



HAL
open science

The deep conservation of the Lepidoptera Z chromosome suggests a non-canonical origin of the W

Christelle Fraïsse, Marion a L Picard, Beatriz Vicoso

► To cite this version:

Christelle Fraïsse, Marion a L Picard, Beatriz Vicoso. The deep conservation of the Lepidoptera Z chromosome suggests a non-canonical origin of the W. *Nature Communications*, 2017, 8 (1), pp.1486. 10.1038/s41467-017-01663-5 . hal-03821394

HAL Id: hal-03821394

<https://hal.science/hal-03821394>

Submitted on 23 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Deep conservation of the Lepidoptera Z-chromosome:**
2 **evidence of a non-canonical origin of the W-chromosome?**

3 Christelle Fraïsse^o

4 Marion A. L. Picard^o

5 Beatriz Vicoso*

6 Institute of Science and Technology Austria, Klosterneuburg, Austria

7 *Corresponding author: E-mail: beatriz.vicoso@ist.ac.at.

8 ^oThese authors contributed equally to this work.

9 Running title: Deep conservation of the Lepidoptera Z-chromosome.

10 Key words: Sex chromosome evolution, ZW species, Lepidoptera, Comparative genomics

11 Statistics:

12 • Word count:

13 • Figure count: 5

14 • Table count: 0

15 • Online supplementary information:

16 – Data: 3

17 – Text: 5

18 – Figures: 4

19 – Tables: 7

20 **Abstract**

21 The canonical model of sex chromosome evolution posits that sex chromosomes originate either
22 when an autosome acquires a sex-determining gene, or when it fuses to an existing sex
23 chromosome. This common autosomal ancestry of XY and ZW pairs is well established in
24 vertebrates, and presumed to be the mode of evolution of most new sex chromosomes.

25 Moths (Lepidoptera) usually have a pair of differentiated ZW sex chromosomes. However, the
26 most basal lineages, as well as the sister order Trichoptera, are Z0 (females lack a W-
27 chromosome). The W-chromosome is therefore thought to have been acquired secondarily, either
28 through: 1. A fusion between an autosome and the Z-chromosome; 2. A non-canonical
29 mechanism (e.g. through the recruitment of a B-chromosome). Our comparative genomics
30 analysis of basal and derived moths supports the secondary acquisition of the W-chromosome
31 and argues against such a fusion, suggesting a non-canonical origin of the Lepidoptera W-
32 chromosome.

33 **Introduction**

34 Despite the near-ubiquity of sexual reproduction in animals, and partial conservation of pathways
35 involved in sex determination, differentiated sex chromosomes have evolved independently in
36 many clades (Graves 2008). Chromosomal sex determination can take two forms: male-
37 heterogamety (XY systems, such as those of mammals), and female-heterogamety (females are
38 ZW, males ZZ, such as in birds); in both cases the sex-specific chromosome (Y or W) is often
39 gene-poor and consists mostly of repetitive DNA. Under the canonical model, such differentiated
40 sex chromosomes initially arise from standard pairs of autosomes: after acquiring a sex-
41 determining gene, newly formed Y (or W) chromosomes accumulate sexually antagonistic
42 mutations that are favorable to the sex they are found in (reviewed in Wright et al. 2016). The
43 presence of such mutations eventually favours the loss of recombination between the sex
44 chromosomes, in order to achieve full linkage between sex and alleles with sex-specific benefits,
45 and sets in motion the degeneration of the sex-specific chromosome (Wright et al. 2016).
46 Additional autosomes can further be added to the sex chromosome component through fusions to
47 either sex chromosome, and become similarly differentiated. Direct support for a shared
48 autosomal origin of the X- and Y-chromosomes (and Z- and W-chromosomes) is found in several
49 vertebrate groups, where Y/W-linked genes are largely homologous to X/Z-linked ones
50 (Livernois et al. 2012). This is lacking in other clades, such as *Drosophila* or Lepidoptera (moths
51 and butterflies), in which no consistent homology is detected between X/Z and Y/W-linked genes
52 (Carvalho 2002; Traut et al. 2013).

53 All butterflies and most moths, including the model silkworm (*Bombyx mori*, Bombycidae) and
54 all species for which there is extensive linkage information, belong to the division Ditrysia. The
55 vast majority of what is known about Lepidoptera sex chromosome evolution concerns only this
56 group, which shares a homologous pair of ZW sex chromosomes. The gene content of the Z-
57 chromosome is well-conserved even between distant species within the Ditrysia (Yasukochi et al.
58 2009), although large-scale chromosomal Z-autosomal rearrangements have occurred in some
59 lineages (Tortricidae: Nguyen et al. 2013; Papilionidae: Nishikawa et al. 2015; Nymphalidae:
60 Mongue et al. 2016). The history of the W-chromosome is hard to assess even within Ditrysia, as
61 the absence of protein coding genes on this chromosome (Traut et al. 2013) makes comparisons
62 between different species difficult, and prevent a direct test of homology with the Z-

63 chromosome. However, extensive cytogenetic analyses have shown that both the gain and loss of
64 W-chromosomes are frequent in Lepidoptera, with closely related species often showing
65 differences in their sex chromosome number (Traut et al. 2007).

66 Since Lepidoptera and Trichoptera are sister orders and the two only major insect lineages with
67 female-heterogamety, this system is thought to have originated in their common ancestor.
68 Interestingly, both Trichoptera and at least one species of Micropterigidae, the most basal lineage
69 of Lepidoptera, are Z0: females only carry one Z-chromosome, and lack a W-chromosome. In
70 Ditrysia, W-chromosomes can usually be detected in interphase spreads as female-specific
71 heterochromatin bodies. The absence of such sex chromatin bodies in other independent basal
72 lineages (Lukhtanov 2000) has led to the suggestion that the Z0 system is ancestral, and that the
73 W-chromosome was acquired secondarily in the ancestor of all Ditrysia and their close outgroup
74 Tischerioidea (Traut and Marec 1996; Traut 1999; Lukhtanov 2000; Sahara et al. 2012).

75 Several models could account for such a secondary acquisition of the W (Figure 1), of which two
76 have been favoured (Lukhtanov 2000; Sahara et al. 2012): (i) A Z-autosome fusion; (ii) The
77 recruitment of a B-chromosome (a dispensable chromosome found in only some individuals of a
78 population) to perform female-specific functions. While such a recruitment of B-chromosomes
79 for sex determination has been proposed for several systems (e.g., *Drosophila*: Hackstein et al.
80 1996; Homoptera: Kuznetsova and Nokkala 1997), it has not been empirically demonstrated. A
81 third model that has so far not been considered is sex chromosome turnover: if the ancestral Z
82 reverted to an autosome and a new sex chromosome pair took over sex determination, this would
83 also result in a neo W-chromosome. Such turnover of ancestral sex chromosomes is unusual, but
84 was recently found to have occurred repeatedly in flies and mosquitoes (Vicoso and Bachtrog
85 2015). In the absence of W-linked genes, testing these different models requires detailed
86 comparisons of the gene content of the Z-chromosome of basal and derived Lepidoptera,
87 something that is still lacking.

88 To investigate this, we sequenced the genomes of three basal Lepidoptera species: *Cameraria*
89 *ohridella* (the horse chestnut moth, family Gracillariidae) is a basal Ditrysonian lineage that has a
90 ZW karyotype (De Prins et al. 2002), while *Nemophora degeerella* (a fairy moth, family

91 Adelidae) and *Triodia sylvina* (the orange swift, family Hepialidae) represent two independent
92 outgroup lineages that are presumed to be lacking a W due to the absence of sex chromatin. We
93 compared male and female genomic coverage to detect Z-derived scaffolds, and showed that the
94 Z-chromosome of the different species is homologous to the Z-chromosome of the model *B.*
95 *mori* (International Silkworm Genome Consortium 2008), and that *C. ohridella* has further
96 acquired a neo-Z chromosome (Chr. 18 of *B. mori*). In order to test for a Z-autosome fusion at
97 the base of the Ditrysia, which could account for the appearance of the W, we investigated gene
98 movement onto and off the Z. We found that very few genes moved onto the Z at the basal
99 branch of the Ditrysia, suggesting that no such fusion occurred. Finally, we obtained further
100 evidence for absence of a W-chromosome in Adelidae by using PCR to exclude female-
101 specificity of *N. degeerella* scaffolds that only have genomic coverage in female samples. These
102 findings shed light on the evolution of the Lepidoptera Z-chromosome, support the view that
103 well-differentiated sex chromosomes are generally extremely stable, and allow us to test several
104 theories that have been proposed for the origin of the W-chromosome.

105 **Results**

106 ***De novo* genome assemblies and genomic coverage**

107 We sequenced the genome of a basal Ditrysian lineage, *C. ohridella*, and two outgroup lineages
108 which are presumed to lack a W chromosome, *N. degeerella* and *T. sylvina*. Scaffolds were
109 assembled using a multi-step procedure combining SOAPdenovo2, GapCloser, SSPACE and
110 Cap3. Assembly results for each species are summarized in Supplementary Table 2. The final
111 assembly consisted of 152,025 *C. ohridella* scaffolds (N50=6,277 bps), 194,957 *N. degeerella*
112 scaffolds (N50=5,350 bps) and 2,353,920 *T. sylvina* scaffolds (N50=1,184 bps). In each species,
113 male and female DNA reads were then mapped separately to the assemblies using Bowtie2, and
114 genomic coverage was estimated with SOAPcoverage. Median coverages were similar in
115 females and males (*C. ohridella*: 61 and 60, *N. degeerella*: 50 and 46, *T. sylvina*: 9.3 and 14.2,
116 respectively), but they were generally lower in *T. sylvina* as this species has a large genome (total
117 assembly length=1.8 Gbps, Supplementary Table 2).

118 **Identification of Z-linked and autosomal genes**

119 Since genomic coverage is proportional to the number of copies of a gene in the genome, we
120 compared the Log₂ of the female over male (F:M) coverage to identify Z-linked and autosomal
121 scaffolds (Supplementary Table 3 and Supplementary Figure 1; Vicoso and Bachtrog 2011). In
122 order to compare the gene content of the Z in the different species, we further mapped *B. mori*
123 genes to the genome assemblies, and used a reciprocal best hit approach to detect 1:1 orthologs
124 for each species (*C. ohridella*: 9,123, *N. degeerella*: 8,361, *T. sylvina*: 5,053). The proportion of
125 Z-linked genes (genes on scaffolds with Log₂(F:M coverage) below 0.6x the median) was
126 similar between *B. mori* (3.71%) and *N. degeerella* (3.72%, p=0.213 with a two-tailed Chi-
127 squared test; Supplementary Table 6) and *T. sylvina* (3.77%, p=0.061), but higher for *C.*
128 *ohridella* (5.32%, p<0.0001). Using more stringent coverage limits (Supplementary Table 3 and
129 Supplementary Figure 2), the number of unclassified genes increased from 0.2% to 7.9% in *C.*
130 *ohridella*, from 0.6% to 14% in *N. degeerella*, and from 5.6% to 11.1% in *T. sylvina*. While the
131 proportion of Z-linked genes remained comparable in *C. ohridella* (5.6%) and *N. degeerella*
132 (4%), it dropped down in *T. sylvina* (1.2%) where increased noise due to lower coverage and a
133 fragmented assembly led many putative Z-linked scaffolds to remain unclassified.

134 Conservation of the Z in Lepidoptera and Trichoptera

135 Scaffolds were assigned to one of the *B. mori* chromosome based on their gene content (using the
136 1:1 orthologs described in the previous section; Supplementary Table 3). We could then evaluate
137 the coverage patterns for scaffolds homologous to the different *B. mori* chromosomes, to test
138 whether the Z-chromosome of Ditrysian (*B. mori* and *C. ohridella*) and non-Ditrysian (*N.*
139 *degeerella* and *T. sylvina*) lineages of Lepidoptera are homologous (Figure 2). In *N. degeerella*
140 and *T. sylvina*, scaffolds mapping to Chr. 1 (*i.e.* Chr. Z) of *B. mori* had a significantly lower
141 female to male coverage ratio than scaffolds mapping to any autosomes of *B. mori* ($p=0$ for all
142 comparisons, with a one-tailed Wilcoxon rank-sum test), confirming the homology of the *non-*
143 Ditrysian and Ditrysian Z-chromosomes. The same was found in *C. ohridella*, with the exception
144 of Chr. 18, which showed patterns of coverage similar to Chr. 1 ($p=0.423$), showing that it was
145 additionally acquired as a neo-Z chromosome (Figure 2A).

146 The proportion of genes classified as Z-linked that also mapped to Chr. 1 was 84.3% in *C.*
147 *ohridella*, 85.4% in *N. degeerella*, and 70.6% in *T. sylvina*. These conserved Z-linked genes were
148 found throughout Chr. 1 (Figure 3), rather than being restricted to a fraction of the chromosome,
149 as might be expected for a secondary acquisition of the W in Ditryisia through a Z-autosome
150 fusion (although such a signal would depend on the conservation of micro-synteny along the
151 chromosome).

152 We extended the analysis to published female DNA-seq data from a Trichoptera (*L. lunatus*:
153 6,646 one-to-one orthologs). In *L. lunatus*, scaffolds mapping to Chr. 1 had a significantly lower
154 female coverage than scaffolds mapping to any autosomes ($p=0$ for all comparisons, with a one-
155 tailed Wilcoxon rank-sum test, Supplementary Figure 3), which further supports a shared Z-
156 chromosome between Lepidoptera and its sister lineage, Trichoptera.

157

158 Gene movement onto and off the Z-chromosome

159 To further investigate whether a Z-autosome fusion occurred in the ancestor of the Ditryisia and
160 Tischerioidea, we combined our classification of Z-linked and autosomal genes in the four
161 Lepidoptera species and assessed gene movement onto and off the Z across the Lepidoptera
162 phylogeny (Figure 4). Such a fusion should be detectable through a large number of genes
163 moving onto the Z at the root of the Ditryisia (genes that are Z-linked in the two Ditryisia species,
164 but not in the outgroups *T. sylvina* and *N. degeerella*). Figure 2 shows that the gene content of the

165 Z-chromosome is broadly conserved across Lepidoptera: of the 77 genes in the sample that are
166 Z-linked in *B. mori*, 62 were present on the ancestral Z-chromosome, and 68 were already Z-
167 linked before the split of Ditrysia and *N. degeerella*. Importantly, only 4 out of 77 genes moved
168 onto the Z at the basal branch of the Ditrysia ($p=0.382$ when this rate is compared to the other
169 lineages (excluding the *C. ohridella* branch) with a two-tailed Chi-squared test; Supplementary
170 Table 7). Since 77 out of 654 (12%) *B. mori* Z-linked genes are represented in the sample, we
171 can rescale the tree, and approximate the total number genes that moved onto the Z at the root of
172 Ditrysia as 34 (the same calculation from the stringent classification yields 35 genes). Even
173 ignoring the fact that some of these will necessarily correspond to individual gene movement
174 onto the Z, 34 is smaller than the number of genes detected on any chromosome of the butterfly
175 *Melitaea cinxia* (Sup. Table. 28 of Ahola et al. 2014), which is thought to represent the ancestral
176 Ditrysonian karyotype of $N=31$ (Ahola et al. 2014). Any fusion would therefore have to involve a
177 chromosome smaller than any described so far in Ditrysia. On the other hand, we were able to
178 detect a strong excess of genes moving onto the Z in the *C. ohridella* lineage (87 genes,
179 $p<0.0001$), consistent with the acquisition of a neo-Z chromosome through a Z-autosome fusion
180 in this lineage. All our conclusions hold when using the stringent classification (Figure 4B).

181 **Genomic support for the absence of W-chromosome in *N. degeerella***

182 Since the W-chromosome is present only in females, we identified candidate W-linked sequences
183 by selecting genomic scaffolds with female-specific read coverage in each species (42 scaffolds
184 in *C. ohridella*, 42 in *N. degeerella* and 2464 in *T. sylvina*). We tested the female-specificity of a
185 subset of these W-candidates by PCR amplification in an additional three male and three female
186 individuals (Figure 5). In the ZW species *C. ohridella*, 70% of all W-candidates yielded female-
187 specific bands (Figure 5 and Supplementary Figure 4). Contrary to this, none of the 42
188 candidates of *N. degeerella* were restricted to females, and corresponded instead to polymorphic
189 sequences found in some male and some female individuals (Figure 5 and Supplementary Figure
190 4). The same was found for *T. sylvina*, although in this case only 24 scaffolds were tested. In
191 each of the two non-Ditrysia species, one spurious female-specific candidate was invalidated
192 after further testing on three other males and females (Supplementary Figure 4). These results
193 generally support the absence of a W-chromosome in *N. degeerella* and *T. sylvina*, although for

194 the latter our evidence is not fully conclusive (as the poorer quality of the assembly makes an
195 exhaustive examination unfeasible).

196 **Discussion**

197 Our comparative genomic analysis demonstrates the conservation of the Z-chromosome over 140
198 Myr of Lepidoptera evolution (Misof et al. 2014), and further supports a shared ancestry of the
199 Lepidoptera and Trichoptera Z-chromosomes (Supplementary Figure S3). The Z-chromosome of
200 Lepidoptera has therefore likely been the sex-determining chromosome for 210 Myr (Misof et al.
201 2014), making it one of the oldest documented sex chromosome (Graves 2016), and supporting
202 to the view that well-differentiated sex chromosomes are generally extremely stable, as has been
203 reported in birds (Pokorná et al. 2011), lizards (Rovatsos et al. 2016), eutherian mammals
204 (Graves 2016), and true bugs (Pal and Vicoso 2015). The gene content of the Z-chromosome has
205 also remained remarkably stable: over 70% of the genes that are Z-linked in the basal *T. sylvina*
206 and 85% of *N. degeerella* Z-linked genes are also Z-linked in *B. mori*. Importantly, we detect no
207 evidence of chromosomal rearrangements involving the Z-chromosome, such as Z-autosome
208 fusions, between non-Ditrysians and Ditrysians (although such fusions have been detected in
209 several Ditrysian lineages, including *C. ohridella* in this study). This is consistent with a
210 previous analysis of karyotypes in Lepidoptera which did not detect a decrease of the male
211 autosome number by one pair at the base of the Ditrysian lineage (Lukhtanov 2000), as would be
212 expected if a Z-autosome fusion had occurred.

213 The deep conservation of the identity and gene content of the Z-chromosome is at odds with the
214 suggestion that either sex chromosome turnover or fusion of an autosome to the Z-chromosome
215 (the two “canonical models” in Figure 1) has led to the acquisition of the W-chromosome in the
216 ancestor of Ditrysia and Tischerioidea, as has been proposed (Lukhtanov 2000). This leaves two
217 possible evolutionary scenarios: 1. The W originated much earlier than currently believed, and is
218 present in some basal lineages (Voleníková 2015); 2. The W had a non-canonical origin, and does
219 not correspond to a degenerated autosome (Lukhtanov 2000). Much of the evidence for the lack
220 of W-chromosomes in basal lineages is indirect, and relies on the frequent formation of sex
221 chromatin bodies when a W-chromosome is present in Ditrysians; such chromatin bodies are
222 lacking in basal lineages. While absence of a W has been confirmed using cytogenetics in the
223 sister order Trichoptera and in Micropterigidae (the most basal Lepidoptera lineage), at least
224 some species of Hepialidae have been found to carry W-chromosomes (Kawazoé 1987;
225 Voleníková 2015, but the later found a ZZW karyotype, suggesting that secondary

226 rearrangements have occurred in this clade). This is further complicated by the fact that, even in
227 Ditrysians, the W is not always involved in sex determination (Yoshido et al. 2016), and has been
228 lost in several lineages (Saturniidae: Yoshido et al. 2006; review Traut et al. 2007). Our genomic
229 approach to detect W-derived scaffolds led to the discovery of many female-specific sequences
230 in the ZW *C. ohridella*, whereas no such female-specific sequences were found in either *N.*
231 *degeerella* or *T. sylvina* (but for the latter only a small subset was tested). These results provide
232 complementary evidence that a W is lacking in these two basal lineages, and suggest that the W
233 was independently acquired in some Hepialidae lineages and in the ancestor of Ditrysia and
234 Tischerioidea.

235 What could then be the origin of the Ditrysiian W-chromosome? The favored hypothesis has been
236 that a B-chromosome may have been recruited to perform female-specific functions, thereby
237 creating a new W-chromosome (Lukhtanov 2000). B-chromosomes are supernumerary
238 chromosomes that are found in many insect species (including several Lepidoptera) and are
239 generally thought to be primarily parasitic. Because of their biological similarity to Y/W
240 chromosomes (they are often highly repetitive and gene poor) and ability in some species to
241 segregate with the X at meiosis, they have been suggested as the source of sex-specific
242 chromosomes in other insects (Nokkala et al. 2003), including *Drosophila* (Carvalho (2002); but
243 in this case no representative X0 species is available for comparison). Another possibility is that
244 a sequence derived from a symbiont or parasite was domesticated for the purpose of sex-
245 determination. Many insects carry endosymbionts, such as *Wolbachia*, which can manipulate the
246 sex ratio of the progeny of infected individuals by interfering with the sex determination pathway
247 (feminizing the progeny gives them an advantage, as they are only transmitted maternally
248 (Leclercq et al. 2016)). The horizontal transfer of a sex-determining gene has recently been
249 reported in pillbugs, where the acquisition by an autosome of a *Wolbachia*-derived female-
250 determining gene has led to the formation of a neo-W-chromosome (Leclercq et al. 2016).
251 Although a W-chromosome fully derived from bacterial DNA has yet to be described, it seems
252 plausible in light of those results. Finally, it is possible that duplication of at least part of a
253 chromosome provided the source material for the evolution of the W-chromosome, although the
254 initial aneuploidy resulting from such a duplication would have been unlikely to stably remain in
255 the population.

256 While we cannot at this point distinguish between these hypotheses, our results argue against a
257 canonical model of W-chromosome evolution, in which a pair of autosomes is recruited to
258 perform sex-determination. Studies of other X0-XY and Z0-ZW pairs of species will shed light
259 on how common such non-canonical origin of sex-specific chromosomes is, and on which of the
260 different hypotheses play a role.

261 **Materials and Methods**

262 **Specimen collection, DNA extraction and sequencing**

263 Male and female individuals of *T. sylvina* and *N. degeerella* were collected on the IST Austria
264 campus between May 2015 and September 2016. *C. ohridella* males and females were collected
265 on horse chestnut trees in Vienna (Austria) in July and August 2016. All individuals were kept at
266 -80°C until DNA extraction. DNA was extracted from either a single male or female individual
267 (*T. sylvina*, *N. degeerella* and *C. ohridella*), or 10 pooled individuals (*C. ohridella*: one library
268 from 10 males and one library from 10 females). Specimens were lysed using the Tissue Lyser II
269 (QUIAGEN) and DNA was isolated using the DNeasy Blood & Tissue Kit (QUIAGEN). DNA
270 was then sheared by sonication with a Bioruptor instrument (Diagenode). Library preparation
271 and sequencing (Hi-seq 2500 Illumina, 125 bps paired-end reads) were performed at the Vienna
272 Biocenter Next Generation sequencing facility (Austria). Library information and statistics are
273 provided in Supplementary Table 1.

274 **Publicly available data**

275 We downloaded one additional DNA-seq library derived from a single adult (Ferguson et al.
276 2014) for *T. sylvina* (accession number: SRR1190481) and several libraries derived from a pool
277 of females for the Trichoptera species *Limnephilus lunatus* (accession numbers: SRR947083 to
278 SRR947088) from <https://www.ncbi.nlm.nih.gov/sra/>.

279 The genome assembly of *L. lunatus* (GCA_000648945.1_Llun_1.0_genomic.fna) was obtained
280 from <https://www.ncbi.nlm.nih.gov/assembly>.

281 ***De novo* genome assemblies**

282 Sequenced reads in BAM format were first converted in Fastq using Bedtools v2.25.0 (Quinlan
283 and Hall 2010) and their quality assessed with FastQC v0.11.2 (Andrews 2010). No further
284 quality filter neither trimming was deemed necessary, as the average Phred score was >25
285 throughout the reads of each library. The genome of each species was assembled from a single
286 paired-end DNA-seq library, as this yielded a more continuous assembly than pooling libraries
287 for each species. Specifically, each library was first individually assembled using SOAPdenovo2
288 (Luo et al. 2012) with default parameters, using a range of kmer lengths from 33 to 93. The least
289 fragmented assembly, as measured by N50 and number of scaffolds, was chosen for further

290 processing (*C. ohridella*: kmer=83, library #44577; *N. degeerella*: kmer=63, library #37850;
291 *T.sylvina*: kmer=63, library #44580). The chosen assembly was then improved using GapCloser
292 v1.12 (<http://soap.genomics.org.cn/index.html>) with default parameters. Scaffolds with a
293 minimum length of 200 bps were selected and reassembled with all libraries in a hierarchical
294 manner using SSPACE v3.0 (Boetzer et al. 2011) with no contig extension option and a
295 minimum of 48 overlapping bases. Finally, *N. degeerella* and *C. ohridella* assemblies were
296 improved using Cap3 v02.10.15 (Huang and Madan 1999) with default parameters. The *T.*
297 *sylvina* assembly was too large for this last step to run without exceeding the memory allowance
298 of our cluster, so it was not performed in this species. Genome assembly statistics are provided in
299 Supplementary Table 2.

300 **Read mapping and estimation of genomic coverage**

301 For all species, the male and female DNA reads were mapped separately to the *de novo*
302 assembled genomic scaffolds using Bowtie2 v2.2.9 (Langmead and Salzberg 2012) with default
303 parameters (--end-to-end --sensitive mode). The resulting alignments were filtered to keep only
304 uniquely mapped reads by selecting lines that did not match “XS:I”, as this is the Bowtie2 tag for
305 the score of the second best alignment of the read, and is not present for reads that only have one
306 match. The male and female coverages were then estimated from the filtered SAM files with
307 SOAPcoverage v2.7.7 (<http://soap.genomics.org.cn/index.html>). The coverage values for each
308 library are provided in Supplementary Data 1.

309 **Detection of 1:1 orthologs, and assignment of scaffolds to the chromosomes of *B. mori***

310 The gene set of *B. mori*, as well as the respective gene chromosomal locations, were obtained
311 from the Silkworm Genome Database (v2.0, <http://silkworm.genomics.org.cn>). For each
312 outgroup species, the *B. mori* gene set was mapped to the genome assembly using Blat (Kent
313 2002) with a translated query and dataset, and a minimum mapping score of 50; only the location
314 with the best score was kept for each gene. When several genes overlapped on the same scaffold
315 by more than 20 base pairs, only the gene that had the highest mapping score was kept for that
316 genomic location. Each scaffold was then assigned to the chromosome on which the majority of
317 *B. mori* genes that mapped to it were located. When a scaffold carried the same number of genes
318 from two or more chromosomes, the mapping scores of all genes from each were added, and the

319 scaffold was assigned to the chromosome that had the largest sum of mapping scores. The final
320 chromosomal assignments are provided in Supplementary Data 2.

321 **Scaffold classification as Z-linked or autosomal - lenient method**

322 For each species, the per-scaffold coverage was summed-up across samples grouped by sex.
323 Only scaffolds longer than 1,500 bps with coverage above 5 and below the 99.5th percentile in
324 the grouped male sample were considered for further analysis. The resulting Log₂ of the female
325 over male coverage (Log₂(F/M)) distribution was expected to be bimodal. The highest peak
326 corresponds to the autosomal mode, which we call “A_mode” and for which males and females
327 are expected to show equal coverage. The second highest peak corresponds to the Z-linked
328 mode. It should be centered around “A_mode – 1”, because Z-linked scaffolds are expected to
329 show a two-fold coverage reduction in female relative to male. We classified all scaffolds having
330 a Log₂(F/M) value less than “A_mode – 0.6” to be Z-linked, and more than “A_mode – 0.4” to
331 be autosomal, based on a *n*-class histogram (*n*=200 for *C. ohridella*; *n*=100 for *N. degeerella* and
332 *T. sylvina*). Scaffolds falling in-between were tagged as “unclassified”. The scaffold
333 classification is provided in Supplementary Table 3 and the R code used is provided in
334 Supplementary Text 1.

335 **Scaffold classification as Z-linked or autosomal - stringent method**

336 We took advantage of the homology of the outgroup and *B. mori* Z-chromosomes to empirically
337 determine stricter coverage cutoffs for Z-linked classification. We reasoned that, as the cut-off
338 chosen for Z-assignment becomes more stringent, the percentage of scaffolds classified as Z-
339 linked that are on the Z-chromosome of *B. mori* (Chr. 1) should increase, but plateau once
340 (almost) only true Z-linked scaffolds are being selected (this final percentage should roughly
341 correspond to the real proportion of conserved genes between the two Z-chromosomes;
342 Supplementary Figure S2). Under this stringent classification, scaffolds were classified as Z-
343 linked if:

- 344 1. they fell below the Log₂(F/M) value at which the proportion of the resulting Z-
345 classified scaffolds mapping to *B. mori* Z-chromosome has reached a plateau, (*C.*
346 *ohridella*: -0.69; *N. degeerella*: -0.57; *T. sylvina*: -1.42);

347 2. they had a female coverage below the female median (as the Z-chromosome is only
348 present in one copy in females).

349 Scaffolds were classified as autosomal if they fell above the Log₂(F/M) corresponding to the
350 10th percentile of the Log₂ of the female over male coverage of all scaffolds mapping to *B. mori*
351 autosomes. Scaffolds falling between the Z-linked and autosomal limits were tagged as
352 “unclassified”. The stringent classification is provided in Supplementary Table 3 and the R code
353 used is provided in Supplementary Text 2.

354 **Detection of gene movement onto and off the Z-chromosome**

355 We combined the Z/autosome classification from the four Lepidoptera species to determine
356 which genes were ancestrally present on the Z-chromosome, and to detect gene movement along
357 the phylogeny (using both the lenient and stringent classifications). The 62 genes that were
358 deemed to be on the ancestral Lepidoptera Z were Z-linked at least in *T. sylvina* and another
359 lineage. These ancestrally Z-linked genes, plus the eight which were Z-linked in *N. degeerella*
360 and in the two Ditrysia species were classified as ancestral to Ditrysia+*N. degeerella*. Finally,
361 genes that were Z-linked only in the two Ditrysia were classified as having moved onto the Z at
362 the root of the Ditrysia, and genes that were Z-linked in only one species (excluding *T. sylvina*)
363 as having moved in the species-specific lineages. Similar arguments apply for movement off the
364 Z, replacing “Z-linked” with autosomal.

365 **In silico identification of W-candidates and PCR testing of female-specificity**

366 W-derived sequences are expected to show no or very low coverage in male samples, and
367 reduced coverage in female samples relative to the autosomal scaffolds (as W-chromosome is
368 only present in one copy in females). Scaffolds larger than 1,000bps that had coverage < 1.1 in
369 each male sample, and above ¼ of the median but below the median in each female sample, were
370 therefore designated as candidate W-derived sequences (42 W-candidates in *C. ohridella* (of
371 which 10 were tested), 42 in *N. degeerella*, and 2,464 in *T. sylvina*). Since many of the *T. sylvina*
372 candidates are likely to result from the lower genomic coverage and fragmented genome
373 assembly, and from the fact that we used a single female, we applied more stringent filters (a
374 minimal scaffold size of 2,000 bps, and null coverage in each male sample) to select the most
375 promising 24 candidates for PCR testing. Fasta sequences of all candidates are provided in

376 Supplementary Data 3 and the R code used to identify W-candidates is provided in
377 Supplementary Text 3.

378 In order to test if these candidates corresponded to female-specific sequences, primers were
379 designed for each, as well as for two autosomal control genes (BGIBMGA013945-TA actin and
380 BGIBMGA009132-TA beta-tubulin *B. mori* orthologs) using Primer 3 plus (Untergasser et al.
381 2012) with default parameters. Primer specificity was assessed by independently mapping each
382 primer against its respective genome assembly using Blat (minimum sequence identity of 100%,
383 tile size of 16, step size of 1, and minimum match of 1). A pair of primer was considered specific
384 if the forward and the reverse primers mapped only once to the target scaffold and never
385 simultaneously to other scaffolds. Primer sequences and expected length of PCR products are
386 provided in Supplementary Table 4.

387 For each species, DNA was extracted from three new males and three new females using the
388 DNeasy Blood & Tissue Kit (QIAGEN). The presence of the sequences was tested in the six
389 individuals using the GoTaq G2 polymerase PCR kit (Promega, Cat#M7841) with an annealing
390 temperature of 52 °C, an elongation time of 1 min, for 35 cycles. Finally, the amplification and
391 size of each product was assessed by electrophoresis on a 1.5% agarose gel, stained with
392 SYBRsafe (ThermoFisher Scientific).

393 **Sex identification of the publicly available *T. sylvina* individual**

394 The published DNA reads were mapped to our *de novo* *T. sylvina* genome assembly and
395 coverage was estimated following the same procedure as for the other samples. The resulting
396 coverage values were then compared between genes classified as Z-linked and genes classified
397 as autosomal in *T. sylvina*. This initial classification, which is provided in Supplementary Table
398 5, was made by only using data from the male and the female we collected, before adding the
399 extra individual. Following the lenient method, we applied an upper Z-linked cut-off value of
400 “A_mode – 0.6” and a lower autosomal cut-off value of “A_mode – 0.4”. The median coverage
401 value of Z-linked scaffolds was not significantly different from that of autosomal scaffolds
402 (median_Z=4.6; median_A=4.6; p=0.383 with a one-tailed Wilcoxon rank-sum test), as expected for
403 a male having two copies of the Z-chromosome. Therefore, we treated the extra individual as a
404 second *T. sylvina* male in our analyses. The R code used is provided in Supplementary Text 4.

405 **Data availability**

406 All reads generated for this study have been deposited on the NCBI Sequence Read Archive
407 (SRA) under BioProject number XXX. The *de novo* genome assemblies generated for this study
408 have been deposited on the NCBI GenBank database, under BioProject number XXX.

409 **Code availability**

410 The full bioinformatic pipeline used in this work is provided in Supplementary Text 5.

411 **Acknowledgements**

412 The authors thank Ariana Macon for DNA extractions and preparation for sequencing, as well as
413 the Vicoso lab and Dominik Schrempf and Chay Graham for taking part in moth collections.
414 This project was funded by an Austrian Science Foundation FWF grant (Project P 28842) to B.V.

415 **Author contributions**

416 Experiments and data analysis: C. F., M. P. and B. V.
417 Writing: C. F., M. P. and B. V.
418 Conceptualization: C. F., M. P. and B. V.
419 Funding acquisition: B. V.

420 **References**

- 421 Ahola V, Lehtonen R, Somervuo P, Salmela L, Koskinen P, Rastas P, Välimäki N, Paulin L, Kvist
422 J, Wahlberg N, et al. 2014. The Glanville fritillary genome retains an ancient karyotype and
423 reveals selective chromosomal fusions in Lepidoptera. *Nat Commun* **5**: 4737.
- 424 Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data.
425 <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (Accessed December 2016).
- 426 Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled
427 contigs using SSPACE. *Bioinformatics* **27**: 578–579.
- 428 Carvalho AB. 2002. Origin and evolution of the Drosophila Y chromosome. *Curr Opin Genet*
429 *Dev* **12**: 664–668.
- 430 International Silkworm Genome Consortium. 2008. The genome of a lepidopteran model insect,
431 the silkworm *Bombyx mori*. *Insect Biochem Mol Biol* **38**: 1036–1045.
- 432 De Prins J, De Prins W, Dall’Asta U. 2002. The karyotype of *Cameraria ohridella* (Lepidoptera:
433 Gracillariidae). *Vlaamse Vereniging Voor Entomologie* **30**: 5–10.

- 434 Ferguson L, Marlétaz F, Carter J-M, Taylor WR, Gibbs M, Breuker CJ, Holland PWH. 2014.
435 Ancient expansion of the hox cluster in Lepidoptera generated four homeobox genes
436 implicated in extra-embryonic tissue formation. *PLoS Genet* 10: e1004698.
- 437 Graves JA. 2008. Weird animal genomes and the evolution of vertebrate sex and sex
438 chromosomes. *Annu Rev Genet* 42: 565–586.
- 439 Graves JAM. 2016. Evolution of vertebrate sex chromosomes and dosage compensation. *Nat*
440 *Rev Genet* 17: 33–46.
- 441 Hackstein JHP, Hochstenbach R, Hauschteck-Jungen E, Beukeboom LW. 1996. Is the Y
442 chromosome of *Drosophila* an evolved supernumerary chromosome? *Bioessays* 18: 317–
443 323.
- 444 Huang X, Madan A. 1999. CAP3: A DNA sequence assembly program. *Genome Res* 9: 868–877.
- 445 Kawazoé A. 1987. The chromosome in the primitive or Microlepidopterous moth-groups. I. Proc
446 *Jpn Acad Ser B Phys Biol Sci* 63: 25–28.
- 447 Kent WJ. 2002. BLAT--the BLAST-like alignment tool. *Genome Res* 12: 656–664.
- 448 Kuznetsova VG, Nokkala S. 1997. Karyotypes, sex chromosome systems, and male meiosis in
449 Finnish psyllids (Homoptera: Psylloidea). *Folia Biol* 45.
- 450 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:
451 357–359.
- 452 Leclercq S, Thézé J, Chebbi MA, Giraud I, Moumen B, Ernenwein L, Grève P, Gilbert C,
453 Cordaux R. 2016. Birth of a W sex chromosome by horizontal transfer of *Wolbachia*
454 bacterial symbiont genome. *Proceedings of the National Academy of Sciences* 113: 15036–
455 15041.
- 456 Livernois AM, Graves JAM, Waters PD. 2012. The origin and evolution of vertebrate sex
457 chromosomes and dosage compensation. *Heredity* 108: 50–58.
- 458 Lukhtanov VA. 2000. Sex chromatin and sex chromosome systems in nonditrysian Lepidoptera

- 459 (Insecta). *J Zoolog Syst Evol Res* **38**: 73–79.
- 460 Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, et al. 2012.
461 SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler.
462 *Gigascience* **1**: 18.
- 463 Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T,
464 Beutel RG, et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution.
465 *Science* **346**: 763–767.
- 466 Mongue AJ, Nguyen P, Volenikova A, Walters JR. 2016. Neo-sex chromosomes in the Monarch
467 butterfly, *Danaus plexippus*. <http://dx.doi.org/10.1101/036483>.
- 468 Nguyen P, Sýkorová M, Šíchová J, Kůta V, Dalíková M, Čapková Frydrychová R, Neven LG,
469 Sahara K, Marec F. 2013. Neo-sex chromosomes and adaptive potential in tortricid pests.
470 *Proc Natl Acad Sci U S A* **110**: 6931–6936.
- 471 Nishikawa H, Iijima T, Kajitani R, Yamaguchi J, Ando T, Suzuki Y, Sugano S, Fujiyama A,
472 Kosugi S, Hirakawa H, et al. 2015. A genetic mechanism for female-limited Batesian
473 mimicry in *Papilio* butterfly. *Nat Genet* **47**: 405–409.
- 474 Nokkala S, Grozeva S, Kuznetsova V, Maryanska-Nadachowska A. 2003. The origin of the
475 achiasmatic XY sex chromosome system in *Cacopsylla peregrina* (Frst.) (Psylloidea,
476 Homoptera). *Genetica* **119**: 327–332.
- 477 Pal A, Vicoso B. 2015. The X Chromosome of Hemipteran Insects: Conservation, Dosage
478 Compensation and Sex-Biased Expression. *Genome Biol Evol* **7**: 3259–3268.
- 479 Pokorná M, Giovannotti M, Kratochvíl L, Kasai F, Trifonov VA, O'Brien PCM, Caputo V, Olmo
480 E, Ferguson-Smith MA, Rens W. 2011. Strong conservation of the bird Z chromosome in
481 reptilian genomes is revealed by comparative painting despite 275 million years divergence.
482 *Chromosoma* **120**: 455–468.
- 483 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic
484 features. *Bioinformatics* **26**: 841–842.

485 Regier JC, Mitter C, Zwick A, Bazinet AL, Cummings MP, Kawahara AY, Sohn J-C, Zwickl DJ,
486 Cho S, Davis DR, et al. 2013. A large-scale, higher-level, molecular phylogenetic study of
487 the insect order Lepidoptera (moths and butterflies). *PLoS One* **8**: e58568.

488 Rovatsos M, Vukić J, Altmanová M, Johnson Pokorná M, Moravec J, Kratochvíl L. 2016.
489 Conservation of sex chromosomes in lacertid lizards. *Mol Ecol* **25**: 3120–3126.

490 Sahara K, Yoshido A, Traut W. 2012. Sex chromosome evolution in moths and butterflies.
491 *Chromosome Res* **20**: 83–94.

492 Traut W-E. 1999. The evolution of sex chromosomes in insects: Differentiation of sex
493 chromosomes in flies and moths. *Eur J Entomol* **96**: 227–235.

494 Traut W, Marec F. 1996. Sex chromatin in Lepidoptera. *Q Rev Biol* **71**: 239–256.

495 Traut W, Sahara K, Marec F. 2007. Sex chromosomes and sex determination in Lepidoptera. *Sex*
496 *Dev* **1**: 332–346.

497 Traut W, Vogel H, Glöckner G, Hartmann E, Heckel DG. 2013. High-throughput sequencing of a
498 single chromosome: a moth W chromosome. *Chromosome Res* **21**: 491–505.

499 Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012.
500 Primer3—new capabilities and interfaces. *Nucleic Acids Res* **40**: e115–e115.

501 Vicoso B, Bachtrog D. 2011. Lack of global dosage compensation in *Schistosoma mansoni*, a
502 female-heterogametic parasite. *Genome Biol Evol* **3**: 230–235.

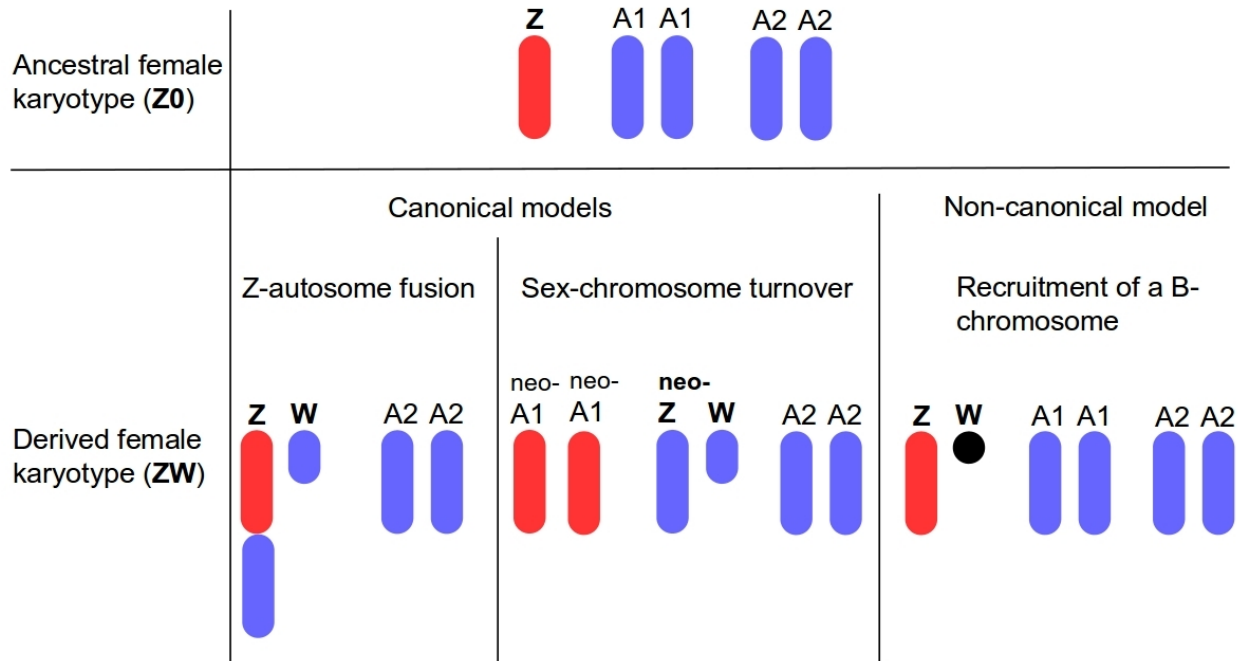
503 Vicoso B, Bachtrog D. 2015. Numerous transitions of sex chromosomes in Diptera. *PLoS Biol*
504 **13**: e1002078.

505 Voleniková A. 2015. Karyotype and sex chromosomes analysis of two species from basal
506 lepidopteran family Hepialidae (Lepidoptera: Hepialidae). ed. P. Nguyen., Magister,
507 University of South Bohemia.

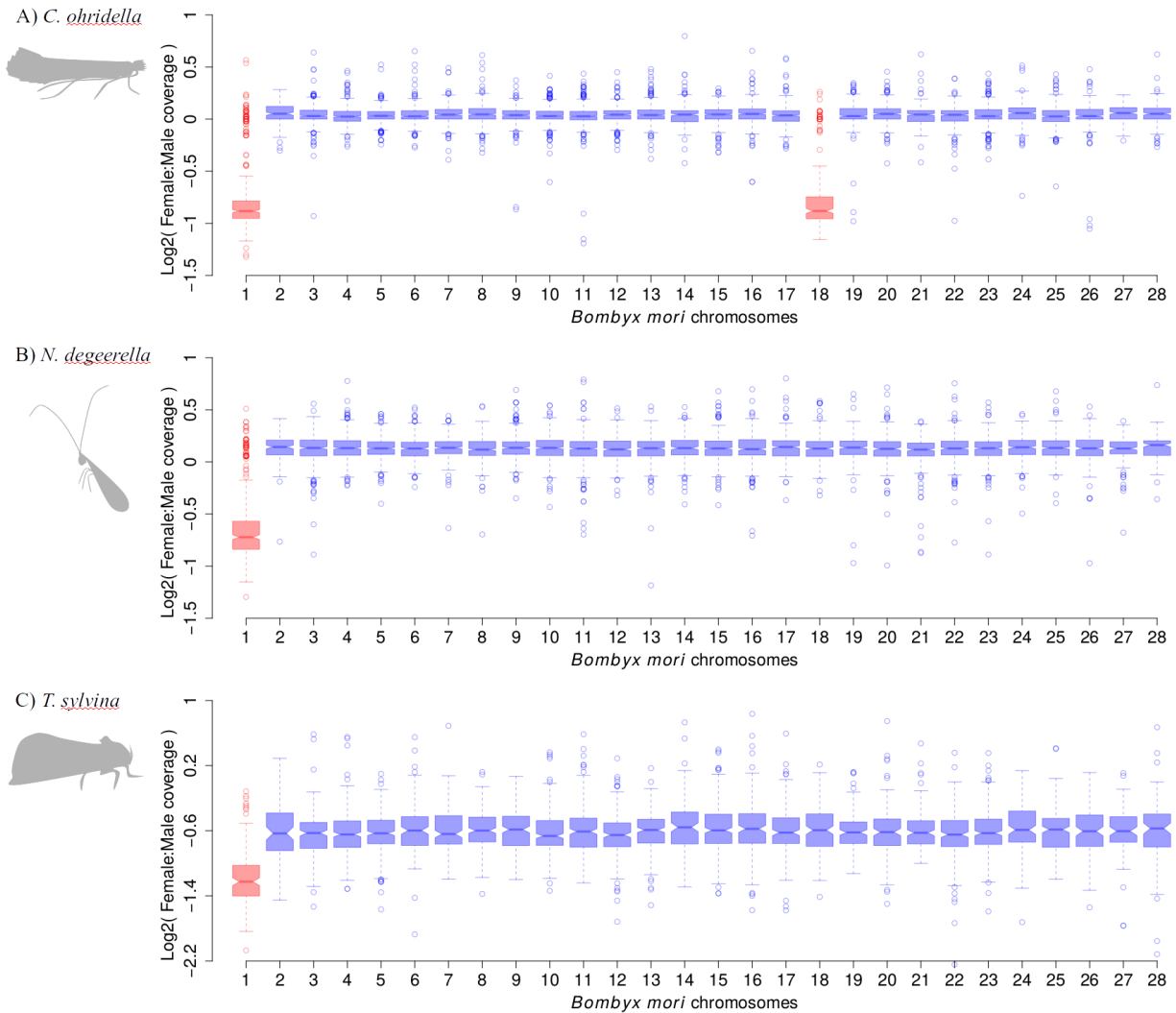
508 Wright AE, Dean R, Zimmer F, Mank JE. 2016. How to make a sex chromosome. *Nat Commun*
509 **7**: 12087.

- 510 Yasukochi Y, Tanaka-Okuyama M, Shibata F, Yoshido A, Marec F, Wu C, Zhang H, Goldsmith
511 MR, Sahara K. 2009. Extensive conserved synteny of genes between the karyotypes of
512 *Manduca sexta* and *Bombyx mori* revealed by BAC-FISH mapping. *PLoS One* **4**: e7465.
- 513 Yoshido A, Marec F, Sahara K. 2016. The fate of W chromosomes in hybrids between wild
514 silkmoths, *Samia cynthia* ssp.: no role in sex determination and reproduction. *Heredity* **116**:
515 424–433.
- 516 Yoshido A, Yamada Y, Sahara K. 2006. The W chromosome detection in several lepidopteran
517 species by genomic in situ hybridization (GISH). *J Insect Biotechnol Sericology* **75**: 147–
518 151.

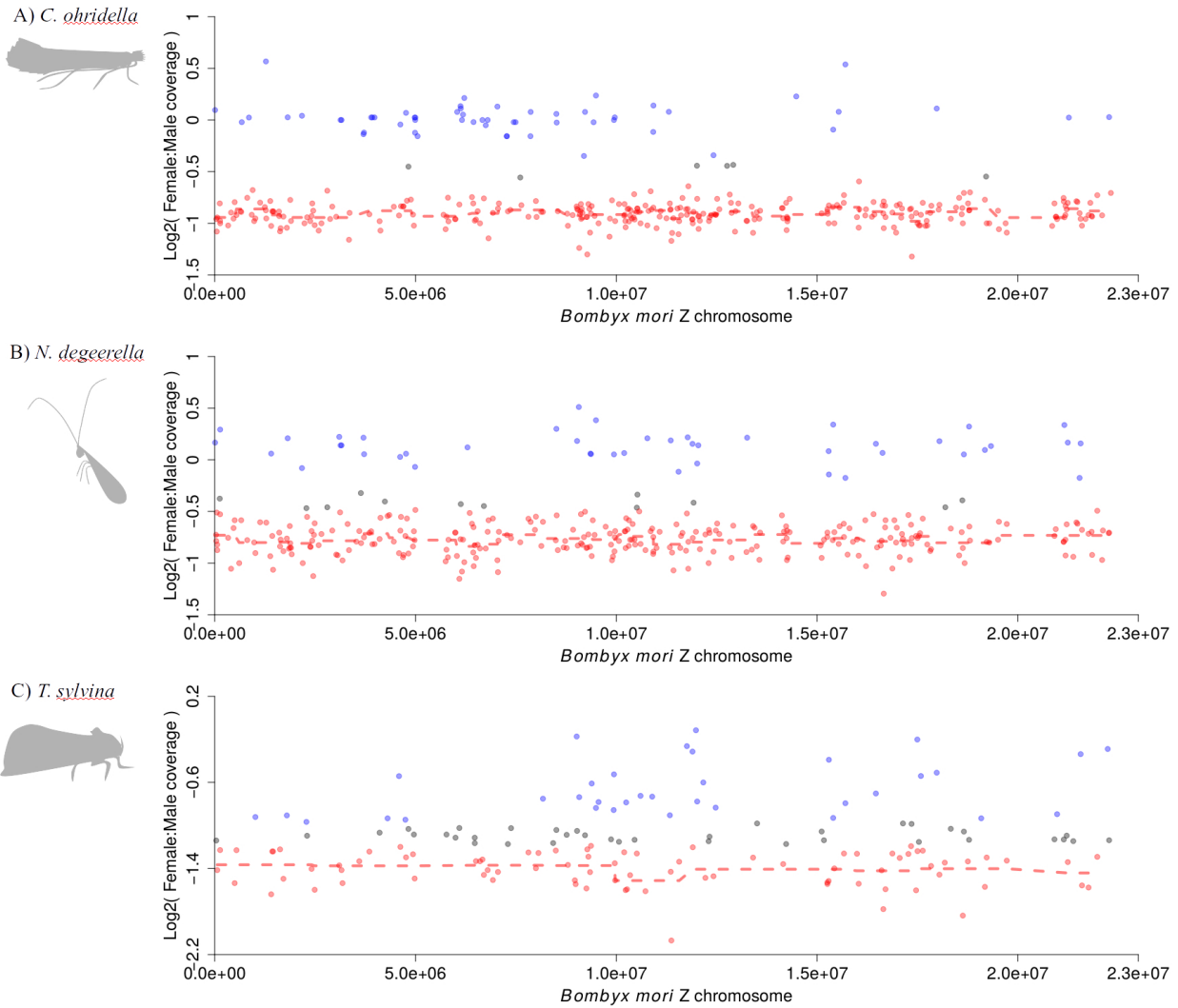
519 **Figures:**



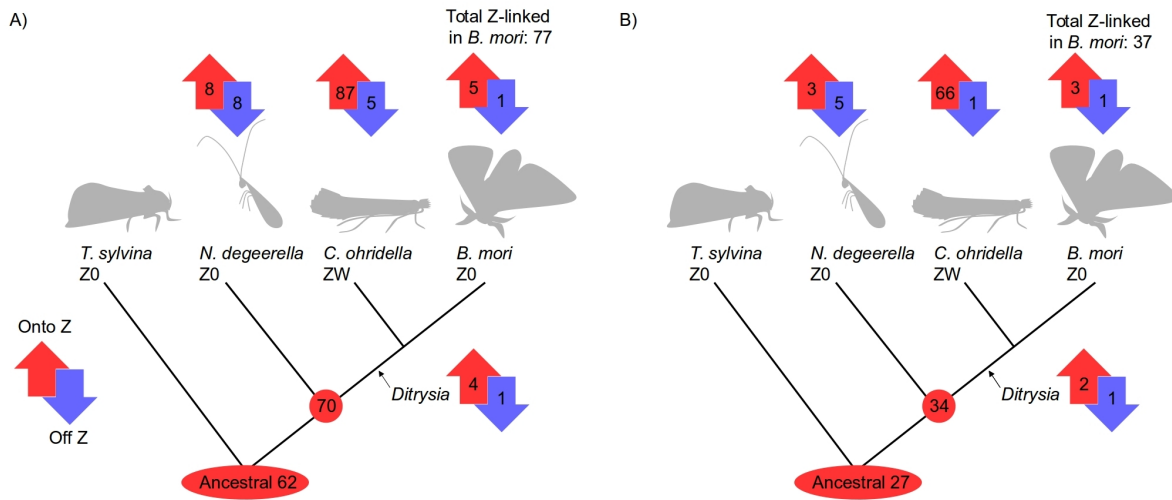
520 **Figure 1.** Hypotheses for the secondary acquisition of the W-chromosome in Lepidoptera.
 521 Ancestrally, females are assumed to lack the W-chromosome. A ZO female karyotype is
 522 represented by the ancestral Z-chromosome (in red) and two pairs of ancestral autosomes (in
 523 blue). In the two canonical models (Z-autosome fusion and sex chromosome turnover), the sex
 524 chromosomes derived from a standard pair of autosomes. In the non-canonical model, the W-
 525 chromosome has a non-autosomal origin, e.g. it involved the recruitment of a B-chromosome.



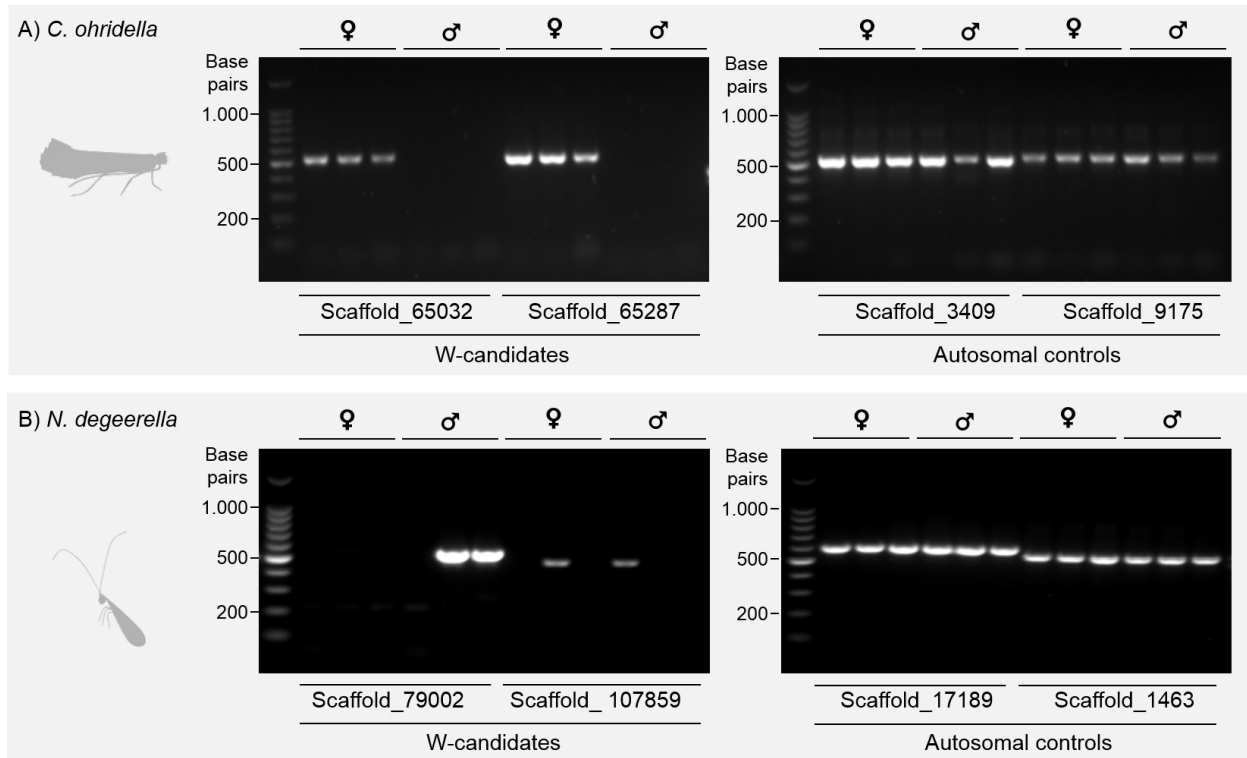
526 **Figure 2.** The Z-chromosome of *B. mori* (Chr. 1) is homologous to the Z-chromosome of basal
 527 Lepidoptera. *C. ohridella* has additionally acquired a neo-Z chromosome (Chr. 18). For each
 528 species, scaffolds were assigned to one of the *B. mori* chromosomes based on their gene content.
 529 The Log₂ of the female to male coverage ratio of each chromosome is shown for A. *C. ohridella*.
 530 B. *N. degeerella*. C. *T. sylvina*. Scaffolds mapping to the *B. mori* Z-chromosomes are in red,
 531 those mapping to the autosomes are in blue.



532 **Figure 3.** Full-length homology between the Z-chromosome of *B. mori* and basal Lepidoptera.
 533 For each species, scaffolds mapping to the Z-chromosome of *B. mori* (Chr. 1) have been
 534 classified as Z-linked (in red), autosomal (in blue), or unclassified (in grey). The Log2 of the
 535 female to male coverage ratio along the *B. mori* Z-chromosome (Chr. 1) is shown for A. *C.*
 536 *ohridella*. B. *N. degeerella*. C. *T. sylvina*. Dashed lines are moving average based on sliding
 537 windows of 10 scaffolds.



538 **Figure 4.** Gene movement onto and off the Z-chromosome does not support a Z-autosome fusion
 539 at the base of the Ditrysia. A. Lenient classification. B. Stringent classification. The phylogenetic
 540 tree is adapted from Regier *et al.* (2013). Red circles show the number of genes present on the
 541 ancestral Z-chromosome. Arrows indicate the number of genes which move onto (up red arrows)
 542 or off (down blue arrows) the Z-chromosome in each lineage after the split with *T. sylvina*.



543 **Figure 5.** Presence of a W-chromosome in *C. ohridella* (A) but not in *N. degeerella* (B).
 544 Amplification patterns of genomic DNA from three females and three males are shown for two
 545 W-candidates, and for two autosomal control genes in each species. The molecular size marker is
 546 indicated on the left (100 bp ladder). Results for all W-candidates are shown in Supplementary
 547 Figure S4.