

# Diverse genetic architectures on the Z chromosome underlie the two rules of speciation in *Papilio* butterfly hybrids

Tianzhu Xiong<sup>1,\*</sup>, Shreeharsha Tarikere<sup>1,4</sup>, Neil Rosser<sup>1</sup>, Xueyan Li<sup>2</sup>, Masaya Yago<sup>3</sup>, and James Mallet<sup>1</sup>

<sup>1</sup>*Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA*

<sup>2</sup>*State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China*

<sup>3</sup>*The University Museum, The University of Tokyo, Tokyo, Japan*

<sup>4</sup>*Current Address: Department of Earth Sciences, Carleton University, Ottawa, ON, Canada*

\*Corresponding Author: txiong@g.harvard.edu

October 28, 2022

## Abstract

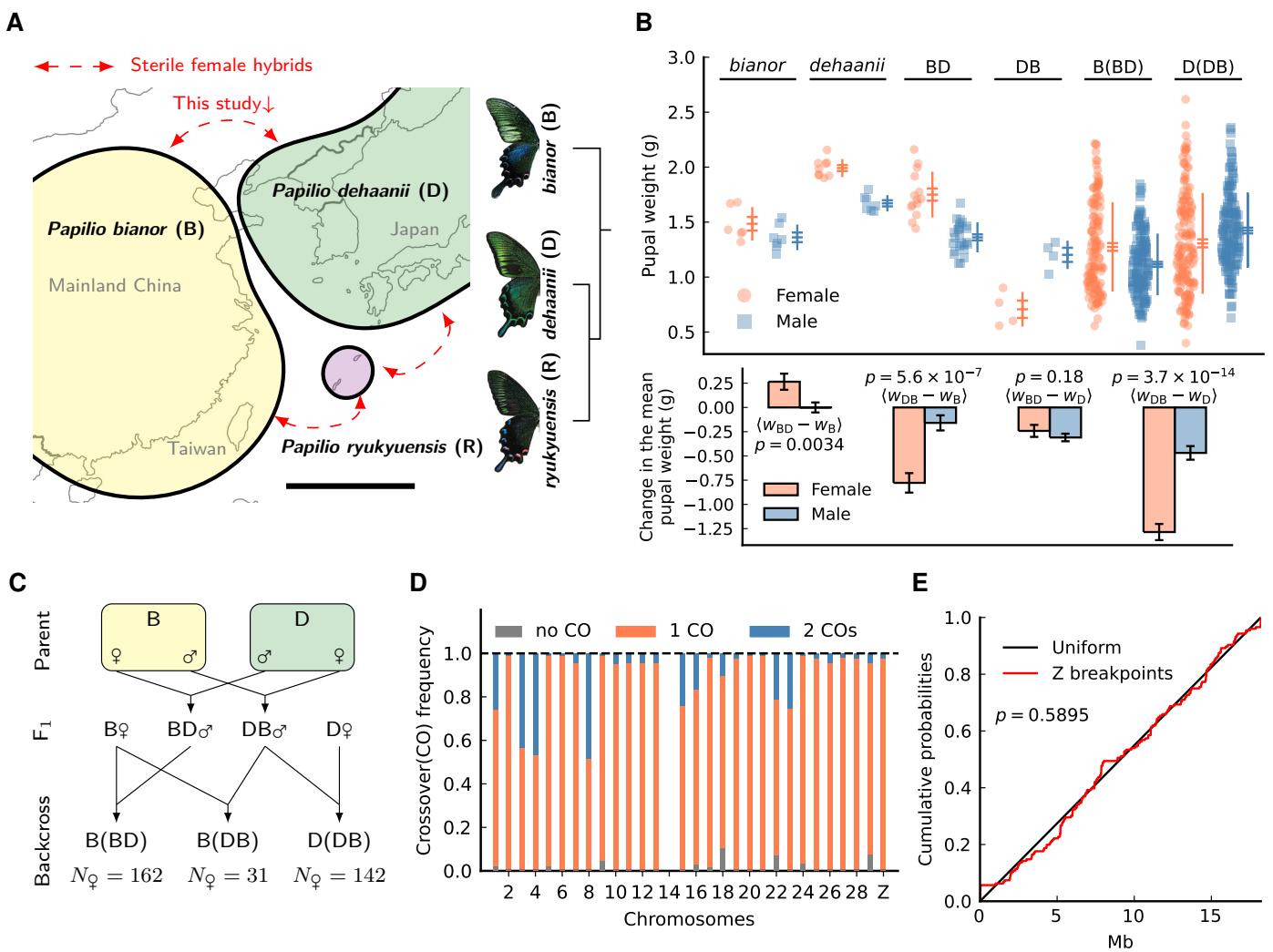
Two empirical rules arise from the incompatibility of interspecific hybrids: Haldane's Rule predicts that the chromosomally heterogametic sex (XY or ZW) is more unfit after hybridization; the large-X/Z effect posits that sex chromosomes play a major role in incompatibility. Classical theories on these two rules rely on evidence mainly from taxa with male heterogamety, while female heterogamety received little investigation. Here, we reveal the genetic architectures of the two rules in hybrids between the butterflies *Papilio bianor* and *Papilio dehaanii*, where the female is the heterogametic sex. In these crosses, hybrid females suffer from both body size abnormality and ovary dysgenesis, while males appear normal and fertile. Curiously, abnormal size in females is mapped to a continuum of Z-linked polygenes, each acting quantitatively with small phenotypic effects. This polygenic system, perhaps spanning the entire Z chromosome, also correctly predicts weaker incompatibility effects in males. For ovary dysgenesis, the underlying genetic architecture can be monogenic or polygenic with different maternal backgrounds. Most peculiarly, when comparing ovary dysgenesis in certain maternal backgrounds between *Papilio* and *Heliconius*, we find that F<sub>1</sub> recombination on the Z chromosome often rescues incompatibilities among backcross individuals, while a non-recombined Z chromosome almost always produces strong ovary defects regardless of ancestry. These results suggest that high fitness in these maternal backgrounds requires a balance between the total quantities of introgression on autosomes and the Z chromosome. Our study highlights that, in addition to incompatibility factors with large effects, genomically dispersed polygenes are also abundant in creating butterfly reproductive isolation.

## 32 1. Introduction

33 Speciation is a complex and a stochastic process, yet it obeys empirical rules across sexually repro-  
34 ducing lineages [1]. One such rule (Haldane's Rule) states that among the offspring between differ-  
35 ent species, the sex with a greater fitness cost is the heterogametic sex [2, 3]. Another rule of spe-  
36 ciation is the so-called large-X/Z effect on hybrid incompatibility [1, 4]. Haldane's Rule is entirely  
37 phenomenological, but it holds firmly across many phylogenetically diverse organisms [5, 6, 7].  
38 Whether its wide pertinence emerges from a common set of genetic mechanisms is an open ques-  
39 tion [8]. The large-X/Z effect has also been suggested as a robust signature, but its evidence is less  
40 direct unless incompatibility factors can be genetically mapped [9, 10, 11, 12, 13].

41 The most compelling evidence on the genetic and molecular basis of Haldane's Rule comes from  
42 studies in mammals and *Drosophila*, both of which have X/Y sex determination, and in which hybrid  
43 males develop more incompatibility [14, 15]. These findings converge upon several genetic mech-  
44 anisms of Haldane's Rule: First, dominance theory posits that the single X chromosome in males  
45 might expose recessive genes that are deleterious in a hybrid genetic background, thus increasing  
46 the likelihood of incompatibility [4, 16]. Second, the heterogametic sex might speed up the evolution  
47 of sex chromosomes, and Haldane's Rule can be produced from potentially a combination of pro-  
48 cesses involving haploid selection (faster-X theory) [17], sex-specific selection (faster-male theory)  
49 [18], and sex chromosome conflict (meiotic drive theory) [19, 20]. Accelerated sex chromosome evo-  
50 lution also provides a natural explanation for the large-X/Z effect. However, as these mechanisms  
51 have direct evidence mostly from male-heterogametic taxa, whether they offer a general explana-  
52 tion for the two rules of speciation or are specific to each system is unknown. For instance, it has  
53 been suggested that the intrinsic sensitivity of spermatogenesis to hybrid disruption can explain a  
54 higher incidence of male sterility [21], but this is not applicable to female-heterogametic taxa.

55 In Lepidoptera (butterflies and moths), the female is the heterogametic sex (with Z/W sex chro-  
56 mosomes), and hybrid females are more prone to abnormal phenotypes than males [22]. Knowing  
57 the genetic basis of this direction of the two rules will allow us to test if proposed genetic mech-  
58 anisms based on male-heterogamety can be extended to explain the same phenomenon in female-  
59 heterogametic taxa. To date, little is known about the genomic basis of hybrid incompatibilities in  
60 Lepidoptera, except for a few studies using sparse genetic markers [23, 24, 25] and one recent whole-  
61 genome QTL study of *Heliconius* ovary dysgenesis [26]. Nonetheless, these studies robustly show  
62 that disruption is often linked to the Z chromosome, conforming to the large-X/Z effect. Here, we  
63 investigate hybrid incompatibility between a pair of Asian swallowtail butterflies, *Papilio bianor* and  
64 *Papilio dehaanii* that produce abnormal females upon hybridization. The two focal species are closely  
65 related and are almost parapatric in NW China (Fig. 1A). All interspecific crosses in this species com-  
66 plex, including the island species *P. ryukyuensis*, produce fertile males but sterile females [27, 28]. In  
67 addition, abnormal body size is also frequently observed in hybrids among related *Papilio* species  
68 [29]. To test for the basis of hybrid incompatibility in this system, we investigate the genetic archi-



**Figure 1:** The study system. **(A)** The distribution and species relationship of the *Papilio bianor* complex. Scale bar=1000km. **(B)** Top: Pupal weight variation among pure species, F<sub>1</sub> hybrids, and backcross individuals. Horizontal bars in each category represent the mean and its standard error, while the vertical line represents the mean +/- standard deviation. Bottom: changes in the mean pupal weight between F<sub>1</sub> and pure species in each sex. Error bars are standard errors of  $\langle W_{F1} - W_{\text{pure}} \rangle$ . We test if changes are significantly different between males and females of the same comparison, and report its *p*-value by calculating Z-score. **(C)** The cross scheme. **(D)** The crossover distribution in F<sub>1</sub> males. Chromosome 14 is excluded due to intrachromosomal assembly problems. **(E)** The cumulative distribution of recombination breakpoints along the Z chromosome is indistinguishable from a uniform distribution (*p*-value is from a Kolmogorov-Smirnov test).

lecture underlying two important traits: body size (pupal weight) and female reproduction (ovary dysgenesis).

## 71 2. Results and Discussion

### 72 2.1 Haldane's Rule between *P. bianor* and *P. dehaanii* involves asymmetrically 73 inherited elements

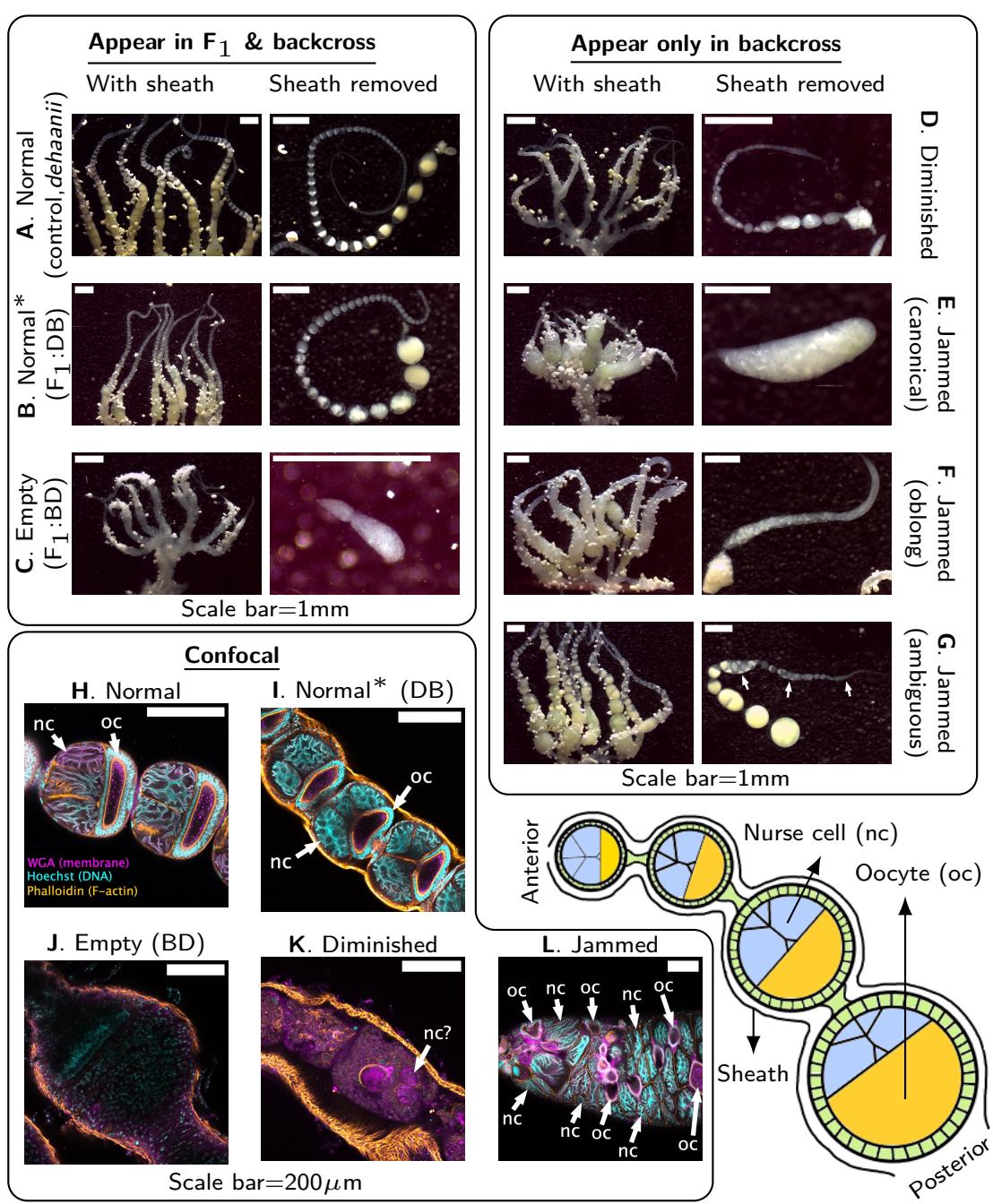
74 To investigate phenotypic abnormalities among interspecific hybrids, we produced reciprocal F<sub>1</sub>  
75 crosses between the two species, and then constructed three types of backcrosses by mating F<sub>1</sub>  
76 males with pure females (Fig. 1C). Cross notations follow the order (female × male). For instance,  
77 “B(BD)” is equivalent to “*bianor* ♀ × (*bianor* ♀ × *dehaanii* ♂) ♂”, where “B” and “D” stand for *bianor*  
78 and *dehaanii*, respectively.

79 To study body size, pupal weight (denoted as W) within three days of pupation is treated as a  
80 proxy for size. We find that F<sub>1</sub> females are significantly smaller than pure females from both species  
81 when the mother is *dehaanii* (cross direction DB), but they are within the range of pure females if  
82 the mother is *bianor* (cross direction BD) (Fig. 1B, top). For F<sub>1</sub> males, the change in pupal weight  
83 (compared to pure males) is smaller or equal to that of F<sub>1</sub> females (Fig. 1B, bottom). This result  
84 suggests that abnormal size in F<sub>1</sub> hybrids is much more prominent in females. We therefore interpret  
85 female-biased F<sub>1</sub> abnormal size in a broad sense as an example of Haldane's Rule. Since reciprocal  
86 crosses affect size in opposite ways, asymmetrically inherited genetic elements must be involved.

87 To study female reproduction in hybrids, we dissected ovaries across the entire pedigree and  
88 determined all major ovary phenotypes (Fig. 2). To our surprise, while F<sub>1</sub> females with a *bianor*  
89 mother have almost empty ovaries and cannot lay eggs (Fig. 2C,J), F<sub>1</sub> females in the reciprocal cross  
90 develop and lay superficially normal eggs (Fig. 2B,I). However, these eggs do not hatch even when  
91 the female is mated, and the distribution of yolk in a follicle is also slightly different from that of  
92 control (compare sheath-removed figures in Fig. 2A and Fig. 2B). As we only determine ovary phe-  
93 notypes rather than female fertility *per se*, ovaries with regularly spaced and spherical follicles are  
94 subsequently all classified as “Normal”, regardless of actual fertility. This is a reasonable decision as  
95 variation in ovary phenotype is often large and can be easily scored under a microscope (Fig. 2A-L).  
96 Overall, ovary phenotypes are strongly asymmetric between the reciprocal crosses, which suggests  
97 the involvement of asymmetrically inherited genetic elements in Haldane's Rule of this system. For  
98 this reason, we map phenotypes conditioning on a fixed maternal background.

### 99 2.2 Male meiotic crossover on the Z chromosome is obligatory, spatially uniform, 100 and with strong interference

101 Recombination determines the ancestry of a hybrid chromosome and is crucial to our subsequent  
102 analysis. To infer haplotypes and crossover patterns across the pedigree, we perform whole-genome  
103 low-coverage sequencing in backcross individuals (~1×), while F<sub>1</sub> and parents are sequenced to  
104 higher depths (>5× and >30×, respectively). Prior to crossover analysis, we use linkage information  
105 from all families to manually correct assembly errors in the original reference genome of *P. bianor* [30]



**Figure 2:** Ovary phenotypes. Monochrome confocal images for each channel are in Fig. S1 and S2 (WGA: membranes; Hoechst: DNA; Phalloidin: F-actin). A schematic diagram of butterfly ovarioles is shown at bottom right. (A,H) Phenotype Normal in a pure individual. The confocal image shows follicles with their sheath removed. (B,I) Phenotype Normal in F<sub>1</sub> females (cross direction DB). Sheath is retained in the confocal image. Follicles are mostly indistinguishable from those of pure females, but the distribution of yolk in late-stage oocytes is overly scattered. It is not classified as a separate phenotype. (C,J) Phenotype Empty in F<sub>1</sub> females (cross direction BD). Sheath is retained in the confocal image, and very little tissue remains inside the sheath. (D,K) Phenotype Diminished. Sheath is retained in the confocal image, and a substantial amount of tissue remains, but without any discernible structures of normal follicles. (E-G,L) Phenotype Jammed (including its subtypes). Sheath is removed in the confocal image, and follicle cells are merged into a tube-like structure with many nurse cells and oocytes. Subtype canonical: the entire ovariole collapses into a single bag of oocytes and nurse cells. Subtype oblong: similar to “canonical”, but the ovariole is elongated. Subtype ambiguous: some parts of the ovariole have merged follicles; other parts contain isolated follicles.

106 except for chromosome 14, where our manual correction cannot resolve all visible errors (Materials  
107 and Methods). As such errors affect the inference of recombination breakpoints, we do not report  
108 crossover patterns on chromosome 14.

109 We infer the crossover pattern in  $F_1$  males by counting estimated recombination breakpoints  
110 across all backcross offspring (female meiosis in Lepidoptera is achiasmatic). Most  $F_1$  males have  
111 at least one crossover per chromosome pair per meiosis, but the degree of crossover interference  
112 varies among chromosomes (Fig. 1D). Double crossovers occur as frequently as over 40% (chromosome 8),  
113 but are almost absent in many chromosomes. Importantly, the Z chromosome in this  
114 system has almost no double crossovers, and its recombination breakpoints are approximately uni-  
115 formly distributed along the chromosomal axis (Fig. 1E). This indicates that recombination on the Z  
116 chromosome can be approximated by a model where the first obligatory crossover occurs uniformly  
117 at random along the chromosome, while the second crossover is permanently suppressed. We apply  
118 this model to the study of abnormal size below.

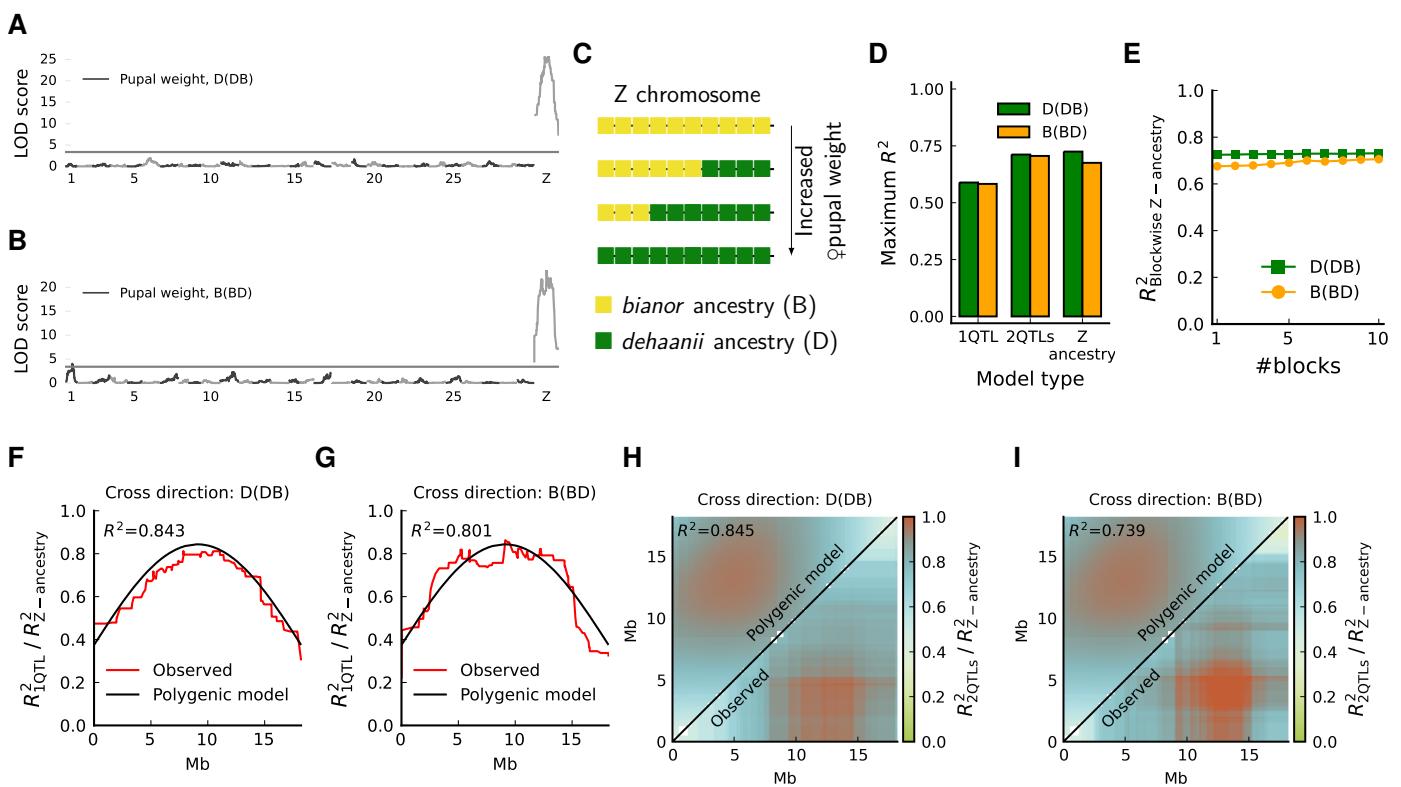
### 119 2.3 Abnormal size is determined by the entire Z chromosome

120 Pupal weight in backcross individuals (measured in backcrosses of type B(BD) and D(DB)) has a  
121 broad distribution (Fig. 1B). At first sight, such a distribution might be easily associated with many  
122 loci across the genome or extreme developmental/environmental stochasticity. Curiously, the Z  
123 chromosome is the only chromosome controlling pupal weight variation among backcross females  
124 in either direction, and no autosomal locus is significantly associated (Fig. 3A,3B). A single-locus  
125 QTL scan (1-QTL) on the Z chromosome can explain over 50% of the phenotypic variance, and a  
126 two-locus additive QTL scan (2-QTL) can explain over 70% (Fig. 3D), so variation of abnormal size  
127 in hybrid females is determined mostly by the Z chromosome.

128 Rather than inferring putative QTL, we argue that this genetic architecture is best described by a  
129 Z-chromosome continuum where many quantitative effects of small magnitude are uniformly dis-  
130 tributed along the chromosomal axis (Fig. 3C). We refer to this as the Z-spanning polygenic system.  
131 Since different parts of the chromosome have identical effects under this assumed architecture, pu-  
132 pal weight will depend only on the average Z chromosome ancestry,  $f_D$  and  $f_B$ , corresponding to the  
133 fraction of the Z chromosome inherited from *dehaanii* and *bianor*, respectively. If effects are additive,  
134 this polygenic model is simply

$$\begin{aligned} W_{B(BD)} &= w_B + \alpha_D f_D + \epsilon \\ W_{D(DB)} &= w_D - \alpha_B f_B + \epsilon \end{aligned} \tag{1}$$

135 where  $W$  is the pupal weight of a backcross female,  $w$  is the intercept,  $\alpha$  is the net effect of Z-linked  
136 introgression, and  $\epsilon$  contains non-genetic fluctuation. We will use  $f$  throughout the paper to refer to  
137 the fraction of introgression, and its subscript specifies the context. The relationship between pupal  
138 weight and average Z chromosome ancestry in our data is shown in Fig. S9.



**Figure 3:** Analysis of abnormal size. **(A, B)** The LOD score of genome-wide 1-QTL scans on female pupal weight. **(C)** The Z-spanning polygenic model posits that every infinitesimal introgression on the Z chromosome has a fixed infinitesimal effect on pupal weight. **(D)** The highest explanatory power ( $R^2$ ) of 1-QTL regressions, 2-QTL regressions, and the regression to the average ancestry on the Z chromosome. **(E)** Dividing the Z chromosome into multiple blocks and performing multiple regression does not strongly improve the explanatory power. **(F, G)** The relative loss of predictive power in a 1-QTL scan on the Z chromosome. **(H, I)** The relative loss of predictive power in a 2-QTL scan on the Z chromosome.

Initial evidence supporting this polygenic architecture is as follows. First, this polygenic model (Equation 1) achieves a similar or even better explanatory power than the best QTL models (Fig. 3D), suggesting that perhaps the entire Z chromosome contributes to the phenotype. Second, to test if different parts of the Z chromosome could have different effects, we divide the Z chromosome into  $K$  equal-sized blocks ( $K = 1, 2, \dots, 10$ ), then perform multiple regression between pupal weight and the vector containing average ancestries on each of the  $K$  blocks. One should expect the total explanatory power to rise significantly with  $K$  if different blocks have very different effects due to increased degrees of freedom in regression. However, we find that using a single block (the entire chromosome) is almost as powerful as using multiple blocks (Fig. 3E). Effects are thus likely distributed evenly on the Z chromosome. This reasoning might seem at odds with the fact that both 1-QTL and 2-QTL scans peak in the interior of the Z chromosome (Fig. S8). However, we show below that such dome-shaped QTL results are exactly predicted by the proposed Z-spanning polygenic system.

## 152 2.4 Strong evidence for the Z-spanning polygenic architecture

153 The reason that 1-QTL scans produce peaks near the center of the chromosome is that the backcross  
154 generation still has long blocks of identical local ancestry. Since the chromosome center is the region  
155 that is the least distant from all other positions on the chromosome, and the crossover process on  
156 the Z chromosome is approximately spatially uniform, local ancestry at the chromosome center will  
157 thus correlate best with average chromosomal ancestry ( $f_D$ ,  $f_B$ ). Consequently, markers near the  
158 center will also be the most informative about any phenotype determined by the proposed poly-  
159 genic model (Equation 1), while markers away from the center will lose phenotypic information.  
160 This is analogous to the “mid-domain effect” in species richness gradients [31, 32]. To quantify how  
161 much phenotypic information is lost by assuming a single marker, we calculate the relative loss of  
162 predictive power between a 1-QTL scan on a specific marker versus the polygenic model (Materials  
163 and Methods):

$$\frac{R_{\text{1QTL}}^2}{R_{\text{Z-ancestry}}^2} = \frac{3}{8} \left[ 1 + 2l \left( 1 - l \right) \right]^2, \quad (2)$$

164 where  $l$  ( $0 \leq l \leq 1$ ) is the relative marker position on the Z chromosome. Since  $R_{\text{Z-ancestry}}^2$  is a  
165 constant, this equation has no degree of freedom besides  $l$ , and it depicts precisely a dome-shaped  
166 predictive power of a 1-QTL scan peaking at the chromosome center ( $l = 0.5$ ). It provides a good  
167 fit to observed 1-QTL results in both cross directions (Fig. 3F,3G). Similarly, for a 2-QTL scan, the  
168 relative loss of predictive power for two markers is

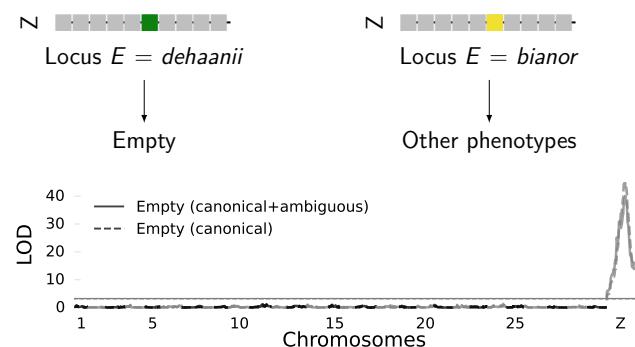
$$\begin{aligned} \frac{R_{\text{2QTL}}^2}{R_{\text{Z-ancestry}}^2} &= \frac{1}{8 - 4|l_1 - l_2|} \left\{ 6 \left| l_1 - l_2 \right| \left( l_1 + l_2 - 1 \right)^2 \right. \\ &\quad \left. + 3 \left[ 1 + 2l_1 \left( 1 - l_1 \right) \right] \left[ 1 + 2l_2 \left( 1 - l_2 \right) \right] \right\}, \end{aligned} \quad (3)$$

169 where  $l_1$  and  $l_2$  are the relative positions of two markers on the Z chromosome. This equation shows  
170 that the two most informative markers are located near  $l_1 \approx 0.27$  and  $l_2 \approx 0.73$ —about a quarter  
171 into the chromosome from both ends. Again, the equation is a good fit to the results of 2-QTL scans  
172 (Fig. 3H,3I). From this, we speculate that, by calculating the same quantity for 3-QTL scans, this  
173 polygenic model will still be a good fit, and it is confirmed in Fig. S10.

174 Despite being polygenic, the proposed architecture is mathematically parsimonious because it is  
175 a linear model with just one predictive variable (mean chromosomal ancestry) (Eq. 1), while still  
176 carrying almost the complete spatial information of QTL scans. Thus, we conclude that abnormal  
177 size is best explained by many Z-linked polygenes, each having small phenotypic effects, evenly  
178 smeared across the entire chromosome.

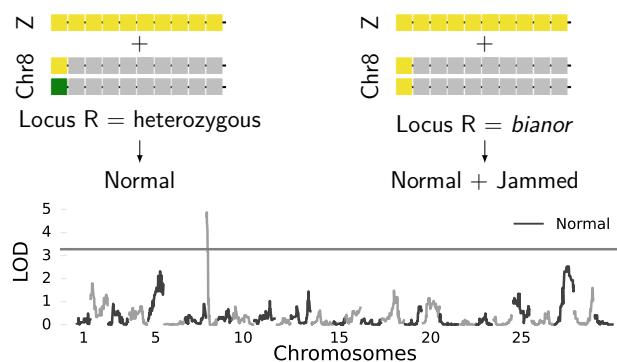
### A Cross direction: B(BD) + B(DB)

**Condition:** All individuals ( $N = 193$ )



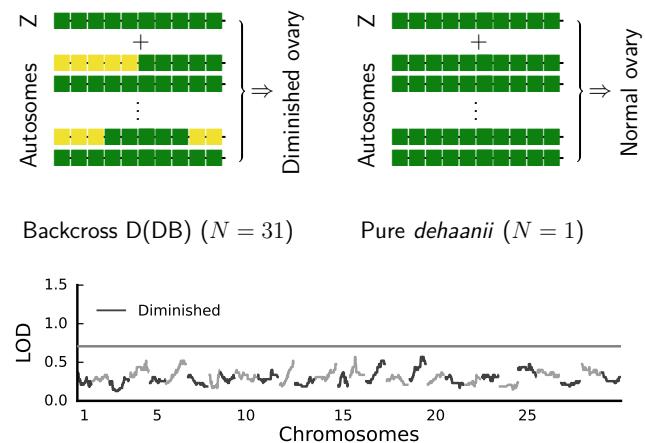
### B Cross direction: B(BD) + B(DB)

**Condition:**  $Z = \text{pure } bianor$  ( $N = 77$ )



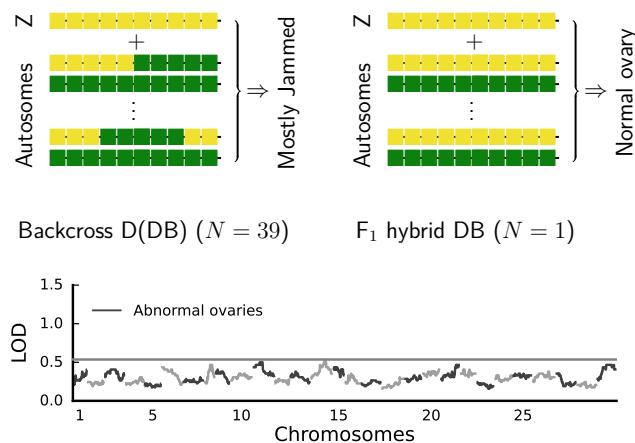
### C Cross direction: D(DB) + D

**Condition:**  $Z = \text{pure } dehaanii$  ( $N = 32$ )



### D Cross direction: D(DB) + DB

**Condition:**  $Z = \text{pure } bianor$  ( $N = 40$ )



**Figure 4:** Analysis of ovary dysgenesis. For maternally *bianor* hybrids: (A) Phenotype Empty is dominantly controlled by Locus *E* on the Z chromosome. The LOD plot shows both the score for the canonical Empty phenotype as well as the score when we include a few ambiguous individuals that are classified as Empty. (B) If the Z chromosome is purely *bianor*, introgression on Locus *R* suppresses abnormal phenotypes. For maternally *dehaanii* hybrids: (C) If the Z chromosome is purely *dehaanii* in backcross females, ovaries develop the Diminished phenotype. (D) If the Z chromosome is purely *bianor* in backcross females, ovaries are mostly Jammed.

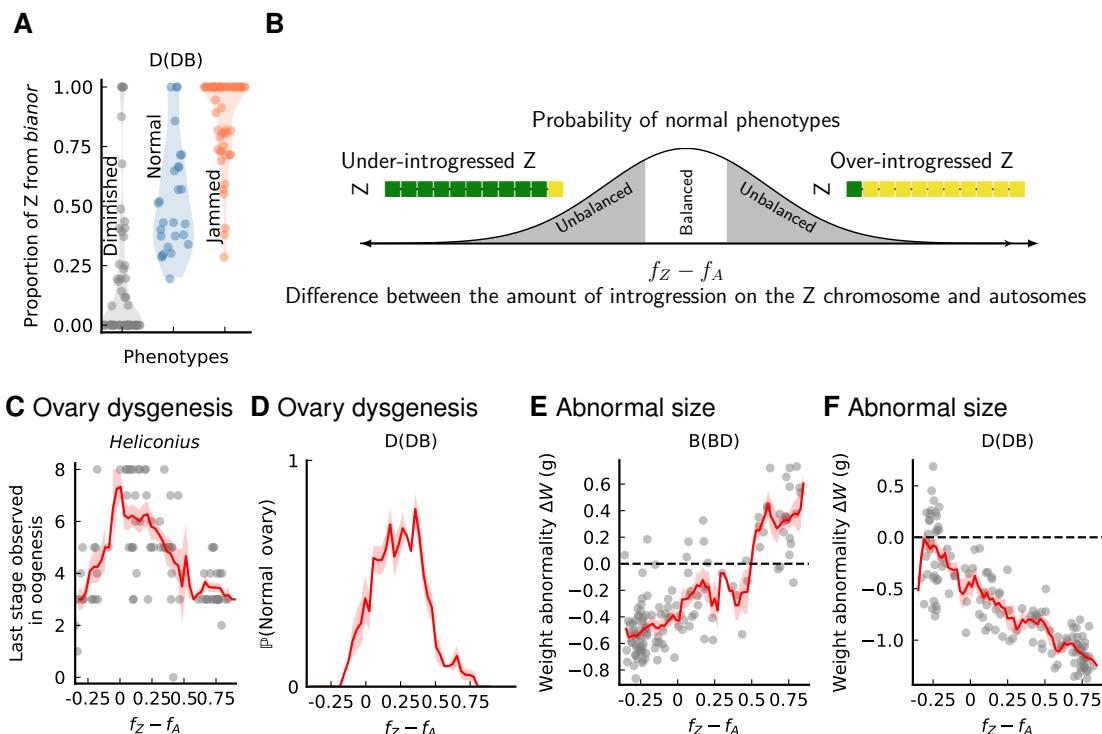
## 2.5 A Z-linked system shapes ovary dysgenesis in maternally *bianor* hybrids

In F<sub>1</sub> females of type BD, ovaries are small and empty (phenotype Empty), and only a trace amount of tissue remains in each ovariole (Fig. 2C,J). The locus (*E*) responsible for Empty is mapped to a 0.8-1Mb interval on the Z chromosome with a simple genetic mechanism (Fig. 4A): the *dehaanii* allele at locus *E* is sufficient to induce Empty regardless of other genetic variation across the genome.

Other ovary phenotypes in this maternal background include Normal and various types of Jammed, where multiple oocytes and a myriad of nurse cells congregate into a single tube-shaped follicle (Fig. 2E-G,L). We remove the effect of locus *E*, together with other possible Z-linked factors by selecting individuals where the Z chromosome has no introgression from *dehaanii*. Conditioning on this subset of individuals, we further identified a locus (*R*) on chromosome 8 that suppresses pheno-

189 type Jammed when it is heterozygous in ancestry (Fig. 4B). Conversely, if locus *R* has a homozygous  
 190 *bianor* genotype, approximately half of the offspring develop Jammed, but the sample size becomes  
 191 too low to map its associated loci. Overall, the genetic basis of ovary dysgenesis in maternally *bianor*  
 192 hybrids is controlled mainly by a small number of genomic regions with large effects.

193 **2.6 Recombination on the Z chromosome rescues ovary dysgenesis in maternally  
 194 *dehaanii* backcrosses**



**Figure 5:** The relationship between phenotypic defect and chromosomal ancestry. Across panels,  $f_Z$  and  $f_A$  are the proportions of introgressed ancestry on the Z chromosome and autosomes, respectively. **(A)** Average Z chromosome ancestry is very informative in ovarian variation in cross direction D(DB). **(B)** The “ancestry-balance” model posits that the probability of observing normal phenotypes is highest when the amount of introgression on autosomes and the Z chromosome reaches a critical balance (not necessarily equal). **(C)** Between *Heliconius pardalinus butleri* and *H. p. sergestus*, oogenesis in maternally *butleri* backcross females is better when Z-linked introgression is at intermediate levels. **(D)** The probability of observing the Normal phenotype conditioning on the difference between  $f_Z$  and  $f_A$  in cross direction D(DB). **(E)** In cross direction B(BD), female pupal weight is less different from that of pure *bianor* females when Z-linked introgression is at intermediate levels. **(F)** In cross direction D(DB), deviation of female pupal weight from that of pure *dehaanii* females is smaller with less amount of Z-linked introgression.

195 In the opposite maternal background ( $F_1$  and backcrosses with *dehaanii* mothers), ovary dysge-  
 196 nesis yields a number of counter-intuitive results. In this maternal background, ovaries fall into  
 197 three phenotypic categories: Diminished, Normal, and Jammed (including its subtypes). Overall,  
 198 the Z chromosome is still the most predictive chromosome with significant LOD peaks (Fig. S11).  
 199 However, two pieces of evidence suggest ovary phenotypes are influenced also by autosomes. First,  
 200 if the Z chromosome has pure *dehaanii* ancestry, backcross females all have the Diminished pheno-

201 type (Fig. 4C,5A). Comparing these individuals to pure *dehaanii* females show that autosome-only  
202 introgression from *bianor* at the level of a backcross individual is sufficient to induce Diminished.  
203 In other words, Diminished is stably caused by the existence of certain levels of introgression on  
204 autosomes, rather than the exact location of introgression, which is evident from autosomal LOD  
205 scores (Fig. 4C). Second, if the Z chromosome has pure *bianor* ancestry, almost all backcross females  
206 develop abnormal ovaries as well, and most of them have the Jammed phenotype (Fig. 4D,5A). Simi-  
207 larly, comparing these individuals against F<sub>1</sub> females of type DB (which have Normal ovaries and a  
208 Z chromosome inherited from *bianor*) leads to the result that abnormality is caused by many autoso-  
209 mal factors (the LOD score in Fig. 4C). Both cases suggest a polygenic role of autosomes in producing  
210 ovary dysgenesis in a pure-Z background. Nonetheless, Normal ovaries do exist among backcross  
211 females but they almost always require a recombined Z chromosome (Fig. 5A). Furthermore, two  
212 individuals with Z chromosomes recombined in opposite ways can be both Normal (Fig. S12). This  
213 quantitative rescue effect of recombination on the Z chromosome is strikingly similar to the only  
214 other genome-wide mapping of ovarian dysgenesis in Lepidoptera (*Heliconius*) [26], which we dis-  
215 cuss below.

## 216 2.7 Z-autosome ancestry balance as a precondition for high fitness

217 In *Heliconius*, F<sub>1</sub> females between *H. pardalinus butleri* and *H. pardalinus sergestus* develop empty  
218 ovaries [26]. When F<sub>1</sub> males are backcrossed to *H. p. butleri*, female offspring have the most nor-  
219 mal ovary phenotypes when the Z chromosome contains about 1/4 introgression from either end  
220 (Fig. 5C). Autosomes must contribute to ovary dysgenesis in D(DB) and *Heliconius* females, because  
221 ovaries are strongly defective with autosome-only introgression ( $f_A \gg 0, f_Z = 0$ ), while increas-  
222 ing introgression on the Z chromosome will ameliorate autosome-induced defects until reaching a  
223 turning point (Fig. 5D). The same phenomenon in abnormal size seems to occur in B(BD) females  
224 (Fig. 5E), but not in D(DB) females (Fig. 5F). If Z chromosomes with recombined ancestry yield  
225 the best phenotype, autosomal introgression must bring in a large quantity of partially dominant  
226 incompatibility factors so that phenotypic defect stably appears when there is no Z-linked intro-  
227 gression. It also implies that Z-linked introgression contains many other factors balancing those on  
228 autosomes quantitatively. The outcome of this assumed process is an “ancestry-balance” precon-  
229 dition between Z and autosomes for the appearance of normal phenotypes (Fig. 5B). Nonetheless,  
230 Z-autosome ancestry balance does not necessarily hold for all traits in all crosses, because some  
231 traits can be masked by Mendelian factors of large effects (Locus E in ovary dysgenesis), and some  
232 traits are only weakly affected by autosome-only introgression (abnormal size in D(DB), although  
233 this particular example can also fit the “ancestry-balance” model by choosing a highly skewed bal-  
234 ance at  $f_Z - f_A = -1/4$ ). In short, if this precondition is valid, phenotypes will segregate by average  
235 ancestries between Z and autosomes and it indicates the involvement of highly dispersed genomic  
236 factors for incompatibility.

## 2.8 Male homogamety ameliorates size abnormality

The converse of the problem underlying Haldane's Rule is to explain weaker incompatibility effects in the homogametic sex. In our system, unlike  $F_1$  males, backcross males in both directions have a significant variation in pupal weight (Fig. 1B), but the spread is visibly smaller than that of females. It is congruent with the homogametic sex being less affected by hybridization, but what could be its genetic mechanism? Here we show analytically that this phenomenon is explicable by the Z-spanning polygenic architecture in male butterflies.

The genetic model outlined by Eq. 1 has a single net effect  $\alpha$  combining both the effect of native alleles ( $\alpha_0$ ) and the effect of introgressed alleles ( $\alpha_1$ ). If we split these contributions into two terms, and let  $\langle \Delta W \rangle$  be the expected deviation from normal pupal weight, then

$$\langle \Delta W \rangle_{\text{Female}} = \underbrace{\alpha_0(1-f)}_{\text{Native}} + \underbrace{\alpha_1 f}_{\text{Introgressed}} \quad (4)$$

In Fig. 5E and 5F, extreme ancestries on the Z chromosome ( $f = 1$  and  $f = 0$ ) will push  $\Delta W$  to two opposite extremes (B(BD)) or to only one extreme (D(DB)). It implies that native and introgressed alleles do not act in the same direction on phenotypes ( $\alpha_0\alpha_1 \leq 0$ ).

Now we apply this equation to backcross males with a slight variation. Backcross males have one copy of the pure Z chromosome and another copy that could be partially introgressed. For regions on the Z chromosome homozygous for native ancestry, the effect is the same as for females ( $\alpha_0$ ), since dosage compensation in Lepidoptera partially and symmetrically suppresses both Z chromosomes in males [33, 34]. For regions heterozygous in ancestry, let  $h_0$  and  $h_1$  be the dominance of native and introgressed alleles, respectively. Again, we assume that dosage compensation exists along regions heterozygous in ancestry (perhaps imperfectly) so that  $h_0$  and  $h_1$  capture not only dominance in the ordinary sense, but also the reduction in effect sizes due to partial expression suppression. Then:

$$\langle \Delta W \rangle_{\text{Male}} = \underbrace{\alpha_0(1-f)}_{\text{Homozygous}} + \underbrace{(h_0\alpha_0 + h_1\alpha_1)f}_{\text{Heterozygous}} \quad (5)$$

The genetic variance in either sex is induced by variation in  $f$ , and its ratio between the two sexes is:

$$\begin{aligned} \frac{V_{g,\text{Male}}}{V_{g,\text{Female}}} &= \left[ \frac{h_1\alpha_1 - (1-h_0)\alpha_0}{\alpha_1 - \alpha_0} \right]^2 \\ &= \left[ \frac{h_1|\alpha_1| + (1-h_0)|\alpha_0|}{|\alpha_1| + |\alpha_0|} \right]^2 \leq 1 \end{aligned} \quad (6)$$

This inequality is simply guaranteed by our observation that alleles of opposite ancestry have mostly opposite phenotypic effects among backcrosses ( $\alpha_0\alpha_1 \leq 0$ ). Even though backcross males were not sequenced, Eq. 6 provides a generic qualitative prediction that the genetic variance of

263 male pupal weight should be smaller than that of females. What we observe is shown in Table 1.  
 264 It clearly shows a weaker effect of Z-linked polygenes on male size. Interestingly, dominance the-  
 265 ory ( $h_0 = h_1 = 0$ ) becomes a particular case of this equation, and if it holds, we can solve for the  
 266 phenotypic effect of introgressed alleles relative to that of native ones:

$$\frac{\alpha_1}{\alpha_0} \Big|_{h_0=h_1=0} = 1 - \sqrt{\frac{V_{g,\text{Female}}}{V_{g,\text{Male}}}} \quad (7)$$

267 If dominance theory does not hold, and effects are fully additive with complete but perhaps asym-  
 268 metric dosage compensation, then  $h_0 + h_1 = 1$ , which leads to a constant ratio of genetic variance,  
 269 and

$$h_1 \Big|_{h_1+h_0=1} = \sqrt{\frac{V_{g,\text{Male}}}{V_{g,\text{Female}}}} \quad (8)$$

270 Under this assumption of complete dosage compensation on additive effects, the ratio now reflects  
 271 which ancestry has a biased expression along heterozygous regions. In the absence of such bias  
 272 ( $h_1 = 1/2$ ), dosage compensation will be symmetric between two haplotypes along regions het-  
 273 erozygous in ancestry, and the ratio of genetic variance will be fixed at 0.25. In our data, assuming  
 274 the additive model, B(BD) males will show more equalized contributions between Z-linked alleles  
 275 of different ancestries ( $h_1$  is closer to 0.5), while D(DB) males are more affected by the contribution  
 276 of introgressed alleles on Z chromosomes. Finally, we can rule out the classical dominance theory in  
 277 D(DB), because native alleles on the Z chromosome have very weak phenotypic effects ( $\alpha_0 \approx 0$ , see  
 278 Fig. 5F), which will force the ratio of  $V_g$  between males and females to be zero if  $h_0 = h_1 = 0$ . Taken  
 279 together, we show that male homogamety can ameliorate size abnormality under quite general con-  
 280 ditions in this particular system. This further validates our proposal of the Z-spanning polygenic  
 architecture.

**Table 1:** The ratio of genetic variance between male and female pupal weight among backcross individuals

Cross direction	D(DB)	B(BD)
Observed ratio of $V_g$	0.37	0.23
95% Confidence interval of $V_g$	(0.18, 0.62)	(0.05, 0.46)
$\alpha_1/\alpha_0$ (dominance theory)	N.A.	-1.09
$h_1$ (complete dosage compensation on additive effects)	0.61	0.48

281

## 282 2.9 Relevance to existing theories

283 We proceed to test if genetic architectures underlying abnormal size and ovary dysgenesis can be  
 284 explained by existing theories of Haldane's rule and the large-X/Z effect. Dominance theory is in

principle compatible with Lepidoptera as it does not rely on sex-specific mechanisms. For abnormal size, we have shown in the previous section that dominance theory is a particular case of the more general reasoning based on the quantitative genetic basis of pupal weight variation. In our reasoning, the critical assumption is the partial suppression ( $h_0, h_1 \in [0, 1)$ ) of mostly opposite allelic effects ( $\alpha_0\alpha_1 \leq 0$ ) between different ancestries on Z chromosomes in the homogametic sex.

For ovary dysgenesis, the effect of autosomal introgression is not completely recessive. For instance, introgression causing Diminished ovaries in maternally *dehaanii* females (yellow tracts in Fig. 4C) is fully heterozygous in ancestry and should contain at least additive factors. In maternally *bianor* females, introgression at locus *R* dominantly suppresses dysgenesis when the Z chromosome contains no introgression (Fig. 4B). We are aware that it is difficult to detect introgressed alleles with recessive effects using this backcross scheme, but our result contrasts with findings in *Drosophila* hybrids, where almost all autosomal introgressions act recessively in hybrid incompatibility [15]. However, it is difficult to judge the recessivity of Z-linked female sterility factors because oogenesis is a female-specific trait, and the female only has a single Z chromosome. Consequently, we cannot rule out dominance theory in ovary dysgenesis.

Since the Z chromosome is involved in almost all cases of hybrid disruption, and its effect is not localized to specific regions except for Locus *E*, can it be explained by increased differentiation on the entire Z chromosome relative to the autosomes? We find no evidence for this faster-Z theory. On the Z chromosome, most sequence windows (50kb) are less divergent between the two species than the average divergence on autosomes (Fig. S13). This echoes existing findings showing mixed support for faster-Z in Lepidoptera [35, 36, 37]. Theories specifically tailored for X/Y systems (the faster-male theory, the spermatogenesis theory, see [18]) do not apply to butterflies as females are heterogametic. Finally, although our data cannot rule out the possibility that female meiotic drive disrupts oogenesis, it is unknown how this process might occur in Lepidoptera females because their chromosomes are truly holokinetic (unlike *C. elegans*, which has “mini-centromeres”) [38, 39], and most documented sex-ratio drives in Lepidoptera are caused by *Wolbachia* infection [40].

## 2.10 Why are incompatibility factors so dispersed? A hypothesis based on epigenetic regulation

One of our key findings is that many factors of perhaps individually small effects are widely dispersed across autosomes or on the Z chromosome. Consequently, average chromosomal ancestry is often more informative of phenotypes than any particular locus. This pattern is similar to the polygenic threshold model of hybrid incompatibility in *Drosophila*, where abnormal phenotypes depend more on the total quantity of introgression than where introgression occurs in the genome [41, 42, 43]. As an attempt to explain our results under a coherent framework, we introduced in Fig. 5B a hypothesis positing that normal phenotypes require average autosomal ancestry and average Z chromosome ancestry to achieve a certain balance. Here, we explain why such balance might involve epigenetic regulation, particularly dosage compensation mechanisms specific to Lep-

322 idoptera.

323 Although we cannot test for mechanisms of Z-autosomes ancestry balance using our current data,  
324 we can think of two extreme scenarios for the identity of underlying factors: First, the balance  
325 might be between identical sequences widely distributed across the genome, perhaps repetitive  
326 DNA and their regulators. But it seems unlikely that this would have a highly uniform effect on the  
327 Z chromosome to additively determine hybrid size. Second, the balance could be entirely epigenetic.  
328 Abnormal size in hybrids that is asymmetric between reciprocal crosses exists in other systems, and  
329 it is often linked to epigenetic regulation such as imprinting [44, 45]. In our case, it must be related  
330 to some process that occurs between autosomes and the Z chromosome with a broad genomic basis.  
331 It is tempting to consider dosage compensation as a candidate mechanism because it affects the  
332 expression of the entire genome as well as specific genes. The details of dosage compensation might  
333 have diverged between the two species (e.g., the absolute level of genomic expression, species-  
334 specific epigenetic factors for dosage compensation etc.), while expression is still fully compensated  
335 between the two sexes within each lineage [34]. In Lepidoptera, dosage compensation is typically  
336 achieved by partial suppression of both Z chromosomes in males, and might also involve some up-  
337 regulation of the Z chromosome in females [33]. To illustrate the effect of hybridization, consider a  
338 hybrid female with only Z-linked introgression. Suppose this Z-linked region introgressed from a  
339 different species is more sensitive to up-regulation than its native homolog, so its presence increases  
340 the total expression of the Z chromosome and ruptures regular levels of dosage compensation. If  
341 hybrids are also fully compensated for dosage, then they might up-regulate autosomes to match  
342 the expression on the Z chromosome, which creates a quantitative effect on expression levels all  
343 over autosomes proportional to the percentage of introgression on the Z chromosome. On the other  
344 hand, if only autosomes have introgression, the expression of the native Z must change accordingly  
345 to achieve dosage compensation in hybrids. If the amount of introgression is balanced between  
346 autosomes and Z so that both are affected in the same way, there is perhaps no effect on the default  
347 dosage compensation program. This process might provide an explanation for the highly dispersed  
348 genetic architecture.

349 This hypothesis makes a number of untested assumptions, but it provides a phenomenological fit  
350 for our observation. In ovary dysgenesis among D(DB) females, since Diminished is a stable ovary  
351 phenotype over many configurations of autosomal introgression (Fig. 4C), it is more likely to be  
352 directly caused by the misexpression of a fixed set of factors, for instance, a set of oogenesis genes on  
353 the pure Z chromosome. The indirect cause is then 15%-30% introgression on autosomes rupturing  
354 the global outcome of dosage compensation when the Z chromosome contains no introgression.  
355 This would explain the dispersed nature of autosomal factors. If recombination occurs on the Z  
356 chromosome, it allows the Z chromosome to achieve a critical level of introgression matching that  
357 of autosomes in backcross individuals. It consequently pushes the system towards normal levels of  
358 dosage compensation, which explains the rescuing effect of recombination.

359 The same logic also applies to size abnormality. If size is affected mainly by the expression of a

360 few essential genes, such as those in the insulin signaling pathways or juvenile hormones [46, 47],  
361 the level of misexpression will be proportional to the difference between autosomal introgression  
362 and Z-linked introgression. Since the Z chromosome is much more variable in mean ancestry com-  
363 pared to that over all autosomes in the backcross generation, it will explain why the Z chromosome  
364 is the most informative of pupal weight, and it will also explain why the entire Z chromosome is  
365 involved rather than its local regions. This hypothesis is also congruent with the reduction of abnor-  
366 mal size in F<sub>1</sub> males: Autosomes and two Z chromosomes in F<sub>1</sub> males inherit half from each species  
367 so that ancestry balance is more likely to be attained; In F<sub>1</sub> females, the only Z chromosome has  
368 pure ancestry, while autosomes have equalized contribution, creating strong ancestry imbalance be-  
369 tween Z and autosomes, which might rupture dosage compensation. Finally, our reasoning for the  
370 reduction of abnormal size in backcross males also involves dosage compensation features unique  
371 to Lepidoptera (partial repression of both Z chromosomes). A future test for this hypothesis should  
372 incorporate genome-wide, allele-specific expression data on hybrids.

### 373 3. Conclusion

374 The two rules of speciation have many genetic mechanisms, and most of them are based on male-  
375 heterogametic systems. Our study reveals that Haldane's Rule and the large-Z effect in this butterfly  
376 system have a highly polygenic basis on *both* autosomes and the Z chromosome, but narrow regions  
377 of large effects do exist in some particular cases. The rescue effect of recombination on the Z chro-  
378 mosome in several traits suggests a balancing process between autosomes and the Z chromosome,  
379 the nature of which awaits to be revealed.

### 380 4. Materials and Methods

#### 381 4.1 Breeding

382 Lineages of *P. dehaanii* were purchased directly from a butterfly farm in Qingdao (Shandong Province,  
383 China), exclusively sourced from a small local population. Lineages of *P. bianor* were collected in  
384 the field from Ningbo (Zhejiang Province, China) for breeding in 2020, 2021, and a few individu-  
385 als were collected in the field from Kunming (Yunnan Province, China) for breeding in 2019. All  
386 crosses were done by hand-pairing. Eggs were collected by putting females in small cages with host  
387 plants under fluorescent light. The following host plants were used throughout the project: *Tetra-*  
388 *dium daniellii*, *Zanthoxylum bungeanum*, *Z. ailanthoides*, *Z. beecheyanum*, *Z. simulans*, *Choisya ternata*,  
389 *Phellodendron amurense*. Juveniles were kept in greenhouse conditions (approximately 20°C~35°C),  
390 with a combination of natural and greenhouse lights to maintain at least 10 hours of illumination  
391 per day. Adults for dissection were immediately put into a 5°C room after eclosion to reduce activ-  
392 ity. Otherwise, they were fed with sugar water once a day, and females were subsequently kept in  
393 the dark, while males were in an illuminated environment to facilitate hand-pairing.

## 394 **4.2 Phenotyping ovaries**

395 Ovaries were dissected from females within five days of eclosion in 1× PBS solution. The ovariole  
396 sheath was manually removed, and most images were taken using the internal camera of a Leica  
397 EZ4 HD microscope (pictures of a few specimens were taken by a cellphone through the eyepiece  
398 of a Zeiss Stemi 2000 microscope). Since ovary phenotypes are categorical, we established all ma-  
399 jor categories by defining the most obvious and the most frequent phenotypes across all dissected  
400 ovaries, and it was confirmed later by confocal imaging that they indeed have large qualitative  
401 differences. Phenotype Jammed is variable in terms of the fraction and the position of Jammed fol-  
402 licles. To reduce subjective opinions in scoring different kinds of Jammed, we lumped them into a  
403 single category in the QTL analysis. A small number of ovaries have ambiguous phenotypes due  
404 to one of the following reasons: 1) Different ovarioles develop different phenotypes; 2) Some part  
405 of the ovary is lost in dissection; 3) Extremely rare phenotypes that are hard to classify into exist-  
406 ing categories. These ambiguous individuals, mostly from cross direction D(DB), were assigned  
407 multiple categories. When such uncertainty affects analyses, we randomly select a phenotype from  
408 previously assigned categories on ambiguous individuals, perform analyses, and repeat the same  
409 procedure many times to control for uncertainty in phenotype.

## 410 **4.3 Staining and confocal imaging of ovaries**

411 Dissected ovaries were fixed in 4% Paraformaldehyde solution in 1× PBS for 20 minutes at room  
412 temperature. The ovaries were washed for 15 minutes each in 0.1% PBTx ( 1× PBS, 0.1% Triton-X  
413 100), 1% PBTx, 2% PBTx, 0.01% Saponin (Sigma Aldrich 47036) in PBS and Block solution (1× PBS,  
414 0.3% Triton-X 100, 0.5% Normal Goat Serum). The ovaries were then stained for 12 hours using the  
415 following reagents at 1:500 dilution in blocking solution: Hoechst 33342 (10 mg/ml, Thermo Fisher  
416 H3570), Wheat Germ Agglutinin-647 (WGA, Thermo Fisher W32466) and Rhodamine Phalloidin  
417 (Thermo Fisher R415). Stained ovaries were washed four times for 15 minutes each in 0.1% PBTx  
418 followed by a final PBS wash. After the washes, ovaries were mounted in equal volumes of PBS and  
419 Vectashield mountant (Vector labs H1900) on a slide.

420 The ovary samples were imaged by acquiring Z-section images on a Zeiss LSM 880 laser scanning  
421 confocal microscope at Harvard Center for Biological Imaging. The microscope was equipped with  
422 an Argon laser and a He/Ne 633 nm laser. Zeiss Plan-APOCHROMAT 10× /0.45 M27 or 20× /0.8 M27  
423 objective lenses were used for imaging. All images were 1024 × 1024 pixels in size and were ac-  
424 quired using PMT detectors. Images were acquired at excitation/emission wavelengths of 405/450  
425 nm for Hoechst 33342, 561/610 nm for Rhodamine Phalloidin, and 633/696 nm for WGA.

## 426 **4.4 DNA extraction and sequencing**

427 Samples from the cross were preserved in either pure ethanol or RNAlater at -20C prior to DNA  
428 extraction. For extraction, we used E.Z.N.A Tissue DNA kits (Omega Bio-tek, Inc.). Whole-genome  
429 library preparation was performed using Illumina DNA 1/4 reactions kits at Harvard University

430 Bauer Core, and subsequently sequenced altogether on a single lane of Illumina NovaSeq S4. Au-  
431 toosomal coverage varies among individuals: backcrosses-1×, parents-5×, grandparents-30×~60×.  
432 Raw reads were trimmed with Cutadapt-3.4 [48] to remove adapters (CTGTCTCTTATACACATCT),  
433 and subsequently mapped to the reference genome of *P. bianor* using the BWA-0.7.17 MEM algo-  
434 rithm. Duplicated reads were marked using Picard-2.25.7 [49]. For samples from the cross, we  
435 used BCFtools-1.9 [50] to pile up reads with very light quality filtering, and to call variants with  
436 associated genotype likelihoods. VCF files produced by the variants caller will be used for linkage  
437 analysis.

#### 438 **4.5 Linkage analysis**

439 For quality control, we first calculated kinship coefficients among individuals using NgsRelate-2  
440 [51] and corrected the pedigree position of a few individuals (Fig. S3). Lep-Map-3 was used for  
441 all subsequent linkage analysis [52]. First, VCF files and the pedigree were combined in module  
442 ParentCall2, and we imputed haplotype structure along the reference genome with module Order-  
443 Markers2. We also generated de novo marker orders using the same module. Comparing de novo  
444 marker orders against the order in the original reference genome revealed some intra-chromosome  
445 assembly problems. We corrected these problems below.

#### 446 **4.6 Linkage-based reference genome correction**

447 For a pre-correction reference genome, we plotted the de novo marker order against genomic order  
448 and discovered some assembly errors that affect the order and orientation among PacBio scaffolds  
449 on chromosomes. Using inferred haplotypes in the grandparent phase, we calculated the corre-  
450 lation of ancestry between each pair of markers, which should be a decreasing function of marker  
451 distance due to recombination. This information helps correct large-scale errors and generate an  
452 intermediate genome. To correct for smaller errors, we inferred the de novo marker order on the  
453 intermediate genome and compared it against the genome itself. This extra step corrected the po-  
454 sition/orientation of several smaller scaffolds. We were able to correct all visible errors except for  
455 those on chromosome 14, where an orientation problem appears to occur *within* a PacBio scaffold  
456 and we were unable to determine its breakpoint. (This may be due to a different population used  
457 here compared to the reference population). The de novo marker order generated on this final  
458 genome yields mostly a collinear relationship to the genomic order (Fig. S4). Finally, haplotypes in  
459 the grandparent phase were re-inferred on the corrected genome for all subsequent analyses.

#### 460 **4.7 Crossover analysis**

461 To infer crossover frequency, we calculated the number of recombination breakpoints in each pa-  
462 ternal haplotype among all backcross individuals (For all paternal haplotypes, see Fig. S5, S6, S7).  
463 No chromosome has more than two breakpoints except for chromosome 14 (one individual). We  
464 excluded chromosome 14 from this crossover analysis due to the aforementioned assembly prob-

465 lem. Let  $n_0$ ,  $n_1$ , and  $n_2$  be the number of haplotypes having 0, 1, or 2 recombination breakpoints for  
466 a given chromosome. The maximum likelihood estimate of crossover frequency is as follows (see  
467 “SI-Crossover analysis” for derivation). First, calculate:

$$\begin{aligned} c_0 &= (n_0 - n_1 + n_2) / (n_0 + n_1 + n_2) \\ c_1 &= (2n_1 - 4n_2) / (n_0 + n_1 + n_2) \\ c_2 &= 4n_2 / (n_0 + n_1 + n_2) \end{aligned} \tag{9}$$

468 If  $c_0, c_1, c_2$  are all nonnegative, they are the frequencies of having 0, 1, or 2 crossovers. If  $c_0 < 0$ , the  
469 adjusted estimate is

$$\begin{aligned} c_0^* &= 0 \\ c_1^* &= (n_0 - n_2) / (n_0 + n_2) \\ c_2^* &= 2n_2 / (n_0 + n_2) \end{aligned} \tag{10}$$

## 470 4.8 QTL analysis

471 One-dimensional QTL analysis was performed with R-package `qtl2` [53]. Ovary phenotypes were  
472 always mapped one at a time using a binary trait logistic mapper (i.e., phenotype of interest = 1,  
473 other phenotypes = 0). This approach is suitable for unordered categorical traits such as ovary  
474 morphology. For pupal weight, we introduced brood as a covariate to control for seasonal variation  
475 in pupal weight due to diet and environmental factors. LOD-scores were always estimated on 1,000  
476 random permutations of phenotypes. To calculate the explanatory power of QTL on pupal weights  
477 ( $R^2_{1\text{QTL}}$ ,  $R^2_{2\text{QTL}}$ ), we performed linear regression of brood-corrected pupal weight on marker ancestry  
478 (one marker for a 1-QTL scan or two additive markers for a 2-QTL scan), and explanatory power is  
479 defined as the  $R^2$  of the regression model.

## 480 4.9 The Z-spanning polygenic system

481 Let  $f \in [0, 1]$  be the fraction of Z chromosome coming from a foreign species, and let  $W$  be the  
482 random variable describing pupal weight. The Z-spanning polygenic system postulates that, on  
483 average, each infinitesimal quantity of introgression on Z has a fixed infinitesimal effect on pupal  
484 weight. In other words,

$$d\langle W \rangle = \alpha df \tag{11}$$

485 In the form of linear regression, this is often written as

$$W = w + \alpha f + \epsilon \tag{12}$$

486 where  $w$  is the average pupal weight without Z-linked introgression, and  $\epsilon$  is a noise variable with  
487 mean zero. Among a group of individuals, the explanatory power ( $R^2$ ) of this model is simply

488  $\alpha^2 \text{Var}(f) / \text{Var}(W)$ . Let  $p(l)$  be the marker ancestry (native=0) at position  $l$  on a chromosome of unit  
 489 length. The Z-spanning polygenic model implies the following Markov Chain:  $\{p(l)\}_{l=0}^1 \rightarrow f \rightarrow$   
 490  $W$ . Thus, ancestry  $p$  at a particular position  $l$  will be much less informative about  $W$  than  $f$ . The  
 491 regression model for a one-marker QTL analysis is  $W \sim p$ , while regression with the polygenic  
 492 model is  $W \sim f$ . The loss of predictive power by using  $W \sim p$  relative to  $W \sim f$  is simply the  
 493 squared coefficient of (multiple) correlation:

$$\frac{R_{W \sim p}^2}{R_{W \sim f}^2} = \rho_{p,f}^2 = \frac{\text{Cov}^2(p, f)}{\text{Var}(p)\text{Var}(f)} \quad (13)$$

494 where  $\rho_{p,f}$  is the correlation coefficient between  $p$  and  $f$ . This result is written as  $R_{\text{1QTL}}^2 / R_{\text{Z-ancestry}}^2$   
 495 in the main text. For a two-marker additive model  $W \sim p_1 + p_2$ , the relative loss is analogously

$$\frac{R_{W \sim p_1 + p_2}^2}{R_{W \sim f}^2} = \begin{bmatrix} \rho_{p_1,f} \\ \rho_{p_2,f} \end{bmatrix}^\top \begin{bmatrix} 1 & \rho_{p_1,p_2} \\ \rho_{p_2,p_1} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_{p_1,f} \\ \rho_{p_2,f} \end{bmatrix} \quad (14)$$

496 These correlation coefficients will depend on marker positions as well as the crossover process. In  
 497 our case, since crossover on the Z chromosome is obligatory, spatially uniform, and with strong in-  
 498 terference, it is straightforward to model one round of crossover for backcross females and calculate  
 499  $\rho$ . Then we get

$$\rho_{p(l),f} = \frac{1}{2} \sqrt{\frac{3}{2}} \left[ 1 + 2l(1-l) \right] \quad (15)$$

$$\rho_{p(l_1),p(l_2)} = 1 - |l_2 - l_1|$$

500 Applying these values to Equations 13 and 14 leads to the final result of Equation 2 and 3

501 For the variance analysis of male pupal weight, we assume that non-genetic fluctuation ( $\epsilon$ ) has  
 502 the same magnitude between males and females, and we estimated  $\text{Var}(\epsilon)$  to be used in estimating  
 503  $V_{g,\text{Male}}$ .

## 504 4.10 Posterior probabilities for ovary phenotypes based on $f_Z - f_A$

505 Let  $f_Z, f_A$  be the fraction of introgression from *bianor* into maternally *dehaanii* individuals on Z and  
 506 autosomes, respectively. Let  $x = f_Z - f_A$  be their difference. For ovary phenotype  $k$  from the set  
 507 {Diminished, Normal, Jammed}, the posterior probability of observing  $k$  conditioning on a given  
 508 range of  $x$  is (by Bayes theorem):

$$\mathbb{P}(k|x_i \leq x \leq x_{i+1}) = \frac{\mathbb{P}(x_i \leq x \leq x_{i+1}|k)\mathbb{P}(k)}{\sum_k \mathbb{P}(x_i \leq x \leq x_{i+1}|k)\mathbb{P}(k)} \quad (16)$$

509 where  $\mathbb{P}(k)$  is calculated from the entire dataset. Due to the uncertainty of ovary phenotypes, we  
510 re-sampled phenotypes for ambiguous individuals and repeated the estimation for 1,000 times and  
511 reported the average value of  $\mathbb{P}(k|x_i \leq x \leq x_{i+1})$  and its standard error.

## 512 **5. Data Availability**

513 Raw reads are deposited into the NCBI Sequence Read Archive (BioProject: PRJNA892033). Scripts  
514 are available from:  
515 [https://github.com/tzxiong/2022\\_Papilio\\_HybridIncompatibilityMapping](https://github.com/tzxiong/2022_Papilio_HybridIncompatibilityMapping)

## 516 **6. Author Contributions**

517 T.X. and J.M. designed the project. T.X. established the cross and performed analysis. S.T. performed  
518 ovary staining, confocal imaging, and contributed to ovary phenotype classification. N.R. provided  
519 and analyzed *Heliconius* data. M.Y. and X.L. provided parental phenotypic data and key genomic  
520 resources. T.X., J.M. and S.T. wrote the original draft. All authors contributed to the improvement  
521 of the manuscript.

## 522 **7. Acknowledgement**

523 T.X. was funded by a studentship from the Department of Organismic and Evolutionary at Har-  
524 vard University, the Quantitative Biology Initiatives at Harvard, the NSF-Simons Center for Math-  
525 ematical and Statistical Analysis of Biology at Harvard (award number #1764269), and a Grant in  
526 Aid of Research from Sigma Xi (The Scientific Research Honor Society). We thank Naomi Pierce  
527 and Adam Cotton for providing background information on the study system; Janet Sherwood,  
528 Shui Xu, Yuchen Zheng, Jinbo Hu, and Anastasios Kougioumoglou for their assistance and knowledge in  
529 breeding/sourcing host plants and butterflies; Cassandra Extavour for discussing oogenesis; John  
530 Wakeley, Robin Hopkins, Liang Qiao, Sarah Dendy, Nathaniel Edelman, Shuzhe Guan, Fernando  
531 Seixas, and Yuttapong Thawornwattana for their support during the project.

## 532 **References**

- 533 [1] Coyne, J. A. Two rules of speciation. *Speciation and its consequences* 180–207 (1989).
- 534 [2] Haldane, J. B. Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics* **12**, 101–109  
535 (1922).
- 536 [3] Orr, H. A. Haldane's rule. *Annual Review of Ecology and Systematics* 195–218 (1997).
- 537 [4] Turelli, M. & Orr, H. A. Dominance, epistasis and the genetics of postzygotic isolation. *Genetics*  
538 **154**, 1663–1679 (2000).
- 539 [5] Presgraves, D. C. & Orr, H. A. Haldane's rule in taxa lacking a hemizygous X. *Science* **282**,  
540 952–954 (1998).

- 541 [6] Schilthuizen, M., Giesbers, M. & Beukeboom, L. Haldane's rule in the 21st century. *Heredity*  
542 107, 95–102 (2011).
- 543 [7] Delph, L. F. & Demuth, J. P. Haldane's rule: genetic bases and their empirical support. *Journal*  
544 *of Heredity* 107, 383–391 (2016).
- 545 [8] Orr, H. A. Haldane's rule has multiple genetic causes. *Nature* 361, 532–533 (1993).
- 546 [9] Presgraves, D. C. Sex chromosomes and speciation in *Drosophila*. *Trends in Genetics* 24, 336–343  
547 (2008).
- 548 [10] Presgraves, D. C. Evaluating genomic signatures of “the large X-effect” during complex speci-  
549 ation. *Molecular Ecology* 27, 3822–3830 (2018).
- 550 [11] Larson, E. L., Keeble, S., Vanderpool, D., Dean, M. D. & Good, J. M. The composite regulatory  
551 basis of the large X-effect in mouse speciation. *Molecular Biology and Evolution* 34, 282–295  
552 (2017).
- 553 [12] Kitano, J. *et al.* A role for a neo-sex chromosome in stickleback speciation. *Nature* 461, 1079–1083  
554 (2009).
- 555 [13] Irwin, D. E. Sex chromosomes and speciation in birds and other ZW systems. *Molecular Ecology*  
556 27, 3831–3851 (2018).
- 557 [14] Davies, B. *et al.* Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice.  
558 *Nature* 530, 171–176 (2016).
- 559 [15] Masly, J. P. & Presgraves, D. C. High-resolution genome-wide dissection of the two rules of  
560 speciation in *Drosophila*. *PLoS Biology* 5, e243 (2007).
- 561 [16] Orr, H. A. & Turelli, M. Dominance and Haldane's rule. *Genetics* 143, 613 (1996).
- 562 [17] Charlesworth, B., Coyne, J. A. & Barton, N. H. The relative rates of evolution of sex chromo-  
563 somes and autosomes. *The American Naturalist* 130, 113–146 (1987).
- 564 [18] Wu, C.-I. & Davis, A. W. Evolution of postmating reproductive isolation: the composite nature  
565 of Haldane's rule and its genetic bases. *The American Naturalist* 142, 187–212 (1993).
- 566 [19] Frank, S. A. Divergence of meiotic drive-suppression systems as an explanation for sex-biased  
567 hybrid sterility and inviability. *Evolution* 45, 262–267 (1991).
- 568 [20] Hurst, L. D. & Pomiankowski, A. Causes of sex ratio bias may account for unisexual sterility  
569 in hybrids: a new explanation of Haldane's rule and related phenomena. *Genetics* 128, 841–858  
570 (1991).
- 571 [21] Tao, Y. & Hartl, D. L. Genetic dissection of hybrid incompatibilities between *Drosophila simu-*  
572 *lans* and *D. mauritiana*.: III. Heterogeneous accumulation of hybrid incompatibilities, degree of  
573 dominance, and implications for Haldane's Rule. *Evolution* 57, 2580–2598 (2003).
- 574 [22] Presgraves, D. C. Patterns of postzygotic isolation in Lepidoptera. *Evolution* 56, 1168–1183  
575 (2002).
- 576 [23] Jiggins, C. D. *et al.* Sex-linked hybrid sterility in a butterfly. *Evolution* 55, 1631–1638 (2001).
- 577 [24] Naisbit, R. E., Jiggins, C. D., Linares, M., Salazar, C. & Mallet, J. Hybrid sterility, Haldane's rule  
578 and speciation in *Heliconius cydno* and *H. melpomene*. *Genetics* 161, 1517–1526 (2002).

- 579 [25] Kost, S., Heckel, D. G., Yoshido, A., Marec, F. & Groot, A. T. A Z-linked sterility locus causes  
580 sexual abstinence in hybrid females and facilitates speciation in *Spodoptera frugiperda*. *Evolution*  
581 **70**, 1418–1427 (2016).
- 582 [26] Rosser, N. *et al.* Complex basis of hybrid female sterility and Haldane's rule in *Heliconius*  
583 butterflies: Z-linkage and epistasis. *Molecular Ecology* **31**, 959–977 (2022).
- 584 [27] Ae, A. S. A study of the *Papilio bianor* Group mainly based on hybridization (Lepidoptera,  
585 Papilionidae). *Tyo Ga* **41**, 13–19 (1990).
- 586 [28] Kitahara, H. & Shirai, K. Crossing experiments with *Papilio okinawensis* Fruhstorfer from Oki-  
587 nawa Island and *P. dehaanii* C. & R. Felder from central Honshu, Japan (Lepidoptera, Papilion-  
588 idae). *Lepidoptera Science* **69**, 85–91 (2018).
- 589 [29] Ae, S. A. A study of hybrids between Japanese and Himalayan *Papilio* butterflies. *Special*  
590 *Bulletin of Lepidopterological Society of Japan* **2**, 75–107 (1966).
- 591 [30] Lu, S. *et al.* Chromosomal-level reference genome of Chinese peacock butterfly (*Papilio bianor*)  
592 based on third-generation DNA sequencing and Hi-C analysis. *GigaScience* **8**, giz128 (2019).
- 593 [31] Lees, D. C., Kremen, C. & Andriamampianina, L. A null model for species richness gradients:  
594 bounded range overlap of butterflies and other rainforest endemics in Madagascar. *Biological*  
595 *Journal of the Linnean Society* **67**, 529–584 (1999).
- 596 [32] Colwell, R. K. & Lees, D. C. The mid-domain effect: geometric constraints on the geography of  
597 species richness. *Trends in Ecology & Evolution* **15**, 70–76 (2000).
- 598 [33] Rosin, L. F., Chen, D., Chen, Y. & Lei, E. P. Dosage compensation in *Bombyx mori* is achieved  
599 by partial repression of both Z chromosomes in males. *Proceedings of the National Academy of*  
600 *Sciences* **119**, e2113374119 (2022).
- 601 [34] Huylmans, A. K., Macon, A. & Vicoso, B. Global dosage compensation is ubiquitous in Lepi-  
602 doptera, but counteracted by the masculinization of the Z chromosome. *Molecular Biology and*  
603 *Evolution* **34**, 2637–2649 (2017).
- 604 [35] Rousselle, M., Faivre, N., Ballenghien, M., Galtier, N. & Nabholz, B. Hemizygosity enhances  
605 purifying selection: lack of fast-Z evolution in two Satyrine butterflies. *Genome Biology and*  
606 *Evolution* **8**, 3108–3119 (2016).
- 607 [36] Pinharanda, A. *et al.* Sexually dimorphic gene expression and transcriptome evolution provide  
608 mixed evidence for a fast-Z effect in *Heliconius*. *Journal of Evolutionary Biology* **32**, 194–204 (2019).
- 609 [37] Mongue, A. J., Hansen, M. E. & Walters, J. R. Support for faster and more adaptive Z chromo-  
610 some evolution in two divergent Lepidopteran lineages. *Evolution* **76**, 332–345 (2022).
- 611 [38] Drinnenberg, I. A., deYoung, D., Henikoff, S. & Malik, H. S. Recurrent loss of CenH3 is associ-  
612 ated with independent transitions to holocentricity in insects. *eLife* **3**, e03676 (2014).
- 613 [39] Senaratne, A. P. *et al.* Formation of the CenH3-deficient holocentromere in Lepidoptera avoids  
614 active chromatin. *Current Biology* **31**, 173–181 (2021).
- 615 [40] Ahmed, M. Z., Araujo-Jnr, E. V., Welch, J. J. & Kawahara, A. Y. *Wolbachia* in butterflies and  
616 moths: geographic structure in infection frequency. *Frontiers in Zoology* **12**, 1–9 (2015).

- 617 [41] Naveira, H. F. Location of X-linked polygenic effects causing sterility in male hybrids of  
618 *Drosophila simulans* and *D. mauritiana*. *Heredity* **68**, 211–217 (1992).
- 619 [42] Liénard, M. A., Araripe, L. O. & Hartl, D. L. Neighboring genes for DNA-binding proteins  
620 rescue male sterility in *Drosophila* hybrids. *Proceedings of the National Academy of Sciences* **113**,  
621 E4200–E4207 (2016).
- 622 [43] Presgraves, D. C. & Meiklejohn, C. D. Hybrid sterility, genetic conflict and complex speciation:  
623 lessons from the *Drosophila simulans* clade species. *Frontiers in Genetics* 1002 (2021).
- 624 [44] Vrana, P. B. *et al.* Genetic and epigenetic incompatibilities underlie hybrid dysgenesis in *Per-*  
625 *omyscus*. *Nature genetics* **25**, 120–124 (2000).
- 626 [45] Brekke, T. D. *et al.* X chromosome-dependent disruption of placental regulatory networks in  
627 hybrid dwarf hamsters. *Genetics* **218**, iyab043 (2021).
- 628 [46] Jindra, M., Palli, S. R. & Riddiford, L. M. The juvenile hormone signaling pathway in insect  
629 development. *Annual Review of Entomology* **58**, 181–204 (2013).
- 630 [47] Okamoto, N. & Yamanaka, N. Nutrition-dependent control of insect development by insulin-  
631 like peptides. *Current Opinion in Insect Science* **11**, 21–30 (2015).
- 632 [48] Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.  
633 *EMBnet.journal* **17**, 10–12 (2011).
- 634 [49] Picard toolkit. <https://broadinstitute.github.io/picard/> (2019).
- 635 [50] Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
- 636 [51] Korneliussen, T. S. & Moltke, I. NgsRelate: a software tool for estimating pairwise relatedness  
637 from next-generation sequencing data. *Bioinformatics* **31**, 4009–4011 (2015).
- 638 [52] Rastas, P. Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequenc-  
639 ing data. *Bioinformatics* **33**, 3726–3732 (2017).
- 640 [53] Broman, K. W. *et al.* R/qltl2: software for mapping quantitative trait loci with high-dimensional  
641 data and multiparent populations. *Genetics* **211**, 495–502 (2019).