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[4,5,6], Matthew R. Gemmell[4], Louise A. Reynolds[4], Ian A. Warren[1], Ian J. Gordon[7], Gregory D. D. Hurst[4], Richard Durbin[3], Simon H. Martin[8], 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. Institute of Infection, Veterinary and Ecological Science, University of Liverpool, Liverpool, United Kingdom
5. Vector Biology, Liverpool School of Tropical Medicine, Liverpool, United Kingdom
6. Department of Biology, University of Oxford, Oxford, UK
7. Centre of Excellence in Biodiversity, University of Rwanda, Huye, Rwanda
8. 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 missipus* and *Hypolimnas bolina)* 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

Sex chromosome evolution, genome assembly, B chromosome, 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).

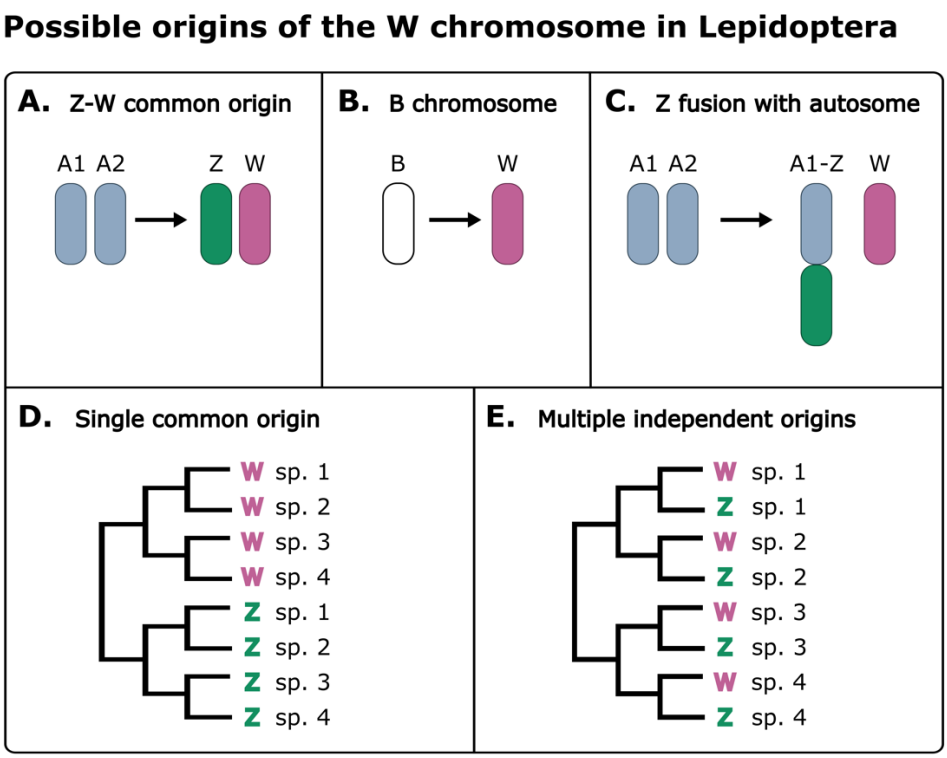


Fig. 1. Possible origins of the W chromosome in Lepidoptera. **A.** One hypothesis on the origin of W chromosomes is that Z and W evolved from the same autosomal pair. **B.** Another hypothesis is that the W chromosome evolved from a B chromosome, which are non-essential chromosomes that vary in number 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).

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

Results

Genome assemblies and synteny between HypMis_v2 and HypBol_v1

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

146 HypMis_v2 was assembled into 32 chromosome-level scaffolds (>6 MB; 30 autosomes and
147 the Z and W chromosomes) and 186 unplaced scaffolds smaller than 1MB (Figure 2B) and
148 has an N50 of 14.6 MB; while HypBol_v1 was assembled into 59 scaffolds all placed onto 31
149 chromosomes, with an N50 of 15.2 MB. Thus, both species have retained the ancestral
150 karyotype of 31 chromosomes. This karyotype is present in other Nymphalinae species such
151 as the painted lady *Vanessa cardui*, which has a comparable genome size (424.8 Mb) to the
152 two *Hypolimnas* species. In general, all chromosomes were slightly larger in HypBol_v1
153 compared to HypMis_v2 with only 4 exceptions (chromosomes 13, 14, 28 and 31; Figure
154 2C). When aligning the two *Hypolimnas* assemblies, there were no fusions or fissions among
155 chromosomes, but multiple large rearrangements were observed (Figure 2A; Supplementary
156 Figure 1). The only caveat to these conclusions is that HypMis_v2 was used to scaffold
157 HypBol_v1 before using the linkage map, which could theoretically have caused us to miss
158 differences in the *H. bolina* chromosomal structure. However, the linkage map had 31 linkage
159 groups, providing an independent line of evidence that the two species have the same
160 karyotype. Moreover, there were no conflicts between the linkage map and the assembly
161 curated with ragout, and therefore no evidence of inter-chromosomal rearrangements (e.g.
162 translocations) between the species. Finally, 12 large inversions on multiple chromosomes
163 are apparent in the alignment of the two assemblies (Supplementary Figure 1).

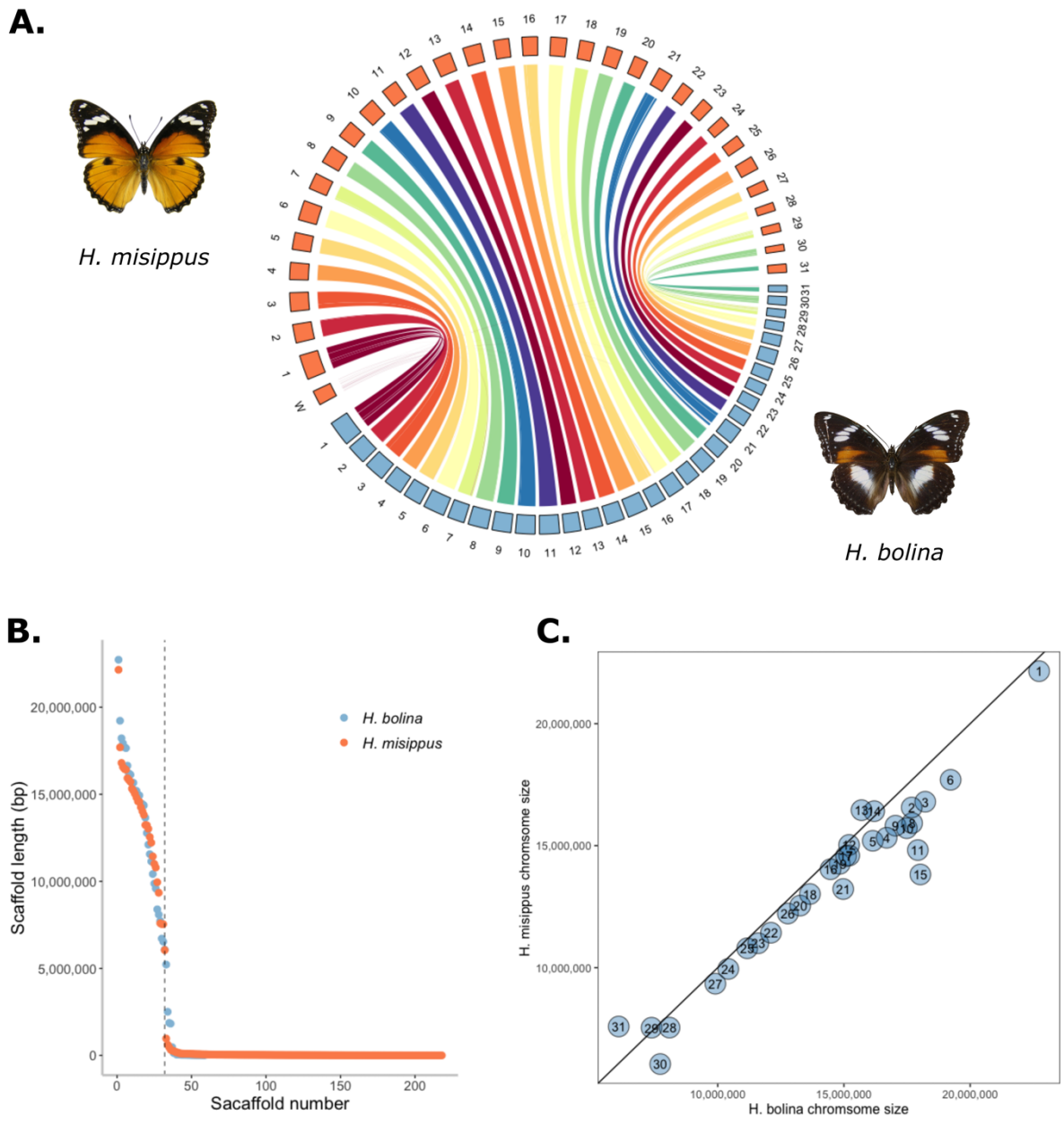


Fig. 2. Chromosome level assemblies for *Hypolimnas misippus* and *H. bolina*. **A.** Chromosomal synteny is conserved in both species. **B.** Chromosome level scaffolds have been assembled for both species. **C.** In general, *H. bolina* chromosomes are larger. Chromosome 1 refers to the Z chromosome. *H. bolina* image modified from (MCZ Harvard University 2023) under CC BY-NC-SA 3.0.

170 **Transposable elements and repeat content**

171 In total, the two *Hypolimnas* genomes contain similar levels of repeats, 43.86% for
172 HypBol_v1 and 42.34% for HypMis_v2, which are slightly higher than in the painted lady
173 *Vanessa cardui* (37.27%) of comparable genome size (424.8 Mb). The composition of
174 repeats is similar in the two *Hypolimnas* genomes, while they differ substantially from the
175 painted lady (Figure 3). The contribution of rolling-circles (also known as *Helitrons*) and
176 DNA-transposons is more than double in *Hypolimnas* than in the painted lady, while
177 retroelements are >1.5 fold higher in the painted lady (Figure 3). This suggests a shift in TE
178 activity, with *Helitrons* and DNA-transposons playing a more important role in *Hypolimnas*
179 species. In both *Hypolimnas*, there has been both a recent and a more ancient expansion of
180 *Helitron* family transposable elements (Figure 3B). Finally, the percentage of unclassified
181 repeats and GC content is broadly the same in the three species.

182

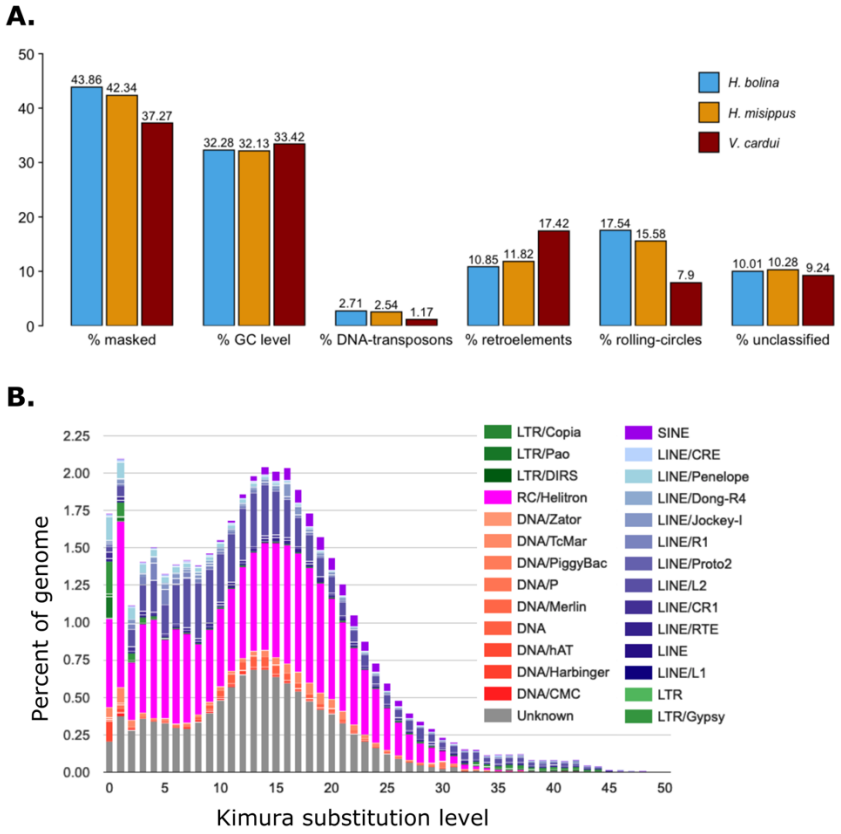


Fig. 3. A. Repeat content of the *Hypolimnas bolina, Hypolimnas misippus* and *Vanessa cardui* assemblies. **B**. Repeat landscape of the *H. misippus* assembly. Helitrons and LTR (Long Terminal Repeat) retrotransposons have undergone a recent expansion, as seen in the higher percentage of the genome 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 were predicted (Supplementary Table 1). 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, Supplementary

195 Table 1). Analysis with BUSCO using the *insecta_odb10* benchmarking set showed that the
196 completeness of the genome and annotation were 98.7% and 98.5% for HypMis_v2, and 94%
197 and 92.4% for HypBol_v1 (Table 1). These scores are comparable to other published
198 Nymphalidae assemblies and annotations such as that of *Danaus chrysippus* (Singh et al.
199 2022) or the small tortoiseshell butterfly, *Aglais urticae* (Bishop et al. 2021), with those for
200 *HypBol_v1* slightly lower possibly due to errors introduced by Nanopore sequencing, which
201 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*	98.5	98.1	0.4	0.7	0.8
	Hypolimnas bolina	92.4	91.7	0.8	2.9	4.7

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

204 Despite having a larger number of genes and mRNAs, the two *Hypolimnas* annotations show
205 a smaller total number of exons and introns than *V. cardui* (Supplementary Figure 2). This is
206 because the two *Hypolimnas* annotations have more single exon genes and their mRNAs

207 have, on average, a smaller number of exons, which might be a difference produced by the
208 distinct annotation pipelines. The total length of mRNAs and genes in the two *Hypolimnas* is
209 shorter then in *V. cardui*, which results in a smaller percentage of the genome covered by
210 them. However, this trend is different for exons, which are on average the same length in the
211 three species, have a comparable total length and cover a similar percentage of the genome.
212 Thus, the longer total length of mRNAs and genes in *V. cardui* is due to a total longer length
213 of introns, due to a higher number of introns per mRNA and an average longer length.

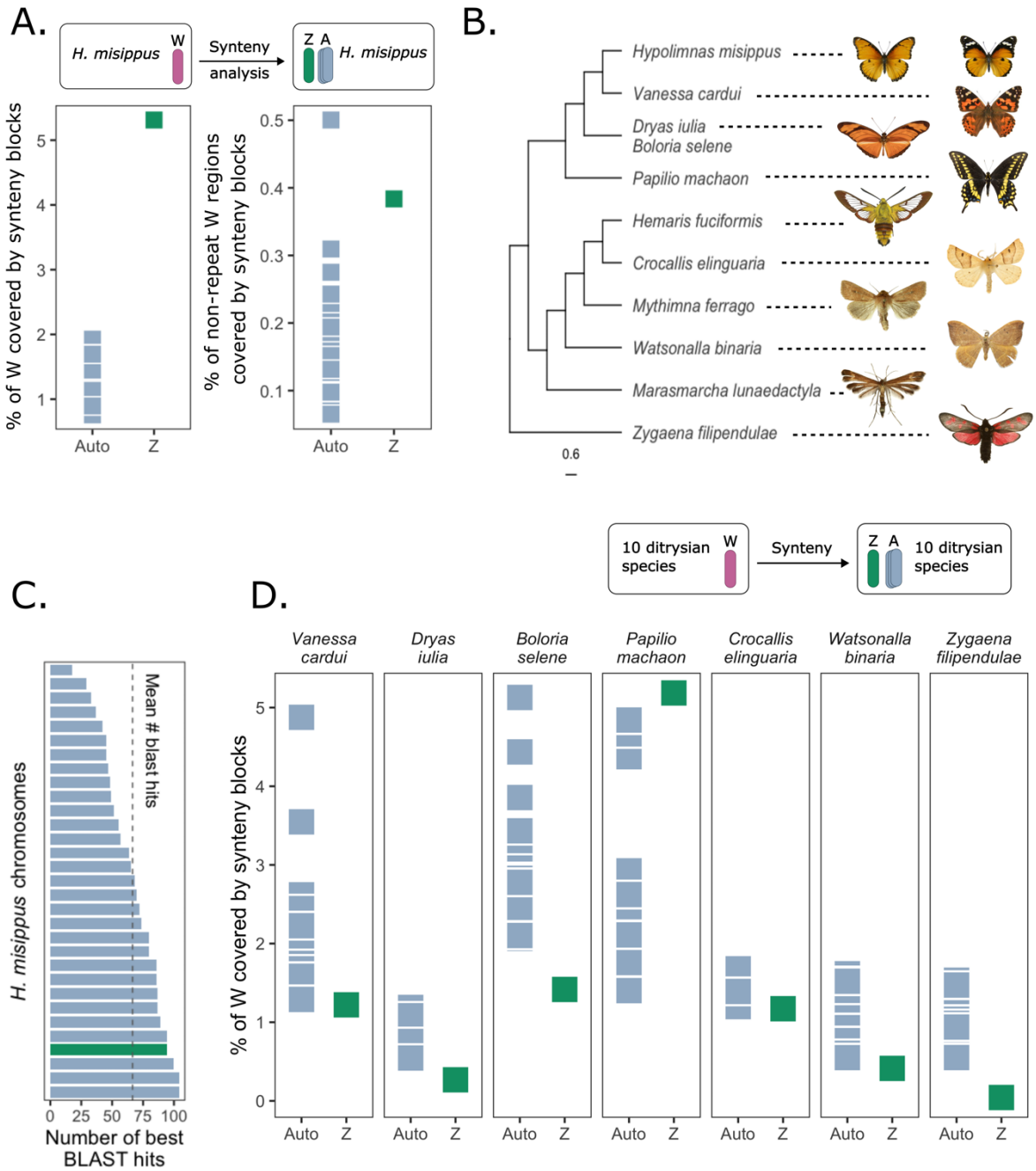
214 **Mixed evidence for the origin of W chromosomes from the same autosome pair as the Z**

215 Using our HypMis_v2 assembly that contains a W chromosome, we set out to explore the
216 evolutionary origins of the W chromosome in Lepidoptera. First, we wanted to clarify
217 whether the W chromosome had originated from a B chromosome or form the same
218 autosomal pair as the Z (Figure 1A-B). To specifically test the hypothesis that the Z and W
219 have a common autosomal origin, we used an approach that searches for syntenic blocks to
220 test for sequence homology between the *H. misippus* W chromosome and the remaining 31
221 chromosomes. We would expect that if the W and the Z share a common autosomal origin,
222 they might share tracts of homologous sequence that date back to their common ancestor. As
223 such, we might expect that the Z and W would each share more syntenic blocks with each
224 other than with any of the autosomes. We found that the Z chromosome indeed shares a
225 higher proportion of non-overlapping syntenic blocks with the W (Figure 4A). However, the
226 W chromosome is the chromosome with the highest percentage of repeat regions (71% in *H.*
227 *misippus*), which could interfere with the analysis (i.e. if the sex chromosomes share many
228 common repetitive sequences that may have accumulated independently on each sex
229 chromosome). Furthermore, the Z chromosome is the longest chromosome in *H. misippus*
230 and has the highest absolute length of repeats (Supplementary Figure 3 and 4). To take this

231 into account, we excluded all matches found in repetitive regions. With this correction, the Z

232 chromosome shared the second highest proportion of syntenic blocks with the W after

233 chromosome 2.



234

235 **Fig. 4.** Genomic synteny of the *H. misippus* W chromosome provides mixed evidence on its

236 origin. **A.** Synteny analysis of the *H. misippus* W chromosome and the autosomes and Z

237 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 chromosome with the Z shown in green. **D.** Synteny analysis of the W chromosomes of several Lepidoptera genome assemblies compared to the autosomes and Z of the same species. A schematic of the method is shown on top. *Hemaris fuciformis* image modified from Didier Descouens under CC BY-SA 4.0. *Dryas iulia* image modified from (Moore 2023) under CC-BY. *Watsonalla binaria* image modified from (Lennuk 2023) under CC BY-SA 4.0.

To further explore the possible homology between the W and Z chromosomes, we extracted the predicted amino acid sequences of all 1,394 annotated proteins in the *H. misippus* W chromosome and searched for homology in the remaining chromosomes using BLASTp. In contrast to the synteny results, we found that the Z chromosome placed 4th in the number of best BLAST protein hits with 94 proteins mapping to it, while 104 mapped to chromosome 2, 3 and 14 (Figure 4C). The average number of best BLAST hits was 66. Interestingly, chromosome 2 showed the highest sequence similarity in non-repeat regions and the highest number of best BLAST hits (Figure 4C).

In view of the differences between these results and previously published homology tests in other species, we decided to expand the analysis to include other Lepidoptera species. We searched for syntenic blocks between the W and the other 31 chromosomes in 6 diverse taxa: *Dryas iulia, Boloria selene, Papilio machaon, Crocallis elinguaria, Watsonalla binaria,* and *Zygaena filipendulae*. These species were chosen as they have a publicly available high-

262 quality assembly containing a W chromosome (Supplementary Figure 5) and cover distinct
263 ditrysian lineages (including 4 superfamilies: *Zyganoidea, Papilonoidea, Drepanoidea* and
264 *Geometroidea*). A low percentage of the W chromosome was covered by synteny blocks
265 when comparing it to the Z in any species (suggesting no homology) except for *P. machaon*,
266 which contrasts with the results in *H. misippus* (Figure 4D). However, these analyses were
267 performed using unmasked genomes and including all regions, which could be interfering
268 with the results.

269 **Comparisons of W chromosomes reveals a possible single common origin in ditrysian**
270 **lineages**

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 compared the *H. misippus* W chromosome
274 to a diverse set of 10 Lepidoptera species, including the 6 species from the previous analysis
275 plus *Vanessa cardui, Hemaris fuciformis, Mythimna farrago* and *Marasmarcha lunaedactyla*.
276 These species cover the breadth of ditrysian lineages, including 8 distinct superfamilies:
277 *Zyganoidea, Papilonoidea, Pterophonoidea, Pyraloidea, Drepanoidea, Noctuoidea,*
278 *Geometroidea* and *Bombycoidea*. We searched for syntenic blocks between the *H. misippus*
279 W chromosome and all the chromosomes of each target species (including the W; Figure
280 5A). This allowed us to test whether W chromosomes are more similar to each other than to
281 any autosome or the Z chromosome, which would provide evidence of a common origin.
282 When comparing the *H. misippus* W chromosome to these species, we found that there is a
283 high proportion of syntenic blocks among W chromosomes (Figure 5A). The highest degree
284 of sharing is seen between W chromosomes of closely related species such as *H. misippus*
285 and *Vanessa cardui* (45 MYa), but is also true even for highly divergent lineages such as

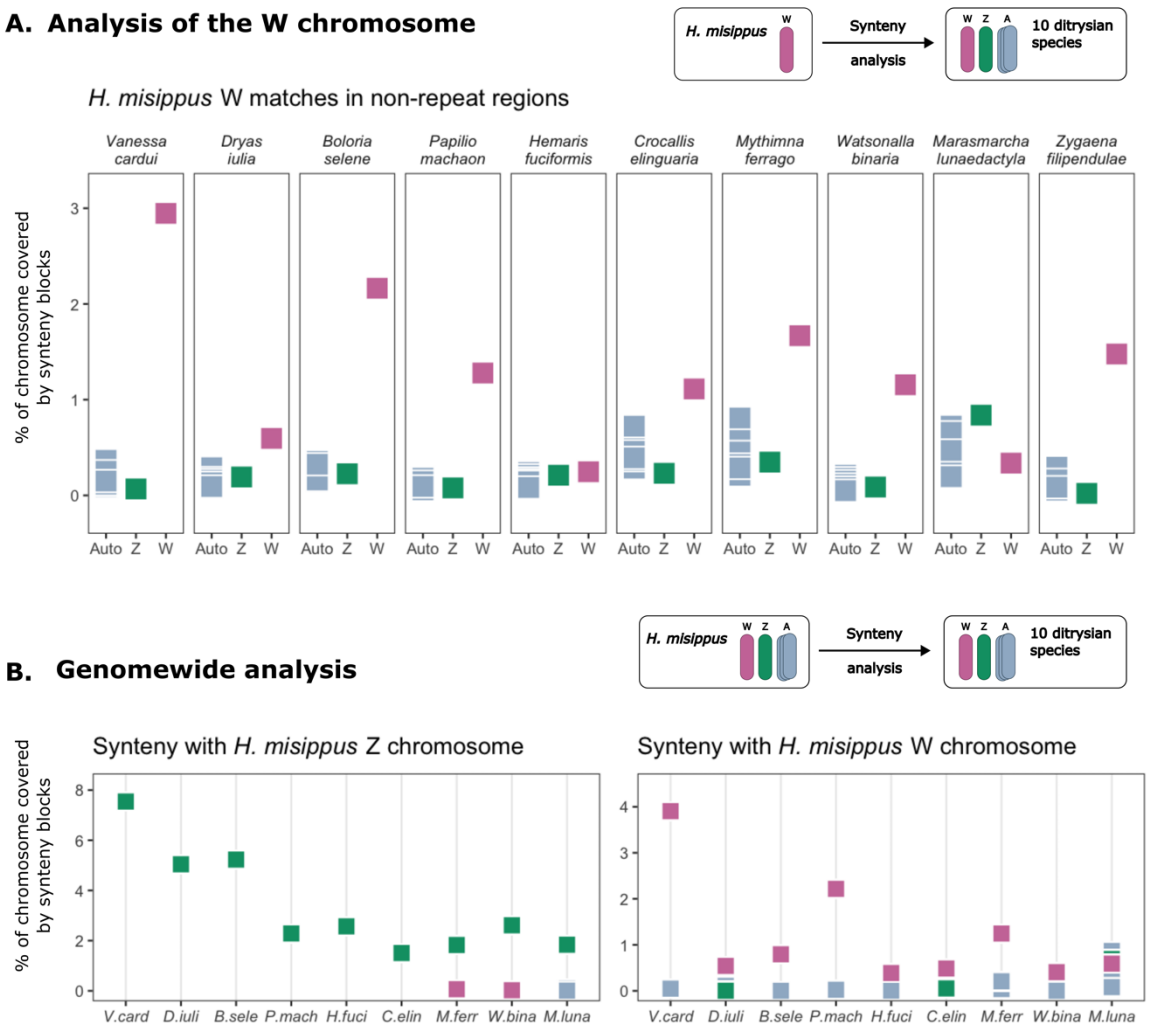
286 *Zygaena filipendulae*, which diverged from *H. misippus* 156 MYa (Kawahara et al. 2019).

287 Interestingly, there is no consistent pattern in the distribution of syntenic blocks along the *H.*

288 *misippus* W chromosome (Supplementary Figure 6), which might suggest that different

289 regions show homology with the *H. misippus* W chromosome across species and

290 chromosome types (W, Z, and autosomes).



291

292 **Fig. 5.** Genome wide comparisons suggests a single origin of ditrysian W chromosomes and

293 shows deep conservation of the Z chromosome. **A**. Percentage of query *H. misippus* W

294 chromosome covered by synteny blocks of the 32 chromosomes of 10 ditrysian species. W *H.*

295 *misippus* chromosome used as query (a schematic of the method is shown on top). **B.** Subset

296 results of genome wide analysis. Percentage of Z (left) and W (right) chromosome covered
297 by synteny blocks to other ditrysian species. W chromosome shown in pink, Z in green and
298 autosomes in grey. A schematic of the method is shown on top. Species names have been
299 abbreviated: *Vanessa cardui, Dryas iulia, Boloria selene, Papilio machaon, Hemaris*
300 *fuciformis, Crocallis elinguaria, Mythimna 60arrago, Watsonalla binaria, Marasmarcha*
301 *lunaedactyla* and *Zygaena filipendulae*.

302 We then performed the same analysis including all chromosomes of *H. misippus* as the query.
303 With this we could test for homology between all autosomes and sex chromsomes of *H.*
304 *misippus* with each chromosome of the target species. Unlike the above method, this
305 approach produces only primary matches for each query chromosome and is thus efficient at
306 finding homologous pairs. This approach yielded the same results as the former, in which the
307 W chromosome of *H. misippus* shares the greatest proportion of syntenic blocks with the W
308 of most species, with *Marasmarcha lunaedactyla* being the one exception (Figure 5B right).
309 Finally, the percentage of the Z chromosome covered by synteny blocks was the highest with
310 other Z chromosomes in all species comparisons (5B left), suggesting that it is highly
311 conserved.

312 **A multi-species comparison reinforces the hypothesis of the single origin of the W**
313 **chromosome and reveals a neo-Z chromosome**

314 In light of the above results suggesting a single origin of the W chromosome in ditrysian
315 lineages, we decided to explore this further by comparing the analysed ditrysian assemblies
316 with each other. We performed pairwise searches for syntenic blocks between species pairs
317 and observed consistent results. First, the Z chromosome is deeply conserved in multiple
318 species pairs and the proportion of syntenic blocks between Z chromosomes of different
319 species decays with species divergence in a similar manner to autosomes (Supplementary

320 Figure 7 and 8). Second, the W chromosome shows a more variable pattern of sequence
321 sharing between species but shows evidence for conservation even between distant species
322 pairs (Supplementary Figure 8). Nonetheless, several of the comparisons found no syntenic
323 blocks between W chromosomes, again indicating that the syntenic blocks shared with the *H.*
324 *misippus* W differed between target species.

325 Finally, the cross-species comparison revealed that the Z chromosome of *Marasmarcha*
326 *lunaedactyla* (chromosome name OV181339.1; identity as the Z chromosome assigned in the
327 public assembly) presented high levels of synteny with chromosome 19 of most other species
328 (Supplementary figures 9 and 10). This could suggest a historical fusion event between the
329 ancestral chromosome 19 and the Z chromosome of *M. lunaedactyla* creating a neo-Z
330 chromosome. The pattern of syntenic blocks seen in the OV181339.1 chromosome was
331 similar for all species compared, with approximately half of the chromosome showing
332 homology to chromosome 19 and half to the Z chromosome (Supplementary Figure 10).
333 Furthermore, the pattern of synetny with other species showed that translocations between the
334 parts belonging to the ancestral Z and ancestral chromosome 19 have resulted in an
335 interleaved pattern of homology. Consistent with these results, BUSCO matches of
336 chromosome OV181339.1 are shared with chromosome 19 and chromosome 1 (the Z
337 chromosome) of *Melitaea cinxia* (Supplementary Figures 9 and 10).

338 **Discussion**

339 *Hypolimnas* species have become a focus for studies of evolutionary genetics, due to their
340 mimetic colouration and co-evolution with *Wolbachia* parasites. Here, we have assembled
341 chromosome level reference genomes for *H. misippus* and *H. bolina* and RNA-informed
342 annotations for both genomes. HypBol_v1 represents the first assembly available for
343 *Hypolimnas bolina,* while HypMis_v2 is a significant improvement on a previously published

344 genome, with higher contiguity and higher N50 and BUSCO scores. Furthermore, the
345 annotation is also improved as we used RNA-seq data from *H. misippus* to inform our
346 annotation, as opposed to only homology methods. We present these two useful resources
347 and use the *H. misippus* assembly to shed light on the evolutionary origins of W
348 chromosomes in Lepidoptera.

349 Several models have been proposed to explain the origin and evolution of the W chromosome
350 in Lepidoptera (Dai et al. 2022; Fraïsse et al. 2017; Lukhtanov 2000; Sahara et al. 2012;
351 Berner et al. 2023). The deep conservation of the Z chromosome and its apparent lack of
352 homology with the W have been suggested to be evidence of a B chromosome origin of the
353 W chromosome (Fraïsse et al. 2017). Outside the Lepidoptera, it has been commonly found
354 that heteromorphic sex chromosomes retain a degree of similarity; evidence of their shared
355 autosomal origin (Wright et al. 2016). Thus, if the W chromosome originated from the same
356 homologous autosome pair as the Z, we might expect to find some residual sequence
357 similarity between the W and Z, as some recent evidence seems to suggest. However,
358 opposing evidence has also been presented, suggesting a lack of similarity between W and Z
359 chromosomes which could point to the evolution of the W from a B chromosome (Fraïsse et
360 al. 2017; Lewis et al. 2021). Furthermore, the lack of similarity among W chromosomes of
361 Lepidoptera species has led to the suggestion of multiple independent recruitments from B
362 chromosomes (Dai et al. 2022; Lewis et al. 2021). Taking this into account, we set out to
363 address two questions. First, did the W chromosome evolve from the same homologous pair
364 of autosomes as the Z or from a B chromosome? Second, do W chromosomes in Lepidoptera
365 share a common origin, that is, did they evolve once or multiple times independently? Our
366 assembly of the W chromosome of HypMis_v2 is an ideal system to explore these questions,
367 as its large size (13.23 Mbp) compared to other W assemblies (e.g. 1.87 Mbp in *Zygaena*

368 *filipendulae* or 2,1 in *Dryas iulia*) increases our chances of detecting homology even in
369 scenarios of extensive degradation.

370 Using tests for homology between the W and remaining 31 chromosomes in *H. misippus* and
371 other Lepidoptera, we show that there is some evidence for homology between the W and the
372 Z chromosome, suggesting a possible common origin. These results differ from the homology
373 tests performed in *Danaus plexipuss* (Lewis et al. 2021), *Kallima inachus* (Yang et al. 2020)
374 and *Dryas iulia* (Lewis et al. 2021), where the evidence was inconsistent with a common
375 origin of the W and the Z chromosome. Instead, our findings are more in line with results in
376 the Asian corn borer, *Ostrinia furnacalis,* a moth belonging to the Pyraloidea, and *Pieris*
377 *mannii* which also suggest a common autosomal origin of the Z and the W chromosome (Dai
378 et al. 2022; Berner et al. 2023). Contrastingly, our results based on protein-coding genes
379 contrast with the analysis of overall homology and suggest that the W chromosome did not
380 evolve from the same homologous autosomal pair as the Z. Nonetheless, as a result of the
381 arrest of recombination and subsequent accumulation of mutations, sex specific
382 chromosomes such as the W are often gene poor and highly degenerated compared to their
383 autosomal counterparts (Wright et al. 2016). Additionally, lepidopteran W chromosomes
384 have been shown to evolve rapidly (Yoshido et al. 2013; Vítková et al. 2007). Thus, the lack
385 of homology between proteins of the W and other chromosomes is expected given the rapid
386 turnover seen in the W chromosome and is in line with results shown in other Lepidoptera
387 species (Yang et al. 2020; Singh et al. 2022; Lewis et al. 2021). Similarly to the protein-
388 coding analysis, the low proportion of synteny blocks between the W and the Z in five of the
389 seven species studied could suggest a B chromosome origin or could be the result of the rapid
390 evolution and degradation of the W, which could impede the detection of ancient homology
391 (Yoshido et al. 2013; Vítková et al. 2007). Taken together, the homology tests between the W
392 and remaining 31 chromosomes in *H. misippus* and other Lepidoptera provide mixed

evidence for the origin of the W from the same homologous pair of autosomes as the Z or from a B chromosome. Alternatively, our finding of synteny blocks between the *H. misippus* W and the Z could point to a fusion of the Z chromosome with an autosome and subsequent formation of the W (Figure 1C). This hypothesis would fit the mounting evidence supporting a Z0 system in non-ditrysian Lepidoptera and early diverging ditrysian lineages (Fraïsse et al. 2017; Hejníčková et al. 2019; Lukhtanov 2000; Sahara et al. 2012).

The second question we wanted to address was whether the W has a single origin in the Lepidoptera or evolved independently multiple times. Previously, the lack of similarity among W chromosomes of Lepidoptera species has been suggested to be the result of multiple independent recruitments from B chromosomes (Lewis et al. 2021). Similarly, phylogenies of W genes in the Asian corn borer could suggest multiple independent origins of the W chromosome from the same autosomal pair as the Z (Dai et al. 2022). Regardless of the specific event that led to the formation of the W chromosome (i.e., from a B chromosome or from an autosome), if the W chromosomes of ditrysian lineages share a common origin, we would expect them to share homologous sequence tracts that date back to their common ancestral chromosome, and therefore we expected to find more syntenic blocks between W chromosomes than between W and and Z or W and autosomes. Comparing the *H. misippus* W chromosome to the genomes of 10 other ditrysian species, our findings matched this expectation in all but one species. These results suggest that W chromosomes of ditrysian lineages may share a single common origin, which contrasts with evidence of rapid turnover of W chromosomes in Lepidoptera (Dai et al. 2022; Lewis et al. 2021). However, it is possible that some species have retained the ancestral W, while some other have recruited other B chromosomes as suggested. For example, in our test for homology between the *H. misippus* W chromosome and each of the chromosomes of *Marasmarcha lunaedactyla*, the W chromosome of *M. lunaedactyla* did not stand out above the Z and some autosomes. Our

418 results also show unambiguous evidence for homology among Z chromosomes of the studied
419 ditrysian species, which is consistent with evidence suggesting deep conservation of the Z
420 chromosome across the Lepidoptera (Fraïsse et al. 2017).

421 Overall, we present evidence of a single common origin of W chromosomes in the
422 Lepidoptera. Furthermore, our results hint at a possible origin from the same autosomal pair
423 as the Z or perhaps to a Z-autosome fusion event that culminated with the formation of the
424 W. Crucially, the rapid evolution of the W chromosome could cause the lack of similarity
425 observed between Z and W chromosomes in some species and hinders the elucidation of its
426 origin. Further studies including more species across the Lepidoptera are necessary to clarify
427 whether the W chromosome evolved from an autosome or from a B chromosome, and to shed
428 light on the possible shared origin of W chromosomes, particularly including non-ditrysian
429 lineages.

430 Materials and methods

431 *Hypolimnas misippus* butterfly rearing and cross preparation

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

440 ***H. misippus* Trio binning genome assembly**

441 One *H. misippus* family was used for sequencing and trio-binning genome assembly. DNA
442 was extracted using MagAttract HMW DNA Kit (QIAGEN). The two parents were
443 sequenced using Illumina short-read sequencing, which resulted in 368.67 and 346.56 million
444 read pairs and a total yield of 55.67 and 52.33 Gbp of data from the mother and father
445 respectively. One of the offspring was sequenced using PacBio long-read sequencing (total
446 yield of 13.46 Gb of data and N50 of 13,493). Trio binning enables the independent assembly
447 of the two parental haplotypes. First, yak-r55 (Li 2022) was used to create a kmer database
448 from the parental Illumina data. Then, hifiasm-0.7-r256 (Cheng et al. 2020) was run in trio
449 mode to assemble the HiFi PacBio long-read data, followed by the purging of haplotypes and
450 overlaps using purgedups v1.2.3 (Guan 2022). Haplotype 1 corresponds to the paternal
451 haplotype, while haplotype 2 corresponds to the maternal haplotype. All subsequent analysis
452 were performed using haplotype 2.

453 **Curation of HypMis_v2 with HiC**

454 To place the assembled scaffolds into chromosomes, Hi-C sequencing was used, which is a
455 chromosome conformation capture technology that provides information about the 3-D
456 interactions between genomic loci. An offspring from an unrelated mating was flash frozen in
457 liquid nitrogen and used for Hi-C sequencing and analysis. 386.99 million read pairs were
458 produced, which represented 58,45 Gb of Hi-C data. Fastq files were converted to cram (two
459 files, one for each read pair end), mapped to haplotype 2 (the maternal haplotype). The *H.*
460 *misippus* assembly was produced after the purgedups stage and processed with Juicer v1.6
461 (Durand, Shamim, et al. 2016). Juicer transforms raw Hi-C data into a list of contacts, which
462 defines pairs of genomic positions that were in close physical contact during the experiment.
463 Then, the main reference assembly was curated using the 3D-DNA pipeline, which corrects

464 misassembles, anchors, orders and orients fragments of DNA based on the Hi-C data
465 (Dudchenko et al. 2017; Durand, Robinson, et al. 2016). 3D-DNA generated assembly
466 heatmaps as part of its workflow, which indicate the frequency of contact between pairs of
467 genomic locations. Obvious errors in the genome assembly such as large genomic inversions
468 were manually edited by examining the Hi-C heatmaps using the Juicebox tool (Durand,
469 Robinson, et al. 2016; Dudchenko et al. 2018). Finally, the edited assembly was exported as
470 an *.assembly* file and converted to a final fasta assembly file using the 'run-asm-pipeline-
471 post-review.sh' script setting –editor-repeat-coverage to 6.

472 **HypBol_v1 genome curation with Ragout**

473 The *H. bolina* reference assembly (HypBol_v1) was produced using a combination of
474 Nanopore long-read sequencing and a linkage map. Pupae were purchased from the Stratford-
475 upon-Avon butterfly farm, which obtains farmed lineages from South East Asia. One female
476 pupa was used for sequencing. High molecular weight DNA was extracted following a
477 phenol chloroform and glass capillary hook protocol following(Quick 2018). A library was
478 prepared for Nanopore sequencing using the Nanopore Ligation Sequencing Kit (SQK-LSK
479 109) following a LSK109 bead free library preparation protocol (Tyson 2020). Post library
480 preparation, a total of 1.78μg DNA remained and was sequenced across 2 R9 chemistry
481 Nanopore MinION flow cells, resulting in 7.35 million reads totalling 14.16 Gb. Adapters
482 were removed using Porechop v0.2.4 (Wick 2022) and reads assembled using redbean (Ruan
483 & Li 2020). Assembled contigs were split into bins using MaxBin2 (Wu et al. 2016) and
484 bacterial contigs and reads were removed from the data using blobtools2 (Challis et al. 2020).
485 This produced an assembly with 13,492 contigs and an N50 of 1.4 kbp.

486 Leveraging the high synteny expected from the two *Hypolimnas* assemblies, Ragout v2.3
487 (Kolmogorov et al. 2014, 2018) was used to improve the initial *HypBol_v1* assembly using

488 HypMis_v2 as a reference. First, the two genomes were soft-masked using RepeatMasker
489 (Smit et al. 2015), creating the repeat library based on the genome being masked. Then,
490 Cactus (Paten et al. 2011) was used to align both genomes using Python 3.8. The resulting
491 HAL alignment was converted to MAF format using the hal2maf utility from the HAL
492 program (Hickey et al. 2013). Finally, Ragout was run using the MAF alignment between
493 HypBol_v1 and HypMis_v2 as input.

494 **Rearing of *H. bolina* individuals**

495 A linkage map was used to improve the HypBol_v1 assembly and place the assembled
496 scaffolds into chromosomes. Two families (178303XX and 182703XX) were reared and
497 sequenced. First, female *H. bolina* purchased from Stratford upon Avon Butterfly Farm of SE
498 Asian origin were mated to wild-caught males from Mo'orea (French Polynesia) at the
499 University of California Berkeley Gump Station research facility. Female SE Asian-Mo'orea
500 F1 hybrid offspring were then mated to pure Mo'orea F1 males. The F2 offspring of one of
501 these crosses is family 178303XX. At the same time, male SE Asia-Moorea F1 hybrids were
502 mated to pure SE Asia F1 females. The F2 offspring of one of these crosses is family
503 182703XX. Butterflies were kept in a large outdoor cage for mating under observation. Any
504 copulating pairs were separated into a small cage. The mated female was then placed in an
505 oviposition cup containing a small plant e.g. *Asystasia sp*. and allowed to lay eggs. Hatched
506 eggs were moved to a rearing box with suitable food plant e.g. Ipomoea sp. and caterpillars
507 reared until pupation. Pupae were moved to individual cups for emergence, adults left to dry
508 for one day and then used for further matings or stored in -80˚C freezer.

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

510 The offspring of the *H. bolina* families (F2) were processed to extract the DNA and prepare
511 the libraries for Illumina sequencing. DNA extractions were carried out using a custom
512 protocol using PureLink buffers and homemade magnetic beads. Briefly, a small piece of
513 thorax tissue (1/10) is placed in an 8-tube PCR strip. Then, 45 μL of PureLink Digestion
514 buffer and 10 μL of Proteinase K (20mg/mL) are added, and the mix is incubated at 58˚C
515 with shaking (500 rpm) for 2-3 hours. Afterwards, 2 μL of RNAseA are added (DNAse free,
516 10mg/mL) and incubated for 10min at room temperature. Then, 45 μL of PureLink Lysis
517 buffer is added to the mix and incubated at 58˚C for 30 minutes with shaking (500 rpm).
518 Following that, a homemade magnetic bead mix is used to extract the DNA from the lysate.
519 First, 37.5 μL of magnetic beads are added together with 75 μL of lysate to a 96-well plate.
520 After mixing, the samples are incubated for 15 minutes at room temperature, then the plate is
521 placed on a magnetic stand for 10 minutes, the supernatant removed, and the beads cleaned
522 with 80% ethanol. After drying out, 50 μL of 10mM Tris (pH=8) are added to elute the
523 sample and incubated at 45˚C for 15 minutes without resuspending. Then, the beads are
524 resuspended and incubated for 20 minutes at room temperature. Finally, the plate is placed on
525 the magnetic stand and, after 10 minutes, the supernatant (the DNA) is transferred to a fresh
526 tube.

527 The F2 were sequenced using a Nextera-based library preparation at intermediate coverage
528 (~11X). A secondary purification using magnetic SpeedBeads™ (Sigma) was performed
529 prior to Nextera-based library preparation. Libraries were prepared following a method based
530 on Nextera DNA Library Prep (Illumina, Inc.) with purified Tn5 transposase (Picelli et al.
531 2014). PCR extension with an i7-index primer (N701–N783) and the N501 and N502 i5-
532 index primers was performed to barcode the samples. Library purification and size selection

533 was done using the same homemade beads as above. Pooled libraries were sequenced by
534 Novogene Cambridge, UK. Libraries of the parental samples were prepared and sequenced to
535 ~20X coverage by Novogene Cambridge, UK.

536 **_H. bolina_ linkage map construction and anchoring of the genome**

537 A linkage map was produced with Lep-Map3 (Rastas 2017) and then used to improve the *H.*
538 *bolina* assembly and place the scaffolds onto chromosomes. First, sequences were mapped to
539 the reference genome using bwa-mem (Li 2013). PCR duplicates were marked using the
540 MarkDuplicates from Picard tools. Sorted BAMs were then created using SAMtools (Li et al.
541 2009) and genotype likelihoods computed. The pedigree of individuals was checked and
542 corrected using IBD (identity-by-descent) with a random subset of 10% of the markers
543 (1,270,024 SNPs) following the IBD pipeline from Lep-Map3. These markers were also used
544 to construct the linkage map. Scaffolds were anchored into chromosomes based on the
545 linkage map using LepAnchor (Rastas 2020).

546 **HypBol_v1 polishing with Pilon**

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

551 **Repeat annotation**

552 Once the two final assemblies had been produced, they were each assessed for repeat content
553 using a custom repeat library. First, a repeat database was built and the repeats of the two
554 finished assemblies modelled using RepeatModeler v. 2.0.2a. Each custom library was then

555 combined with the Lepidoptera library extracted from Dfam (Storer et al. 2021). This merged
556 library was used to soft mask the genome using RepeatMasker v 4.1.0 with the cut-off score
557 set to 250 and skipping the bacterial insertion element check. The resulting soft masked
558 assemblies were used for gene annotation.

559 **RNA-seq sample preparation**

560 To assist with genome annotation, RNA-seq data was obtained from 4 *H. misippus*
561 individuals (2 adults and 2 pupae) and 17 *H. bolina* individuals (6 adults and 11 pupae).
562 Butterflies were purchased from Stratford-upon-Avon Butterfly Farm and kept at room
563 temperature until dissection. 4 tissues were dissected out from *H. misippus* pupae and placed
564 in RNA-later (Sigma): wing discs, thorax, head, and abdomen; while *H. bolina* pupae 3
565 tissues were dissected: wing discs, thorax-head, and abdomen. Only abdomen, head and
566 thorax samples were dissected from adults of either species. Two pooling strategies were
567 followed: 1) 2 *H. bolina* pupae were pooled by individual, pooling head, thorax, abdomen,
568 and wing discs together and sequenced at high coverage (50M reads). And 2) 4 *H. misippus*
569 and 15 *H. bolina* adults and pupae were dissected into tissues and pooled by species and
570 tissue. Each pooled sample, 4 for *H. misippus* and 3 for *H. bolina,* was sequenced to 20M
571 reads. RNA was extracted using a modified Trizol protocol using the same protocol as in
572 (Brien et al. 2022).

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

574 First, low quality ends and adaptors from the RNA-seq data were trimmed using TrimGalore!
575 v 0.5.0 (Krueger 2015). Then, the reads were mapped to the soft masked genomes using
576 STAR v 2.5.0a (Dobin et al. 2013). Two rounds of mapping (2pass) were performed,
577 including all the splice junction files in the second round. Then the resulting mapped reads

578 and the soft masked genome assembly were used to generate a gene annotation using
579 BRAKER v 2.1.5 (Brůna et al. 2021), running it a second time to add UTR annotations with
580 options addUTR=on and skipAllTraining (Keilwagen et al. 2019). A single isoform per gene
581 was selected and completeness of the annotation assessed using BUSCO (Seppey et al. 2019;
582 Simão et al. 2015) using the Insecta_odb10 (n=1367) set of genes.

583 **Homology with Merian units**

584 A growing practice in Lepidoptera genetics is to name chromosomes of reference assemblies
585 by their homology to the ancestral lepidopteran chromosomes, also known as Merian units in
586 honour of the scientist Maria Sibylla Merian (Wulf 2016; Wright et al. 2023) and allow for
587 the tracking of conserved synteny blocks across the phylogeny of the Lepidoptera. To assign
588 homology to the Merian units, lepidopteran_odb10 BUSCO assignment of the *Melitaea*
589 *cinxia* genome was compared to the lepidopteran_odb10 BUSCO assignment of *HypBol_v2*
590 and *HypBol_v1*. *Melitaea cinxia* present 31 Merian units, which represents the ancestral state
591 of Ditrysia (Wright et al. 2023). *M. cinxia* belongs to the *Nymphalinae* which is the same
592 subfamily as *Hypolimnas*, and the two taxa diverged about 30 MYa (Espeland et al. 2018).
593 Merian units in *HypBol_v1* and *H. missipus* were assigned by choosing the chromosome
594 sharing the highest BUSCO hits with a *M. cinxia* chromosome.

595 **Searches of synteny blocks**

596 Synteny of the two assemblies was assessed using two methods. First, the two final genome
597 assemblies were aligned using D-GENIES (Cabanettes & Klopp 2018), which produced a paf
598 file that was used to detect candidate inversions between the two genome assemblies
599 (Supplementary Figure 1). Second, synteny between the two assemblies was evaluated using

600 Satsuma2 Synteny (Grabherr et al. 2010) and a circos plot generated using the circlize
601 package v 0.4.14 (Gu et al. 2014) in R v 4.1.2 (Figure 2).

602 To examine the origin and evolution of the W chromosome in Lepidoptera, first, synteny-
603 based homology between the *H. misippus* W chromosome, the Z chromosome and the
604 autosomes was evaluated using Satsuma2. Satsuma2 is an aligner of whole genome
605 assemblies intended to find homology based on sequence similarity. Satsuma2 first maps all
606 genomic windows of the query genome to the target genome with a percentage of identity
607 higher than 45%, and then filters those hits based on large scale synteny, that is, it keeps only
608 matches that are concordant with each other. Thus, Satsuma2 is not only aligning genomic
609 windows, but also evaluating synteny between those blocks. Finally, Satsuma2 focusses on
610 regions with a high number of hits, to exhaustively evaluate the region around those hits, a
611 strategy analogous to the battleship game. To evaluate synteny between the W and the
612 remaining 31 chromosomes in *H. misippus,* the W chromosome was used as a query and the
613 31 assembled chromosomes as targets for Satstuma2. Resulting mapped regions were then
614 filtered to keep only non-overlapping regions using the package GenomicRanges in R v4.1.2
615 (Lawrence et al. 2013). Satsuma2 does not require input genome assemblies to be masked.
616 Because of its algorithm and filtering steps, repeat regions mapping to multiple places in the
617 genome have decreased score and may be filtered out. However, to evaluate only the synteny
618 of non-repeat regions, the RepeatMasker output, which details the coordinates of repeats in
619 the genome, was used to filter out repeat regions of the W chromosome. Finally, the effect of
620 sequence identity was evaluated by performing the analysis with no identity filter and
621 applying a threshold of 70% similarity.

622 The synteny between the W and the rest of the genome was also evaluated in 6 other
623 Lepidoptera species: *Boloria selene* (GCA_905231865.2), *Crocallis elinguaria*

624 (GCA_907269065.1), *Hemaris fuciformis* (GCA_907164795.1), *Papilio machaon*
625 (GCA_912999745.1), *Watsonalla binaria* (GCA_929442735.1), *Zygaena filipendulae*
626 (GCA_907165275.2) and *Dryas iulia* (GCA_019049465.1). In all cases, the W chromosome
627 was used as the query, results were converted to non-overlapping regions and percentage of
628 the W covered by matches calculated. No filter based on repeats was applied, as no repeat
629 library was available for these species.

630 After that, Satsuma2 was used to compare the synteny of the *H. misippus* W (used as query)
631 and Z chromosomes to 10 Lepidoptera species, the 6 from the previous analysis and also
632 *Hemaris fuciformis* (GCA_907164795.1), *Marasmarcha lunaedactyla* (GCA_923062675.1),
633 *Vanessa cardui* (GCF_905220365.1), and *Mythimna farrago* (GCA_910589285.2).
634 Chromosomes of all the species analysed were re-named by their homology to *M. cinxia* as
635 above. Again, results were converted to non-overlapping regions, repeat regions of the W
636 chromosome filtered and percentage of the W chromosome covered by matches calculated.
637 Using only the *H. misippus* W chromosome as query ensures that secondary and tertiary
638 matches are also reported. A more specific analysis was produced by using the *H. misippus*
639 assembly as query to search for synteny in the 10 Lepidoptera target species. Using whole
640 assemblies as input limits the results to only primary matches, pairing all homologous
641 autosomes.

642 **BLAST of *H. misippus* W chromosome protein-coding genes**

643 To further explore the degree of homology between the *H. misippus* W chromosome and the
644 autosomes and Z chromosome, protein sequence homology was assessed using BLASTp
645 v2.4.0+ (BLAST 2013; Kent 2002). First a protein BLAST database was built using the
646 protein sequences contained in the *H. misippus* autosomes and the Z chromosome in the
647 HypMisi_v2 annotation. Then, protein sequences found in the W chromosome were extracted

648 and blasted to the database setting the minimum e-value to 1e-10 and the maximum number
649 of target sequences to be reported to 5,000. The best match for each protein sequence was
650 selected based on e-value score, and if two matched had the same value percentage of identity
651 was used. Finally, genome coordinates were extracted from the HypMisi_v2 annotation.

652 **Acknowledgements**

653 This work was supported by the Natural Environment Research Council [grant number
654 NE/L002507/1 to A.O; NE/N010434/1 to GH]; the Cambridge Trust [European Research
655 Scholarship to A.O] and St. John's College [Benefactors' Scholarship to A.O]. SHM was
656 supported by a Royal Society University Research Fellowship [grant number
657 URF\R1\180682].

658 **Data availability**

659 Genome assemblies and annotations for *Hypolimnas misippus* and *Hypolimnas bolina* are
660 available NCBI BioProject accession PRJEB64523. HiFi PacBio and Illumina raw reads of
661 the *H. misippus* trio binning samples are available NCBI BioProject accession PRJEB64523.
662 The data underlying this article are available in 10.5281/zenodo.8172459. Code used for
663 analyses can be found in
664 https://github.com/annaorteu/Hypolimnas_genome_Wchr_evolution.

665 **References**

666 Bachtrog D et al. 2011. Are all sex chromosomes created equal? Trends in Genetics.
667 27:350–357. doi: 10.1016/j.tig.2011.05.005.

668 Bachtrog D et al. 2014. Sex Determination: Why So Many Ways of Doing It? PLOS Biology.
669 12:e1001899. doi: 10.1371/journal.pbio.1001899.

Berner D, Ruffener S, Blattner LA. 2023. Chromosome-Level Assemblies of the Pieris mannii Butterfly Genome Suggest Z-Origin and Rapid Evolution of the W Chromosome. Genome Biology and Evolution. 15:evad111. doi: 10.1093/gbe/evad111.

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

Bishop G, Ebdon S, Lohse K, Vila R. 2021. The genome sequence of the small tortoiseshell butterfly, Aglais urticae (Linnaeus, 1758). doi: 10.12688/wellcomeopenres.17197.1.

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

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

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

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

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

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

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

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

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

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

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

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

704 Durand NC, Robinson JT, et al. 2016. Juicebox Provides a Visualization System for Hi-C
705 Contact Maps with Unlimited Zoom. Cell Systems. 3:99–101. doi:
706 10.1016/j.cels.2015.07.012.

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

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

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

714 Fraïsse C, Picard MAL, Vicoso B. 2017. The deep conservation of the Lepidoptera Z
715 chromosome suggests a non-canonical origin of the W. Nat Commun. 8:1486. doi:
716 10.1038/s41467-017-01663-5.

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

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

721 Guan D. 2022. Purge_Dups. https://github.com/dfguan/purge_dups (Accessed September
722 29, 2022).

723 Hejníčková M et al. 2019. Absence of W Chromosome in Psychidae Moths and Implications
724 for the Theory of Sex Chromosome Evolution in Lepidoptera. Genes. 10:1016. doi:
725 10.3390/genes10121016.

726 Hickey G, Paten B, Earl D, Zerbino D, Haussler D. 2013. HAL: a hierarchical format for storing
727 and analyzing multiple genome alignments. Bioinformatics. 29:1341–1342. doi:
728 10.1093/bioinformatics/btt128.

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

731 Kawahara AY et al. 2019. Phylogenomics reveals the evolutionary timing and pattern of
732 butterflies and moths. Proc Natl Acad Sci U S A. 116:22657–22663. doi:
733 10.1073/pnas.1907847116.

734 Keilwagen J, Hartung F, Grau J. 2019. GeMoMa: Homology-Based Gene Prediction Utilizing
735 Intron Position Conservation and RNA-seq Data. In: Gene Prediction: Methods and
736 Protocols. Kollmar, M, editor. Methods in Molecular Biology Springer: New York, NY pp.
737 161–177. doi: 10.1007/978-1-4939-9173-0_9.

738 Kent WJ. 2002. BLAT - The BLAST-like alignment tool. Genome Research. doi:
739 10.1101/gr.229202. Article published online before March 2002.

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

742 Kolmogorov M, Raney B, Paten B, Pham S. 2014. Ragout—a reference-assisted assembly
743 tool for bacterial genomes. Bioinformatics. 30:i302–i309. doi:
744 10.1093/bioinformatics/btu280.

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

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

749 Lennuk L. 2023. Estonian Museum of Natural History Department of Zoology. doi:
750 10.15468/98CXTC.

751 Lewis JJ et al. 2021. The Dryas iulia Genome Supports Multiple Gains of a W Chromosome
752 from a B Chromosome in Butterflies. Genome Biology and Evolution. 13. doi:
753 10.1093/gbe/evab128.

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

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

757 Li H et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics.
758 25:2078–2079. doi: 10.1093/bioinformatics/btp352.

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

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

765 MCZ Harvard University. 2023. Museum of Comparative Zoology, Harvard University. doi:
766 10.15468/P5RUPV.

767 Moore R. 2023. Auckland Museum Entomology Collection. doi: 10.15468/CADE8J.

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

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

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

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

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

779 Ruan J, Li H. 2020. Fast and accurate long-read assembly with wtdbg2. Nat Methods.
780 17:155–158. doi: 10.1038/s41592-019-0669-3.

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

783 Sahoo RK et al. 2018. Evolution of Hypolimnas butterflies (Nymphalidae): Out-of-Africa
784 origin and Wolbachia-mediated introgression. Molecular Phylogenetics and Evolution.
785 123:50–58. doi: 10.1016/j.ympev.2018.02.001.

786 Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: Assessing Genome Assembly and
787 Annotation Completeness. In: Gene Prediction: Methods and Protocols. Kollmar, M, editor.
788 Methods in Molecular Biology Springer: New York, NY pp. 227–245. doi: 10.1007/978-1-
789 4939-9173-0_14.

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

793 Singh KS et al. 2022. Genome assembly of Danaus chrysippus and comparison with the
794 Monarch Danaus plexippus. G3 Genes|Genomes|Genetics. 12:jkab449. doi:
795 10.1093/g3journal/jkab449.

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

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

800 Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. 2021. The Dfam community resource of
801 transposable element families, sequence models, and genome annotations. Mobile DNA.
802 12:2. doi: 10.1186/s13100-020-00230-y.

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

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

808 Vane-Wright RI, Ackery PR, Smiles RL. 1977. The polymorphism, mimicry, and host plant
809 relationships of Hypolimnas butterflies. Biological Journal of the Linnean Society. 9:285–297.
810 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.

Wright CJ, Stevens L, Mackintosh A, Lawniczak M, Blaxter M. 2023. Chromosome evolution in Lepidoptera. 2023.05.12.540473. doi: 10.1101/2023.05.12.540473.

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.