

1 **Whole-chromosome hitchhiking driven by a male-killing endosymbiont**

2 Simon H. Martin*^{1,2}, Kumar Saurabh Singh³, Ian J. Gordon⁴, Kennedy Saitoti Omufwoko^{5,6},
3 Steve Collins⁷, Ian A Warren¹, Hannah Munby¹, Oskar Brattström¹, Walther Traut⁸, Dino J.
4 Martins^{5,6}, David A. S. Smith⁹, Chris D. Jiggins¹, Chris Bass³, Richard H. ffrench-Constant³

5 ¹Department of Zoology, University of Cambridge, Cambridge, CB2 3EJ, United Kingdom

6 ²Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3FL, United
7 Kingdom

8 ³Centre for Ecology and Conservation, University of Exeter, Penryn Campus, Penryn, TR10
9 9EZ, United Kingdom

10 ⁴BirdLife International Kigali Office, Box 2527, Kigali Post Office, Kigali, Rwanda

11 ⁵Mpala Research Centre, P O Box 555, Nanyuki, 10400 Kenya.

12 ⁶Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544,
13 USA

14 ⁷African Butterfly Research Institute, Box 14308-0800, Nairobi, Kenya

15 ⁸Institut für Biologie, Zentrum für medizinische Struktur- und Zellbiologie, Universität Lübeck,
16 Ratzeburger Allee 160, 23538 Lübeck, Germany

17 ⁹Natural History Museum, Eton College, Windsor SL4 6DW, UK

18 *Correspondence to: simon.martin@ed.ac.uk

19 **Abstract**

20 Neo-sex chromosomes are found in many taxa, but the forces driving their emergence and
21 spread are poorly understood. The female-specific neo-W chromosome of the African monarch
22 (or queen) butterfly *Danaus chrysippus* presents an intriguing case study because it is restricted
23 to a single ‘contact zone’ population, involves a putative colour patterning supergene, and co-
24 occurs with infection by the the male-killing endosymbiont *Spiroplasma*. We investigated the
25 origin and evolution of this system using whole genome sequencing. We first identify the ‘BC
26 supergene’, a large region of suppressed recombination that links two colour patterning loci.
27 Association analysis suggests that the genes *yellow* and *arrow* control the forewing colour
28 pattern differences between *D. chrysippus* subspecies. We then show that the same chromosome
29 has recently formed a neo-W that has spread through the contact zone within ~2200 years. We
30 also assembled the genome of the male-killing *Spiroplasma*, and find that it shows perfect
31 genealogical congruence with the neo-W, suggesting that the neo-W has hitchhiked to high
32 frequency as the male killer has spread through the population. The complete absence of female
33 crossing-over in the Lepidoptera causes whole-chromosome hitchhiking of a single neo-W
34 haplotype, carrying a single allele of the BC supergene, and dragging multiple non-synonymous
35 mutations to high frequency. This has created a population of infected females that all carry the
36 same recessive colour patterning allele, making the phenotypes of each successive generation
37 highly dependent on uninfected male immigrants. Our findings show how hitchhiking can occur
38 between the unlinked genomes of host and endosymbiont, with dramatic consequences.

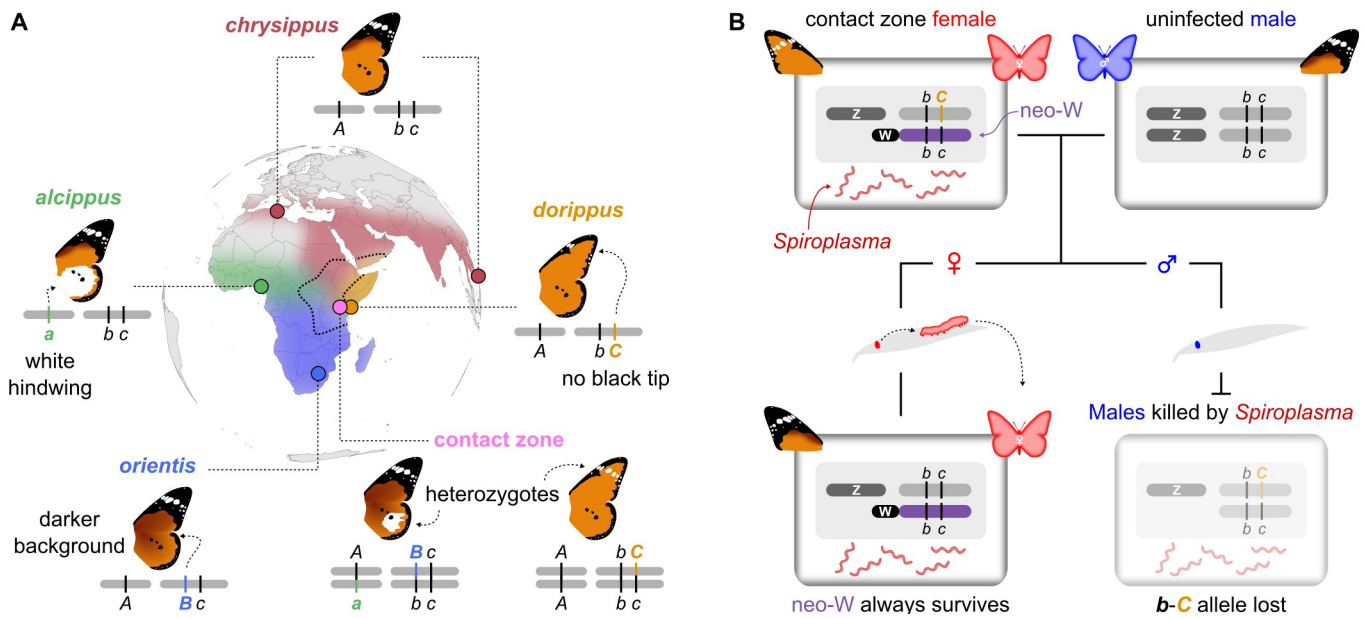
39 Introduction

40 Structural changes to the genome play an important role in evolution by altering the extent
41 of recombination among loci. This is best studied in the context of chromosomal inversions that
42 cause localised recombination suppression, and can be favoured by selection if they help to
43 maintain clusters of co-adapted alleles (or ‘supergenes’) in the face of genetic mixing [1–4]. A
44 greater extent of recombination suppression occurs in the formation of heteromorphic sex
45 chromosomes, which can link sex-specific alleles in a similarly to supergenes [5]. However,
46 suppressed recombination can also have costs. In particular, male-specific Y and female-specific
47 W chromosomes can be entirely devoid of recombination, making them vulnerable to genetic
48 hitchhiking and the accumulation of deleterious mutations through ‘Muller’s ratchet’, which may
49 explain their deterioration over time [6–8]. These contrasting benefits and costs of recombination
50 suppression are of particular interest in the evolution of neo-sex chromosomes, which can form
51 through fusion of autosomes to existing sex chromosomes. There is accumulating evidence that
52 neo-sex chromosomes are common in animals [9–15], but the processes underlying their
53 emergence, spread and subsequent evolution have not been widely studied. In particular, there
54 are few studied examples of recently-formed neo-sex chromosomes that are not yet fixed in a
55 species.

56 The African monarch (or queen) butterfly *Danaus chrysippus*, provides a unique test case
57 for the causes and consequences of changes in genome architecture. Like its American cousin
58 (*D. plexippus*), it feeds on milkweeds and has bright colour patterns that warn predators of its
59 distastefulness. However, within Africa *D. chrysippus* is divided into four subspecies with
60 distinct colour patterns and largely distinct ranges (Fig. 1A). Predator learning should favour the
61 maintenance of a single monomorphic warning in any single area. For this reason, researchers
62 have long been puzzled by the large polymorphic contact zone in East and Central Africa, where
63 all four *D. chrysippus* subspecies meet and interbreed [16–18] (Fig. 1A). Crosses have shown
64 that colour pattern differences between the subspecies are controlled by Mendelian autosomal
65 loci, including the tightly linked ‘B’ and ‘C’ loci (putatively a ‘BC supergene’ [19]) that define
66 three common forewing patterns [20,21] (Fig. 1B). However, crosses with females from the
67 contact zone revealed that the BC chromosome has become sex linked, forming a neo-W that is
68 unique to this population [19,22]. Since female meiosis is achiasmatic (it lacks crossing-over) in
69 the Lepidoptera, the formation of a neo-W would instantaneously cause perfect linkage, not just

70 of the B and C loci, but of an entire non-recombining chromosome, along with other maternally-
71 inherited DNA.

72 What is particularly striking is that the presence of the neo-W coincides with infection by a
73 maternally-inherited ‘male killer’ endosymbiont related to *Spiroplasma ixodetis*, which kills male
74 offspring and leads to highly female-biased populations in the contact zone [22–24]. The
75 combination of neo-W and male killing is expected to dramatically alter the inheritance and
76 evolution of the BC chromosome [22,25]: Infected females typically give rise to all-female
77 broods who should always inherit the same colour patterning allele on their neo-W, along with
78 the male-killer, while the other maternal allele is systematically eliminated in the dead sons (Fig.
79 1C), forming a genetic sink for all colour pattern alleles not on the neo-W. It has been suggested
80 that the restriction of male killing to females with the neo-W, and only in the region in which
81 hybridisation occurs between subspecies, may not be a coincidence [19,22,25–27]. However, the
82 genomic underpinnings of this system, including the genetic controllers of colour pattern, the
83 source and spread of the neo-W, and its relationship with the male killer, have until now
84 remained a mystery. We generated a reference genome for *D. chrysippus* and used whole genome
85 sequencing of population samples to uncover the interconnected evolution of the BC supergene,
86 neo-W and *Spiroplasma*. Our findings reveal a recent whole-chromosome selective sweep
87 caused by hitchhiking between the host and endosymbiont genomes.



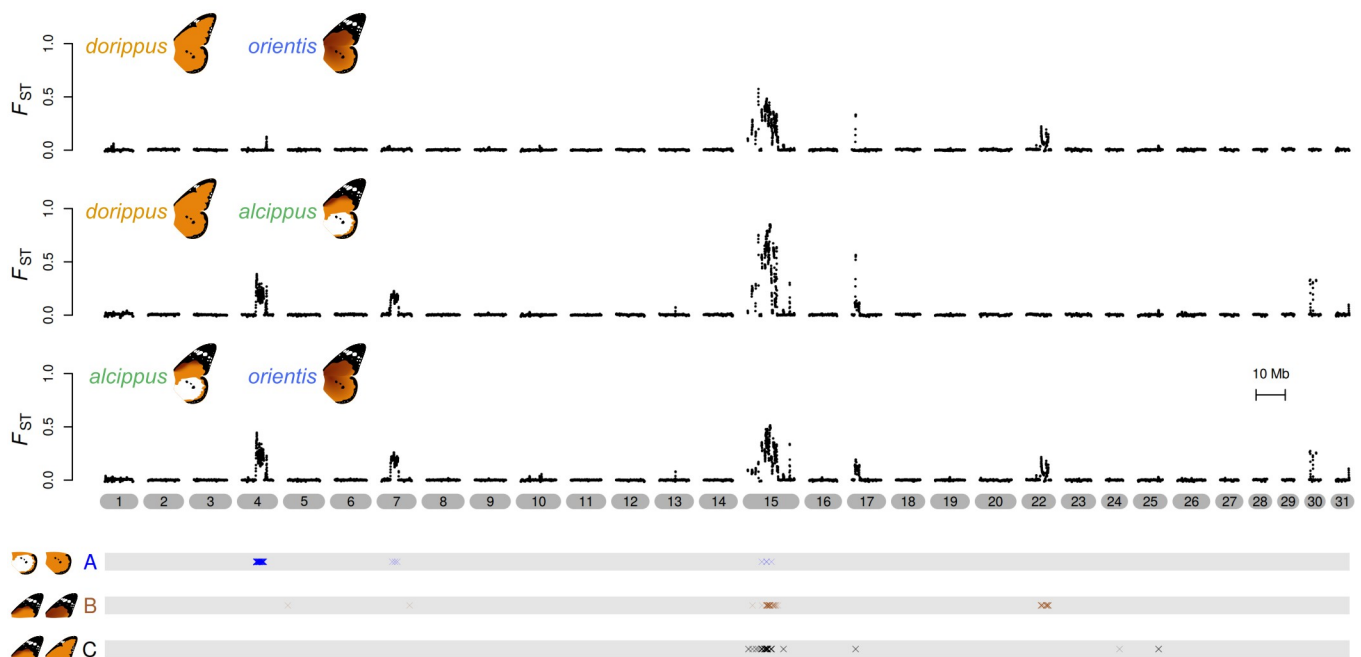
88 **Fig. 1. Geography and genetics of colour pattern.** (A) Approximate ranges of the four
 89 subspecies of *D. chrysippus*, with the contact zone outlined. Sampling locations for each of the
 90 subspecies and the contact zone are indicated. Cartoon chromosomes show the genotypes of each
 91 subspecies at the A (white hindwing patch), B (brown background colour) and C (forewing tip)
 92 colour patterning loci, based on previous crosses [20]. Note the linkage of B and C, putatively
 93 forming a ‘BC supergene’ [19]. Two examples of heterozygotes that can be found in the
 94 contact zone are shown. Note that *Cc* heterozygotes can exhibit the *transiens* phenotype with
 95 white markings on the forewing with ~50% penetrance. (B) Model showing how fusion of the BC
 96 autosome to the W chromosome has produced a neo-W (purple) in contact zone females (top
 97 left), while males have two autosomal copies of the BC chromosome (top right). Daughters
 98 inherit the neo-W, while sons inherit the other BC chromosome haplotype from their mother. The
 99 latter allele is then lost due to male killing by *Spiroplasma*.

100 Results and Discussion

101 Identification of the BC supergene

102 We assembled a high quality draft genome for *D. chrysippus*, with a total length of 322
 103 Megabases (Mb), a scaffold N50 length of 0.63 Mb, and a BUSCO [28] completeness score of
 104 94% (Table S1-S8). We then further scaffolded the genome into a pseudo-chromosomal assembly
 105 based on homology with the *Heliconius melpomene* genome [29–31] accounting for known
 106 fusions that differentiate these species [9,30,32] (Fig. S1). We also re-sequenced 42 individuals
 107 representing monomorphic populations of each of the four subspecies and a polymorphic
 108 population from a known male-killing hotspot near Nairobi, in the contact zone (Fig 1A, Table
 109 S9).

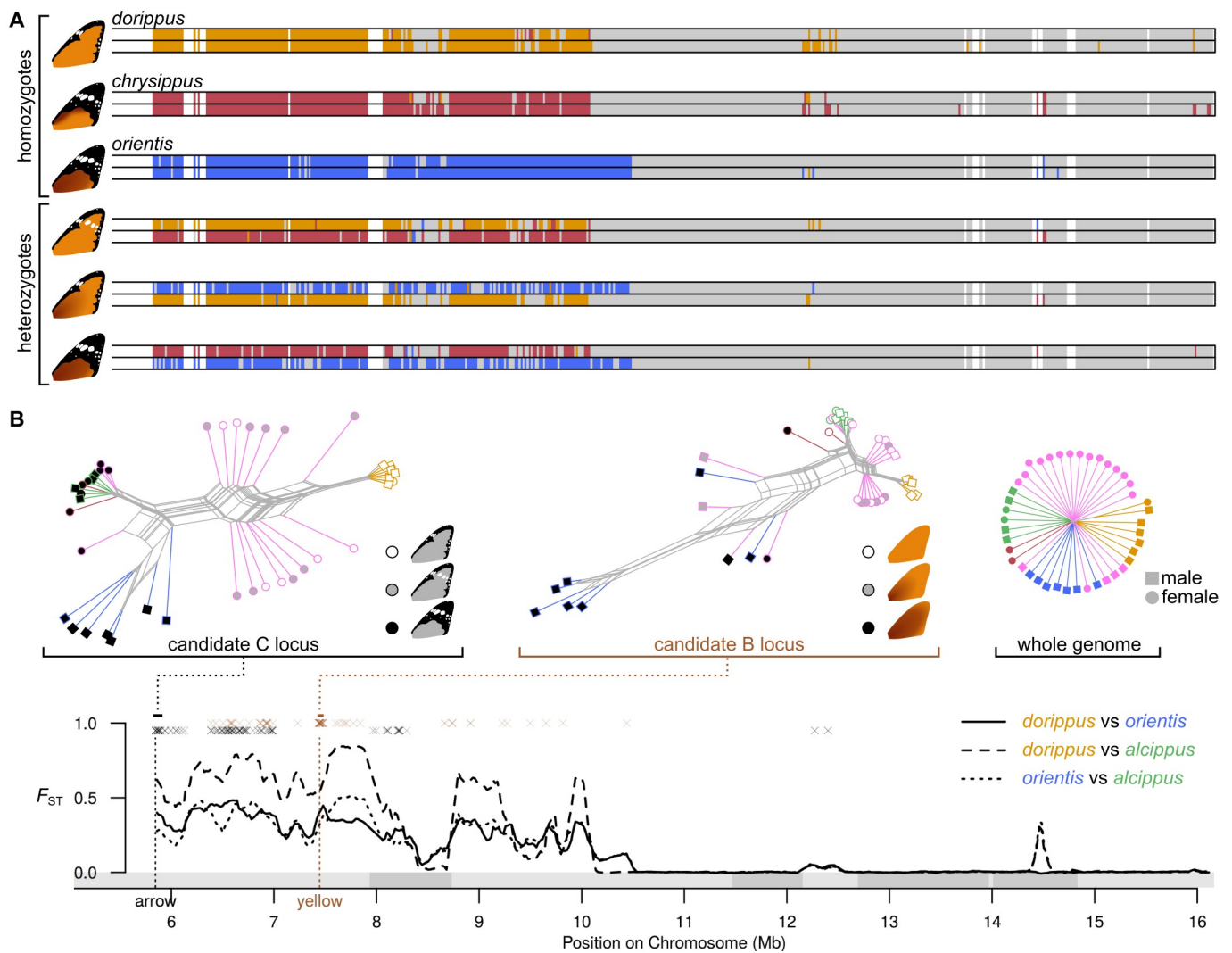
110 To identify the putative BC supergene, we scanned for genomic regions showing high
111 differentiation between the subspecies and an association with colour pattern. Genetic
112 differentiation (F_{ST}) is largely restricted to a handful of broad peaks, with a background level of
113 approximately zero (Fig. 2, S2). This low background level implies a large and nearly panmictic
114 population. Indeed, average genome-wide diversity at putatively neutral 4-fold degenerate 3rd
115 codon positions is 0.042, which is the highest value reported for any arthropod to our knowledge
116 [33], indicating an extremely large effective population size. The islands of differentiation that
117 stand out from this background imply selection for local adaptation maintaining particular
118 differences between the subspecies, similar to patterns seen between geographic races of
119 *Heliconius* butterflies [34]. However, here the peaks of differentiation are broad, covering
120 several Mb, implying some mechanism of recombination suppression such as inversions that
121 differentiate the subspecies.



122 **Fig. 2. Genetic differentiation and associations with colour pattern.** Pairwise genetic
123 differentiation (F_{ST}), plotted in 100 kb sliding windows with a step size of 20 kb across all
124 chromosomes. Three different pairs of subspecies for which sample sizes were ≥ 6 are shown.
125 F_{ST} of ~ 0 indicates a lack of genetic differentiation between populations, and peaks indicate
126 strong differentiation. Below, locations of SNPs most strongly associated with the A, B and C
127 loci are shown (Wald test, 99.99% quantile). See Fig. S2 for a more detailed plot.

128 The inclusion of the polymorphic contact-zone samples, and the fact that three of the
129 subspecies each carry a unique colour pattern allele (Fig. 1A), allowed us to identify particular
130 differentiated regions associated with the three major colour pattern traits. A ~3 Mb region on
131 chromosome 4 is associated with the white hindwing patch (A locus) and a ~5 Mb region on
132 chromosome 15 (hereafter chr15) is associated with both background colour (B locus) and the
133 forewing black tip (C locus) (Fig. 2, S2). Below, we refer to this region on chr15, which spans
134 over 200 protein-coding genes, as the BC supergene [19], although we note that additional
135 associated SNPs on chromosome 22 suggest that background wing melanism may also be
136 influenced by other loci.

137 Clustering analysis based on genetic distances reveals three clearly distinct alleles at the
138 BC supergene (Fig 3A). This further supports the hypothesis of recombination suppression,
139 although a number of individuals show mosaic patterns consistent with rare recombination (Fig
140 S3). The three main alleles correspond to the three common forewing phenotypes, so we term
141 these $BC^{chrysippus}$ (orange background with black forewing tip, formerly *bbcc*), $BC^{dorippus}$ (orange
142 without black tip, formerly *bbCC*), and $BC^{orientis}$ (brown background with black forewing tip,
143 formerly *BBcc*) (Fig 3A). Fifteen of the twenty contact zone individuals are heterozygous,
144 carrying two distinct BC alleles, and a few show evidence for recombination, as is also seen in
145 some of the southern African *orientis* individuals (Fig. S3). As shown previously, $BC^{dorippus}$
146 (which includes the dominant C allele) and $BC^{orientis}$ (which includes the dominant B allele) are
147 both dominant over the recessive $BC^{chrysippus}$ (Fig. 3A, S3).



148 **Fig. 3. Identification of the BC supergene on chromosome 15.** (A) Allelic clustering on
149 chromosome 15 (chr15) in six representative individuals (see Fig. S3 for all individuals, and see
150 panel B for chromosome positions). Coloured blocks indicate 20 kb windows in which sequence
151 haplotypes could be assigned to one of three clusters based on pairwise genetic distances (see
152 Methods for details). Windows in grey show insufficient relative divergence to be assigned to a
153 cluster and white indicates missing data. (B) Genetic differentiation (F_{ST} ; bottom) among three
154 subspecies pairs across chr15. Note that the first ~6 Mb of the chromosome is not shown due to
155 complex structural variation (see main text). Scaffolds are indicated below the plot in alternating
156 shades of grey. Above the plot, locations of SNPs most strongly associated with the B and C loci
157 (Wald test, 99.99% quantile) are shown in brown and black, respectively. Distance-based
158 phylogenetic networks constructed for candidate regions for the B and C loci are also shown,
159 along with a corresponding network for the whole genome for contrast. Colours indicate
160 subspecies as in Fig. 1A, and shapes indicate sex. Phenotypes for B and C are coded black and
161 white for putative homozygotes and gray for putative heterozygotes. The locations of our most
162 likely candidate genes for B (*yellow*) and C (*arrow*) are indicated.

163 Although it can be challenging to identify particular functional mutations in regions of
164 suppressed recombination, the presence of some recombinant individuals allowed us to narrow
165 down candidate regions for the B and C loci. A cluster of SNPs most strongly associated with
166 background colour (B locus) is found just upstream of the gene *yellow*, and a phylogenetic
167 network for a 30 kb region around *yellow* groups individuals nearly perfectly by phenotype (Fig.
168 3B). Differential expression of Yellow in *Drosophila* is associated with different levels of
169 melanism [35] and *yellow* knockouts in other butterflies show reduced melanin pigmentation
170 [36], making this a compelling candidate for the background colour polymorphism.

171 Associations with forewing tip (C locus) are more dispersed across the supergene region,
172 but the most strongly-associated SNPs are found toward the proximal end, and a phylogenetic
173 network for this 60 kb region similarly clusters individuals by phenotype (Fig. 3B). Within this
174 region, two of the most strongly associated SNPs fall in the gene *arrow*, which encodes a low-
175 density lipoprotein receptor-related protein (LRP). In *Drosophila*, Arrow is essential for Wnt
176 signalling in wing development [37]. Wnt signalling is known to underlie variation in colour
177 pattern in *Heliconius* butterflies [38] and knock-out mutants for the Wnt ligand gene *WntA* in *D.*
178 *plexippus* show a loss of pigmentation [39]. This makes *arrow* a promising candidate for the C
179 locus gene. These putative colour patterning loci will be investigated in future studies by
180 narrowing down the associated regions and performing functional manipulation.

181 Irrespective of their precise mode of action, the patterns of association imply that the B and
182 C loci are ~1.6 Mb apart, and would therefore be fairly loosely linked under normal
183 recombination. This physical distance translates to around 7.6 cM, assuming crossover rates
184 similar to those in *Heliconius* [31,40], whereas the estimated recombination distance between B
185 and C based on crosses is 1.9 cM [41]. Theory predicts that recombination suppression can be
186 favoured if it maintains linkage disequilibrium (LD) between co-adapted alleles in the face of
187 gene flow [1–4], but convincing empirical cases in which distinct functional loci are maintained
188 in LD are rare [5,42]. To our knowledge, ours is the first example of a butterfly supergene in
189 which the data strongly support the existence of two distinct genes that independently affect
190 colour pattern maintained in LD by suppressed recombination.

191 It is likely that chromosomal rearrangements contribute to recombination suppression at
192 the BC supergene. Although our short-read data do not allow us to test directly for inversions,
193 they do reveal dramatic variation in sequencing coverage over the proximal end of the

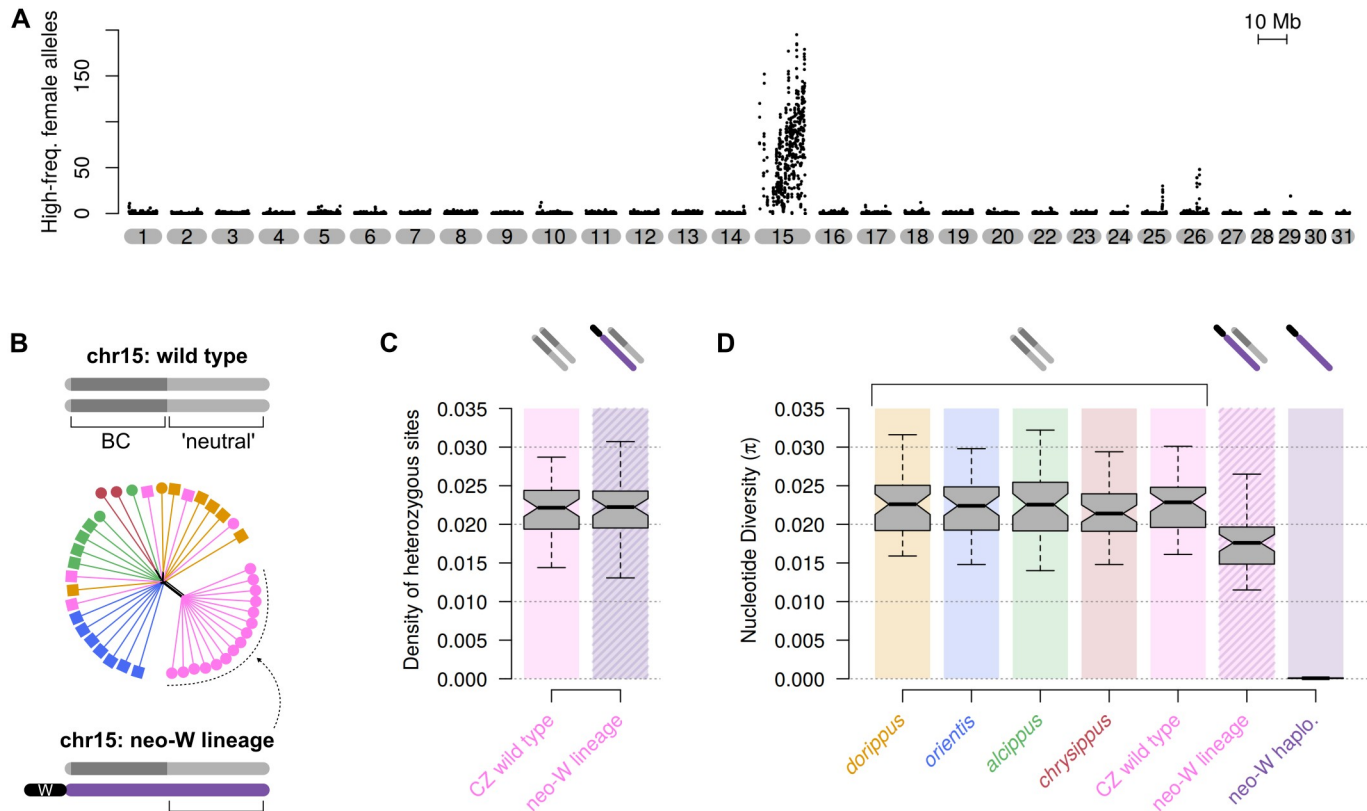
194 chromosome. Comparison of coverage among individuals suggests a large (~5 Mb) polymorphic
195 insertion in this region (Fig. S4A). Synteny comparison with *H. melpomene* reveals that this
196 insertion involves an expansion in copy number of a region of several hundred kb. Comparison
197 of copy numbers for two of the genes in this insertion with several other species confirms that
198 the insertion is derived in *D. chrysippus*, and unique to the $BC^{dorippus}$ allele (Fig. S5B, S5C). The
199 expansion appears to occur just a few kb from the coding region of *arrow* (Fig. S4B), and is also
200 perfectly associated with the presence of the dominant *C* phenotype (absence of black forewing
201 tip) (Fig. S5A). It is possible that it has a causal effect on the phenotype by influencing the
202 expression of *arrow*, but it might also be simply linked to the causative mutation. Either way, we
203 suggest that this large structural change, which increases the length of the chromosome by nearly
204 a third, contributes to recombination suppression between the $BC^{dorippus}$ allele and other supergene
205 alleles by interfering with chromosome pairing in heterozygotes.

206 **A neo-W chromosome traps a single haplotype of chromosome 15 in contact zone females**

207 Previous crossing experiments indicated that the BC chromosome has become sex-linked
208 in contact zone females [22]. To confirm this hypothesis using genetic tools, we created a ‘cured
209 line’ by treating a female from an all-female brood with tetracycline to eliminate *Spiroplasma*
210 and allow the survival of male offspring [23]. A cross using this female confirms perfect sex-
211 linkage of forewing phenotype (Fig. S6A). We then used PCR assays on a subsequent sibling
212 cross from the cured line to confirm that maternal alleles for chr15 segregate with sex while
213 paternal alleles segregate randomly (Fig. S6B). These results exactly match the model (Fig. 1B)
214 in which the BC supergene has become linked to the W chromosome in females, but continues to
215 segregate as an autosome in males.

216 Although we were unable to definitively identify any scaffolds from the ancestral W
217 chromosome, which is likely to be highly repetitive, we can test whether chr15 shows the
218 expected hallmarks of a young neo-W, hypothesised to have formed through fusion to the
219 ancestral W [22]. Due to complete absence of recombination in females, we expect that a single
220 fused haplotype of chr15 would be spreading in the population. Any unique mutations specific to
221 this haplotype should therefore occur at high frequency in females and be absent in males. We
222 scanned for such high-frequency female-specific mutations, and found them to be abundant
223 across the entire length of chr15 and nearly absent throughout the rest of the genome (Fig. 4A).
224 At the individual level, we can clearly identify 15 females (14 collected in the contact zone and

225 the single ‘cured line’ female) that consistently share these high-frequency mutations (Fig. S7).
 226 Genetic distance among these females in the ‘neutral’ region of chr15 (outside the BC supergene)
 227 is reduced, indicating that they all share a similar haplotype of the fused chromosome (Fig. 4B).



228 **Fig. 4. Recent sweep of a young neo-W.** (A) The number of high-frequency female-specific
 229 mutations ($> 20\%$ in females and absent in males) in 100 kb sliding windows (20 kb step size).
 230 (B) Distance-based phylogenetic network for the distal ‘neutral’ region of chr15 (lighter grey
 231 portion), outside of the BC supergene reveals the 15 females that carry the conserved neo-W
 232 haplotype. Cartoons show how the homologous neutral region of chr15 is outside of the BC
 233 supergene but would still capture reduced divergence among individuals carrying a shared non-
 234 recombining neo-W. (C) Boxplot comparing the density of heterozygous sites in 100 kb
 235 windows in the neutral region of chr15 between wild type individuals from the contact zone (CZ)
 236 and those carrying the neo-W. Cartoon chromosomes above the plot match those shown in panel
 237 B. A relative lack of elevated heterozygosity in the neo-W lineage indicates a lack of divergence
 238 of the fused neo-W haplotype, consistent with the fusion being recent. (D) Boxplot of nucleotide
 239 diversity (π) within each population for the same neutral region of chr15. On the far right, π is
 240 shown for the haploid neo-W haplotype specifically, based on partial sequences isolated from
 241 this haplotype (see Methods and Fig. S8 for details). The near absence of genetic diversity
 242 reflects a very rapid spread of the neo-W through the population.

243 **The neo-W formed recently and spread rapidly**

244 Genetic variation accumulated in the neo-W lineage since its formation can tell us about its
245 age. Sequence divergence between the neo-W and autosomal copies of chr15 (inferred from the
246 density of heterozygous sites in the neutral region of chr15 in females carrying the neo-W) is not
247 significantly different from that between the autosomal copies in ‘wild type’ individuals that lack
248 the fusion (Fig. 4C, Wilcoxon signed rank test, $p=0.36$, $n=48$ 100 kb windows). This implies that
249 insufficient time has passed since the fusion event for significant accumulation of new mutations.
250 The limited divergence of the neo-W haplotype from the autosomal copy of chr15 in each female
251 makes it challenging to isolate. Nonetheless, by identifying diagnostic mutations that are unique
252 to, and fixed in the neo-W lineage, we were able to isolate sequencing reads from the shared
253 haplotype and reconstruct a partial neo-W sequence for each female (Fig. S8). A dated genealogy
254 based on these sequences places the root of the neo-W lineage at ~2200 years (26,400
255 generations) ago (posterior mean = 2201, std. dev. = 318).

256 The neo-W is present in all but one of the contact zone females, implying a rapid spread
257 since its formation. This process is similar to a selective sweep of a beneficial mutation, except
258 that complete recombination suppression in females means that the sweep affects the entire
259 chromosome equally. Unlike a conventional sweep, it is not expected to eliminate genetic
260 diversity from the population as these females will also carry an autosomal copy of chr15
261 inherited from their father (Fig. 1B). Indeed, we see only a 20% reduction in overall nucleotide
262 diversity (π) on chr15 in females of the neo-W lineage (Fig. 4D). However, when we consider
263 only the neo-W haplotype in each of these females we see a nearly complete absence of genetic
264 variation, with a π value of 0.00007, two orders of magnitude lower than for autosomal copies of
265 chr15 (0.0228) (Fig. 4D), further supporting a very recent and rapid spread.

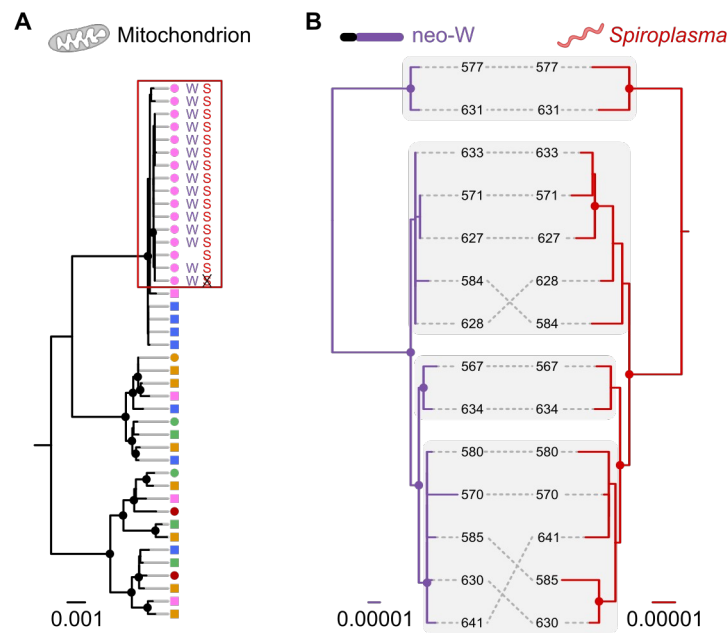
266 The neo-W haplotype carries the recessive $BC^{chrysippus}$ allele at the BC supergene (Fig. S3).
267 However, many of the individuals carrying the neo-W express the *dorippus* phenotype because
268 they carry the dominant $BC^{dorippus}$ allele on their autosomal chr15 chromatid. This highlights the
269 question of what selective driver might have caused the spread of a recessive colour patterning
270 allele.

271 **Hitchhiking between the neo-W and *Spiroplasma***

272 We hypothesised that the neo-W has spread not due to direct natural selection on warning
273 pattern but rather through co-inheritance with the male-killing *Spiroplasma*, which is itself
274 spreading through the population as a selfish element. Experiments have shown that infected
275 female offspring may show increased survival relative to those from uninfected broods due to
276 reduced competition for resources [43]. For this to drive the spread of the neo-W would require
277 strict vertical inheritance of *Spiroplasma* down the female line, such that it is always co-inherited
278 with the neo-W.

279 We identified nine scaffolds making up the 1.75 Mb *Spiroplasma* genome in our *D.*
280 *chrysippus* assembly (Fig. S9A). Infected individuals are clearly identifiable by mapping
281 resequencing reads to the *Spiroplasma* scaffolds (Fig. S9B). As predicted, all females in the neo-
282 W lineage are infected (with the exception of the cured line female, in which *Spiroplasma* had
283 been eliminated). Moreover, all infected females fall into the same mitochondrial clade (Fig.
284 5A). To confirm that the *Spiroplasma* is strictly vertically inherited and always associated with a
285 single female lineage, we used PCR assays for *Spiroplasma* and mitochondrial haplotype and
286 expanded our sample size to 158 individuals, including samples used in previous studies going
287 back two decades [19,23] (Data S1, Fig. S10). This confirms the perfect association: 100% of
288 infected individuals (n=42) carry the same mitochondrial haplotype, and this haplotype is
289 otherwise rare, occurring in 8% of uninfected individuals (n=116) (Fig. S10).

290 Like the neo-W, the *Spiroplasma* genomes carry limited variation among individuals ($\pi =$
291 0.0005), consistent with a single and recent outbreak of the endosymbiont. Although the lack of
292 variation makes it challenging to infer genealogies, our inferred maximum likelihood
293 genealogies for the neo-W and *Spiroplasma* are strikingly congruent (Fig 4B). In a permutation
294 test for congruence [44] the observed level exceeds all 100,000 random permutations, indicating
295 strong support for co-inheritance of the neo-W and *Spiroplasma* [45].



296 **Fig. 5. Matrilineal inheritance causes coupling between neo-W and *Spiroplasma*.** (A)
297 Maximum likelihood phylogeny for the whole mitochondrial genome. Individuals are coloured
298 according to population of origin (see Fig. 1A), and those carrying the neo-W ('W') and
299 *Spiroplasma* ('S') are indicated (including one cured individual in which *Spiroplasma*
300 was eliminated). Females are indicated by circles and males by squares. (B) Maximum likelihood
301 phylogenies for the neo-W haplotype and *Spiroplasma* genome isolated from infected females.
302 Corresponding clades are shaded to indicate congruence. Note that two samples are excluded in
303 panel B: the cured sample which lacked *Spiroplasma* due to tetracycline treatment, and one
304 infected female found to lack the neo-W. Whether the latter represents an ancestral state or
305 secondary loss requires further investigation. In all trees, nodes supported by more than 700 of
306 1000 bootstrap replicates are indicated by circles.

307 The combined spread of three unlinked DNA molecules – the mitochondrial genome, neo-
308 W and *Spiroplasma* genome – constitutes a form of genetic hitchhiking, but is facilitated by their
309 strict matrilineal inheritance rather than physical linkage. While it is well known that selfish
310 maternally-inherited endosymbionts can drive the spread of a single mitochondrial haplotype
311 through a population and even across species boundaries [46,47], our findings show how an
312 entire chromosome can be captured in the same way. Hitchhiking with selfish elements may be
313 of general importance in driving the spread of neo-sex chromosomes.

314 In *D. chrysippus*, it is currently unclear whether the neo-W or male-killer emerged first. It
315 is also unclear whether their co-occurrence in a single ancestor was simply a coincidence or

316 instead reflects some functional connection, such as the suggestion that the neo-W might confer
317 susceptibility to the male killer [22]. It is important to note that this is not the first time a neo-sex
318 chromosome has formed in this taxon. A fusion of chromosome 21 to the ancestral Z
319 chromosome occurred in an ancestor of all *Danaus* species, producing a neo-Z [9,32,48]. It is
320 speculated that a fusion of chromosome 21 to the ancestral W also occurred [9,48] but this is
321 difficult to conclusively verify due to degradation of W chromosome over longer timescales. If
322 this hypothesis of an ancient neo-W is correct, then the neo-W we describe (W-chr15) might in
323 fact be better described as a neo-neo-W (W-chr21-chr15). It is possible that the spread of the
324 original W-chr21 was also driven by hitchhiking with a selfish endosymbiont.

325 **Genetic and phenotypic consequences of recombination suppression**

326 Sex chromosome evolution in many other taxa involves the progressive spread of
327 recombination suppression outward from the a sex-determining locus [49]. By contrast, the
328 absence of crossing over in female meiosis means that a lepidopteran neo-W experiences
329 complete and immediate recombination suppression over its entire length. The young age of the
330 *D. chrysippus* neo-W therefore provides a rare opportunity to study the early evolutionary
331 consequences of recombination suppression across an entire chromosome. Two related processes
332 could shape its evolution: hitchhiking of pre-existing deleterious mutations that were initially
333 rare in the population [6], and accumulation of novel deleterious mutations due to reduced
334 purging through recombination and selection (i.e. Muller's Ratchet) [7].

335 As a proxy for the 'genetic load' of deleterious mutations in the population, we considered
336 P_n/P_s , the ratio of non-synonymous to synonymous polymorphisms. Due to purifying selection,
337 non-synonymous polymorphisms are typically rare, and where they do occur the mutant allele
338 typically occurs at low frequency in the population [50]. When considering all polymorphisms in
339 the neo-W lineage, we see no overall excess of P_n/P_s for chr15 (excluding genes within the BC
340 supergene excluded to avoid bias) (Fig. S11A). However, when we partition polymorphisms by
341 allele frequency, we see a two 2-3 fold increase in P_n/P_s for mutations at high frequency,
342 specifically those at exactly 50% frequency, on chr15 (Fig. S11B, S11C). This implies that
343 hitchhiking has indeed led to the inadvertent spread of pre-existing deleterious alleles that were
344 initially rare in the population but happened to be present on the neo-W haplotype, and are
345 therefore now found in all females in this lineage. The lack of an overall excess of genetic load

346 probably reflects insufficient time for accumulation of new deleterious mutations since the
347 formation of the neo-W.

348 At the phenotypic level, perhaps counter-intuitively, the spread of a single supergene allele
349 on the neo-W has not caused homogenization of warning pattern among contact zone females,
350 and might in fact have the opposite effect. In locations where the neo-W and *Spiroplasma* are
351 nearly fixed, such as our sampling site near Nairobi, the high incidence of male killing implies
352 that the population is strongly shaped by immigrant males. Since the *BC^{chrysippus}* allele on the neo-
353 W is universally recessive, daughters will tend to match the phenotype of their immigrant father.
354 However, because the neo-W is always transmitted to daughters, the paternal chr15 copy will be
355 lost to male killing after one generation, creating a genetic sink for immigrant male genes
356 [22] (Fig. S12B). This combination of processes makes for a female population that is highly
357 sensitive to the source of immigrants, which is known to fluctuate seasonally with monsoon
358 winds [16,51] (Fig. S12A). This model leads to the testable prediction that seasonal fluctuations
359 in female phenotypes should be most dramatic where male killing is most abundant.

360 **Future evolutionary trajectories**

361 The future of the neo-W and *Spiroplasma* outbreak is uncertain. A lack of males could lead
362 to local extinctions [27], but extinction of the entire infected lineage is unlikely given the high
363 dispersal ability of males. Since the spread of the male killer is dependent on reduced sibling
364 competition in all-female broods, the spread might also be limited in environments where
365 competition is reduced, or where oviposition behaviour reduces opportunities for competition
366 [43]. In other systems, sex-ratio distortion has driven adaptive responses by the host, including
367 changes to the mating system [52] and the evolution of resistance to male-killing [53,54]. The
368 absence of evidence for these phenomena in *D. chrysippus* might simply reflect the recency of
369 the male killing outbreak. Eventually, we also expect the non-recombining neo-W to begin to
370 degenerate through further hitchhiking, gene loss and the spread of repetitive elements [8,49].
371 This young system provides a rare opportunity to study how these phenomena unfold through
372 time and space.

373 **Methods**

374 **Reference genome sequencing, assembly and annotation**

375 Detailed methods for generation of the *D. chrysippus* reference genome are provided in
376 Supplementary Methods. Briefly, a draft assembly was generated using SPAdes [55] from a
377 combination of paired-end and mate-pair libraries of various insert sizes. Scaffolding and
378 resolution of haplotypes was performed using Redundans [56] and Haplomerger2 [57]. The
379 assembly was annotated using a combination of de-novo gene predictors yielding 16,654 protein
380 coding genes. Mitochondrial genomes were assembled using NOVOplasty [58].

381 Although we currently lack linkage information for further scaffolding, we generated a
382 pseudo-chromosomal assembly based on homology with the highly contiguous *Heliconius*
383 *melpomene* genome [30,31,59], adjusted for known karyotypic differences [9,30–32,48]. In total,
384 282 Mb (87% of the genome) could be confidently assigned to chromosomes (Fig. S1).

385 Scaffolds representing the *Spiroplasma* genome were identified based on read depth of re-
386 mapped reads (Fig. S9A) and homology to other available *Spiroplasma* genomes. Annotation was
387 performed using the RAST server pipeline [60,61].

388 **Population sample resequencing and genotyping**

389 This study made use of 42 newly sequenced *D. chrysippus* individuals, as well as
390 previously sequenced individuals of the sister species, *D. petilia* (n=1), and next closest outgroup
391 *D. gilippus* (n=2) [62] (Table S9). Details of DNA extraction, sequencing and genotyping are
392 provided in Supplementary Methods. Briefly, DNA was extracted from thorax tissue and
393 sequenced (paired-end, 150 bp) to a mean depth of coverage 20x or greater. Reads were mapped
394 to the *D. chrysippus* reference assembly using Stampy [63] v1.0.31 and genotyping was
395 performed using GATK version 3 [64,65]. Genotype calls were required to have an individual
396 read depth ≥ 8 , and heterozygous and alternate allele calls were further required to have an
397 individual genotype quality (GQ) ≥ 20 for downstream analyses.

398 **Genomic differentiation and associations with wing pattern**

399 We used the fixation index (F_{ST}) to examine genetic differentiation across the genome
400 among the three subspecies for which we had six or more individuals sequenced. F_{ST} was
401 computed using the script popgenWindows.py (github.com/simonhmartin/genomics_general)

402 with a sliding windows of 100 kb, stepping in increments of 20 kb. Windows with fewer than
403 20,000 genotyped sites after filtering (see above) were ignored.

404 To identify SNPs associated with the three Mendelian colour pattern traits (i.e. the A B and
405 C loci) (Fig. 1A), we used PLINK v1.9 [66] with the '--assoc' option, and provided quantitative
406 phenotypes of 0, 1 or 0.5 for assumed heterozygotes. In addition to the quality and depth filters
407 above, SNPs used for this analysis were required to have genotypes for at least 40 individuals, a
408 minor allele count of at least 2, and to be heterozygous in no more than 75% of individuals.
409 SNPs were also thinned to a minimum distance of 100 bp.

410 To examine relationships among diploid individuals in specific regions of interest, we
411 constructed phylogenetic networks using the Neighbor-Net [67] algorithm, implemented in
412 SplitsTree [68]. Pairwise distances used for input were computed using the script distMat.py
413 (github.com/simonhmartin/genomics_general).

414 **Haplotype cluster assignment**

415 To assign haplotypes to clusters in the BC supergene region, we first phased genotypes
416 using SHAPEIT2 [69,70] using SNPs filtered as for association analysis above, except with a
417 minor allele count of at least 4, and no thinning. Default parameters were used for phasing except
418 that the effective population size was set to 3×10^6 . To minimise phasing switch errors, we
419 analysed each 20 kb window separately. Cluster assignment for both haplotypes from each
420 individual was based on average genetic distance to all haplotypes from each of three reference
421 groups: *D. c. dorippus*, *D. c. orientis*, or *D. c. alcippus* (which is representative of *chrysippus* as
422 they share the same alleles at the BC supergene). A haplotype was assigned to one of the three
423 groups if its average genetic distance to members of that group was less than 80% the average
424 distance to the other two groups, otherwise it was left as unassigned. Genetic distances were
425 computed using the script popgenWindows.py (github.com/simonhmartin/genomics_general).

426 **Identification of neo-W specific sequencing reads**

427 To identify females carrying the neo-W chromosome, we visualised the distribution of
428 female-specific derived mutations that occur at high-frequency. Allele frequencies were
429 computed using the script freq.py (github.com/simonhmartin/genomics_general). Due to the
430 absence of female meiotic crossing over in Lepidoptera, all females carrying the neo-W fusion
431 should share a conserved chromosomal haplotype for the entire fused chromosome. To isolate

432 this shared fused haplotype from the autosomal copy, we first identified diagnostic mutations as
433 those that are present in a single copy in each member of the ‘neo-W lineage’ and absent from all
434 other individuals and outgroups. We then isolated the sequencing read pairs from each of these
435 females that carry the derived mutation (Fig. S8). This resulted in a patchy alignment file, with a
436 stack of read pairs over each diagnostic mutation. Based on these aligned reads, we genotyped
437 each individuals as described above, except here setting the ploidy level to 1, and requiring a
438 minimum read depth of 3.

439 **Diversity and divergence of the neo-W**

440 The lack of recombination across the neo-W makes it possible to gain insights into its age.
441 Over time, mutations will arise that differentiate the neo-W from the recombining autosomal
442 copies of the chromosome. We estimated this divergence based on average heterozygosity in
443 females carrying the neo-W, and compared it to heterozygosity from contact-zone individuals
444 not carrying the neo-W. Heterozygosity was computed using the Python script
445 `popgenWindows.py` (github.com/simonhmartin/genomics_general) focusing only on the ‘neutral’
446 portion of the chromosome (i.e. the distal portion from 11 Mb onwards), which is outside of the
447 BC supergene. Heterozygosity was computed in 100 kb windows, and windows were discarded
448 if they contained fewer than 20,000 sites genotyped in at least two individuals from each
449 population.

450 A recent spread of the neo-W through the population should also be detectable in the form
451 of strong conservation of the neo-W haplotype in all females that carry it (i.e. reduced genetic
452 diversity). We therefore computed nucleotide diversity (π) in 100 kb windows as above.
453 Reported values of π and heterozygosity represent the mean \pm standard deviation across 100 kb
454 windows.

455 **Genealogical analyses**

456 We produced maximum likelihood trees for the mitochondrial genome, neo-W and
457 *Spiroplasma* genome, using PhyML v3 [71] with the GTR substitution model. Given the small
458 number of SNPs in both the neo-W and *Spiroplasma* genome, regions with inconsistent coverage
459 across individuals were excluded manually. Only sites with no missing genotypes were included.

460 We estimated the root node age for the neo-W using BEAST2 [72,73] version 2.5.1 with a
461 fixed clock model and an exponential population growth prior. For all other priors we used the

462 defaults as defined by BEAUti v2.5.1. We assumed a mutation rate of 2.9×10^{-9} per generation
463 based on a direct estimate for *Heliconius* butterflies [74] and 12 generations per year [75].
464 BEAST2 was run for 500,000,000 iterations, sampling every 50,000 generations, and we used
465 Tracer [76] version 1.7.1 to check for convergence of posterior distributions and compute the
466 root age after discarding a burn-in of 10%.

467 We tested for congruence between the neo-W and *Spiroplasma* trees using PACo [44],
468 which assesses the goodness-of-fit between host and parasite distance matrices, with 100,000
469 permutations. Distance matrices were computed using the script distMat.py
470 (github.com/simonhmartin/genomics_general).

471 **Analysis of synonymous and non-synonymous polymorphism**

472 We computed P_n/P_s as the ratio of non-synonymous polymorphisms per non-
473 synonymous site to synonymous polymorphisms per synonymous site. Synonymous and non-
474 synonymous sites were defined conservatively as 4-fold degenerate and 0-fold degenerate codon
475 positions, respectively, with the requirement that the other two codon positions are invariant
476 across the entire dataset. Only sites genotyped in all 15 females in the neo-W lineage were
477 considered, and counts were stratified by minor allele frequency using the script sfs.py
478 (github.com/simonhmartin/genomics_general).

479 **Butterfly rearing and molecular diagnostics**

480 To generate a stock line that is cured of *Spiroplasma* infection, we treated caterpillars from
481 all-female broods with Tetracycline, following Jiggins et al. [23]. A ‘cured line’ was initiated
482 from a single treated female that had the heterozygous *Cc* ‘*transiens*’ phenotype (Fig 1A). This
483 female was crossed to a wild male (homozygous *cc*) to test for sex linkage of phenotype. The
484 cured line was using sibling crosses and the persistence of males indicated that *spiroplasma* had
485 been eliminated.

486 We then applied a molecular test for sex linkage of chr15 using the F5 brood from the
487 cured line. We designed two separate PCR diagnostics based on SNPs segregating on chr15 to
488 distinguish between the two chromosomes of the male and the female parents (Table S10). PCR
489 was performed using the Phusion HF Master Mix and HF Buffer (New England Biosciences).

490 To screen for *Spiroplasma* infection, we designed a PCR assay targeting the
491 glycerophosphoryl diester phosphodiesterase (GDP) gene (Table S10). PCR was performed as

492 above. We confirmed the the sensitivity of this diagnostic by analyzing individuals of known
493 infection status based on whole genome sequencing.

494 To investigate whether *Spiroplasma* infection was always associated with a single
495 mitochondrial haplotype, we designed a PCR RFLP for the Cytochrome Oxidase Subunit I (COI)
496 that differentiates the infected ‘K’ lineage from uninfected (Table S10). PCR was performed as
497 above. A subset of products were verified by Sanger sequencing after purification using the
498 QIAquick PCR Purification Kit (Qiagen, USA).

499 **References**

- 500 1. Kirkpatrick M, Barton N. Chromosome inversions, local adaptation and speciation.
501 *Genetics*. 2006;173: 419–34. doi:10.1534/genetics.105.047985
- 502 2. Bürger R, Akerman A. The effects of linkage and gene flow on local adaptation: A two-
503 locus continent–island model. *Theor Popul Biol. Academic Press*; 2011;80: 272–288.
504 doi:10.1016/J.TPB.2011.07.002
- 505 3. Guerrero RF, Rousset F, Kirkpatrick M. Coalescent patterns for chromosomal inversions
506 in divergent populations. *Philos Trans R Soc B Biol Sci*. 2012;367: 430–438.
507 doi:10.1098/rstb.2011.0246
- 508 4. Charlesworth D, Charlesworth B. Selection on recombination in clines. *Genetics*. 1979;91:
509 575–580.
- 510 5. Charlesworth D. The status of supergenes in the 21st century: Recombination suppression
511 in Batesian mimicry and sex chromosomes and other complex adaptations. *Evol Appl*.
512 John Wiley & Sons, Ltd (10.1111); 2016;9: 74–90. doi:10.1111/eva.12291
- 513 6. Rice WR. Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex
514 chromosome. *Genetics*. 1987;116: 161–167.
- 515 7. Charlesworth B. Model for evolution of Y chromosomes and dosage compensation. *Proc*
516 *Natl Acad Sci U S A*. 1978;75: 5618–22.
- 517 8. Bachtrog D, Charlesworth B. The temporal dynamics of processes underlying Y
518 chromosome degeneration. *Genetics*. *Genetics*; 2008;179: 1513–25.
519 doi:10.1534/genetics.107.084012
- 520 9. Mongue AJ, Nguyen P, Volenikova A, Walters JR. Neo-sex chromosomes in the Monarch
521 butterfly, *Danaus plexippus*. *bioRxiv*. 2017; doi:10.1534/g3.117.300187

- 522 10. Bracewell RR, Bentz BJ, Sullivan BT, Good JM. Rapid neo-sex chromosome evolution
523 and incipient speciation in a major forest pest. *Nat Commun.* Springer US; 2017;8.
524 doi:10.1038/s41467-017-01761-4
- 525 11. Pala I, Naurin S, Stervander M, Hasselquist D, Bensch S, Hansson B. Evidence of a neo-
526 sex chromosome in birds. *Heredity.* Nature Publishing Group; 2012;108: 264–272.
527 doi:10.1038/hdy.2011.70
- 528 12. Kitano J, Ross JA, Mori S, Kume M, Jones FC, Chan YF, et al. A role for a neo-sex
529 chromosome in stickleback speciation. *Nature.* Nature Publishing Group; 2009;461:
530 1079–1083. doi:10.1038/nature08441
- 531 13. Nguyen P, Sykorova M, Sichova J, Kuta V, Dalikova M, Capkova Frydrychova R, et al.
532 Neo-sex chromosomes and adaptive potential in tortricid pests. *Proc Natl Acad Sci.*
533 2013;110: 6931–6936. doi:10.1073/pnas.1220372110
- 534 14. Carabajal Paladino LZ, Provazníková I, Berger M, Bass C, Aratchige NS, López SN, et al.
535 Sex Chromosome Turnover in Moths of the Diverse Superfamily Gelechioidea. Barluenga
536 M, editor. *Genome Biol Evol.* Narnia; 2019;11: 1307–1319. doi:10.1093/gbe/evz075
- 537 15. Steinemann M, Steinemann S. Enigma of Y chromosome degeneration: neo-Y and neo-X
538 chromosomes of *Drosophila miranda* a model for sex chromosome evolution. *Genetica.*
539 1998;102–103: 409–20.
- 540 16. Smith DAS, Owen DF, Gordon IJ, Lowis NK. The butterfly *Danaus chrysippus* (L .) in
541 East Africa : polymorphism and morph-ratio clines within a complex , extensive and
542 dynamic hybrid zone. *Zool J Linn Soc.* 1997;120: 51–78.
- 543 17. Smith DAS. Heterosis, epistasis and linkage disequilibrium in a wild population of the
544 polymorphic butterfly *Danaus chrysippus* (L .). *Zool J Linn Soc.* 1980;69: 87–109.
- 545 18. Smith D a S, Owen DF, Gordon IJ, Owiny AM. Polymorphism and evolution in the
546 butterfly *Danaus chrysippus* (L.) (Lepidoptera: Danainae). *Heredity.* 1993;71: 242–251.
547 doi:10.1038/hdy.1993.132
- 548 19. Smith DAS, Gordon IJ, Allen JA. Reinforcement in hybrids among once isolated
549 semispecies of *Danaus chrysippus* (L.) and evidence for sex chromosome evolution. *Ecol*
550 *Entomol.* 2010;35: 77–89. doi:10.1111/j.1365-2311.2009.01143.x
- 551 20. Smith DAS. Genetics of Some Polymorphic Forms of the African Butterfly *Danaus*
552 *chrysippus* L. (Lepidoptera: Danaidae). *Insect Syst Evol.* Brill; 1975;6: 134–144.
553 doi:10.1163/187631275X00235
- 554 21. Clarke CA, Sheppard PM, Smith AG. genetics of fore and hindwing colour in crosses
555 between *Danaus chrysippus* from Australia and from Sierra Leone (Danidae). *Lepid Soc*
556 *J.* 1973;

- 557 22. Smith DAS, Gordon IJ, Traut W, Herren J, Collins S, Martins DJ, et al. A neo-W
558 chromosome in a tropical butterfly links colour pattern, male-killing, and speciation. *Proc*
559 *R Soc B Biol Sci.* 2016;283: 20160821. doi:10.1098/rspb.2016.0821
- 560 23. Jiggins FM, Hurst GD, Jiggins CD, v d Schulenburg JH, Majerus ME. The butterfly
561 *Danaus chrysippus* is infected by a male-killing *Spiroplasma* bacterium. *Parasitology.*
562 2000;120 (Pt 5: 439–46. doi:10.1017/S0031182099005867
- 563 24. Herren JK, Gordon I, Holland PWH, Smith D. The butterfly *Danaus chrysippus*
564 (Lepidoptera: Nymphalidae) in Kenya is variably infected with respect to genotype and
565 body size by a maternally transmitted male-killing endosymbiont (*Spiroplasma*). *Int J*
566 *Trop Insect Sci.* 2007;27: 62. doi:10.1017/S1742758407818327
- 567 25. Gordon IJ, Ileri P, Smith DAS. Hologenomic speciation: Synergy between a male-killing
568 bacterium and sex-linkage creates a “magic trait” in a butterfly hybrid zone. *Biol J Linn*
569 *Soc.* 2014;111: 92–109. doi:10.1111/bij.12185
- 570 26. Lushai G, Allen JA, Goulson D, Maclean N, Smith DAS. The butterfly *Danaus chrysippus*
571 (L.) in East Africa comprises polyphyletic, sympatric lineages that are, despite behavioural
572 isolation, driven to hybridization by female-biased sex ratios. *Biol J Linn Soc.* 2005;86:
573 117–131. doi:10.1111/j.1095-8312.2005.00526.x
- 574 27. Idris E, Saeed M, Hassan S. Biased sex ratios and aposematic polymorphism in African
575 butterflies: A hypothesis. *Ideas Ecol Evol.* 2013;6: 5–16. doi:10.4033/iee.2013.6.2.n
- 576 28. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. BUSCO:
577 assessing genome assembly and annotation completeness with single-copy orthologs.
578 *Bioinformatics.* Oxford University Press; 2015;31: 3210–3212.
579 doi:10.1093/bioinformatics/btv351
- 580 29. Heliconius Genome Consortium T. Butterfly genome reveals promiscuous exchange of
581 mimicry adaptations among species. *Nature.* 2012;487: 94–8. doi:10.1038/nature11041
- 582 30. Davey JW, Chouteau M, Barker SL, Maroja L, Baxter SW, Simpson F, et al. Major
583 Improvements to the *Heliconius melpomene* Genome Assembly Used to Confirm 10
584 Chromosome Fusion Events in 6 Million Years of Butterfly Evolution. *G3.* 2016;6: 695–
585 708. doi:10.1534/g3.115.023655
- 586 31. Davey JW, Barker SL, Rastas PM, Pinharanda A, Martin SH, Durbin R, et al. No evidence
587 for maintenance of a sympatric *Heliconius* species barrier by chromosomal inversions.
588 *Evol Lett.* 2017;1: 138–154. doi:10.1002/evl3.12
- 589 32. Ahola V, Lehtonen R, Somervuo P, Salmela L, Koskinen P, Rastas P, et al. The *Glanville*
590 fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions
591 in Lepidoptera. *Nat Commun.* 2014;5: 1–9. doi:10.1038/ncomms5737

- 592 33. Leffler EM, Bullaughey K, Matute DR, Meyer WK, Ségurel L, Venkat A, et al. Revisiting
593 an old riddle: what determines genetic diversity levels within species? PLoS Biol.
594 2012;10: e1001388. doi:10.1371/journal.pbio.1001388
- 595 34. Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, et al.
596 Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. Genome
597 Res. 2013;23: 1817–1828. doi:10.1101/gr.159426.113
- 598 35. Wittkopp PJ, Vaccaro K, Carroll SB. Evolution of yellow Gene Regulation and
599 Pigmentation in *Drosophila*. Curr Biol. Cell Press; 2002;12: 1547–1556.
600 doi:10.1016/S0960-9822(02)01113-2
- 601 36. Zhang L, Martin A, Perry MW, van der Burg KRL, Matsuoka Y, Monteiro A, et al.
602 Genetic Basis of Melanin Pigmentation in Butterfly Wings. Genetics. Genetics; 2017;205:
603 1537–1550. doi:10.1534/genetics.116.196451
- 604 37. Rives AF, Rochlin KM, Wehrli M, Schwartz SL, DiNardo S. Endocytic trafficking of
605 Wingless and its receptors, Arrow and DFrizzled-2, in the *Drosophila* wing. Dev Biol.
606 Academic Press; 2006;293: 268–283. doi:10.1016/j.ydbio.2006.02.006
- 607 38. Martin A, Papa R, Nadeau NJ, Hill RI, Counterman BA, Halder G. Diversification of
608 complex butterfly wing patterns by repeated regulatory evolution of a Wnt ligand. Proc
609 Natl Acad Sci U S A. 2012;109: 12632–12637. doi:doi.org/10.1073/pnas.1204800109
- 610 39. Mazo-Vargas A, Concha C, Livraghi L, Massardo D, Wallbank RWR, Zhang L, et al.
611 Macroevolutionary shifts of WntA function potentiate butterfly
612 wing-pattern diversity. Proc Natl Acad Sci. 2017;114: 10701 LP – 10706.
- 613 40. Van Belleghem SM, Rastas P, Papanicolaou A, Martin SH, Arias CF, Supple MA, et al.
614 Complex modular architecture around a simple toolkit of wing pattern genes. Nat Ecol
615 Evol. 2017;1: 0052. doi:10.1038/s41559-016-0052
- 616 41. Smith DAS. Evidence for autosomal meiotic drive in the butterfly *Danaus chrysippus* L.
617 Heredity. 1976;36: 139–142.
- 618 42. Coughlan JM, Willis JH. Dissecting the role of a large chromosomal inversion in life
619 history divergence throughout the *Mimulus guttatus* species complex. Mol Ecol. John
620 Wiley & Sons, Ltd (10.1111); 2019;28: 1343–1357. doi:10.1111/mec.14804
- 621 43. Gordon IJ, Ileri P, Smith DAS. Preference for isolated host plants facilitates invasion of
622 *Danaus chrysippus* (Linnaeus, 1758) (Lepidoptera: Nymphalidae) by a bacterial male-
623 killer *Spiroplasma*. Austral Entomol. 2015;54: 210–216. doi:10.1111/aen.12113
- 624 44. Balbuena JA, Míguez-Lozano R, Blasco-Costa I. PACo: A Novel Procrustes Application
625 to Cophylogenetic Analysis. Moreau CS, editor. PLoS One. Public Library of Science;
626 2013;8: e61048. doi:10.1371/journal.pone.0061048

- 627 45. Richardson MF, Weinert LA, Welch JJ, Linheiro RS, Magwire MM, Jiggins FM, et al.
628 Population Genomics of the Wolbachia Endosymbiont in *Drosophila melanogaster*. PLoS
629 Genet. 2012;8. doi:10.1371/journal.pgen.1003129
- 630 46. Jiggins FM. Male-killing Wolbachia and mitochondrial DNA: selective sweeps, hybrid
631 introgression and parasite population dynamics. Genetics. 2003;164: 5–12.
- 632 47. Hurst GDD, Jiggins FM. Problems with mitochondrial DNA as a marker in population,
633 phylogeographic and phylogenetic studies: The effects of inherited symbionts. Proc R Soc
634 B Biol Sci. 2005;272: 1525–1534. doi:10.1098/rspb.2005.3056
- 635 48. Traut W, Ahola V, Smith DAS, Gordon IJ, French-Constant RH. Karyotypes versus
636 Genomes: The Nymphalid Butterflies *Melitaea cinxia*, *Danaus plexippus*, and *D.*
637 *chrysippus*. Cytogenet Genome Res. 2018;153: 46–53. doi:10.1159/000484032
- 638 49. Wright AE, Dean R, Zimmer F, Mank JE. How to make a sex chromosome. Nature
639 Communications. Nature Publishing Group; 2016. p. 12087. doi:10.1038/ncomms12087
- 640 50. Fay JC, Wyckoff GJ, Wu C-I. Positive and Negative Selection on the Human Genome.
641 Genetics. 2001;158: 1227–1234.
- 642 51. Smith DAS, Owen DF. Colour genes as markers for migratory activity: The butterfly
643 *Danaus chrysippus* in Africa. Oikos. 1997;78: 127–135.
- 644 52. Jiggins FM, Hurst GDD, Majerus MEN. Sex-ratio-distorting Wolbachia causes sex-role
645 reversal in its butterfly host. Proc R Soc B Biol Sci. 2000;267: 69–73.
646 doi:10.1098/rspb.2000.0968
- 647 53. Hornett EA, Charlat S, Duplouy AMR, Davies N, Roderick GK, Wedell N, et al.
648 Evolution of male-killer suppression in a natural population. PLoS Biol. 2006;4: 1643–
649 1648. doi:10.1371/journal.pbio.0040283
- 650 54. Wilfert L, Jiggins FM. The dynamics of reciprocal selective sweeps of host resistance and
651 a parasite counter-adaptation in *Drosophila*. Evolution. Wiley/Blackwell (10.1111);
652 2013;67: 761–773. doi:10.1111/j.1558-5646.2012.01832.x
- 653 55. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A
654 New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J
655 Comput Biol. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY
656 10801 USA ; 2012;19: 455–477. doi:10.1089/cmb.2012.0021
- 657 56. Prysycz LP, Gabaldón T. Redundans: an assembly pipeline for highly heterozygous
658 genomes. Nucleic Acids Res. Oxford University Press; 2016;44: e113–e113.
659 doi:10.1093/nar/gkw294

- 660 57. Huang S, Kang M, Xu A. HaploMerger2: rebuilding both haploid sub-assemblies from
661 high-heterozygosity diploid genome assembly. Berger B, editor. Bioinformatics. Oxford
662 University Press; 2017;33: 2577–2579. doi:10.1093/bioinformatics/btx220
- 663 58. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: *de novo* assembly of organelle
664 genomes from whole genome data. Nucleic Acids Res. Oxford University Press; 2016;45:
665 gkw955. doi:10.1093/nar/gkw955
- 666 59. Dasmahapatra KK, Walters JR, Briscoe AD, Davey JW, Whibley A, Nadeau NJ, et al.
667 Butterfly genome reveals promiscuous exchange of mimicry adaptations among species.
668 Nature. Nature Publishing Group; 2012;487: 94–8. doi:10.1038/nature11041
- 669 60. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the
670 Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic
671 Acids Res. Oxford University Press; 2014;42: D206–D214. doi:10.1093/nar/gkt1226
- 672 61. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server:
673 Rapid Annotations using Subsystems Technology. BMC Genomics. BioMed Central;
674 2008;9: 75. doi:10.1186/1471-2164-9-75
- 675 62. Zhan S, Zhang W, Niitpõld K, Hsu J, Haeger JF, Zalucki MP, et al. The genetics of
676 monarch butterfly migration and warning colouration. Nature. 2014;514: 317–21.
677 doi:10.1038/nature13812
- 678 63. Lunter G, Goodson M. Stampy: a statistical algorithm for sensitive and fast mapping of
679 Illumina sequence reads. Genome Res. 2011;21: 936–9. doi:10.1101/gr.111120.110
- 680 64. DePristo M a, Banks E, Poplin R, Garimella K V, Maguire JR, Hartl C, et al. A framework
681 for variation discovery and genotyping using next-generation DNA sequencing data. Nat
682 Genet. 2011;43: 491–8. doi:10.1038/ng.806
- 683 65. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A,
684 et al. From fastQ data to high-confidence variant calls: The genome analysis toolkit best
685 practices pipeline. Curr Protoc Bioinforma. Hoboken, NJ, USA: John Wiley & Sons, Inc.;
686 2013;43: 11.10.1-11.10.33. doi:10.1002/0471250953.bi1110s43
- 687 66. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation
688 PLINK: rising to the challenge of larger and richer datasets. Gigascience. Narnia; 2015;4:
689 7. doi:10.1186/s13742-015-0047-8
- 690 67. Bryant D, Moulton V. Neighbor-Net: An Agglomerative Method for the Construction of
691 Phylogenetic Networks. Mol Biol Evol. 2003;21: 255–265. doi:10.1093/molbev/msh018
- 692 68. Huson DH, Bryant D. Application of Phylogenetic Networks in Evolutionary Studies. Mol
693 Biol Evol. 2006;23: 254–267. doi:10.1093/molbev/msj030

- 694 69. Delaneau O, Zagury J-F, Marchini J. Improved whole-chromosome phasing for disease
695 and population genetic studies. *Nat Methods*. 2013;10: 5–6. doi:10.1038/nmeth.2307
- 696 70. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of
697 genomes. *Nat Methods*. 2012;9: 179–181. doi:10.1038/nmeth.1785
- 698 71. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms
699 and methods to estimate maximum-likelihood phylogenies: assessing the performance of
700 PhyML 3.0. *Syst Biol*. 2010;59: 307–21. doi:10.1093/sysbio/syq010
- 701 72. Bouckaert RR. DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics*.
702 2010;26: 1372–3. doi:10.1093/bioinformatics/btq110
- 703 73. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, et al. BEAST 2: A
704 Software Platform for Bayesian Evolutionary Analysis. Prlic A, editor. *PLoS Comput*
705 *Biol*. Public Library of Science; 2014;10: e1003537. doi:10.1371/journal.pcbi.1003537
- 706 74. Keightley PD, Pinharanda A, Ness RW, Simpson F, Dasmahapatra KK, Mallet J, et al.
707 Estimation of the Spontaneous Mutation Rate in *Heliconius melpomene*. *Mol Biol Evol*.
708 2015;32: 239–243. doi:10.1093/molbev/msu302
- 709 75. Owen DF, Chanter DO. Population biology of tropical African butterflies. Sex ratio and
710 genetic variation in *Acraea encedon*. *J Zool*. 1969;157: 345–374. doi:10.1111/j.1469-
711 7998.1969.tb01707.x
- 712 76. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in
713 Bayesian Phylogenetics Using Tracer 1.7. Susko E, editor. *Syst Biol*. 2018;67: 901–904.
714 doi:10.1093/sysbio/syy032

715 **Acknowledgements**

716 We are grateful to Godfrey Amoni Etelej, Laura Hebberecht-Lopez and Glennis Julian for
717 support with butterfly rearing. We thank Jenny York, Frank Jiggins and Deborah Charlesworth
718 helpful comments. This work was funded by European Research Council (ERC) European Union
719 Horizon 2020 research and innovation programme grant 646625 (CB), ERC grant 339873 (CDJ,
720 SHM), National Geographic Society Research Grant WW-138R-17 (IJG), and a Royal Society
721 University Research Fellowship (SHM).