the W; Figure
280 5A) but using only the W chromosome of *H. misippus* as query. With this we can assess if W
281 chromosomes are more similar to each other than any autosome or the Z chromosome and
282 clarify if they share a common origin. When comparing the *H. misippus* W chromosome to
283 these species, we found that there is a high level of similarity among W chromosomes,
284 defined again as the percentage of non-repeat regions of the W chromosome covered by

285 matches to target chromosomes (Figure 5A). The highest degree of similarity is seen between
286 W chromosomes of closely related species such as *H. misippus* and *Vanessa cardui* (45
287 MYa), but is also true even for highly divergent lineages such as *Zygaena filipendulae,* which
288 diverged from *H. misippus* 156 MYa (Kawahara et al. 2019). Interestingly, there is no
289 consistent pattern in the distribution of homologous regions along the *H. misippus* W
290 chromosome, which means that different regions show homology to the *H. misippus* W
291 chromosome across species and chromosome types (W, Z and autosomes; Supplementary
292 Figure 5).

293 We then performed the same analysis including all chromosomes of *H. misippus* as the query.
294 With this we could assess similarity of all autosomes and sex chromsomes of *H. misippus* to
295 each of the target species. This approach produces only primary matches for each query
296 chromosome and is thus efficient at finding the homologous pair. However, using a single
297 chromosome as query as previously done ensures that secondary and tertiary matches, that is
298 matches that have low score, are also produces and is thus more exhaustive at describing
299 similarities of *one* chromosome has with any autosome or sex chromosome of another
300 assembly. This approach yielded the same results as the former, in which the W chromosome
301 of *H. misippus* share most similarities with the W of most species (Figure 5B right). Finally,
302 the Z chromosome does not show more similarity to the W than the autosomes and is highly
303 conserved across species (Figure 5B left).

304

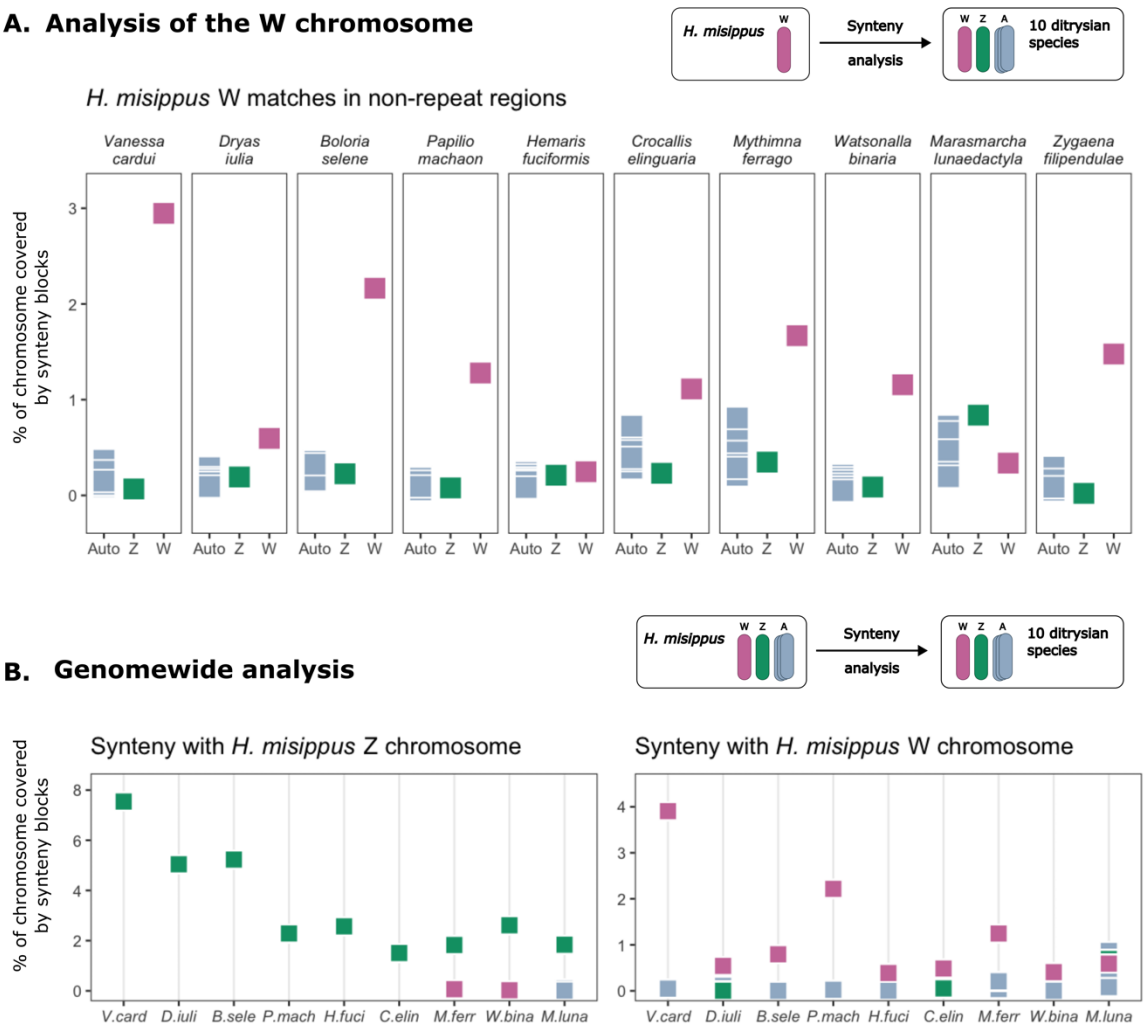


Fig. 5 Genome wide synteny analysis suggests a single origin of ditrysian W chromosomes and shows deep conservation of the Z chromosome. **A**. Percentage of query *H. misippus* W chromosome matched to the 32 chromosomes of 10 ditrysian species. W *H. misippus* chromosome used as query (a schematic of the method is shown on top). **B.** Subset results of genome wide analysis. Percentage of Z (left) and W (right) chromosome covered by synteny blocks to other ditrysian species. W chromosome shown in pink, Z in green and autosomes in grey. A schematic of the method is shown on top. Species names have been abbreviated: *Vanessa cardui*, *Dryas iulia*, *Boloria selene*, *Papilio machaon*, *Hemaris fuciformis*, *Crocallis elinguaria*, *Mythimna ferrago*, *Watsonalla binaria*, *Marasmarcha lunaedactyla* and *Zygaena filipendulae*.

316 **A multi-species synteny comparison reinforces the hypothesis of the single origin of the**
317 **W chromosome and reveals a neo-Z chromosome**

318 In light of the results of the *H. misippus* synteny analyses suggesting a single origin of the W
319 chromosome in ditrysian lineages, we decided to explore this further by comparing the
320 analysed ditrysian assemblies with each other. We performed pairwise synteny analyses
321 between all species pairs and observed consistent results. First, the Z chromosome is deeply
322 conserved in multiple species pairs and the homology between Z chromosomes of different
323 species decays with species divergence in a similar manner to autosomes (Supplementary
324 Figure 6). Second, the W chromosome shows a more variable pattern of homology between
325 species but is conserved even with distant species pairs (Supplementary Figure 6).
326 Nonetheless, several of the comparisons found no similarity between W chromosomes.
327 Finally, the cross-species comparison revealed that the Z chromosome of *Marasmarcha*
328 *lunaedactyla* (chromosome name OV181339.1; identity as the Z chromosome assigned in the
329 public assembly) presented high levels of synteny with chromosome 19 of most other species
330 (Supplementary figures 7 and 8). This could suggest a historical fusion event between the
331 ancestral chromosome 19 and the Z chromosome of *M. lunaedactyla* creating a neo-Z
332 chromosome. The pattern of homologous regions seen in the OV181339.1 chromosome was
333 similar for all species compared, with approximately half of the chromosome showing
334 homology to chromosome 19 and half to the Z chromosome (Supplementary Figure 8).
335 Furthermore, the pattern of homology with other species showed that recombination events
336 between the parts belonging to the ancestral Z and ancestral chromosome 19 have resulted in
337 an interleaved pattern of homology blocks. Consistent with these results, BUSCO matches of
338 chromosome OV181339.1 are shared with chromosome 19 and chromosome 1 (the Z
339 chromosome) of *Melitaea cinxia* (Supplementary Figures 7 and 8).

340 **Discussion**

341 *Hypolimnas* species have become a focus for studies of evolutionary genetics, due to their
342 mimetic colouration and co-evolution with *Wolbachia* parasites. Here, we have assembled
343 chromosome level reference genomes for *H. misippus* and *H. bolina* and RNA-informed
344 annotations for both genomes. HypBol_v1 represents the first assembly available for
345 *Hypolimnas bolina,* while HypMis_v2 is a significant improvement on a previously published
346 genome, with higher contiguity and higher N50 and BUSCO scores. Furthermore, the
347 annotation is also improved as we used RNA-seq data from *H. misippus* to inform our
348 annotation, as opposed to only homology methods. We present these two useful resources
349 and use the *H. misippus* assembly to shed light on the evolutionary origins of W
350 chromosomes in Lepidoptera.

351 Several models have been proposed to explain the origin and evolution of the W chromosome
352 in Lepidoptera (Dai et al. 2022; Fraïsse et al. 2017; Lukhtanov 2000; Sahara et al. 2012). The
353 deep conservation of the Z chromosome and its apparent lack of homology with the W have
354 been suggested to be evidence of a B chromosome origin of the W chromosome (Fraïsse et
355 al. 2017). Outside the Lepidoptera, it has been commonly found that heteromorphic sex
356 chromosomes retain a degree of similarity; evidence of their shared autosomal origin (Wright
357 et al. 2016). Thus, if the W chromosome originated from the same homologous autosome pair
358 as the Z, we might expect to find some residual sequence similarity between the W and Z, as
359 some recent evidence seems to suggest (Dai et al. 2022). However, opposing evidence has
360 also been presented, suggesting a lack of similarity between W and Z chromosomes which
361 could point to the evolution of the W from a B chromosome (Fraïsse et al. 2017; Lewis et al.
362 2021). Furthermore, the lack of similarity among W chromosomes of Lepidoptera species has
363 led to the suggestion of multiple independent recruitments from B chromosomes (Dai et al.

364 2022; Lewis et al. 2021). Taking this into account, we set out to elucidate two questions.
365 First, did the W chromosome evolve from the same homologous pair of autosomes as the Z
366 or from a B chromosome? Second, do W chromosomes in Lepidoptera share a common
367 origin, that is, did they evolve once or multiple times independently?

368 Using homology tests between the W and remaining 31 chromosomes in *H. misippus* and
369 other Lepidoptera we show that there is a degree of similarity between the W and the Z
370 chromosome, suggesting a possible common origin. These results differ from the homology
371 tests performed in *Danaus plexipuss* (Lewis et al. 2021), *Kallima inachus* (Yang et al. 2020)
372 and *Dryas iulia* (Lewis et al. 2021), where no similarity was found between the W and the Z
373 chromosome and are more in line with results in the Asian corn borer, *Ostrinia furnacalis,* a
374 moth belonging to the Pyraloidea, which suggests a common autosomal origin of the Z and
375 the W chromosome (Dai et al. 2022). Contrastingly, the lack of similarity between the W and
376 the Z in five of the seven species studied could suggest a B chromosome origin or could be
377 the result of the rapid evolution of the W, which could be interfering with the inference of
378 homology (Yoshido et al. 2013; Vítková et al. 2007). Taken together, the homology tests
379 between the W and remaining 31 chromosomes in *H. misippus* and other Lepidoptera provide
380 mixed evidence for the origin of the W from the same homologous pair of autosomes as the Z
381 or from a B chromosome. Alternatively, our results showing that there is a degree of synteny
382 between the *H. misippus* W and the Z could point to a fusion of the Z chromosome with an
383 autosome and subsequent formation of the W (Figure 1C). This hypothesis would fit the
384 mounting evidence supporting a Z0 system in non-ditrysian Lepidoptera and early diverging
385 ditrysian lineages (Fraïsse et al. 2017; Hejníčková et al. 2019; Lukhtanov 2000; Sahara et al.
386 2012).

387 The second question we wanted to clarify was whether the W has a single origin in the
388 Lepidoptera or evolved independently multiple times. Previously, the lack of similarity
389 among W chromosomes of Lepidoptera species has been suggested to be the result of
390 multiple independent recruitments from B chromosomes (Lewis et al. 2021). Similarly,
391 phylogenies of W genes in the Asian corn borer could suggest multiple independent origins
392 of the W chromosome from the same autosomal pair as the Z (Dai et al. 2022). Regardless of
393 the specific event that led to the formation of the W chromosome (i.e., from a B chromosome
394 or from an autosome), if the W chromosomes of ditrysian lineages share a common origin,
395 we would expect them to have a higher level of similarity among them than expected by
396 chance and thus higher than the similarity to the autosomes or Z chromosome. Using a
397 synteny analysis of the *H. misippus* W chromosome and the genomes of 10 other ditrysian
398 species, we found that there is a high level of similarity between the W chromosomes. These
399 results suggest that W chromosomes of ditrysian lineages may share a single common origin,
400 which contrasts with evidence of rapid turnover of W chromosomes in Lepidoptera (Dai et al.
401 2022; Lewis et al. 2021). However, it is possible that some species have retained the ancestral
402 W, while some other have recruited other B chromosomes as suggested. Furthermore, our
403 synteny results show clear homology among Z chromosomes of the studied ditrysian species,
404 which is consistent with evidence suggesting deep conservation of the Z chromosome across
405 the Lepidoptera (Fraïsse et al. 2017).

406 Overall, we present evidence of a single common origin of W chromosomes in the
407 Lepidoptera. Furthermore, our results hint at a possible origin from the same autosomal pair
408 as the Z or perhaps to a Z-autosome fusion event that culminated with the formation of the
409 W. Crucially, the rapid evolution of the W chromosome could cause the lack of similarity
410 observed between Z and W chromosomes in some species and hinders the elucidation of its
411 origin. Further studies including more species across the Lepidoptera are necessary to clarify

412 whether the W chromosome evolved from an autosome or from a B chromosome and to shed
413 light on the possible shared origin of W chromosomes, particularly including non-ditrysian
414 lineages.

415 Materials and methods

416 Hypolimnas *misippus* butterfly rearing and cross preparation

417 A trio-binning approach was used for sequencing of the *Hypolimnas misippus* assembly,
418 which consists of sequencing the parents and offspring from one family. First, butterflies
419 were obtained from Stratford-upon-Avon Butterfly Farm, UK, and reared in greenhouses in
420 Madingley, Cambridge, UK. Larvae were fed *Portulaca oleracea* and *Portulaca quadrifida*.
421 Adult butterflies were kept in a large cage (1.5 meters x 1.5 meters x 2 meters) and observed
422 until mating occurred when mating pairs were transferred to separate smaller cages during
423 copulation. One mated pair was used for trio-based genome sequencing, and their offspring
424 were reared until pupation and then flash frozen as pupae in liquid nitrogen.

425 *H. misippus* Trio binning genome assembly

426 One *H. misippus* family was used for sequencing and trio-binning genome assembly. DNA
427 was extracted using MagAttract HMW DNA Kit (QIAGEN). The two parents were
428 sequenced using Illumina short-read sequencing, which resulted in 368.67 and 346.56 million
429 read pairs and a total yield of 55.67 and 52.33 Gbp of data from the mother and father
430 respectively. One of the offspring was sequenced using PacBio long-read sequencing (total
431 yield of 13.46 Gb of data and N50 of 13,493). Trio binning enables the independent assembly
432 of the two parental haplotypes. First, yak-r55 (Li 2022) was used to create a kmer database
433 from the parental Illumina data. Then, hifiasm-0.7-r256 (Cheng et al. 2020) was run in trio

434 mode to assemble the HiFi PacBio long-read data, followed by the purging of haplotypes and
435 overlaps using purgedups v1.2.3 (Guan 2022). Haplotype 1 corresponds to the paternal
436 haplotype, while haplotype 2 corresponds to the maternal haplotype. All subsequent analysis
437 were performed using haplotype 2.

438 **Curation of *HypMis_v2* with HiC**

439 To place the assembled scaffolds into chromosomes, Hi-C sequencing was used, which is a
440 chromosome conformation capture technology that provides information about the 3-D
441 interactions between genomic loci. An offspring from an unrelated mating was flash frozen in
442 liquid nitrogen and used for Hi-C sequencing and analysis. 386.99 million read pairs were
443 produced, which represented 58,45 Gb of Hi-C data. Fastq files were converted to cram (two
444 files, one for each read pair end), mapped to haplotype 2 (the maternal haplotype). The *H.*
445 *misippus* assembly was produced after the purgedups stage and processed with Juicer v1.6
446 (Durand, Shamim, et al. 2016). Juicer transforms raw Hi-C data into a list of contacts, which
447 defines pairs of genomic positions that were in close physical contact during the experiment.
448 Then, the main reference assembly was curated using the 3D-DNA pipeline, which corrects
449 misassembles, anchors, orders and orients fragments of DNA based on the Hi-C data
450 (Dudchenko et al. 2017; Durand, Robinson, et al. 2016). 3D-DNA generated assembly
451 heatmaps as part of its workflow, which indicate the frequency of contact between pairs of
452 genomic locations. Obvious errors in the genome assembly such as large genomic inversions
453 were manually edited by examining the Hi-C heatmaps using the Juicebox tool (Durand,
454 Robinson, et al. 2016; Dudchenko et al. 2018). Finally, the edited assembly was exported as
455 an *.assembly* file and converted to a final fasta assembly file using the 'run-asm-pipeline-
456 post-review.sh' script setting –editor-repeat-coverage to 6.

457 ***HypBol_v1* genome curation with Ragout**

458 The *H. bolina* reference assembly (*HypBol_v1*) was produced using a combination of
459 Nanopore long-read sequencing and a linkage map. Pupae originating from the Philippines
460 were ordered from Stratford-upon-Avon butterfly farm. One female pupa was used for
461 sequencing. High molecular weight DNA was extracted following a phenol chloroform and
462 glass capillary hook protocol following(Quick 2018). A library was prepared for Nanopore
463 sequencing using the Nanopore Ligation Sequencing Kit (SQK-LSK 109) following a
464 LSK109 bead free library preparation protocol (Tyson 2020). Post library preparation, a total
465 of 1.78μg DNA remained and was sequenced across 2 R9 chemistry Nanopore MinION flow
466 cells, resulting in 7.35 million reads totalling 14.16 Gb. Adapters were removed using
467 Porechop v0.2.4 (Wick 2022) and reads assembled using redbean (Ruan & Li 2020).
468 Assembled contigs were split into bins using MaxBin2 (Wu et al. 2016) and bacterial contigs
469 and reads were removed from the data using blobtools2 (Challis et al. 2020). This produced
470 an assembly with 13,492 contigs and an N50 of 1.4 kbp.

471 Leveraging the high synteny expected from the two *Hypolimnas* assemblies, Ragout v2.3
472 (Kolmogorov et al. 2014, 2018) was used to improve the initial *HypBol_v1* assembly using
473 *HypMis_v2* as a reference. First, the two genomes were soft-masked using RepeatMasker
474 (Smit et al. 2015), creating the repeat library based on the genome being masked. Then,
475 Cactus (Paten et al. 2011) was used to align both genomes using Python 3.8. The resulting
476 HAL alignment was converted to MAF format using the hal2maf utility from the HAL
477 program (Hickey et al. 2013). Finally, Ragout was run using the MAF alignment between
478 *HypBol_v1* and *HypMis_v2* as input.

479 **Rearing of *H. bolina* individuals**

480 A linkage map was used to improve the *HypBol_v1* assembly and place the assembled
481 scaffolds into chromosomes. Two families (178303XX and 182703XX) were reared and
482 sequenced. First, female *H. bolina* purchased from Stratford upon Avon Butterfly Farm
483 (originating from the Philippines) were mated to wild-caught males from Mo'orea (French
484 Polynesia) at the University of California Berkeley Gump Station research facility. Female
485 Philippines-Mo'orea F1 hybrid offspring were then mated to pure Mo'orea F1 males. The F2
486 offspring of one of these crosses is family 178303XX. At the same time, male Philippines-
487 Moorea F1 hybrids were mated to pure Philippine F1 females. The F2 offspring of one of
488 these crosses is family 182703XX. Butterflies were kept in a large outdoor cage for mating
489 under observation. Any copulating pairs were separated into a small cage. The mated female
490 was then placed in an oviposition cup containing a small plant e.g. *Asystasia sp*. and allowed
491 to lay eggs. Hatched eggs were moved to a rearing box with suitable food plant e.g. Ipomoea
492 sp. and caterpillars reared until pupation. Pupae were moved to individual cups for
493 emergence, adults left to dry for one day and then used for further matings or stored in -80˚C
494 freezer.

495 **Illumina library preparation of *H. bolina* family samples**

496 The offspring of the *H. bolina* families (F2) were processed to extract the DNA and prepare
497 the libraries for Illumina sequencing. DNA extractions were carried out using a custom
498 protocol using PureLink buffers and homemade magnetic beads. Briefly, a small piece of
499 thorax tissue (1/10) is placed in an 8-tube PCR strip. Then, 45 μL of PureLink Digestion
500 buffer and 10 μL of Proteinase K (20mg/mL) are added, and the mix is incubated at 58˚C
501 with shaking (500 rpm) for 2-3 hours. Afterwards, 2 μL of RNAseA are added (DNAse free,
502 10mg/mL) and incubated for 10min at room temperature. Then, 45 μL of PureLink Lysis

503 buffer is added to the mix and incubated at 58°C for 30 minutes with shaking (500 rpm).
504 Following that, a homemade magnetic bead mix is used to extract the DNA from the lysate.
505 First, 37.5 μL of magnetic beads are added together with 75 μL of lysate to a 96-well plate.
506 After mixing, the samples are incubated for 15 minutes at room temperature, then the plate is
507 placed on a magnetic stand for 10 minutes, the supernatant removed, and the beads cleaned
508 with 80% ethanol. After drying out, 50 μL of 10mM Tris (pH=8) are added to elute the
509 sample and incubated at 45°C for 15 minutes without resuspending. Then, the beads are
510 resuspended and incubated for 20 minutes at room temperature. Finally, the plate is placed on
511 the magnetic stand and, after 10 minutes, the supernatant (the DNA) is transferred to a fresh
512 tube.

513 The F2 were sequenced using a Nextera-based library preparation at intermediate coverage
514 (~11X). A secondary purification using magnetic SpeedBeads™ (Sigma) was performed
515 prior to Nextera-based library preparation. Libraries were prepared following a method based
516 on Nextera DNA Library Prep (Illumina, Inc.) with purified Tn5 transposase (Picelli et al.
517 2014). PCR extension with an i7-index primer (N701–N783) and the N501 and N502 i5-
518 index primers was performed to barcode the samples. Library purification and size selection
519 was done using the same homemade beads as above. Pooled libraries were sequenced by
520 Novogene Cambridge, UK. Libraries of the parental samples were prepared and sequenced to
521 ~20X coverage by Novogene Cambridge, UK.

522 ***H. bolina* linkage map construction and anchoring of the genome**

523 A linkage map was produced with Lep-Map3 (Rastas 2017) and then used to improve the *H.*
524 *bolina* assembly and place the scaffolds onto chromosomes. First, sequences were mapped to
525 the reference genome using bwa-mem (Li 2013). PCR duplicates were marked using the
526 MarkDuplicates from Picard tools. Sorted BAMs were then created using SAMtools (Li et al.

527 2009) and genotype likelihoods computed. The pedigree of individuals was checked and
528 corrected using IBD (identity-by-descent) with a random subset of 10% of the markers
529 (1,270,024 SNPs) following the IBD pipeline from Lep-Map3. These markers were also used
530 to construct the linkage map. Scaffolds were anchored into chromosomes based on the
531 linkage map using LepAnchor (Rastas 2020).

532 ***HypBol_v1* polishing with Pilon**

533 After anchoring with the linkage map, three iterations of Pilon v1.24 (Walker et al. 2014) in
534 diploid mode were run to correct the draft assembly by correcting bases, filling gaps and
535 fixing mis-assemblies. Samples used for the Pilon correction were CAM035727,
536 CAM035728, CAM035186, CAM035187, CAM035188, CAM035189.

537 **Repeat annotation**

538 Once the two final assemblies had been produced, they were each assessed for repeat content
539 using a custom repeat library. First, a repeat database was built and the repeats of the two
540 finished assemblies modelled using RepeatModeler v. 2.0.2a. Each custom library was then
541 combined with the Lepidoptera library extracted from Dfam (Storer et al. 2021). This merged
542 library was used to soft mask the genome using RepeatMasker v 4.1.0 with the cut-off score
543 set to 250 and skipping the bacterial insertion element check. The resulting soft masked
544 assemblies were used for gene annotation.

545 **RNA-seq sample preparation**

546 To assist with genome annotation, RNA-seq data was obtained from 4 *H. misippus*
547 individuals (2 adults and 2 pupae) and 17 *H. bolina* individuals (6 adults and 11 pupae).
548 Butterflies were purchased from Stratford-upon-Avon Butterfly Farm and kept at room

549 temperature until dissection. 4 tissues were dissected out from *H. misippus* pupae and placed
550 in RNA-later (Sigma): wing discs, thorax, head, and abdomen; while *H. bolina* pupae 3
551 tissues were dissected: wing discs, thorax-head, and abdomen. Only abdomen, head and
552 thorax samples were dissected from adults of either species. Two pooling strategies were
553 followed: 1) 2 *H. bolina* pupae were pooled by individual, pooling head, thorax, abdomen,
554 and wing discs together and sequenced at high coverage (50M reads). And 2) 4 *H. misippus*
555 and 15 *H. bolina* adults and pupae were dissected into tissues and pooled by species and
556 tissue. Each pooled sample, 4 for *H. misippus* and 3 for *H. bolina,* was sequenced to 20M
557 reads. RNA was extracted using a modified Trizol protocol using the same protocol as in
558 (Brien et al. 2022).

559 **RNA-seq mapping and gene annotation**

560 First, low quality ends and adaptors from the RNA-seq data were trimmed using TrimGalore!
561 v 0.5.0 (Krueger 2015). Then, the reads were mapped to the soft masked genomes using
562 STAR v 2.5.0a (Dobin et al. 2013). Two rounds of mapping (2pass) were performed,
563 including all the splice junction files in the second round. Then the resulting mapped reads
564 and the soft masked genome assembly were used to generate a gene annotation using
565 BRAKER v 2.1.5 (Brůna et al. 2021), running it a second time to add UTR annotations with
566 options addUTR=on and skipAllTraining (Keilwagen et al. 2019). Completeness of the
567 annotation was assessed using BUSCO (Seppey et al. 2019; Simão et al. 2015) using the
568 Insecta_odb10 (n=1367) set of genes.

569 **Homology with Merian units**

570 A growing practice in Lepidoptera genetics is to name chromosomes of reference assemblies
571 by their homology to *Melitaea cinxia*, which is the first assembled lepidopteran genome that

572 retains the ancestral 31 chromosome karyotype (Ahola et al. 2014). The numbered
573 chromosomes are known as Merian units in honour of the scientist Maria Sibylla Merian
574 (Wulf 2016) and allow for the tracking of conserved synteny blocks across the phylogeny of
575 the Lepidoptera. *Melitaea cinxia* belongs to the *Nymphalinae* which is the same subfamily as
576 *Hypolimnas*, and the two taxa diverged about 30 MYa (Espeland et al. 2018). To assign
577 homology to the Merian units, lepidopteran_odb10 BUSCO assignment of the *Melitaea*
578 *cinxia* genome was compared to the lepidopteran_odb10 BUSCO assignment of *HypBol_v2*
579 and *HypBol_v1*. Merian units in *HypBol_v1* and *H. missipus* were assigned by choosing the
580 chromosome sharing the highest BUSCO hits with a *M. cinxia* chromosome.

581 **Synteny analyses**

582 Synteny of the two assemblies was assessed using two methods. First, the two final genome
583 assemblies were aligned using D-GENIES (Cabanettes & Klopp 2018), which produced a paf
584 file that was used to detect candidate inversions between the two genome assemblies
585 (Supplementary Figure 1). Second, synteny between the two assemblies was evaluated using
586 Satsuma2 Synteny (Grabherr et al. 2010) and a circos plot generated using the circlize
587 package v 0.4.14 (Gu et al. 2014) in R v 4.1.2 (Figure 2).

588 To examine the origin and evolution of the W chromosome in Lepidoptera, first, synteny-
589 based homology between the *H. misippus* W chromosome, the Z chromosome and the
590 autosomes was evaluated using Satsuma2. Satsuma2 is an aligner of whole genome
591 assemblies intended to find homology based on sequence similarity. Satsuma2 first maps all
592 genomic windows of the query genome to the target genome with a percentage of identity
593 higher than 45%, and then filters those hits based on large scale synteny, that is, it keeps only
594 matches that are concordant with each other. Thus, Satsuma2 is not only aligning genomic
595 windows, but also evaluating synteny between those blocks. Finally, Satsuma2 focusses on

596 regions with a high number of hits, to exhaustively evaluate the region around those hits, a
597 strategy analogous to the battleship game. To evaluate synteny between the W and the
598 remaining 31 chromosomes in *H. misippus,* the W chromosome was used as a query and the
599 31 assembled chromosomes as targets for Satstuma2. Resulting mapped regions were then
600 filtered to keep only non-overlapping regions using the package GenomicRanges in R v4.1.2
601 (Lawrence et al. 2013). Satsuma2 does not require input genome assemblies to be masked.
602 Because of its algorithm and filtering steps, repeat regions mapping to multiple places in the
603 genome have decreased score and may be filtered out. However, to evaluate only the synteny
604 of non-repeat regions, the RepeatMasker output, which details the coordinates of repeats in
605 the genome, was used to filter out repeat regions of the W chromosome. Finally, the effect of
606 sequence identity was evaluated by performing the analysis with no identity filter and
607 applying a threshold of 70% similarity.

608 The synteny between the W and the rest of the genome was also evaluated in 6 other
609 Lepidoptera species: *Boloria selene* (GCA_905231865.2), *Crocallis elinguaria*
610 (GCA_907269065.1), *Hemaris fuciformis* (GCA_907164795.1), *Papilio machaon*
611 (GCA_912999745.1), *Watsonalla binaria* (GCA_929442735.1), *Zygaena filipendulae*
612 (GCA_907165275.2) and *Dryas iulia* (GCA_019049465.1). In all cases, the W chromosome
613 was used as the query, results were converted to non-overlapping regions and percentage of
614 the W covered by matches calculated. No filter based on repeats was applied, as no repeat
615 library was available for these species.

616 After that, Satsuma2 was used to compare the synteny of the *H. misippus* W (used as query)
617 and Z chromosomes to 10 Lepidoptera species, the 6 from the previous analysis and also
618 *Hemaris fuciformis* (GCA_907164795.1), *Marasmarcha lunaedactyla* (GCA_923062675.1),
619 *Vanessa cardui* (GCF_905220365.1), and *Mythimna farrago* (GCA_910589285.2).

620 Chromosomes of all the species analysed were re-named by their homology to *M. cinxia* as
621 above. Again, results were converted to non-overlapping regions, repeat regions of the W
622 chromosome filtered and percentage of the W chromosome covered by matches calculated.
623 Using only the *H. misippus* W chromosome as query ensures that secondary and tertiary
624 matches are also reported. A more specific analysis was produced by using the *H. misippus*
625 assembly as query to search for synteny in the 10 Lepidoptera target species. Using whole
626 assemblies as input limits the results to only primary matches, pairing all homologous
627 autosomes.

628 **BLAST of *H. misippus* W chromosome genes**

629 To further explore the degree of homology between the *H. misippus* W chromosome and the
630 autosomes and Z chromosome, protein sequence homology was assessed using BLASTp
631 v2.4.0+ (BLAST 2013; Kent 2002). First a protein BLAST database was built using the
632 protein sequences contained in the *H. misippus* autosomes and the Z chromosome in the
633 HypMisi_v2 annotation. Then, protein sequences found in the W chromosome were extracted
634 and blasted to the database setting the minimum e-value to 1e-10 and the maximum number
635 of target sequences to be reported to 5,000. The best match for each protein sequence was
636 selected based on e-value score, and if two matched had the same value percentage of identity
637 was used. Finally, genome coordinates were extracted from the HypMisi_v2 annotation.

638 **Acknowledgements**

639 This work was supported by the Natural and Environmental Research Council [grant number
640 NE/L002507/1 to A.O]; the Cambridge Trust [European Research Scholarship to A.O] and
641 St. John's College [Benefactors' Scholarship to A.O].

642 **References**

643 Bachtrog D et al. 2011. Are all sex chromosomes created equal? Trends in Genetics. 27:350–
644 357. doi: 10.1016/j.tig.2011.05.005.

645 Bachtrog D et al. 2014. Sex Determination: Why So Many Ways of Doing It? PLOS Biology.
646 12:e1001899. doi: 10.1371/journal.pbio.1001899.

647 Beukeboom LW, Perrin N. 2014. *The Evolution of Sex Determination*. Oxford University Press
648 doi: 10.1093/acprof:oso/9780199657148.001.0001.

649 BLAST. 2013. *BLAST Basic Local Alignment Search Tool*. doi: 10.1006/jmbi.1990.9999.

650 Brien MN et al. 2022. Colour polymorphism associated with a gene duplication in male
651 wood tiger moths. 2022.04.29.490025. doi: 10.1101/2022.04.29.490025.

652 Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic
653 eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein
654 database. NAR Genomics and Bioinformatics. 3:lqaa108. doi: 10.1093/nargab/lqaa108.

655 Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an interactive, efficient
656 and simple way. PeerJ. 6:e4958. doi: 10.7717/peerj.4958.

657 Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. 2020. BlobToolKit – Interactive Quality
658 Assessment of Genome Assemblies. G3 Genes|Genomes|Genetics. 10:1361–1374. doi:
659 10.1534/g3.119.400908.

660 Charlat S et al. 2009. The joint evolutionary histories of Wolbachia and mitochondria in
661 Hypolimnas bolina. BMC Evolutionary Biology. 9:64. doi: 10.1186/1471-2148-9-64.

662 Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2020. Haplotype-resolved de novo assembly
663 with phased assembly graphs. arXiv:2008.01237 [q-bio]. http://arxiv.org/abs/2008.01237
664 (Accessed August 27, 2020).

665 Dai W, Mank JE, Ban L. 2022. Repeated origin of the W chromosome from the Z
666 chromosome in Lepidoptera. 2022.10.21.512844. doi: 10.1101/2022.10.21.512844.

667 Dalíková M et al. 2017. New Insights into the Evolution of the W Chromosome in
668 Lepidoptera. Journal of Heredity. 108:709–719. doi: 10.1093/jhered/esx063.

669 Dobin A et al. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29:15–21. doi:
670 10.1093/bioinformatics/bts635.

671 Dudchenko O et al. 2017. De novo assembly of the Aedes aegypti genome using Hi-C yields
672 chromosome-length scaffolds. Science. 356:92–95. doi: 10.1126/science.aal3327.

673 Dudchenko O et al. 2018. The Juicebox Assembly Tools module facilitates de novo assembly
674 of mammalian genomes with chromosome-length scaffolds for under $1000. bioRxiv.
675 254797. doi: 10.1101/254797.

676 Durand NC, Robinson JT, et al. 2016. Juicebox Provides a Visualization System for Hi-C
677 Contact Maps with Unlimited Zoom. Cell Systems. 3:99–101. doi:
678 10.1016/j.cels.2015.07.012.

679 Durand NC, Shamim MS, et al. 2016. Juicer Provides a One-Click System for Analyzing Loop-
680 Resolution Hi-C Experiments. cels. 3:95–98. doi: 10.1016/j.cels.2016.07.002.

681 Dyson EA, Kamath MK, Hurst GDD. 2002. Wolbachia infection associated with all-female
682 broods in Hypolimnas bolina (Lepidoptera: Nymphalidae): evidence for horizontal
683 transmission of a butterfly male killer. Heredity. 88:166–171. doi: 10.1038/sj.hdy.6800021.

684 Espeland M et al. 2018. A Comprehensive and Dated Phylogenomic Analysis of Butterflies.
685 Current Biology. 28:770-778.e5. doi: 10.1016/j.cub.2018.01.061.

686 Fraïsse C, Picard MAL, Vicoso B. 2017. The deep conservation of the Lepidoptera Z
687 chromosome suggests a non-canonical origin of the W. Nat Commun. 8:1486. doi:
688 10.1038/s41467-017-01663-5.

689 Grabherr MG et al. 2010. Genome-wide synteny through highly sensitive sequence
690 alignment: Satsuma. Bioinformatics. 26:1145–1151. doi: 10.1093/bioinformatics/btq102.

691 Gu Z, Gu L, Eils R, Schlesner M, Brors B. 2014. circlize implements and enhances circular
692 visualization in R. Bioinformatics. 30:2811–2812. doi: 10.1093/bioinformatics/btu393.

693 Guan D. 2022. Purge_Dups. https://github.com/dfguan/purge_dups (Accessed September
694 29, 2022).

695 Hejníčková M et al. 2019. Absence of W Chromosome in Psychidae Moths and Implications
696 for the Theory of Sex Chromosome Evolution in Lepidoptera. Genes. 10:1016. doi:
697 10.3390/genes10121016.

698 Hickey G, Paten B, Earl D, Zerbino D, Haussler D. 2013. HAL: a hierarchical format for storing
699 and analyzing multiple genome alignments. Bioinformatics. 29:1341–1342. doi:
700 10.1093/bioinformatics/btt128.

701 Hornett EA et al. 2006. Evolution of male-killer suppression in a natural population. PLoS
702 Biology. 4:1643–1648. doi: 10.1371/journal.pbio.0040283.

703 Kawahara AY et al. 2019. Phylogenomics reveals the evolutionary timing and pattern of
704 butterflies and moths. Proc Natl Acad Sci U S A. 116:22657–22663. doi:
705 10.1073/pnas.1907847116.

706 Keilwagen J, Hartung F, Grau J. 2019. GeMoMa: Homology-Based Gene Prediction Utilizing
707 Intron Position Conservation and RNA-seq Data. In: Gene Prediction: Methods and
708 Protocols. Kollmar, M, editor. Methods in Molecular Biology Springer: New York, NY pp.
709 161–177. doi: 10.1007/978-1-4939-9173-0_9.

710 Kent WJ. 2002. BLAT - The BLAST-like alignment tool. Genome Research. doi:
711 10.1101/gr.229202. Article published online before March 2002.

712 Kolmogorov M et al. 2018. Chromosome assembly of large and complex genomes using
713 multiple references. Genome Res. doi: 10.1101/gr.236273.118.

714 Kolmogorov M, Raney B, Paten B, Pham S. 2014. Ragout—a reference-assisted assembly
715 tool for bacterial genomes. Bioinformatics. 30:i302–i309. doi:
716 10.1093/bioinformatics/btu280.

717 Krueger F. 2015. Trim galore. A wrapper tool around Cutadapt and FastQC to consistently
718 apply quality and adapter trimming to FastQ files.

719 Lawrence M et al. 2013. Software for Computing and Annotating Genomic Ranges. PLOS
720 Computational Biology. 9:e1003118. doi: 10.1371/journal.pcbi.1003118.

721 Lewis JJ et al. 2021. The Dryas iulia Genome Supports Multiple Gains of a W Chromosome
722 from a B Chromosome in Butterflies. Genome Biology and Evolution. 13. doi:
723 10.1093/gbe/evab128.

724 Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
725 arXiv:1303.3997 [q-bio]. http://arxiv.org/abs/1303.3997 (Accessed August 27, 2020).

726 Li H. 2022. lh3/yak. https://github.com/lh3/yak (Accessed September 25, 2022).

727 Li H et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics.
728 25:2078–2079. doi: 10.1093/bioinformatics/btp352.

729 Lukhtanov VA. 2000. Sex chromatin and sex chromosome systems in nonditrysian
730 Lepidoptera (Insecta). Journal of Zoological Systematics and Evolutionary Research. 38:73–
731 79. doi: 10.1046/j.1439-0469.2000.382130.x.

732 Marsh NA, Clarke CA, Rothschild M, Kellett DN. 1977. Hypolimnas bolina (L.), a mimic of
733 danaid butterflies, and its model Euploea core (Cram.) store cardioactive substances.
734 Nature. 268:726–728. doi: 10.1038/268726a0.

735 Paten B et al. 2011. Cactus: Algorithms for genome multiple sequence alignment. Genome
736 Res. 21:1512–1528. doi: 10.1101/gr.123356.111.

737 Picelli S et al. 2014. Tn5 transposase and tagmentation procedures for massively scaled
738 sequencing projects. Genome Res. 24:2033–2040. doi: 10.1101/gr.177881.114.

739 Quick J. 2018. Ultra-long read sequencing protocol for RAD004.
740 https://www.protocols.io/view/ultra-long-read-sequencing-protocol-for-rad004-mrxc57n
741 (Accessed December 13, 2022).

742 Rastas P. 2020. Lep-Anchor: automated construction of linkage map anchored haploid
743 genomes. Bioinformatics. 36:2359–2364. doi: 10.1093/bioinformatics/btz978.

744 Rastas P. 2017. Lep-MAP3: robust linkage mapping even for low-coverage whole genome
745 sequencing data. Bioinformatics. 33:3726–3732. doi: 10.1093/bioinformatics/btx494.

Ruan J, Li H. 2020. Fast and accurate long-read assembly with wtdbg2. Nat Methods. 17:155–158. doi: 10.1038/s41592-019-0669-3.

Sahara K, Yoshido A, Traut W. 2012. Sex chromosome evolution in moths and butterflies. Chromosome Res. 20:83–94. doi: 10.1007/s10577-011-9262-z.

Sahoo RK et al. 2018. Evolution of Hypolimnas butterflies (Nymphalidae): Out-of-Africa origin and Wolbachia-mediated introgression. Molecular Phylogenetics and Evolution. 123:50–58. doi: 10.1016/j.ympev.2018.02.001.

Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: Assessing Genome Assembly and Annotation Completeness. In: Gene Prediction: Methods and Protocols. Kollmar, M, editor. Methods in Molecular Biology Springer: New York, NY pp. 227–245. doi: 10.1007/978-1-4939-9173-0_14.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 31:3210–3212. doi: 10.1093/bioinformatics/btv351.

Singh KS et al. 2022. Genome assembly of Danaus chrysippus and comparison with the Monarch Danaus plexippus. G3 Genes|Genomes|Genetics. 12:jkab449. doi: 10.1093/g3journal/jkab449.

Smit A, Hubley R, Green P. 2015. RepeatMasker Open-4.0. 2013–2015.

Smith DAS. 1976. Phenotypic diversity, mimicry and natural selection in the African butterfly Hypolimnas misippus L. (Lepidoptera: Nymphalidae). Biological Journal of the Linnean Society. 8:183–204. doi: 10.1111/j.1095-8312.1976.tb00245.x.

Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. 2021. The Dfam community resource of transposable element families, sequence models, and genome annotations. Mobile DNA. 12:2. doi: 10.1186/s13100-020-00230-y.

Turner JRG, Sheppard PM. 1975. Absence of crossing-over in female butterflies (Heliconius). Heredity. 34:265–269. doi: 10.1038/hdy.1975.29.

Tyson J. 2020. Bead-free long fragment LSK109 library preparation. https://www.protocols.io/view/bead-free-long-fragment-lsk109-library-preparation-7eshjee (Accessed December 13, 2022).

Vane-Wright RI, Ackery PR, Smiles RL. 1977. The polymorphism, mimicry, and host plant relationships of Hypolimnas butterflies. Biological Journal of the Linnean Society. 9:285–297. doi: 10.1111/j.1095-8312.1977.tb00271.x.

Vítková M, Fuková I, Kubíčková S, Marec F. 2007. Molecular divergence of the W chromosomes in pyralid moths (Lepidoptera). Chromosome Res. 15:917–930. doi: 10.1007/s10577-007-1173-7.

Walker BJ et al. 2014. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLOS ONE. 9:e112963. doi: 10.1371/journal.pone.0112963.

Wick R. 2022. Porechop. https://github.com/rrwick/Porechop (Accessed September 28, 2022).

Wright AE, Dean R, Zimmer F, Mank JE. 2016. How to make a sex chromosome. Nat Commun. 7:12087. doi: 10.1038/ncomms12087.

Wu Y-W, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics. 32:605–607. doi: 10.1093/bioinformatics/btv638.

Wulf A. 2016. The Woman Who Made Science Beautiful. The Atlantic. https://www.theatlantic.com/science/archive/2016/01/the-woman-who-made-science-beautiful/424620/ (Accessed July 7, 2022).

Yang J et al. 2020. Chromosome-level reference genome assembly and gene editing of the dead-leaf butterfly Kallima inachus. Molecular Ecology Resources. 20:1080–1092. doi: 10.1111/1755-0998.13185.

Yen EC et al. 2020. A haplotype-resolved, de novo genome assembly for the wood tiger moth (Arctia plantaginis) through trio binning. Gigascience. 9. doi: 10.1093/gigascience/giaa088.

Yoshida K et al. 2011. B Chromosomes Have a Functional Effect on Female Sex Determination in Lake Victoria Cichlid Fishes. PLOS Genetics. 7:e1002203. doi: 10.1371/journal.pgen.1002203.

Yoshido A, Šíchová J, Kubíčková S, Marec F, Sahara K. 2013. Rapid turnover of the W chromosome in geographical populations of wild silkmoths, Samia cynthia ssp. Chromosome Res. 21:149–164. doi: 10.1007/s10577-013-9344-1.