The genetic basis of hybrid female sterility and Haldane's rule in a butterfly: Z-linkage and epistasis

Neil Rosser[1,2*], Nathaniel B. Edelman[1*], Lucie M. Queste[2], Michaela Nelson[2], Fernando Seixas[1], Kanchon K. Dasmahapatra[2] and James Mallet[1]

[1]Department of Organismic and Evolutionary Biology, Harvard University, Mass., USA

[2]Department of Biology, University of York, York, UK

[*]Authors for correspondence

Abstract

Hybrids between diverging populations are often sterile or inviable. These phenotypes are frequently more severe in the heterogametic sex – a pattern known as Haldane's rule. While the genetic basis of Haldane's Rule has been extensively studied in systems where the male is the heterogametic sex, the causes of Haldane's Rule in taxa where the female is heterogametic, such as Lepidoptera and birds, are largely unknown. Here, we study hybrid sterility between allopatric populations of the butterfly *Heliconius pardalinus*. Although these populations mate freely in captivity, female F1 hybrids in both directions of cross are sterile, in accordance with Haldane's rule. We find that sterility occurs because development of F1 female oocytes is arrested after the oocyte is differentiated from nurse cells, but before yolk deposition. We crossed fertile male F1 hybrids to females from the parental populations, and used the resulting backcrosses to map sterility loci (QTLs) and identify genes differentially expressed in the ovary as a function of oocyte development. We find a

major effect of the Z chromosome on sterility, similar to the "large X effect" in *Drosophila*, with strong epistatic interactions between loci at either end of the Z chromosome. In addition, we provide evidence of Dobzhansky-Muller incompatibilities between Z chromosome and autosomal loci, as well as between autosomal loci on chromosomes 8 and 20. Among loci differentially expressed between females with arrested vs. non-arrested ovary development, we identify six candidate genes known from *Drosophila melanogaster* and *Parage aegeria* oogenesis. This study represents the first in-depth characterization of hybrid sterility using modern genetic tools in the Lepidoptera.

Keywords— Speciation, Haldane's Rule, hybrid sterility, ZW sex determination, Dobzhansky-Muller incompatibilities, Lepidoptera

1 Introduction

Hybrids between diverging populations are often sterile or inviable (Darwin, 1859; Presgraves, 2010). Because these phenotypes are a major source of reproductive isolation, elucidating their genetic basis is seen as key to understanding speciation (Coughlan & Matute, 2020; Nosil & Schluter, 2011; Castillo & Barbash, 2017; Butlin *et al.*, 2012). Hybrid dysfunction is usually thought to result from epistatic interactions among genes, which appear to solve what has been called "Darwin's dilemma" (Coyne & Orr, 2004): the question of how incompatibilities between species can evolve without negative consequences from new substitutions within species (Bateson, 1909; Dobzhansky, 1936; Muller, 1942). Under the Dobzhansky-Muller model, diverging populations acquire different novel alleles at loci across their genomes. In hybrids, previously untested combinations of alleles are brought together and interact to reduce fitness (Orr, 1995; Brideau *et al.*, 2006; Tang & Presgraves, 2009; Presgraves, 2007; Maheshwari & Barbash, 2011).

Although Dobzhansky-Muller incompatibilities (DMIs) may involve only a pair of genes (Sweigart *et al.*, 2006), they are likely to be more complex, even early in speciation (Kalirad & Azevedo, 2017; Phadnis, 2011; Schumer *et al.*, 2014). The expected number of two-locus DMIs is predicted to increase approximately as the square of the number of divergent substitutions between two species; the "snowball effect" (Orr, 1995; Orr & Turelli, 2001; Matute *et al.*, 2010). DMIs involving more than two loci should accumulate even more rapidly, because as the number of interacting loci increase, so too does the number of potentially negative combinations (Orr, 1995). In keeping with these predictions, widespread DMIs across the genomes of a number of species have been inferred from genetic association data (e.g. Schumer *et al.*, 2014; Good *et al.*, 2008). DMIs are often assumed to be acquired in allopatry. However, they may also evolve in the presence of gene flow as pleiotropic by-products of divergent selection, if it is strong enough to outweigh selection against hybrids (Bank *et al.*, 2012). There is also evidence that polymorphic alleles

with negative epistatic interactions are common even within species (Corbett-Detig *et al.*, 2013). For populations diverging with gene flow, we might therefore expect to see traits under divergent selection as causes of incompatibility (Wright *et al.*, 2013; Ferris *et al.*, 2017; Feng *et al.*, 2020).

One of the few rules of speciation is Haldane's Rule: among hybrids, when one sex is absent, rare, or sterile, it is usually the heterogametic sex (males in XX/XY systems and females in ZZ/ZW systems (Haldane, 1922)). Greater sterility of the heterogametic sex has been found in 213 out of 223 pairs (>95%) of a diverse array of taxa, and has at least 10 phylogenetically independent origins (Schilthuizen *et al.*, 2011; Delph & Demuth, 2016). The ubiquity of Haldane's rule therefore suggests that postzygotic incompatibilities evolve with some predictability across a wide range of taxa (Coyne, 1992). Hybrid sterility of the heterogametic sex also evolves early in the evolution of postzygotic isolation, typically before hybrid inviability (Coyne & Orr, 1989; 1997; Presgraves, 2010). It may therefore have a disproportionate effect in reducing gene flow, and as such is of particular interest for understanding speciation (Ramsey *et al.*, 2003).

Most explanations for Haldane's rule depend on DMIs. The hypothesis to have received the most support is dominance theory, in which hybrid sterility and inviability are produced by interactions between sex chromosomes and autosomes (Coyne & Orr, 2004). If sex-linked alleles causing incompatibilities have a largely recessive effect, incompatibilities will disproportionately affect the heterogametic sex (Turelli & Orr, 1995; Orr, 1997; Turelli & Moyle, 2007). Nonetheless, male heterogametic species without strongly differentiated sex chromosomes also conform to Haldane's Rule (Presgraves & Orr, 1998), suggesting that other forces also contribute, such as "faster-male" evolution (Wu & Davis, 1993) and faster evolution of the sex chromosome (Sackton *et al.*, 2014). The genetic and molecular mechanisms of hybrid sterility have been identified in some cases (Brideau *et al.*, 2006; Tang & Presgraves, 2009; Schartl, 2008; Mihola *et al.*, 2009; Bayes & Malik, 2009), but this work has been primarily carried out in organisms with XX/XY sex determination, in which males

hybrids are sterile or inviable.

Lepidoptera (butterflies and moths) yielded the first example of a sex linked trait (Doncaster & Raynor, 1906), even before *Drosophila* (Morgan, 1910; 1911), and the correlation of inheritance with cytological discovery of chromosomes led to the chromosomal theory of inheritance. Lepidoptera are also among the groups of taxa that Haldane considered when formulating his eponymous rule (Haldane, 1922). They have ZW/ZZ sex determination, where females are the heterogametic sex and are more susceptible to hybrid dysfunction. Lepidoptera were also critical in evaluating the relative impacts of dominance and faster male evolution in Haldane's rule (Presgraves, 2002), and have provided evidence that faster-Z evolution may contribute to the phenomenon in female heterogametic systems (Prowell Pashley, 1998; Sackton *et al.*, 2014).

Several examples of Haldane's Rule have been reported in *Heliconius* butterflies (Nymphalidae) which comprise about 48 species that occur throughout much of tropical America (Jiggins, 2017). Female hybrids between *Heliconius melpomene* and *Heliconius cydno* (*sensu lato*) are sterile (Naisbit *et al.*, 2002; Salazar *et al.*, 2005; Sánchez *et al.*, 2015), and there is evidence for genetic linkage between sterility loci and a trait (mimetic wing colour pattern) under divergent ecological selection (Merrill *et al.*, 2011). As well as between species, female hybrid sterility is known within species, between some populations of *Heliconius melpomene* (Jiggins *et al.*, 2001). Here, we investigate the genetic and molecular basis of Haldane's rule in two named subspecies of *Heliconius pardalinus*: *H. pardalinus butleri* and *H. pardalinus sergestus*. These largely allopatric subspecies are strongly genetically differentiated, with *H. p. butleri* more closely related to its sympatric relative *Heliconius elevatus*, thereby rendering *H. pardalinus* paraphyletic (Rosser *et al.*, 2019). They inhabit different habitats, with *H. p. sergestus* restricted to dry forests in the Huallaga/Mayo valleys of the Andes, and *H. p. butleri* inhabiting lowland rainforest across the adjacent Amazon basin (Fig. 1). Although they mate freely when brought together in captivity, they rarely co-occur in nature, and F1 hybrid females in both directions of cross are completely sterile

(Rosser *et al.*, 2019). We here characterize the ovary phenotype in parental populations, F1 hybrids and backcrosses. We then use backcrosses to *H. p. butleri* to generate a QTL map and intersect these data with genes differentially expressed between fertile and sterile individuals to identify candidate genes and epistatic interactions responsible for hybrid sterility.

2 Materials and Methods

2.1 Butterfly rearing, Nucleic acid preservation and ovary dissection

Butterfly stocks were collected in the Departments of San Martín, Loreto and Ucayali, Peru, and captive populations of *H. p. sergestus* and *H. p. butleri* were established in insectaries in Tarapoto, Peru, as previously described (Rosser *et al.*, 2019). Female butterflies were collected from insectaries 15 days after eclosion, allowing time for eggs to develop fully (Dunlap-Pianka *et al.*, 1977; Naisbit *et al.*, 2002). Wings were removed and stored in glassine envelopes as vouchers. Thorax and head were removed and stored in NaCl-saturated dimethyl sulfoxide at –20°C for DNA extraction and processing. Approximately half of the ovaries were dissected immediately, and for the remainder, abdomens were stored in 96% ethanol and transferred to Harvard University, Cambridge, MA for fine dissection. In all cases, ovaries were dissected from the abdomen in ice-cold phosphate buffered saline (PBS) using fine forceps and insect pins. Tracheae and fat bodies were removed manually, and images were taken at 8X, 12.5X, and 20X magnification for phenotyping. Of the ovaries dissected in the field, six backcrosses and two pure *H. pardalinus butleri* were stored in RNALater solution for RNAseq (ThermoFisher AM7020).

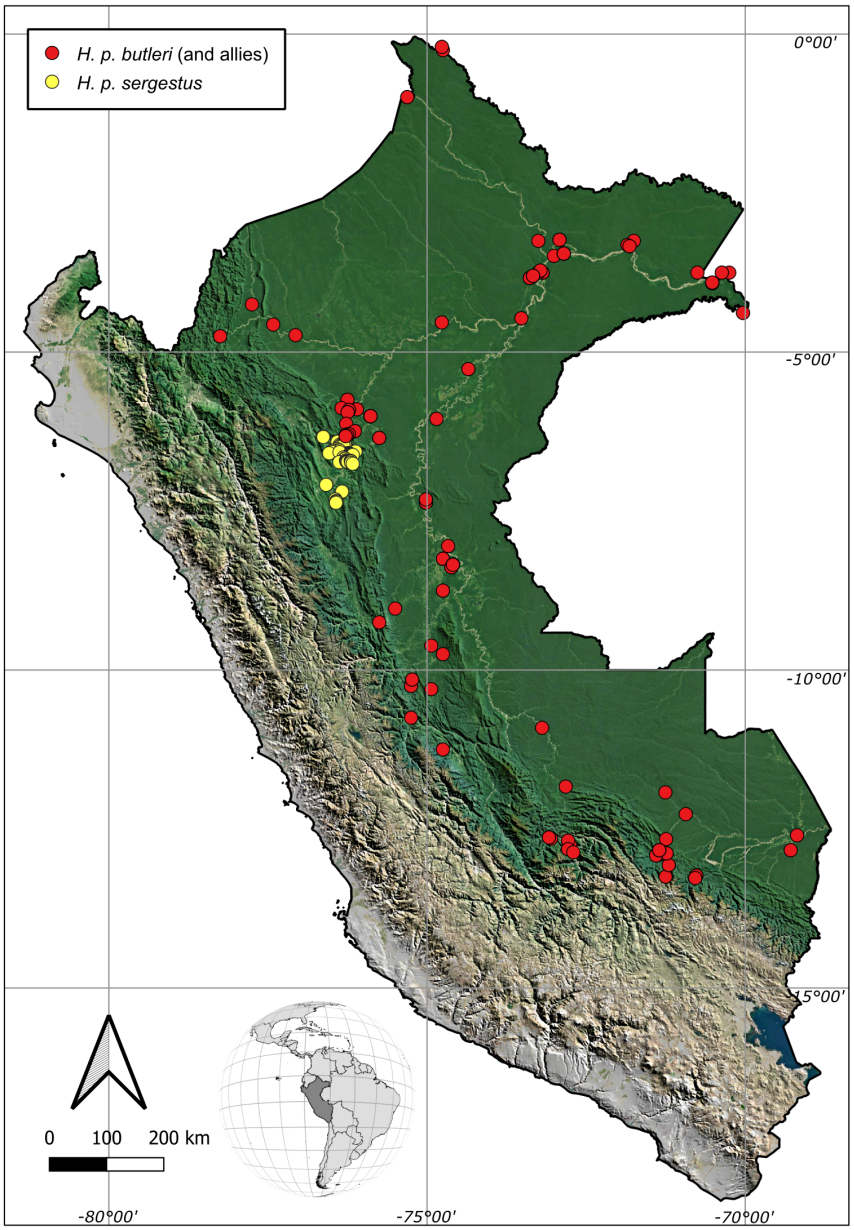

Figure 1: Distribution of *H. pardalinus* in Peru. The yellow dots correspond to *H. pardalinus sergestus* and the red dots *H. pardalinus butleri*, which intergrades with other lowland subspecies in central and southern Peru and elsewhere in the Amazon basin.

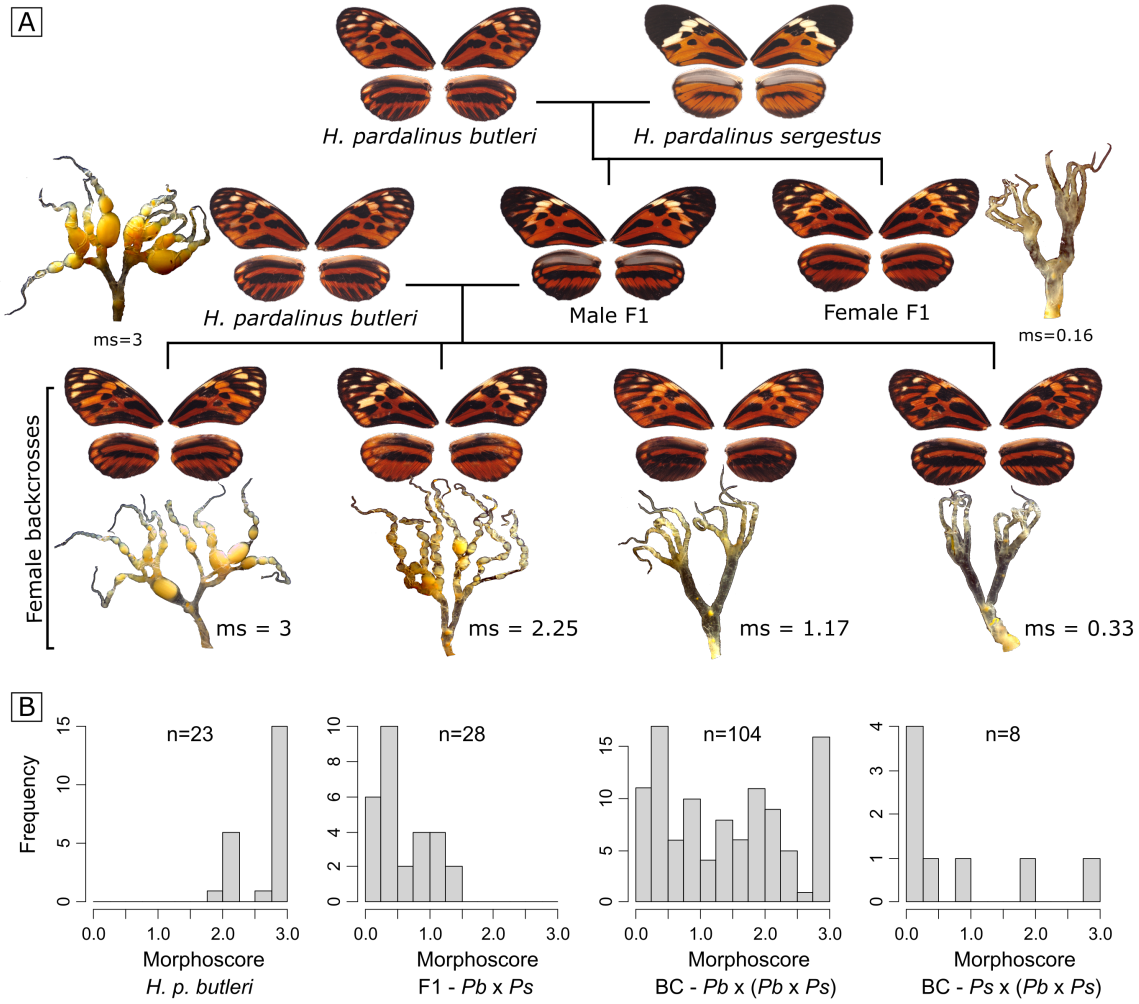

Figure 2: Crossing scheme and distribution of phenotypes. A Crossing *H. p. butleri* with *H. p. sergestus* in either direction produces sterile female F1s, but male F1s are fertile. Backcrossing these males in either direction produces females with variable fertility. Example wing phenotypes and dissected ovaries for backcrosses to *H. p. butleri* are shown, with fertile individuals to the left and sterile individuals to the right; ms = morphoscore assigned to the dissected ovary. **B** Histograms of ovary morphoscores for i) *H. p. butleri* females, ii) F1s produced by mating a female *H. p. butleri* (*Pb*) with a male *H. p. sergestus* (*Ps*), iii) backcrosses produced by mating fertile male F1 (*Pb* x *Ps*) with female *H. p. butleri* (*Pb*), and finally, backcrosses produced by mating fertile male F1 (*Pb* x *Ps*) with female *H. p. sergestus* (*Ps*)

2.2 Ovary staining and phenotyping

Individual ovarioles from alcohol-stored ovaries were removed and rehydrated by 15 minute incubations in serial dilutions of ethanol in 0.1% tween 20 in 1X phosphate-buffered saline (PBT) (Ethanol concentrations: 95%, 90%, 80%, 60%, 40%, 20%, 0%). Once fully re-hydrated, ovaries were incubated in acridine orange solution (ThermoFisher A1301; 5 µg/mL in PBT) to visualize cytoplasm. They were then washed in PBT before being stained with DAPI (1 µL/mL in PBT), washed once more in PBT, and mounted on slides with VectaShield (Vector Labs). Slides with stained ovarioles were scanned with a Zeiss Axio Scan Z1, and high-magnification images were taken with a Zeiss LSM 880 upright confocal microscope. The most highly developed follicle in each ovariole was staged through visual comparison to oocyte development stages described in *Bombyx mori* (Yamauchi & Yoshitake, 1984).

Due to field storage conditions, we were unable to perform staining and determine the latest developmental stage for all backcross females. We therefore developed a scoring scheme based on gross ovary morphology (Fig. S1). Three images from each ovary were scored blind by two independent scorers. The resulting six scores per ovary were averaged to yield a single "morphoscore" for each individual. Using the 42 individuals for which we could confidently assign the latest developmental stage and morphoscore, we verified that the two metrics were highly correlated (logistic regression, $p = 2.45 \times 10^{-11}$, Supplementary Fig. S2).

2.3 DNA extraction and sequencing

RNA-free genomic DNA was extracted from individuals used in QTL mapping (see below) using a Qiagen DNeasy Blood and Tissue Kit and following the manufacturer's standard protocol. Restriction site Associated DNA (RADSeq) libraries were prepared using a protocol modified from Etter et al. (Etter *et al.*, 2011; Hoffman *et al.*, 2014), using a *PstI* restriction enzyme, sixteen 6bp P1 barcodes and eight indexes. DNA was Covaris sheared

and gel size selected to 300-700bp. 128 individuals were sequenced per lane, with 125bp paired-end reads, on an Illumina HiSeq 2500.

2.4 RNA extraction and sequencing

Ovaries stored in RNALater were further dissected into pre-vitellogenic (i.e., before yolk deposition) follicles, vitellogenic follicles, and choriogenic follicles + chorionated eggs (the chorion is the proteinaceous "eggshell" of an insect egg). Each of these three subsets was processed separately. Tissue was blotted dry with Kimwipes to remove excess RNALater solution, transferred to TRIZOL and homogenized with the PRO200 tissue homogenizer (PRO Scientific). RNA was extracted with the Direct-zol RNA miniprep kit (Zymo R2051). The mRNA libraries were prepared by the Harvard University Bauer Core with the KAPA mRNA HyperPrep kit, with mean fragment insert sizes of 200-300bp, and were sequenced on a NovaSeq S2, producing an average of 49 million paired-end, 50 bp reads per library (Table S2).

RNASeq reads were mapped to the *H. melpomene* v2.5 transcriptome (Pinharanda *et al.*, 2019) using kallisto (Bray *et al.*, 2016). Approximately 70% of reads were mapped to the transcriptome per sample, and that value did not differ between the *H. pardalinus butleri* samples and the backcrosses (Table S2). Aligned reads were normalized to account for sequencing coverage, transcript length, and RNA composition using sleuth (Pimentel *et al.*, 2017). Raw counts were log-transformed, and expression differences were calculated by comparing the likelihood of the model: $ln(counts) \sim 1$ to the model $ln(counts) \sim 1 + binaryscore$ (Pimentel *et al.*, 2017).

In order to identify conserved genes expressed in butterfly oogenesis, we used BLAST to identify *H. melpomene* transcripts orthologous to genes expressed in the ovarian transcriptome of the Speckled Wood butterfly *Parage aegeria* (Carter *et al.*, 2013). In addition, we

used OrthoFinder (Emms & Kelly, 2019) to identify transcripts with orthologous genes in *Drosophila melanogaster*, and then filtered this list with the keywords "oogenesis" OR "follicle" OR "nurse" OR "oocyte" using the phenotypic data on Flybase (http://flybase.org).

2.5 SNP calling

FastQ files from each RAD library were demultiplexed using process_radtags from Stacks (Catchen *et al.*, 2013), and BWA-MEM (Li, 2013) was used with default parameters to map the reads both to the *H. melpomene* genome (Hmel2.5) (Davey *et al.*, 2017) and to the *H. pardalinus* genome (Hpar) (Seixas *et al.*, 2020). BAM files were then sorted and indexed with SAMtools (Li *et al.*, 2009), and Picard-tools v 1.119 (https://github.com/broadinstitute/picard) was used to add read groups and mark PCR duplicates. To check for incorrectly labelled samples, we estimated the sex of a sample by dividing the mean number of reads per kilobase on the Z chromosome by the mean value for autosomes. This returned a value close to 1 in males and 0.7 in females, which can then be compared with the recorded sex of the sample. To further check for labelling errors, confirm pedigrees, and assign samples with unrecorded pedigree to families, we used Plink 1.9 (Chang *et al.*, 2015) to estimate the fraction of the genome that is identical by descent (IBD; $\widehat{\pi}$) between all pairwise combinations of samples (siblings and parent-offspring comparisons should yield $\widehat{\pi}$ values close to 0.5). In addition, for specimens that were sequenced multiple times in order to improve coverage, we checked that samples were derived from the same individual (with $\widehat{\pi}$ values close to one). We then merged these samples, using the MergeSamFiles command from Picard-tools, and used Samtools' mpileup command to call single nucleotide polymorphisms (SNPs) for linkage map construction.

2.6 Linkage map construction

Linkage maps were built using reads aligned to each of the reference genomes using LepMAP3 (Rastas, 2017). The ParentCall2 module was used to correct erroneous or missing parental genotypes, and call sex-linked markers using a log-odds difference of >2. We used Filtering2 to remove SNPs showing segregation distortion, specifying a P-value limit of 0.01 (i.e., there is a 1:100 chance that a randomly segregating marker is discarded). Because we genotyped only female offspring, we did not filter sex-linked markers for segregation distortion. We then used SeparateChromosomes2 to cluster markers to linkage groups, specifying zero recombination in females and joining pairs of markers with LOD-score greater than 14. To obtain recombination distances between markers, we fixed the order of the markers to their order on the Hmel2.5 or Hpar genome assemblies, and then evaluated this order, again using paternally and dual informative markers.

2.7 QTL mapping

Genetic data were analysed as backcrosses (Fig. 2) using the paternally inherited allele. We used R/QTL (Broman *et al.*, 2003) to estimate genotype probabilities at 1 cM intervals, using the Haldane mapping function and an assumed genotyping error rate of 0.001. Loci with inferred genotypes were labelled using the chromosome and the centimorgan position. We used Haley-Knott (H-K) regression to test for associations between the estimated genotype probabilities at each marker and morphoscore (Haley & Knott, 1992). BB genotypes were coded as 0.5 and BS genotypes were coded as -0.5, where B is the *H. p. butleri* allele and S is the *H. p. sergestus* allele.

We first built a single locus additive QTL model at each position in the genome (H_1; $y = \mu_1 + \beta_1 q_1 + \varepsilon$) and calculated the $\log_{10}$ likelihood ratio (LOD score) comparing (H_1) with the null hypothesis of no QTL (H_0; $y = \mu_1 + \varepsilon$). To identify loci that act in combination

to produce the phenotype, we then estimated LOD scores using all pairwise combinations of typed markers and inferred genotypes at 1 cM intervals across the genome, while allowing interactions between them (H_f; $y = \mu_1 + \beta_1 q_1 + \beta_2 q_2 + \beta_3 q_1 q_2 + \varepsilon$). The difference between LOD scores for (H_f) and the corresponding two locus additive model (H_a; $y = \mu_1 + \beta_1 q_1 + \beta_2 q_2 + \varepsilon$) gives the improvement in fit attributable purely to interactive effects (H_{int}). The difference between LOD_f and the maximum LOD score obtained from single QTL locus models at either marker indicates the presence of a second QTL, allowing for epistasis (H_{fv1}). We also performed these analyses while controlling for kinship. To do this, we used LepMap to estimate $\widehat{\pi}$ (IBD) between all individuals. We then created a variance-covariance matrix of genetic relatedness, and included this in our models as a random effect. Significance of QTL scans was assessed by permuting the phenotypes relative to the genotypes (10,000 permutations). Because we analysed only female offspring, the degrees of freedom for QTL models at sex-linked and autosomal loci are the same, and so we set a single genome-wide significance threshold for each scan.

2.8 Population genomics

To examine genomic differentiation between the *H. p. sergestus*, *H. p. butleri* and *H. elevatus*, previously published whole genome re-sequencing data (four individuals each taxon) were used (NCBI accession numbers: ERS070236; ERS977673; ERS977674; ERS070238; ERS4368504; SRS329822; SRS329823; SRS329824; SRS329825; SRS329826; SRS3298233; SRS1247739; ERS235668; ERS977715; ERS977716; ERS977717). Raw reads were filtered for Illumina adapters using cutadapt (Martin, 2011) and mapped to the Hmel2.5 (Davey *et al.*, 2017) and Hpar (Seixas *et al.*, 2020) genomes using BWA MEM v0.7.15. Duplicate reads were removed using sambamba v0.6.8 (Tarasov *et al.*, 2015) and the Genome Analysis Toolkit (GATK) v3.8 RealignerTargetCreator and IndelRealigner modules (DePristo *et al.*, 2011; McKenna *et al.*, 2010) were used to realign reads around indels. Genotype calling was

performed for each taxon separately with bcftools (Li *et al.*, 2009) mpileup and call modules (Li, 2011), using the multiallelic and rare-variant calling option (-m) and requiring a minimum mapping quality and base quality of 20. Genotype calls were required to have a minimum quality score (QUAL) of 20, RMSMappingQuality (MQ) $\geq$ 20, genotype quality (GQ) $\geq$ 20 and a minimum individual depth of coverage (DP) $\geq$ 8 (or DP $\geq$ 4 for the Z chromosome of females). Genotypes within 5 bp of an indel were recorded as missing data.

Levels of differentiation (F_{ST}), and pairwise genetic distances (D_{XY}), between the three taxa studied here were estimated along the genome in overlapping 25 kb windows (with 5 kb steps) using the popgenWindows.py script (available from https://github.com/simonhmartin/genomics_general).

3 Results

We reared 143 F1 hybrid offspring of *H. p. butleri* and *H. p. sergestus*. Female F1s in both directions of cross were sterile. To investigate the molecular and genetic basis of hybrid sterility between the two populations, we backcrossed fertile F1 hybrid males to both parental species, rearing 320 offspring. F1 and backcross broods eclosed with approximately equal sex ratios (69 females:73 males and 164 females:156 males, respectively), indicating no differences in viability of immature stages.

3.1 Ovary phenotype

To characterize the sterility phenotype, we dissected ovaries from parental, F1, and backcross virgin females 15 days after eclosion, well after they were expected to have reached full development (Dunlap-Pianka *et al.*, 1977). After dissection, we scored the developmental progress of ovaries based on gross morphology (Morphoscore, Fig. S1). Almost all in-

dividuals from parental populations contained developing follicles which reached the final stages of vitellogenesis, and most had fully developed eggs (n=11/12). However, ovaries of F1 hybrids seemed to be devoid of developing oocytes (Fig. 2). Female backcrosses with *H. p. butleri* mothers yielded an approximately bimodal distribution of ovary phenotypes (Fig. 2B), while backcrosses to *H. p. sergestus* exhibited a skewed distribution, with mostly sterile individuals (Fig. 2B). However, our sample size in this direction was very small (n=8). In a subset of samples, we characterized the developmental stages of oocytes through nuclear staining with DAPI, using the stages described in the silkmoth as a reference *Bombyx mori* (Fig. 3). These scores correlated strongly with our morphoscores (Fig. S2) and revealed that all F1 and backcross individuals had early-stage follicles, but that sterile individuals showed arrested development after oocytes reached approximately stage 3. This stage marks a developmental timepoint after oocyte vs. nurse cell differentiation and follicle formation, but before vitellogenesis (yolk deposition) (Yamauchi & Yoshitake, 1984; Büning, 1994).

3.2 QTL mapping

We analysed 87 females from 7 families produced by backcrossing F1 males to *H. p. butleri* females. Using RADSeq reads aligned to Hmel2.5 reference genome, the linkage map for these individuals comprised 124,456 markers across 21 chromosomes, with a total map length of 1106.95 cM (Supplementary table S1 and figure S3).For further details on linkage maps see the supplementary results.

When scanning the genome for additive, single locus QTLs associated with sterility (H_1), we detected a broad significant region on the Z chromosome (Figures 4, S8, Table S3). The maximum LOD score was observed at 29.2 cM (Figure 4B, C), with mean predicted morphoscores of 1.81 for the *H. p. butleri* allele and 0.93 for the *H. p. sergestus* allele (R^2 = 0.20). The *H. p. butleri* allele had higher predicted morphoscores than the *H. p. sergestus*

allele all along the Z chromosome, but the difference declined to nearly zero towards the distal end of the chromosome.

When scanning for interacting QTLs we identified a negative interaction between a pair of loci at opposing ends (∼5 cM and ∼55 cM) of the sex chromosome, with the full epistatic model explaining 54% of the variance in morphoscore (Figure 5, Table S3). This pair of loci was highly significant (P<0.001) irrespective of whether we tested the combined additive effects and interaction (H_f), the additive effect of the second locus plus the interaction (H_{fv1}) or the interaction alone (H_{int}), and was robust to family-specific effects (figure S8). Recombinant Z chromosomes (Z_{BS} or Z_{SB}) had higher fitness (i.e., greater average morphoscores) than either non-recombinant, (Z_{BB} or Z_{SS}) (Figure 5). In addition to the interacting loci on the sex chromosome, we further identified significant pairs of QTLs between the Z chromosome and chromosomes 4, 12 and 15 (Table S3, figure 5). We then tested for the single QTL at 29.2 cM on the sex chromosome while controlling for the epistatically interacting pair of QTLs at either end. It remained significant, but its position shifted slightly to 33.86 cM. Bringing these three QTLs together in a single model ($y = \mu_1 + \beta_1 q_1 + \beta_2 q_2 + \beta_3 q_3 + \beta_4 q_1 q_2 + \varepsilon$) explained 62% of the variance in ovary morphoscore.

To further elucidate these results, we divided individuals into four groups depending on their genotypes at the two interacting loci on the Z chromosome (Z_{BB}, Z_{SS}, Z_{BS}, Z_{SB}). For each of these groups, we then plotted fertility against the fraction of the autosomes homozygous for *H. p. butleri* alleles (*B/B*). We hypothesised that if sterility is driven by interactions between the Z chromosome and autosomes, this fraction should be positively correlated with morphoscore for those individuals holding a Z_{BB}. As expected, for Z_{BB} individuals, we found a significant positive correlation between the proportion of the autosome derived from *H. p. butleri* (Figure 6A). Interestingly, we also found a significant negative correlation for Z_{SB} individuals. We then conducted QTL mapping on each of these groups. For individuals holding a Z_{SB}, we identified a significant interaction (LOD_{int} = 6.97, P<0.01, $R^2 = 0.79$) between loci at 9.3 cM on chromosome 8 and 11.9 cM on chromo-

some 20 (Figure 6B-D). No significant QTLs were detected for the other subgroups.

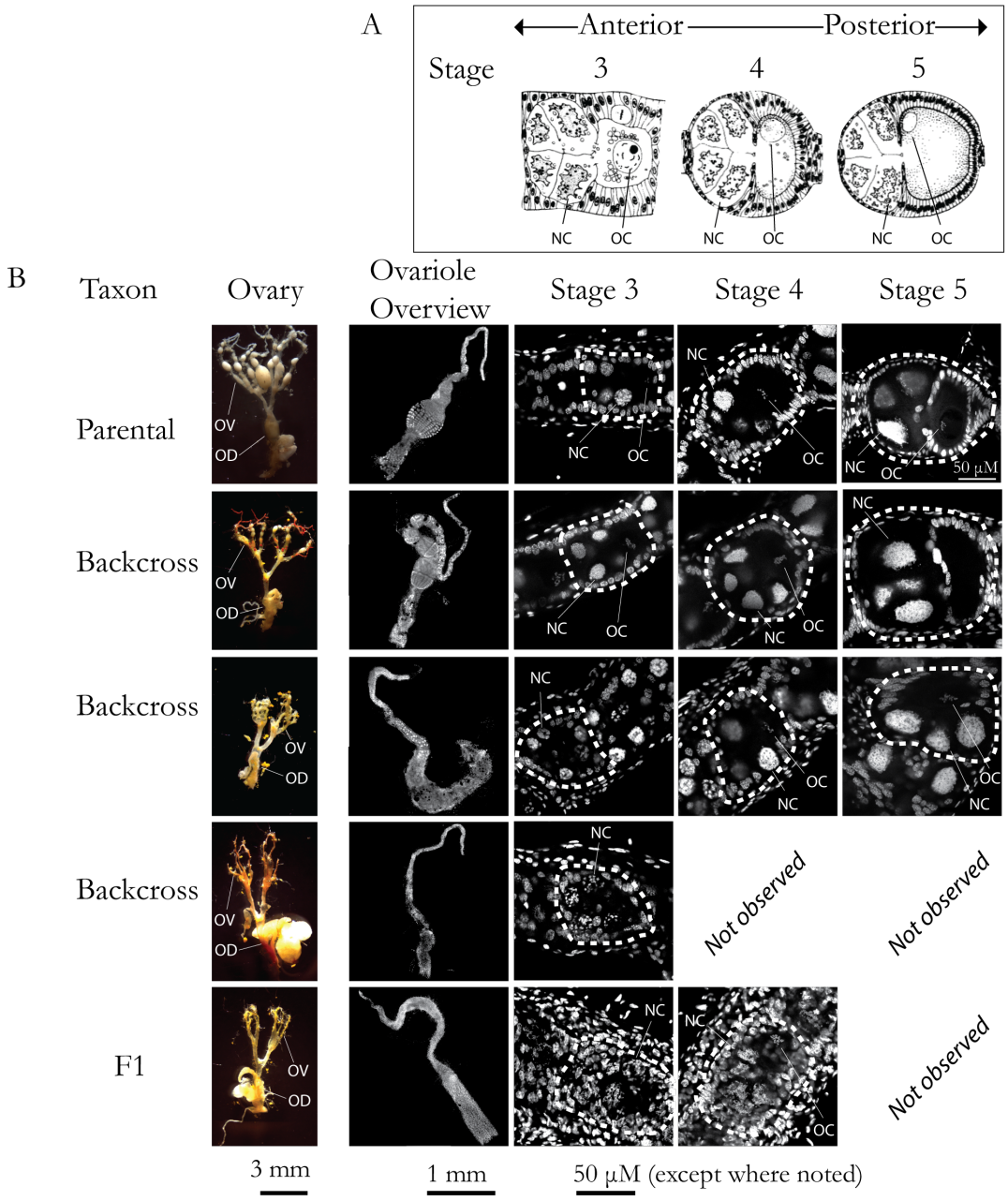

Figure 3: Developing oocytes. A. Idealized developing follicle stages, from (Yamauchi & Yoshitake, 1984) **B.** Brightfield and confocal images of DAPI-stained ovaries. Each row displays an overview image, as well as individual follicles at indicated stages from the same ovary. Scale bars for ovariole overviews are shown below the relevant column. Scale bar for stages 3-5 is shown below the stage 3 column, except where indicated in image. "Not observed" represent stages not present in the illustrated ovariole. In ovary images, one ovariole and the oviduct are indicated. Individual follicles are encircled by dashed lines. Where visible, one nurse cell nucleus (NC) and the oocyte cell nucleus (OC) in the highlighted follicle are indicated.

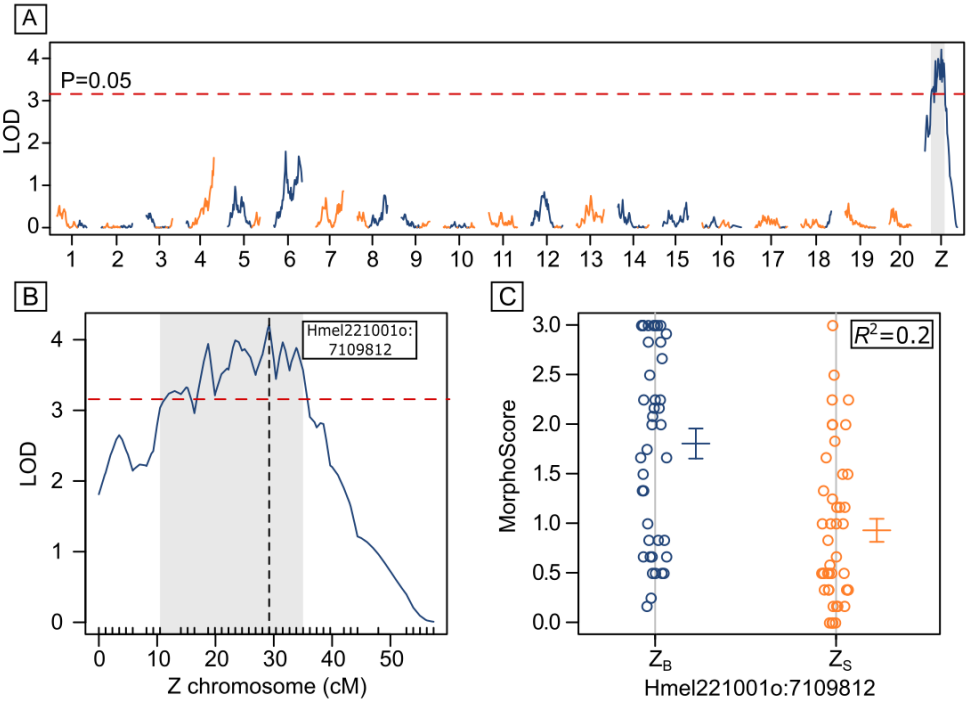

Figure 4: Single QTL analysis. A. LOD scores at each marker across the genome, calculated using H-K regression and with reads aligned to Hmel2.5. The red line indicates the genome-wide significance threshold ($p<0.05$; 10,000 permutations), and the grey shaded area the Bayesian credible intervals for the peak on Z. Lines are coloured depending on whether the *H. p. butleri* allele (blue) or the *H. p. sergestus* allele (yellow) had higher fertility. **B.** Enlargement of Z chromosome. **C.** Morphoscores at the QTL peak (29.21 cM, indicated by the vertical dashed line, and coresponding to physical position Hmel221001o:7109812). Z markers are hemizygous and coded by a single letter (B = *H. p. butleri* and S = *H. p. sergestus*), and explain 20% of the variance in morphoscore. Errors bars are standard errors.

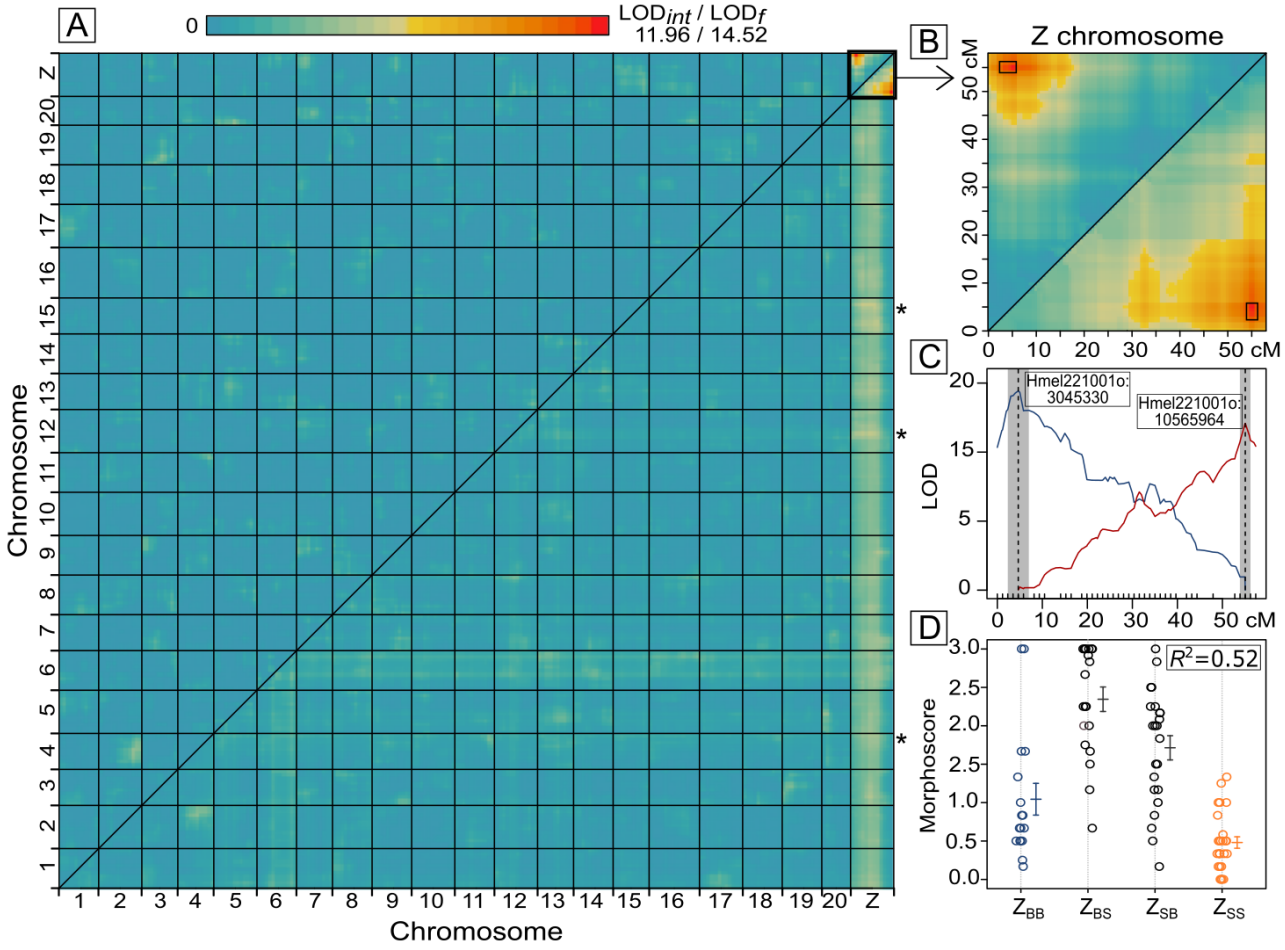

Figure 5: Multiple QTL analysis. A. Heat map for LOD_f scores (the full model; lower right triangle) and LOD_{int} scores (the interaction component; upper left triangle) between pairwise combinations of markers across the genome, using H-K regression and reads aligned to Hmel2.5. Blues indicate low scores, reds indicate high scores (maximum observed $\text{LOD}_{int} = 11.96$, maximum observed $\text{LOD}_f = 14.52$). Statistically significant LOD_f scores between the Z chromosome and the autosomes are marked with an asterisk. **B.** Enlargement of the Z chromosome, with the Bayesian credible intervals of the significant interaction shown as black boxes. **C.** Profile LOD curves for the epistatic QTL on Z chromosome. The dotted line gives the positions of the QTL peaks, and the grey shaded errors indicate the Bayesian credible intervals. The physical positions of the markers at the QTL peaks are shown in the text boxes. **D.** Morphoscores for 87 backcross individuals grouped by their haplotypes at the two interacting markers on the Z chrom (Hmel221001o:3045330 and Hmel221001:10565964). These 4 haplotypes explain 52% of the variance in morphoscore. Unrecombined pairs of markers inherited from *H. p. butleri* (Z_{BB}) or *H. p. sergestus* (Z_{SS}) are coloured blue and orange, respectively. Errors bars are standard errors.

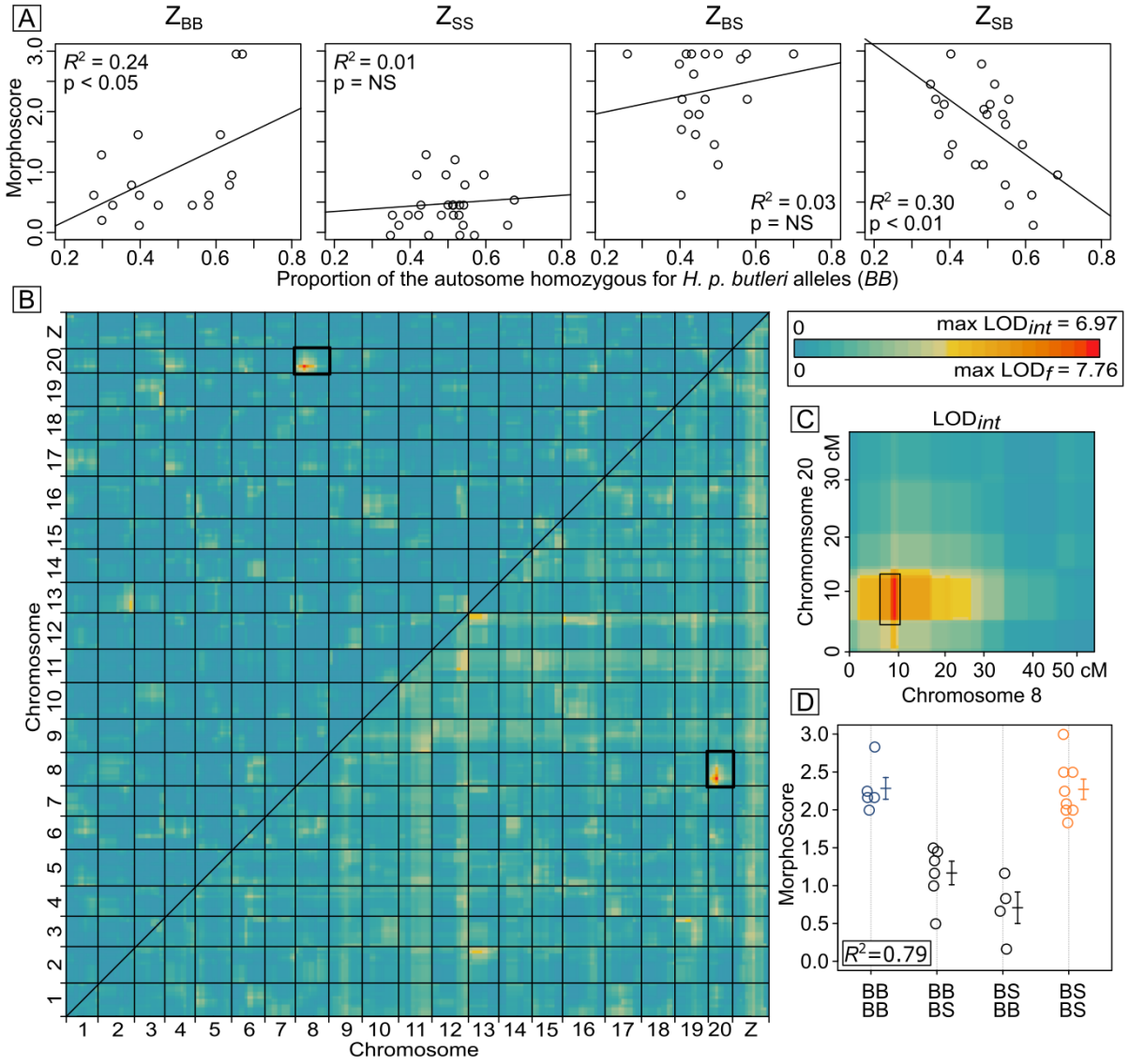

Figure 6: Analysis of Z linked epistatic markers. A. For each Z chromosome haplotype (Z_{BB}, Z_{SS}, Z_{BS}, Z_{SB}), the proportion of the autosome that is homozygous for *H. p. butleri* alleles was plotted against morphoscore. **B.** Heat map for two dimensional QTL scan using only Z_{SB} individuals. LOD_f scores are shown in the lower right triangle) and LOD_{int} scores in the upper left triangle. **C.** Enlargement of LOD_{int} between chromosome 8 and chromosome 20 using just Z_{SB} individuals, with the Bayesian credible intervals of the QTLs shown as black boxes. **D.** Morphoscores for the four autosomal genotypes of Z_{SB} individuals, with the genotype at chromosome 8 (Hmel208001o:1005579) written above, and the genotype at chromosome 20 (Hmel220003o:5817143) written below. These genotypes explain 79% of the variance in morphoscore. Errors bars are standard errors.

LOD_f	chr	QTL1 marker	cM	limits (cM)	limits (physical)	chr	QTL2 marker	cM	limits (cM)	limits (physical)	μ_1	$\beta_1 q_1$	$\beta_2 q_2$	$\beta_3 q_1 q_2$	R^2
4.21**	Z	Hmel221001o: 7109812	29.21	10.49-35.02	4710269-7966755						1.370.09	0.88± 0.19***			0.2
6.72+	4	Hmel204001o: 9208991	49.03	25.72-49.03	6102423-7036†	Z	Hmel221001o: 7109812	29.21	15-36.19	5201659-7976401	1.32± 0.09***	-0.56± 0.18**	0.81± 0.18***	-0.59± 0.37	0.3
6.65+	12	c12.loc25	25	17.47-55.93	5100627-16319705	Z	Hmel221001o: 5752071	18.75	3.49-33.86	2210592-7862353	1.39± 0.09***	0.44± 0.18*	0.87± 0.18***	1± 0.36**	0.3
7.87**	15	Hmel215003o: 8041294	38.48	10-44.3	4230382-9907975	Z	Hmel221001o: 5752071	18.75	14.07-30	4795563-7180089	1.34± 0.09***	0.29± 0.18	0.7± 0.18***	1.47± 0.36***	0.34
14.52***	Z	Hmel221001o: 3045330	4.65	2.33-5.81	2179644-4463341	Z	Hmel221001o: 10565964	55.08	53.91-56.24	8861371-13311117	1.4± 0.07***	0.6± 0.15***	-0.03± 0.15	-2.53± 0.3***	0.54
7.76*	8	Hmel208001o: 1005579	9.3	8.14-10	618400-1390337	20	Hmel220003o: 5817143	11.86	6-12	3012264-6466723	1.61± 0.08***	0.24± 0.16	-0.22± 0.16	2.68± 0.32***	0.79

Table 1: Summary of significant single locus (H_1) and two locus (H_f) QTL models, using reads aligned to Hmel2.5. The first column gives the LOD score of the full model (H_f), with the significance estimated by permutation (+P<0.1, *P<0.05, **P<0.01, ***P<0.001). The next columns are the chromosome and QTL marker (scaffold and median physical position within the peak). The centimorgan limits are the Bayesian credible intervals, and the physical limits are the nearest typed flanking markers of that interval (all physical limits were on the same scaffold as the QTL peak, except for the chromosome 4 interaction † with Z, which was on scaffold Hmel204003 of Hmel2.5). The final five columns give the parameter estimates and R^2 of the model. $\beta_1 q_1$ and $\beta_2 q_2$ are the estimated additive effects for the QTLs, i.e. the difference between the phenotypic averages for the alternative genotypes, and $\beta_3 q_1 q_2$ is the coefficient for the interaction between the 2 loci. Model coefficients comprise the estimated value, the standard error, and the significance (thresholds as above). The significant interaction between chromosome 8 and 20 was detected using individuals holding a Z_{SB} chromosome only.

3.3 Population genomics of the Z chromosome

Because *H. p. sergestus* and *H. p. butleri* are parapatric and probably mate occasionally in the wild, we expect that hybrid sterility loci should be under strong divergent selection in the two populations. Consistent with this hypothesis, we find that genomic divergence, between *H. p. sergestus* and *H. p. butleri*, as measured by F_{ST} and D_{xy}, is strongly elevated over a 0.3 Mb region in the center of the Z chromosome, overlapping the center of the additive QTL peak (Fig. 7). F_{ST} is generally elevated at the ends of the Z chromosome as well, in line with the interactive QTL. However, these regions have very low recombination, which can lead to high F_{ST} values even in the absence of selection Burri (2017). More work must be done to verify if the F_{ST} values observed at the ends of the chromosome are in fact a result of hybrid sterility.

In the central region of high divergence between *H. p. sergestus* and *H. p. butleri*, we also observed a reduction in D_{xy} between *H. p. butleri* and its sister taxa, *H. elevatus* (Fig. 7). *H. p. butleri* and *H. elevatus* are known to hybridize in the wild, and their F1 hybrid offspring are fully fertile Rosser *et al.* (2019). This raises questions about the relative timings of species divergence, subspecies divergence, hybridization, and evolution of sterility loci that will be expanded upon in the discussion.

3.4 Differential expression analysis

Since we localized the dysgenic phenotype to early stage, pre-vitellogenic oocytes (Fig 3), we focused on this region in quantifying RNA expression differences among backcross individuals. We dissected the pre-vitellogenic (approx. stage 3 and earlier) follicles from six backcross ovaries, two of which were assigned a morphoscore of 0-1, two of 1-2, and two of 2-3. Because we micro-dissected pre-vitellogenic follicles to investigate the specific phenotype of developmental failure in early-stage oocytes, we further classified the phenotypes

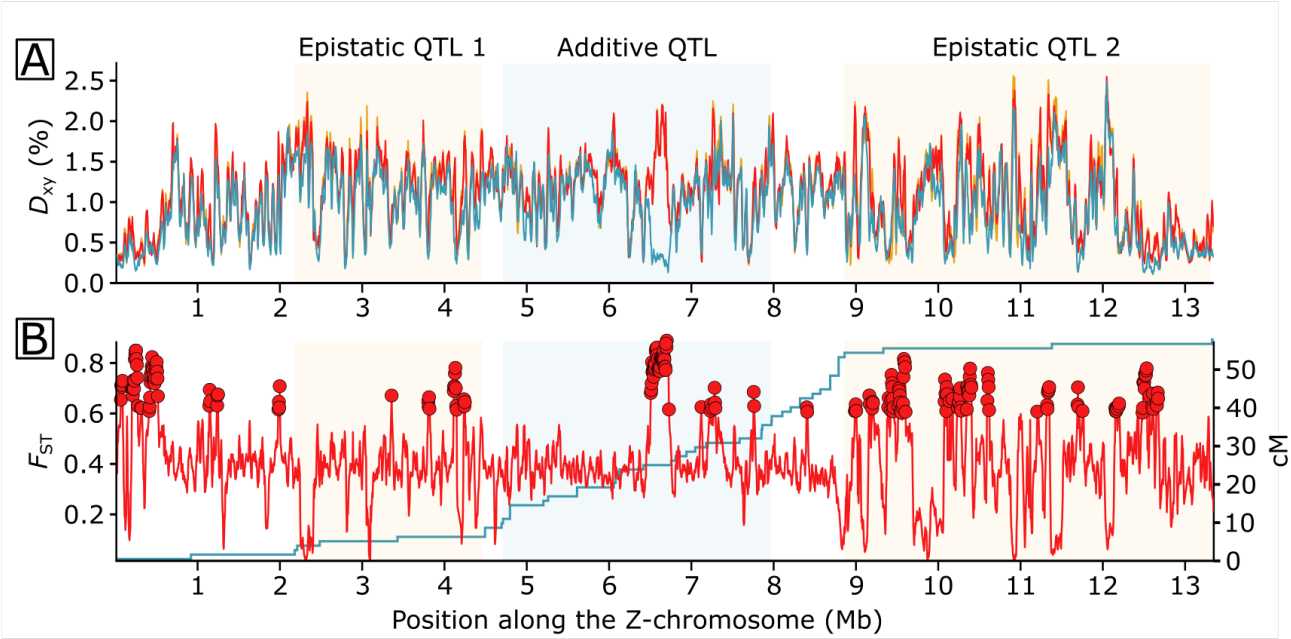

Figure 7: Population genetic summary statistics and recombination rate along the Z chromosome. A. Pairwise absolute genetic distance (D_{xy}) between *H. p. butleri* and *H. p. sergestus* (red), *H. p. butleri* and *H. elevatus* (blue), and *H. p. sergestus* and *H. elevatus* (yellow). **F.** Genetic differentiation (F_{ST}; red line) between *H. p. butleri* and *H. p. sergestus*, with genome-wide B_{ST} outliers as points, based on Z-scores > 3. The blue line shows genetic distance (cM) plotted against physical distance (Mb). Shaded areas correspond to the Bayesian credible intervals for the two epistatic QTL at 4.65 and 55 cM, and the single additive QTL at 29.21 cM. D_{xy} and F_{ST} were calculated in sliding windows of 25 kb (with 5 kb increments).

with a binary scheme: 0 for absence of vitellogenic follicles, 1 for presence of any vitellogenic follicles. In each case, tissue was dissected from all four ovarioles of a single ovary, and we acquired approximately 49 million reads per individual. After filtering our data for sequencing and mapping quality, we quantified expression of 16,774 transcripts (Fig. 5A).

We then carried out a principal components analysis of these expression data. PC1, which explains over 50% of variance, separates the three morphoscore categories in order (Fig. 8B). We performed a Wald test to evaluate the effect of change in expression of each transcript to the sterility phenotype in the backcrosses (Chen *et al.*, 2011). After correcting for multiple comparisons, a total of 14%, or 2315 transcripts showed significant effects of expression on binary phenotype ($q < 0.05$) (Fig. 5C,D). Of these, 941 displayed a positive association with development, meaning that the transcript was expressed at a higher level in more highly developed ovaries. The remaining 1386 differentially expressed transcripts displayed a negative association with development. To narrow our list of candidate genes, we filtered our differentially expressed transcripts for genes implicated in butterfly oogenesis. We identified 1,771 transcripts in the *H. melpomene* transcriptome that were strong BLAST hits of genes expressed in *Pararge aegeria* eggs and ovaries (Carter *et al.*, 2013). As expected, these genes showed generally high expression levels in the sampled transcriptomes relative to other genes (Fig. 8A,B). 306 (17%) of these genes were also differentially expressed in backcrosses with different developmental phenotypes. One of the transcripts, *Trailer hitch* (*tral*), has high overall expression, strong differential expression, and is known to be involved in oogenesis in *D. melanogster* and *P. aegeria* (Wilhelm *et al.*, 2005; Carter *et al.*, 2013) (Fig. 8).

We then searched within the Bayesian credible intervals of the QTLs for differentially expressed transcripts with orthologs implicated in oogenesis in either *D. melanogaster* or *P. aegeria* (Fig. 8). Applying this approach to the two interacting QTLs on the Z chromosome, we identified one candidate gene (*magu*) in the first QTL 4.65 cM, and eight in the second QTL at 55 cM (*Egfr*, *fax*,*Gs2*, *Nedd8*, *parvin*, *Prm*, *sls*, *Syx7*). Within the cen-

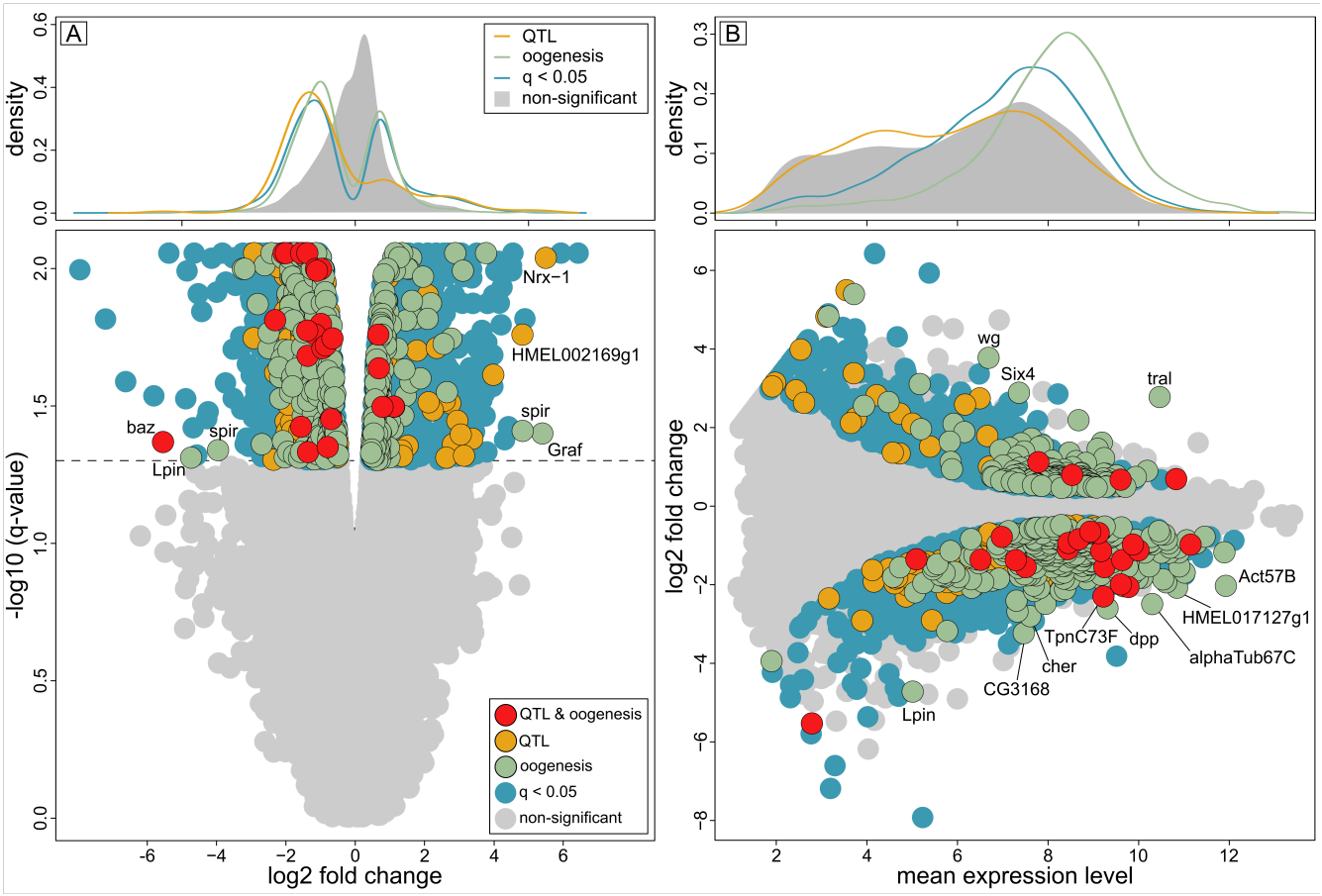

Figure 8: Backcross differential expression. A. Volcano plot. The change in expression between fertile (morphoscore > 1.5) and sterile (morphoscore ≤ 1.5) ovaries is plotted against the q-value. Positive values of fold change imply higher expression in fertile ovaries. Significantly differentially expressed transcripts are shown in blue, of which those with orthologues implicated in oogenesis in either *D. melanogaster* or *P. aegeria* are in green, those within the Bayesian credible intervals of QTLs on chromosomes 8, 20 and Z are shown in orange, and those fitting all these criteria are in red. QTL and/or oogenesis outliers (those with absolute fold change values in the top 1% of all transcripts) are labelled. Non-sigificant transcripts are in grey. The density plot shows the distribution of fold change values i) implicated in oogenesis, ii) found within QTLs, and iii) significantly differentially expressed in fertile/sterile individuals, with the density of all transcripts shaded in grey. Interestingly, most QTL transcripts are overexpressed in the sterile ovaries. **B.** Mean expression of transcripts plotted against fold change in expression. Symbols as in A, except the density plot shows the distribution of mean expression levels. Labelled QTL and/or oogenesis outliers were defined as those falling in the top 1% of transcripts ranked using mean expression $\times$ fold change.

tral additive QTL on the Z chromosome at 29.2 cM, we found two candidate genes (*trol* and *csw*).Within the QTL at 11.86 cM on chromosome 20, we identified 11 candidates (*baz*, *CG12104*, *CG1572*, *CrebB*, *Ect4*, *Eip75B*, *ine*, *mys*, *Pitslre*, *Ran*, *TpnC73F*). In the QTL at 9.3 cM on chromosome 8, there were only 3 differentially expressed transcripts, only one of which had an orthologue known to be involved in oogenesis (*Art1*). However, one transcript (HMEL037834g1.t2, with the orthologue *Nrx-1*) stood out due to very high fold change ($\beta = 5.51$) between sterile and fertile individuals, and its physical position (997,675 - 1,074,578 Mb) falls in the centre of the peak of the QTL (844,849 - 1,232,231 Mb). Genes involved in oogenesis and those significantly differentially expressed between ovaries of varying development were skewed towards being overexpressed in undeveloped ovaries. This pattern was even more extreme among all genes in QTL intervals, regardless of their function (Fig. 8A). This could mean that the high expression is due to a general phenomenon such as increased chromatin availability, or derepression of a broad transcriptional regulator.

4 Discussion

4.1 Genetics of hybrid incompatibility in *Heliconius pardalinus*

Here, we show that crossing *H. p. butleri* and *H. p. sergestus* in both directions results in F1 hybrid females that are sterile due to disrupted oocyte development, and QTL analysis of backcrosses shows that sterility is sex-linked. We identify a strong epistatic interaction between loci at opposite ends of the Z chromosome, and a broader additive QTL towards the centre. In addition, we identify significant QTLs linking the Z chromosome and autosomes (4, 12 and 15), and an epistatic interaction between chromosomes 8 and 20. By intersecting these with the results of differential expression analysis, we identify a number of candidate genes.

Our results conform to the "two rules of speciation" (Coyne, 2018): (1) the "large X effect" (in Lepidoptera, this is a large effect of the Z chromosome) on hybrid incompatibility, and (2) the tendency for greater sterility/inviability in the heterogametic sex (Haldane's rule) than in the homogametic sex. There is general consensus that Haldane's rule can be explained in part by "dominance theory", where interactions between recessive X– (or Z)–linked alleles from one species and a hybrid autosomal genetic background cause incompatibilities (Coyne & Orr, 2004). The large-X effect is expected under dominance theory, but faster X evolution, X chromosome meiotic drive or incorrect dosage compensation of X-linked genes in the heterogametic germline may also play a role (Presgraves, 2008). For example, in some *Drosophila*, the sex chromosome has a somewhat higher density of hybrid male sterility factors than the autosomes, suggesting faster X evolution, and it also seems likely that male sexual traits evolve faster than those in females due to sexual selection (the so called "faster male" hypothesis) (Wu & Davis, 1993; Masly & Presgraves, 2007). In Lepidoptera, the Z chromosome has long been implicated in a large Z effect (Prowell Pashley, 1998), although there are many caveats, because the Z chromosome is also expected to have a lower effective population size, and so relative divergence of sex chromosomes versus autosomes may be overestimated (Martin & Jiggins, 2017; Presgraves, 2018). In contrast to the case in *Drosophila*, however, it is the females that show hybrid sterility, so sexual selection among males and the "faster male" theory cannot contribute to Haldane's Rule here.

Geographically distant populations of *Heliconius melpomene* show hybrid female sterility in only one direction of cross (Jiggins *et al.*, 2001). Asymmetries in hybrid sterility are predicted when Dobzhansky-Muller incompatibilities are relatively few, due to recent divergence (Muller, 1942; Turelli & Moyle, 2007). In our case, crosses in both directions between *H. p. sergestus* and *H. p. butleri* produce sterile hybrid females, suggesting more complex, multilocus causes of sterility. Furthermore, if hybrid sterility arises due to interactions between the Z chromosome and autosomes, it implies that there are sterility factors at some autosomal loci where *H. p. butleri* alleles are dominant, and others at which *H. p. sergestus*

alleles are dominant. The observation that individuals with unrecombined Z chromosomes have low fertility supports this (Figure 5D). A Z chromosome inherited from *H. p. sergestus* will have deleterious interactions with any autosomal loci where *H. p. butleri* alleles are dominant, and so in a backcross to *H. p. butleri*, individuals carrying such a chromosome should never have full fitness. Similarly, individuals with a Z chromosome inherited from *H. p. butleri* should also have reduced fertility, because of deleterious interactions with autosomal loci with a dominant *H. p. sergestus* allele. However, some fraction of backcrosses with unrecombined *H. p. butleri* Z chromosomes should be fully fertile, when they happen to be homozygous for *H. p. butleri* alleles at all *H. p. sergestus* autosomal dominant loci in the interaction. Indeed, two individuals do; these are clearly visible as outliers in Figure 5D and have the highest proportion of their autosomes homozygous for *H. p. butleri* alleles (*B/B*) (Figure 6A). This result is therefore consistent with epistatic interactions between the autosomes and sex chromosomes, and additionally suggests that a limited number of autosomal loci are involved. In contrast, individuals holding a recombined Z_{BS} or Z_{SB} chromosome have higher fertility. Z_{SB} individuals are also notable for high variance in fertility (Figure 5D) and they show a negative correlation between fertility and the proportion of their autosomes that is *B/B* (Figure 6A). QTL mapping for Z_{SB} individuals (Figure 6B) shows that this variation in fertility can be explained almost entirely by an interaction between chromosome 8 and chromosome 20, which explains 79% of the variance in morphoscore (Table S3). Individuals which are either homozygous or heterozygous at both loci are fully fertile, but individuals which are homozygous at one locus and heterozygous at the other are infertile (Figure 6D).

4.2 Candidate genes

Oocyte development fails in sterile hybrid females at Lepidoptera stages 3-4 (Fig 3) (homologous with stages 8-9 of oogenesis in *D. melanogaster*), a period characterised by bor-

der follicle cell migration (Yamauchi & Yoshitake, 1984). Within the Bayesian credibility intervals of the sex-linked QTLs and the interacting QTLs on chromosomes 8 and 20, we identified 24 transcripts that were differentially expressed between sterile and fertile individuals and had orthologs known to be involved in oogenesis in either *D. melanogaster* or the Speckled Wood butterfly (*P. aegeria*). Three of these are known to be associated with border follicle cells.

Within the proximal epistatic Z-linked QTL at 4.65 cM, we identified only one candidate gene, *magu*, mutants of which are known to cause defective border cell migration in *D. melanogaster* (Raza *et al.*, 2019). Within the distal Z-linked QTL at 55.08 cM, 8 candidate genes were identified. One of these, *Epidermal growth factor receptor* (*Egfr*), guides dorsal migration of border cells during *Drosophila* oogenesis stage 9 (Duchek & Rørth, 2001), and is also expressed in the ovarian transcriptome of *P. aegeria* (Carter *et al.*, 2013). We found 11 candidates involved in oogenesis within the QTL on chromosome 20. One of these, the multi-PDZ domain protein *bazooka* (*baz*), regulates border cell migration (Pinheiro & Montell, 2004), is expressed in the *P. aegeria* ovarian transcriptome, and furthermore is notable for being highly overexpressed in sterile individuals ($\log_2$ fold change =-5.52 for transcript HMEL016161g1.t3, the sixth lowest value in the dataset [Figure 8A]). On chromosome 8, HMEL037834g1.t2, with ortholog *Neurexin 1* (*Nrx-1*), stood out as having the third highest $\log_2$ fold change (5.51) in the dataset. While not known to be involved in oogenesis, *Neurexin 1* is known to influence expression of *gurken* (*grk*) (Geng & Macdonald, 2007). The asymmetrical localization of *gurken* mRNA is key for its function during oogenesis, to establish anterior-posterior and dorso-ventral axes in the egg and embryo, and *gurken* encodes a TGFα family signaling ligand that activates the intracellular MAP kinase pathway via the the product of *Egfr*.

Differentially expressed transcripts located within QTL intervals, such as those discussed above, represent candidate regions for *cis*-acting differences between the two species. Looking at differential expression alone, we can also identify putative trans-acting effects, or

downstream consequences of the QTL identified here. *Trailer-hitch* (*tral* has strong differential expression, high overall expression in ovaries, and is known to be involved in *Drosophila* oogenesis at stages 8-9 (Figure 8, Figure 9) (Wilhelm *et al.*, 2005; Snee & Macdonald, 2009). Like *Nrx-1*, *tral* is involved in specifying the localization of the dorso-ventral patterning gene *grk*.

We also noticed that alternative splices of transcript HMEL015815g1, orthologous to gene *spire* (*spir*) stood out as outliers in figure 8A. Although mapping to chromosome 1 and not in a QTL, HMEL037834g1.t2 was significantly underexpressed in sterile individuals ($\log_2$ fold change = 4.85), and HMEL015815g1.t6 significantly overexpressed ($\log_2$ fold change = -3.94). *spire* is involved specifically in stages 8-9 of oogenesis in *D. melanogaster*, where it affects the dorsal-ventral and anterior-posterior axes of the egg (Dahlgaard *et al.*, 2007; Wellington *et al.*, 1999).

4.3 Comparison to *Drosophila* hybrid incompatibility

The fruit fly *Drosophila melanogaster* and its relatives have long been used as models to study developmental phenomena, including the genetic and molecular basis of hybrid sterility. Some classical Dobzhansky-Muller incompatibilities have been identified and characterized in the genus (Brideau *et al.*, 2006; Tang & Presgraves, 2009; Bayes & Malik, 2009). Because *Drosophila* has XY sex determination, it is normally males that show hybrid sterility (Haldane, 1922). However, hybrid female dysgenesis has been observed in *D. melanogaster* in so-called P-M hybrids, in which oogenesis arrests at a very early stage (Kidwell *et al.*, 1977; Schaefer *et al.*, 1979; Bingham *et al.*, 1982). This phenotype is due to a loss of control of P element transposition, which is typically suppressed through the *Drosophila* piRNA pathway in P strain flies (Evgen'ev *et al.*, 1997; Kelleher *et al.*, 2012). Superficially, the *Heliconius* sterility phenotype described in this study parallels the *Drosophila* case. The hypothesis that transposon silencing through the piRNA pathway is mis-regulated

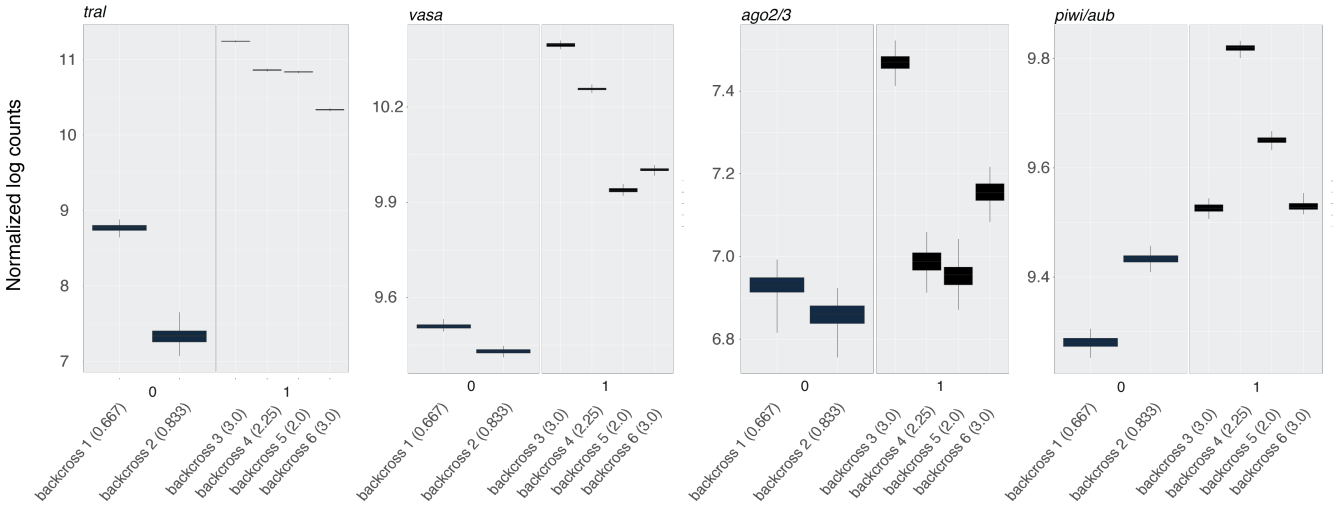

Figure 9: mRNA expression of piRNA-related transcripts. Expression levels for *tral*, *vasa*, *AGO2/3*, and *piwi/aubergine* are shown. Individuals are categorized based on whether they contained vitellogenic follicles (1) or not (0). In addition, each individual's morphoscore is indicated in parentheses. In each case, abundance is lower in undeveloped ovaries, but only significantly so for *tral* and *vasa*. Width of bars represents the interquartile range of bootstrap-resampled reads for each individual. Lower whisker is the smallest observation greater than or equal to lower edge of bar - 1.5 * IQR. Upper whisker is the largest observation less than or equal to upper edge of bar + 1.5 * IQR.

in sterile female hybrids has been explicitly tested in a different *Heliconius* system, *H. melpomene* and *H. cydno*. A subset of transposable elements were indeed derepressed in F1 hybrids, but there was no evidence that piRNAs themselves or three proteins involved in the piRNA pathway were misexpressed (Pinharanda, 2017). In the current study, *H. pardalinus* female hybrids with low fertility scores expressed three proteins in the piRNA pathway (*piwi/aubergine*, *AGO2/3*, and *vasa*) at lower levels than in well-developed individuals, though only *vasa* expression differences were significant (Fig. 9). In addition, one of our candidate genes, *tral*, is known to form a complex with piRNA proteins which inhibits transposition Liu *et al.* (2011). A *Drosophila*-like transposon derepression mechanism is therefore plausible in this system, but more work must be done to test this hypothesis.

4.4 Evolution of hybrid incompatibilities

Heliconius p. sergestus is endemic to the dry forests of upper Huallaga valley in the Andes, and is separated from *H. p. butleri* in the Amazonian lowlands by the intervening Cordillera Escalera (Figure 1). Nonetheless, the two subspecies are known to come into contact and some putative wild hybrids exist (Michel Cast *pers. comm.*, Rosser *et al.*, 2019; Brown, 1976). Theory predicts that in the face of gene flow, DMIs are more likely to be maintained when they are linked to traits involved in divergent adaptations (Bank *et al.*, 2012), and *Heliconius* provide a possible example of this (Merrill *et al.*, 2011). Divergent selection to the dry habitats might therefore have driven the evolution of sterility between *H. p. sergestus* and *H. p. butleri*.

However, previous phylogenetic analyses found *H. pardalinus* to be paraphyletic, with *H. p. butleri* more closely related to the widespread Amazonian species *H. elevatus* than to *H. p. sergestus* (*Heliconius* Genome Consortium, 2012; Rosser *et al.*, 2019). For example, within the F_{ST} peak between *H. p. sergestus* and *H. p. butleri* on the Z chromosome (Figure 7B), *H. p. butleri* and *H. elevatus* are barely differentiated (Figure 7A). Nonetheless, the ancestry of several genomic regions recapitulated traditional species relationships, in which the two *H. pardalinus* subspecies were sisters (Kryvokhyzha, 2014). These data are consistent with a model where *H. p. butleri* and *H. elevatus* exchanged genes through hybridization in the Amazon basin, to the exclusion of *H. p. sergestus* (Kryvokhyzha, 2014). Crosses between *H. p. butleri* and *H. elevatus* are fully fertile, but crosses between *H. p. sergestus* and *H. elevatus* are sterile (Rosser *et al.*, 2019), with similar phenotypes to those identified here. It is therefore plausible that sterility could have evolved after the initial speciation event between *H. pardalinus* and *H. elevatus*, meaning that at one time hybrids between *H. p. butleri* and *H. elevatus* were sterile as well. This sterility phenotype may have been lost due to gene flow in the Amazon, a hypothesis that merits further study.

5 Conclusions

The genetics of Haldane's Rule and Dobzhansky-Muller incompatibilities has been extensively studied in *Drosophila* and a few other male heterogametic systems, but hitherto there has been little genetic work on female heterogametic systems such as Lepidoptera and birds. Our work with *Heliconius* butterflies represents the first such study in Lepidoptera. Two previous papers had investigated similar hybrid female sterility between *Heliconius melpomene* and *H. cydno* (Jiggins *et al.*, 2001; Naisbit *et al.*, 2002). In these previous studies, the Z chromosome was shown to be involved, but the paucity of genetic markers precluded mapping hybrid sterility traits.

Here we employ a QTL analysis involving thousands of markers across the genome, and show that multiple regions are involved in hybrid female sterility, with an especially large effect of the Z chromosome. We demonstrate strong epistatic interactions between loci at opposite ends of the Z chromosome. We use RNASeq to investigate differential gene expression between backcross females with ovaries that developed versus those that did not, and by combining these results with those of QTL mapping we identify six candidate genes (*magu*, *Egfr*, *baz*, *Nrx-1*, *tral*, and *spir*) that may be involved in the sterility phenotype. In addition, we provide evidence of putative Dobzhansky-Muller incompatibilities between the Z chromosome and autosomes, and between chromosomes 8 and 20. Many questions remain unanswered, and functional genetic studies will be required to understand the mechanisms of ovariole development failure in hybrids. Nonetheless, we were able to show that many of the findings from studies of Haldane's Rule in *Drosophila* male sterility (for example, multilocus effects, epistasis, involvement of the sex chromosome), are replicated in female sterile hybrids in a female heterogametic system, *Heliconius* butterflies. Future work should address the exact genetic basis of sterility, and the potential for a tie-in with selfish genetic elements and the genes that act to defend the genome against their replication.

6 Acknowledgements

This work was funded by NERC grant NE/K012886/1 to KKD and Harvard University. We also thank SERFOR, the Peruvian Ministry of Agriculture, and the Área de Conservación Regional Cordillera Escalera for collecting permits (0289-2014-MINAGRI-DGFFS/DGEFFS, 020-014/GRSM/PEHCBM/DMA/ACR-CE, 040–2015/GRSM/PEHCBM/DMA/ACR-CE). We are extremely grateful to the following people for help and support with field work in Peru: Ronald Mori Pezo, Corita Cordova, Mario Tuanama, Jarreth Caldwell, Christian Perez, Cesar Lopez, Stephanie Galluser and Gerardo Lamas. We are also very grateful to Michael Turelli and Daven Presgraves for helpful discussions regarding QTL results, and Shreeharsha Tarikere for commenting on the manuscript.

References

Bank, C., Bürger, R. & Hermisson, J. (2012) The limits to parapatric speciation: Dobzhansky–Muller incompatibilities in a continent–island model. *Genetics* **191**, 845–863.

Bateson, W. (1909) Heredity and variation in modern lights. *Darwin and Modern Science* (ed. A.C. Seward), pp. 85–101, Cambridge University Press, Cambridge, UK.

Bayes, J.J. & Malik, H.S. (2009) Altered heterochromatin binding by a hybrid sterility protein in *Drosophila* sibling species. *Science* **326**, 1538–1541.

Bingham, P.M., Kidwell, M.G. & Rubin, G.M. (1982) The molecular basis of PM hybrid dysgenesis: the role of the P element, a P-strain-specific transposon family. *Cell* **29**, 995–1004.

Bray, N.L., Pimentel, H., Melsted, P. & Pachter, L. (2016) Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* **34**, 525–527.

Brideau, N.J., Flores, H.A., Wang, J., Maheshwari, S., Wang, X. & Barbash, D.A. (2006) Two Dobzhansky-Muller genes interact to cause hybrid lethality in *Drosophila*. *Science* **314**, 1292–1295.

Broman, K.W., Wu, H., Sen, Ś. & Churchill, G.A. (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890.

Brown, K.S. (1976) An illustrated key to the silvaniform *Heliconius* (lepidoptera: Nymphalidae) with descriptions of new subspecies. *Transactions of the American Entomological Society* **102**, 373–484.

Büning, J. (1994) *The Insect Ovary: Ultrastructure, Previtellogenic Growth and Evolution.* Springer Science & Business Media, Berlin.

Burri, R. (2017) Interpreting differentiation landscapes in the light of long-term linked selection. *Evolution Letters* **1**, 118–131.

Butlin, R., Debelle, A., Kerth, C., Snook, R.R., Beukeboom, L.W., Castillo, R.C., Diao, W., Maan, M.E., Paolucci, S., Weissing, F.J. *et al.* (2012) What do we need to know about speciation? *Trends in Ecology & Evolution* **27**, 27–39.

Carter, J.M., Baker, S.C., Pink, R., Carter, D.R., Collins, A., Tomlin, J., Gibbs, M. & Breuker, C.J. (2013) Unscrambling butterfly oogenesis. *BMC Genomics* **14**, 283.

Castillo, D.M. & Barbash, D.A. (2017) Moving speciation genetics forward: modern techniques build on foundational studies in *Drosophila*. *Genetics* **207**, 825–842.

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A. & Cresko, W.A. (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22**, 3124–3140.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. & Lee, J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, s13742–015.

Chen, Z., Liu, J., Ng, H.K.T., Nadarajah, S., Kaufman, H.L., Yang, J.Y. & Deng, Y. (2011) Statistical methods on detecting differentially expressed genes for RNA-seq data. *BMC Systems Biology* **5**, S1.

Corbett-Detig, R.B., Zhou, J., Clark, A.G., Hartl, D.L. & Ayroles, J.F. (2013) Genetic incompatibilities are widespread within species. *Nature* **504**, 135–137.

Coughlan, J.M. & Matute, D.R. (2020) The importance of intrinsic postzygotic barriers throughout the speciation process. *Philosophical Transactions of the Royal Society B* **375**, 20190533.

Coyne, J.A. (1992) Genetics and speciation. *Nature* **355**, 511–515.

Coyne, J.A. (2018) "Two rules of speciation" revisited. *Molecular Ecology* **27**, 3749–3752.

Coyne, J.A. & Orr, H.A. (1989) Patterns of speciation in *Drosophila*. *Evolution* **43**, 362–381.

Coyne, J.A. & Orr, H.A. (1997) "Patterns of speciation in *Drosophila*" revisited. *Evolution* **51**, 295–303.

Coyne, J.A. & Orr, H.A. (2004) *Speciation*. Sinauer Associates, Sunderland, MA.

Dahlgaard, K., Raposo, A.A., Niccoli, T. & St Johnston, D. (2007) Capu and Spire assemble a cytoplasmic actin mesh that maintains microtubule organization in the *Drosophila* oocyte. *Developmental Cell* **13**, 539–553.

Darwin, C. (1859) *On the Origin of Species by Means of Natural Selection*. John Murray, London.

Davey, J.W., Barker, S.L., Rastas, P.M., Pinharanda, A., Martin, S.H., Durbin, R., McMillan, W.O., Merrill, R.M. & Jiggins, C.D. (2017) No evidence for maintenance of a sympatric *Heliconius* species barrier by chromosomal inversions. *Evolution Letters* **1**, 138–154.

Delph, L.F. & Demuth, J.P. (2016) Haldane's rule: genetic bases and their empirical support. *Journal of Heredity* **107**, 383–391.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., Del Angel, G., Rivas, M.A., Hanna, M. *et al.* (2011) A framework for variation discovery and genotyping using next-generation dna sequencing data. *Nature Genetics* **43**, 491.

Dobzhansky, T. (1936) Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics* **21**, 113.

Doncaster, L. & Raynor, G.H. (1906) Breeding experiments with Lepidoptera. *Proceedings of the Zoological Society of London* **1**, 125–133.

Duchek, P. & Rørth, P. (2001) Guidance of cell migration by EGF receptor signaling during *Drosophila* oogenesis. *Science* **291**, 131–133.

Dunlap-Pianka, H., Boggs, C.L. & Gilbert, L.E. (1977) Ovarian dynamics in heliconiine butterflies: programmed senescence versus eternal youth. *Science* **197**, 487–490.

Emms, D.M. & Kelly, S. (2019) Orthofinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**, 1–14.

Etter, P.D., Preston, J.L., Bassham, S., Cresko, W.A. & Johnson, E.A. (2011) Local *de novo* assembly of RAD paired-end contigs using short sequencing reads. *PloS One* **6**.

Evgen'ev, M.B., Zelentsova, H., Shostak, N., Kozitsina, M., Barskyi, V., Lankenau, D.H. & Corces, V.G. (1997) *Penelope*, a new family of transposable elements and its possible

role in hybrid dysgenesis in *Drosophila virilis*. *Proceedings of the National Academy of Sciences* **94**, 196–201.

Feng, C., Yi, H., Yang, L. & Kang, M. (2020) The genetic basis of hybrid male sterility in sympatric *Primulina* species. *BMC Evolutionary Biology* **20**, 1–12.

Ferris, K.G., Barnett, L.L., Blackman, B.K. & Willis, J.H. (2017) The genetic architecture of local adaptation and reproductive isolation in sympatry within the *Mimulus guttatus* species complex. *Molecular Ecology* **26**, 208–224.

Geng, C. & Macdonald, P. (2007) Identification of genes that influence *gurken* expression. *Fly* **1**, 259–267.

Good, J.M., Dean, M.D. & Nachman, M.W. (2008) A complex genetic basis to X-linked hybrid male sterility between two species of house mice. *Genetics* **179**, 2213–2228.

Gullan, P.J. & Cranston, P.S. (2014) *The Insects: An Outline of Entomology*. John Wiley & Sons.

Haldane, J.B.S. (1922) Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics* **12**, 101–109.

Haley, C.S. & Knott, S.A. (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**, 315.

Heliconius Genome Consortium (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **487**, 94.

Hoffman, J.I., Simpson, F., David, P., Rijks, J.M., Kuiken, T., Thorne, M.A., Lacy, R.C. & Dasmahapatra, K.K. (2014) High-throughput sequencing reveals inbreeding depression in a natural population. *Proceedings of the National Academy of Sciences* **111**, 3775–3780.

Jiggins, C.D. (2017) *The Ecology and Evolution of Heliconius Butterflies*. Oxford University Press, Oxford.

Jiggins, C.D., Linares, M., Naisbit, R.E., Salazar, C., Yang, Z.H. & Mallet, J. (2001) Sex-linked hybrid sterility in a butterfly. *Evolution* **55**, 1631–1638.

Kalirad, A. & Azevedo, R.B. (2017) Spiraling complexity: a test of the snowball effect in a computational model of RNA folding. *Genetics* **206**, 377–388.

Kelleher, E.S., Edelman, N.B. & Barbash, D.A. (2012) *Drosophila* interspecific hybrids phenocopy piRNA-pathway mutants. *PLoS Biology* **10**.

Kidwell, M.G., Kidwell, J.F. & Sved, J.A. (1977) Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility and male recombination. *Genetics* **86**, 813–833.

Kryvokhyzha, D. (2014) *Whole genome resequencing of Heliconius butterflies revolutionizes our view of the level of admixture between species*. Master's thesis, Uppsala Universitët, Uppsala.

Li, H. (2011) A statistical framework for snp calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993.

Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997* .

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079.

Liu, L., Qi, H., Wang, J. & Lin, H. (2011) PAPI, a novel TUDOR-domain protein, complexes with AGO3, ME31B and TRAL in the nuage to silence transposition. *Development* **138**, 1863–1873.

Maheshwari, S. & Barbash, D.A. (2011) The genetics of hybrid incompatibilities. *Annual Review of Genetics* **45**, 331–355.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12.

Martin, S. & Jiggins, C. (2017) Interpreting the genomic landscape of introgression. *Current Opinion in Genetics and Development* **47**, 69–74.

Masly, J.P. & Presgraves, D.C. (2007) High-resolution genome-wide dissection of the two rules of speciation in *Drosophila*. *PLoS Biology* **5**, e243.

Matute, D.R., Butler, I.A., Turissini, D.A. & Coyne, J.A. (2010) A test of the snowball theory for the rate of evolution of hybrid incompatibilities. *Science* **329**, 1518–1521.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297–1303.

Merrill, R.M., Van Schooten, B., Scott, J.A. & Jiggins, C.D. (2011) Pervasive genetic associations between traits causing reproductive isolation in *Heliconius* butterflies. *Proceedings of the Royal Society B: Biological Sciences* **278**, 511–518.

Mihola, O., Trachtulec, Z., Vlcek, C., Schimenti, J.C. & Forejt, J. (2009) A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* **323**, 373–375.

Morgan, T.H. (1910) Sex-limited inheritance in *Drosophila*. *Science N.S.* **32**, 125–133.

Morgan, T.H. (1911) Random segregation versus coupling in Mendelian inheritance. *Science N.S.* **34**, 384.

Muller, H. (1942) Isolating mechanisms, evolution, and temperature. *Biological Symposia*, vol. 6. Temperature and Evolution. Isolating Mechanisms. Genetic Control of Embryonic Development., pp. 71–125.

Naisbit, R.E., Jiggins, C.D., Linares, M., Salazar, C. & Mallet, J. (2002) Hybrid sterility, Haldane's rule and speciation in *Heliconius cydno* and *H. melpomene*. *Genetics* **161**, 1517–1526.

Nosil, P. & Schluter, D. (2011) The genes underlying the process of speciation. *Trends in Ecology & Evolution* **26**, 160–167.

Orr, H.A. (1995) The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics* **139**, 1805–1813.

Orr, H.A. (1997) Haldane's rule. *Annual Review of Ecology and Systematics* **28**, 195–218.

Orr, H.A. & Turelli, M. (2001) The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. *Evolution* **55**, 1085–1094.

Phadnis, N. (2011) Genetic architecture of male sterility and segregation distortion in *Drosophila pseudoobscura* Bogotá–USA hybrids. *Genetics* **189**, 1001–1009.

Pimentel, H., Bray, N.L., Puente, S., Melsted, P. & Pachter, L. (2017) Differential analysis of RNA-seq incorporating quantification uncertainty. *Nature Methods* **14**, 687.

Pinharanda, A., Rousselle, M., Martin, S.H., Hanly, J.J., Davey, J.W., Kumar, S., Galtier, N. & Jiggins, C.D. (2019) Sexually dimorphic gene expression and transcriptome evolution provide mixed evidence for a fast-Z effect in *Heliconius*. *Journal of evolutionary biology* **32**, 194–204.

Pinharanda, A.L.P. (2017) *The genomic basis of species barriers in Heliconius butterflies*. Ph.D. thesis, University of Cambridge, Cambridge, UK.

748 Pinheiro, E.M. & Montell, D.J. (2004) Requirement for *Par-6* and *Bazooka* in *Drosophila*
749 border cell migration. *Development* **131**, 5243–5251.

750 Presgraves, D.C. (2002) Patterns of postzygotic isolation in Lepidoptera. *Evolution* **56**,
751 1168–1183.

752 Presgraves, D.C. (2007) Speciation genetics: epistasis, conflict and the origin of species.
753 *Current Biology* **17**, R125–R127.

754 Presgraves, D.C. (2008) Sex chromosomes and speciation in *Drosophila*. *Trends in Genetics*
755 **24**, 336–343.

756 Presgraves, D.C. (2010) Darwin and the origin of interspecific genetic incompatibilities. *The*
757 *American Naturalist* **176**, S45–S60.

758 Presgraves, D.C. (2018) Evaluating genomic signatures of "the large x-effect" during com-
759 plex speciation. *Molecular ecology* **27**, 3822–3830.

760 Presgraves, D.C. & Orr, H.A. (1998) Haldane's rule in taxa lacking a hemizygous X. *Sci-*
761 *ence* **282**, 952–954.

762 Prowell Pashley, D. (1998) Sex linkage and speciation in Lepidoptera. *Endless Forms.*
763 *Species and Speciation* (eds. D.J. Howard & S.H. Berlocher), pp. 309–319, Oxford Uni-
764 versity Press, New York.

765 Ramsey, J., Bradshaw Jr, H. & Schemske, D.W. (2003) Components of reproductive isola-
766 tion between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolu-*
767 *tion* **57**, 1520–1534.

768 Rastas, P. (2017) Lep-MAP3: robust linkage mapping even for low-coverage whole genome
769 sequencing data. *Bioinformatics* **33**, 3726–3732.

770 Raza, Q., Choi, J.Y., Li, Y., O'Dowd, R.M., Watkins, S.C., Chikina, M., Hong, Y., Clark,
771 N.L. & Kwiatkowski, A.V. (2019) Evolutionary rate covariation analysis of *E-cadherin*

identifies *Raskol* as a regulator of cell adhesion and actin dynamics in *Drosophila*. *PLoS Genetics* **15**, e1007720.

Rosser, N., Queste, L.M., Cama, B., Edelman, N.B., Mann, F., Mori Pezo, R., Morris, J., Segami, C., Velado, P., Schulz, S. *et al.* (2019) Geographic contrasts between pre- and postzygotic barriers are consistent with reinforcement in *Heliconius* butterflies. *Evolution* **73**, 1821–1838.

Sackton, T.B., Corbett-Detig, R.B., Nagaraju, J., Vaishna, L., Arunkumar, K.P. & Hartl, D.L. (2014) Positive selection drives faster-Z evolution in silkmoths. *Evolution* **68**, 2331–2342.

Salazar, C.A., Jiggins, C.D., Arias, C., Tobler, A., Bermingham, E. & Linares, M. (2005) Hybrid incompatibility is consistent with a hybrid origin of *Heliconius heurippa* Hewitson from its close relatives, *Heliconius cydno* Doubleday and *Heliconius melpomene* Linnaeus. *Journal of Evolutionary Biology* **18**, 247–256.

Sánchez, A.P., Pardo-Diaz, C., Enciso-Romero, J., Muñoz, A., Jiggins, C.D., Salazar, C. & Linares, M. (2015) An introgressed wing pattern acts as a mating cue. *Evolution* **69**, 1619–1629.

Schaefer, R.E., Kidwell, M.G. & Fausto-Sterling, A. (1979) Hybrid dysgenesis in *Drosophila melanogaster*: morphological and cytological studies of ovarian dysgenesis. *Genetics* **92**, 1141–1152.

Schartl, M. (2008) Evolution of *Xmrk*: an oncogene, but also a speciation gene? *Bioessays* **30**, 822–832.

Schilthuizen, M., Giesbers, M. & Beukeboom, L. (2011) Haldane's rule in the 21st century. *Heredity* **107**, 95–102.

Schumer, M., Cui, R., Powell, D.L., Dresner, R., Rosenthal, G.G. & Andolfatto, P. (2014)

High-resolution mapping reveals hundreds of genetic incompatibilities in hybridizing fish species. *Elife* **3**, e02535.

Seixas, F.A., Edelman, N.B. & Mallet, J. (2020) Synteny-based genome assembly for 16 species of *Heliconius* butterflies, and an assessment of structural variation across the genus. *bioRxiv* p. 10.1101/2020.10.29.359505v1.

Snee, M.J. & Macdonald, P.M. (2009) *Bicaudal C* and *trailer hitch* have similar roles in *gurken* mRNA localization and cytoskeletal organization. *Developmental Biology* **328**, 434–444.

Sweigart, A.L., Fishman, L. & Willis, J.H. (2006) A simple genetic incompatibility causes hybrid male sterility in *Mimulus*. *Genetics* **172**, 2465–2479.

Tang, S. & Presgraves, D.C. (2009) Evolution of the *Drosophila* nuclear pore complex results in multiple hybrid incompatibilities. *Science* **323**, 779–782.

Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J. & Prins, P. (2015) Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034.

Turelli, M. & Moyle, L.C. (2007) Asymmetric postmating isolation: Darwin's corollary to Haldane's rule. *Genetics* **176**, 1059–1088.

Turelli, M. & Orr, H.A. (1995) The dominance theory of Haldane's Rule. *Genetics* **140**, 389–402.

Wellington, A., Emmons, S., James, B., Calley, J., Grover, M., Tolias, P. & Manseau, L. (1999) Spire contains actin binding domains and is related to ascidian posterior end mark-5. *Development* **126**, 5267–5274.

Wilhelm, J.E., Buszczak, M. & Sayles, S. (2005) Efficient protein trafficking requires *trailer hitch*, a component of a ribonucleoprotein complex localized to the ER in *Drosophila*. *Developmental Cell* **9**, 675–685.

820 Wright, K.M., Lloyd, D., Lowry, D.B., Macnair, M.R. & Willis, J.H. (2013) Indirect evo-
821 lution of hybrid lethality due to linkage with selected locus in *Mimulus guttatus*. *PLoS*
822 *Biology* **11**, e1001497.

823 Wu, C.I. & Davis, A.W. (1993) Evolution of postmating reproductive isolation: the com-
824 posite nature of Haldane's Rule and its genetic bases. *The American Naturalist* **142**,
825 187–212.

826 Yamauchi, H. & Yoshitake, N. (1984) Developmental stages of ovarian follicles of the silk-
827 worm, *Bombyx mori* L. *Journal of Morphology* **179**, 21–31.

Supplementary information

Linkage Map

For the reads aligned to Hpar, the linkage map comprised 159,952 markers (a 29% improvement on Hmel2.5, with a total map length of 1157.27 cM). Marey maps plotting physical distances against genetic distance showed that linkage maps created using Hmel2.5 aligned reads and Hpar aligned reads were broadly similar (Supplementary figures S4 and S5). However, with Hpar some additional large regions with low recombination were apparent (for example, on the distal end of chromosome 15). Plots of estimated recombination rates between all pairs of markers for both linkage maps showed no evidence of misplaced markers (Supplementary figure S6). To validate our genotypic data and linkage map, we performed a QTL analysis on a wing colour pattern trait (presence / absence of yellow apical dots on the forewing). As expected, this showed a significant QTL peak encompassing the gene *cortex* (Supplementary figure S7), which is known to be involved in yellow colour pattern elements in *Heliconius* (Nadeau et al. (2016) The gene *cortex* controls mimicry and crypsis in butterflies and moths. Nature 534, 106–110).

Chromosome	Hpar		Hmel2.5	
	n markers	cM	n markers	cM
1	11627	54.84	9353	53.57
2	4923	55.98	3537	55.98
3	5558	49.52	4342	46.57
4	4153	50.2	3038	49.04
5	5698	60.46	4618	57.29
6	8715	51.22	6853	50.03
7	8383	48.84	6914	48.84
8	5820	59.53	4549	53.66
9	5395	62.06	4193	50.32
10	11760	55.84	8840	55.87
11	6682	51.8	5333	50.44
12	9825	57.4	7818	55.93
13	10806	49.1	8278	48.95
14	4570	54.94	3452	54.96
15	6474	44.27	4919	45.46
16	6037	88.75	5041	69.48
17	10856	57.01	8272	57.04
18	9377	57.04	7427	53.81
19	10134	53.57	7847	53.63
20	8637	37.64	6616	38.68
Z	4527	57.29	3216	57.4
TOTALS	159957	1157.27	124456	1106.95

Table S1: Numbers of markers on each chromosome and chromosome length in centimorgans for linkage maps using reads aligned to Hpar (left) and Hmel2.5 (right) reference genomes.

Specimen	population	tissue	reads ($\times 10^6$)	pct Q$\geq$30	mapped ($\times 10^6$)	pct mapped
NR15-459	backcross	EGG	53.74	94.35	33.70	63
NR15-459	backcross	TIP	55.64	94.64	38.68	70
NR15-459	backcross	VIT	50.93	95.04	35.50	70
NR15-461	backcross	EGG	42.73	94.86	29.07	68
NR15-461	backcross	TIP	47.19	94.90	33.06	70
NR15-461	backcross	VIT	52.75	94.46	34.29	65
NR15-465	backcross	TIP	44.82	95.16	31.15	69
NR15-465	backcross	VIT	39.70	93.90	25.80	65
NR15-474	backcross	EGG	48.49	94.86	34.02	70
NR15-474	backcross	TIP	39.77	95.45	28.54	72
NR15-474	backcross	VIT	49.50	94.64	34.89	70
NR15-475	backcross	TIP	53.07	94.84	37.04	70
NR15-475	backcross	VIT	47.44	94.82	32.28	68
NR15-483	backcross	TIP	46.67	94.80	32.21	69
NR15-483	backcross	VIT	42.97	94.50	29.09	68
NR15-473	butleri	EGG	46.92	94.64	32.00	68
NR15-473	butleri	TIP	48.52	94.58	32.32	67
NR15-473	butleri	VIT	50.60	95.11	34.67	69
NR15-488	butleri	EGG	57.83	94.57	39.05	68
NR15-488	butleri	TIP	55.19	94.50	37.71	68
NR15-488	butleri	VIT	50.63	94.51	35.38	70
NE19-08	cydno	EGG	50.02	93.24	37.20	74
NE19-08	cydno	TIP	37.13	94.73	28.22	76
NE19-08	cydno	VIT	56.74	95.15	42.88	76
NE19-10	cydno	EGG	47.44	94.55	33.28	70
NE19-10	cydno	TIP	51.48	94.89	39.10	76
NE19-10	cydno	VIT	44.53	94.76	32.45	73
NE19-04	melpomene	TIP	47.09	95.14	31.57	67
NE19-04	melpomene	VIT	65.44	95.42	42.63	65
NE19-06	melpomene	EGG	48.09	95.03	32.28	67
NE19-06	melpomene	TIP	49.99	94.45	35.03	70
NE19-06	melpomene	VIT	45.57	94.91	26.82	59

Table S2: RNA sequencing and read mapping statistics to the Hmel2.5 reference.

LOD_f	chr	QTL1 marker	cM	limits (cM)	limits (physical)	chr	QTL2 marker	cM	limits (cM)	limits (physical)	μ_1	$\beta_1 q_1$	$\beta_2 q_2$	$\beta_3 q_1 q_2$	R^2
4.21**	Z	hpar210001: 7007245	29.15	10.49-35	4649965-7821836						1.37± 0.09***	0.88± 0.19***			0.2
6.93*	4	hpar040001: 13734678	50.2	29.21-50.2	10014197-13750774	Z	hpar210001: 7007245	29.15	12.85-36.13	4663108-7827537	1.32± 0.09***	-0.6± 0.18**	0.81± 0.18***	-0.52± 0.36	0.31
6.61+	12	hpar120001: 7636661	27.16	18.94-57.4	5248779-17866482	Z	hpar210001: 5687707	18.69	10-33.81	4649939-7712239	1.39± 0.09***	0.44± 0.18*	0.88± 0.18***	0.98± 0.36**	0.3
7.87**	15	hpar150002: 10183589	37.29	8.17-43.11	5304875-13942281	Z	hpar210001: 5687707	18.69	12.85-29.15	4663108-7059083	1.34± 0.09***	0.29± 0.18	0.7± 0.18***	1.47± 0.36***	0.34
14.52***	Z	hpar210001: 3084795	4.65	2.33-5.81	2035613-4420376	Z	hpar210001: 10667602	54.96	53.8-56.13	8793636-13440534	1.4± 0.07***	0.6± 0.15***	-0.03± 0.15	-2.53± 0.3***	0.54
7.79*	8	c8.loc13	13	11-13	485709-1483261	20	hpar200003: 6960802	13.02	6-13.02	3402889-6961037	1.61± 0.08***	0.24± 0.16	-0.22± 0.16	2.68± 0.32***	0.79

Table S3: Summary of significant single locus (H_1) and two locus (H_f) QTL modelsusing reads aligned to Hpar. The first column gives the LOD score of the full model (H_f), with the significance estimated by permutation (+P<0.1, *P<0.05, **P<0.01, ***P<0.001). The next columns are the chromosome and QTL marker (scaffold and median physical position within the peak). The centimorgan limits are the Bayesian credible intervals, and the physical limits are the nearest typed flanking markers of that interval (all physical limits were on the same scaffold as the QTL peak, except for the chromosome 4 interaction † with Z, which was on scaffold Hmel204003 of Hmel2.5). The final five columns give the parameter estimates and R^2 of the model. $\beta_1 q_1$ and $\beta_2 q_2$ are the estimated additive effects for the QTLs, i.e. the difference between the phenotypic averages for the alternative genotypes, and $\beta_3 q_1 q_2$ is the coefficient for the interaction between the 2 loci. Model coefficients comprise the estimated value, the standard error, and the significance (thresholds as above). The significant interaction between chromosome 8 and 20 was detected using individuals holding a Z_{SB} chromosome only.

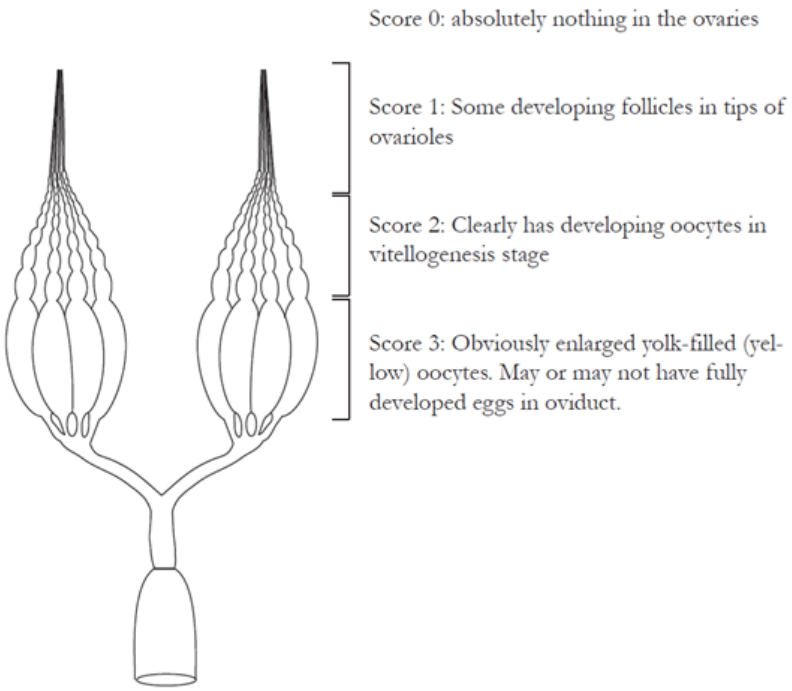

Figure S1: Ovary score guide. This scheme was used to characterize hybrid ovarioles in terms of gross developmental phenotype (Gullan & Cranston, 2014).

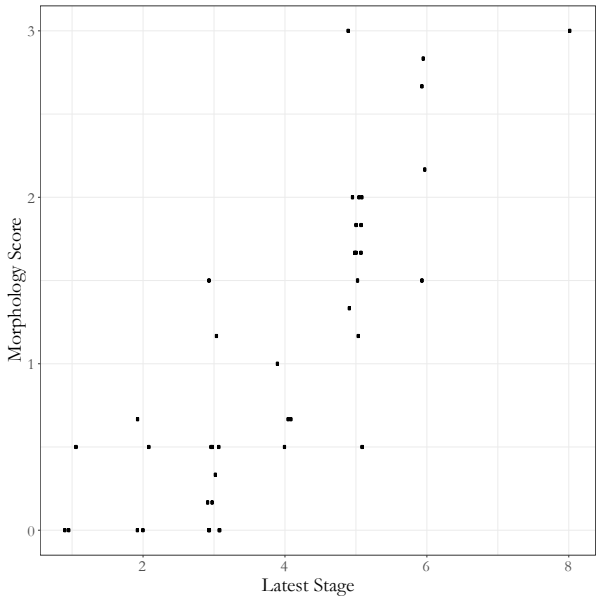

Figure S2: Comparison of scoring schemes. Among individuals for whom we were able to score both latest developmental stage and morphoscore, the values were highly correlated. Note that the "Latest Stage" measurement is the latest stage observed in the sampled ovariole. Fully-developed eggs (stage 12, (Yamauchi & Yoshitake, 1984)) may have been present in the oviduct but were not observed in ovarioles themselves.

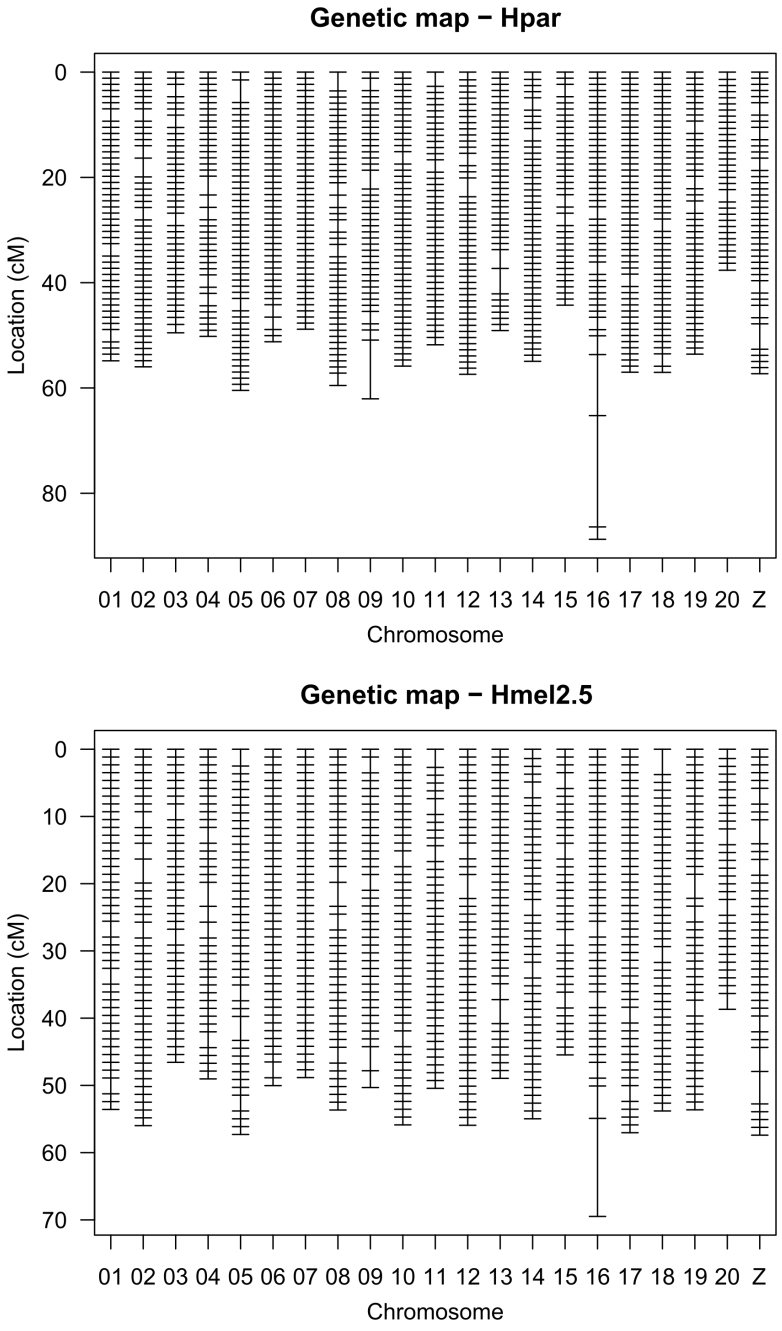

Figure S3: Marker locations for linkage maps using reads aligned to Hpar and Hmel2.5.

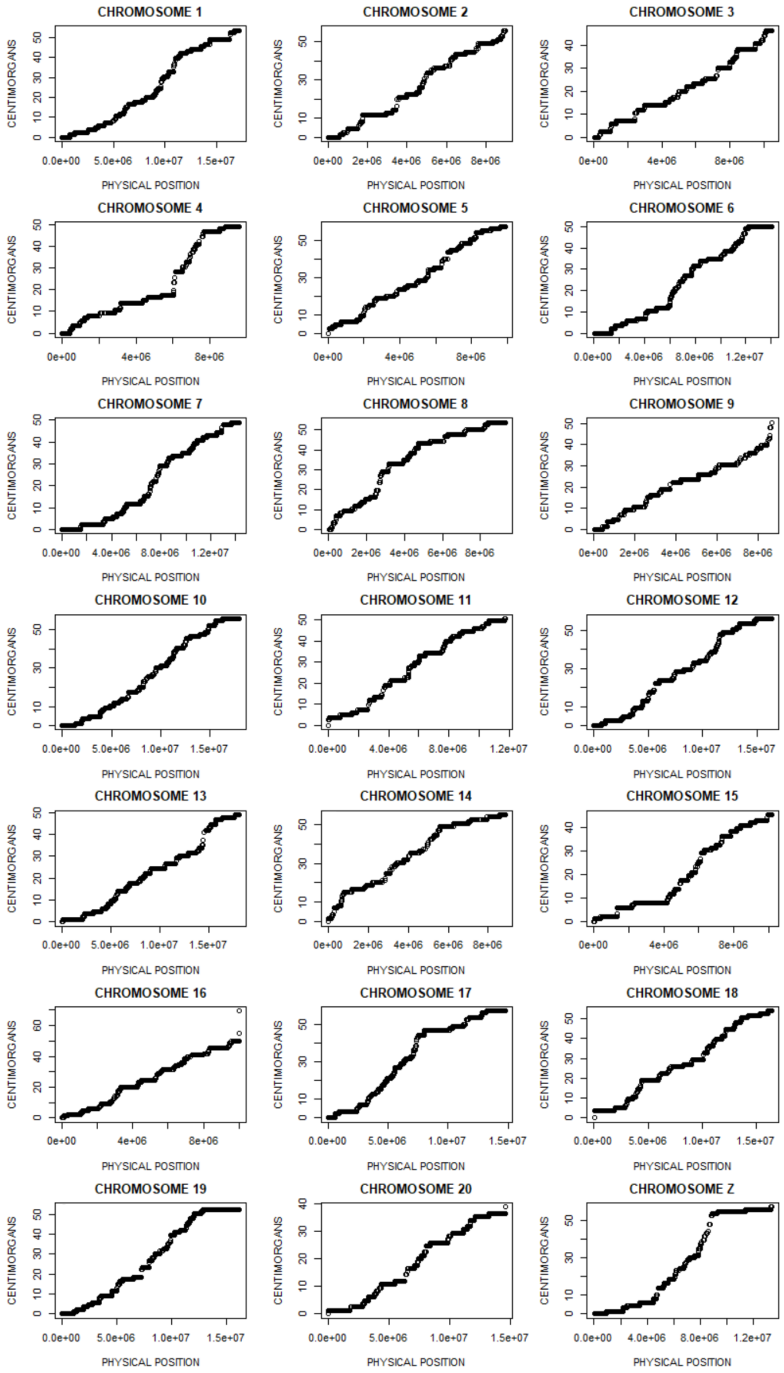

Figure S4: Marey maps using reads aligned to Hmel2.5. Markers conspicuously out of place (e.g. on chromosome 4) have been joined from a second scaffold.

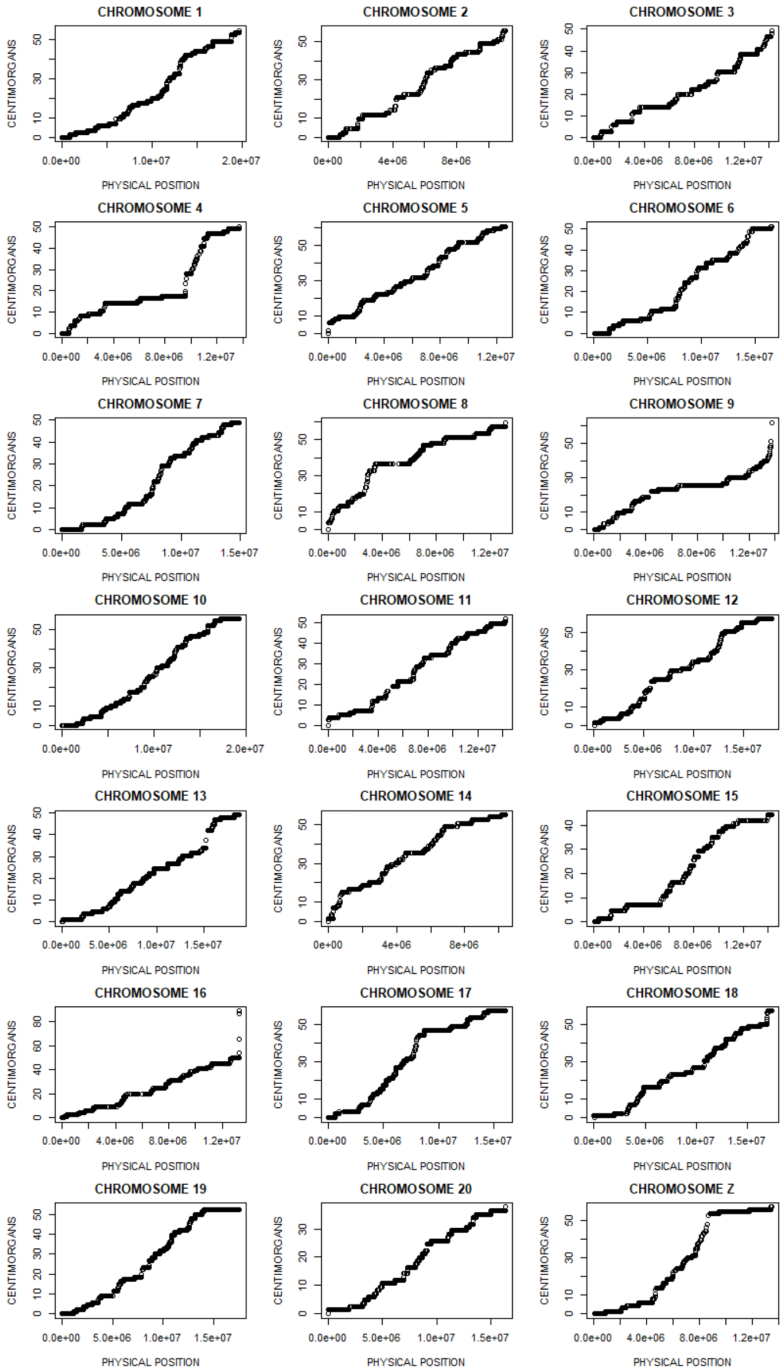

Figure S5: Marey maps using reads aligned to Hpar. Markers conspicuously out of place (e.g. on chromosome 19) have been joined from a second scaffold.

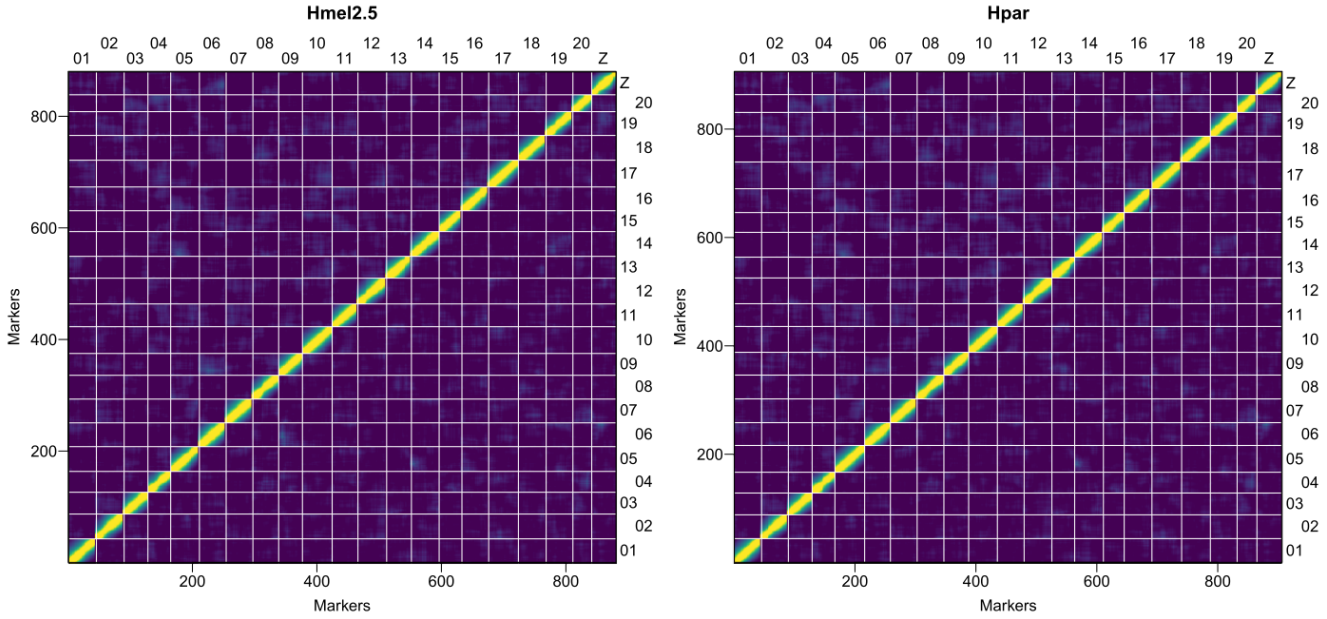

Figure S6: Pairwise recombination fractions (upper left) and LOD scores for the test of $r = \frac{1}{2}$ (lower right), using the Hmel2.5 and Hpar linkage maps. Yellow indicates linkage, blue indicates no linkage.

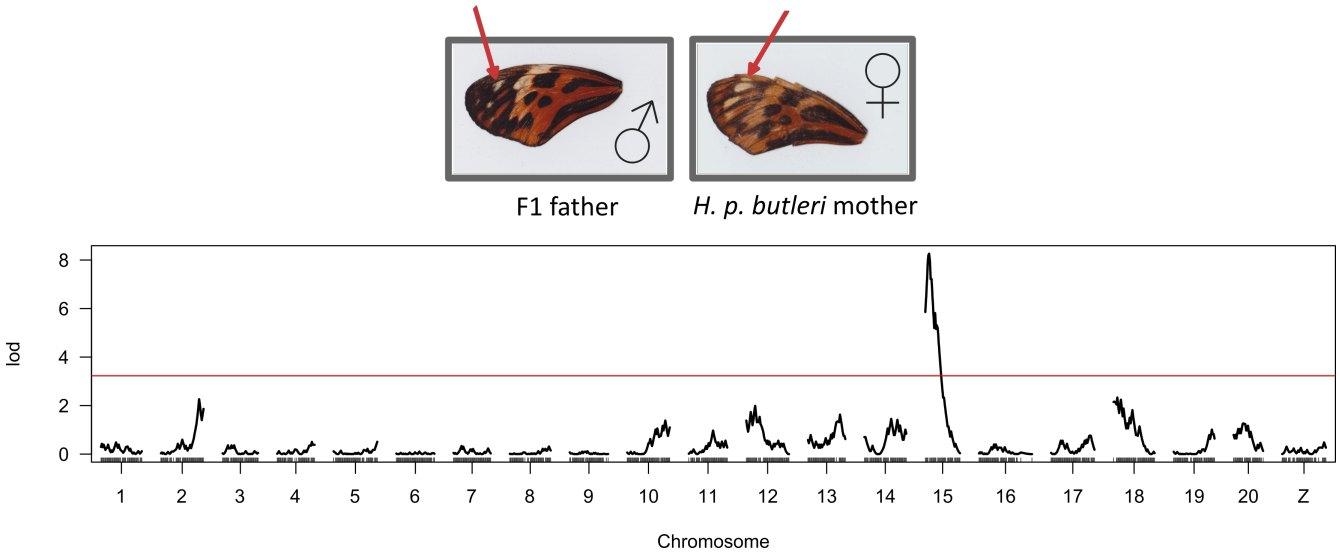

Figure S7: Colour pattern QTL. Backcross individuals produced by crossing a male F1 with a *H. p. butleri* female where scored by eye for large/reduced apical dots on the forewing (indicated by the red arrows in the figure). This trait was then analysed using Haley-Knott regression, and showed a significant QTL on chromosome 15, in the region of the gene *cortex*.

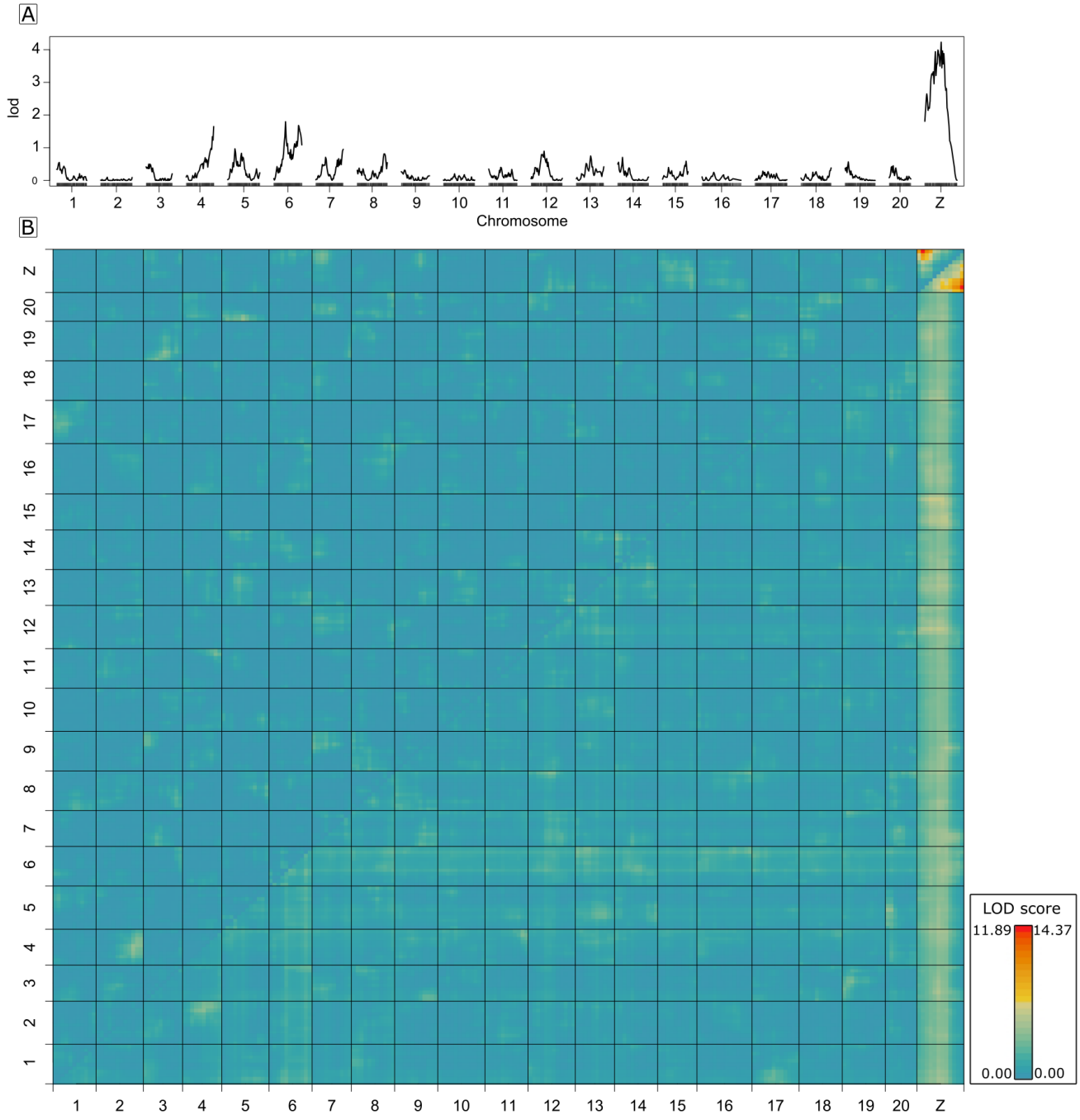

Figure S8: QTL genome scans controlling for kinship, using reads aligned to Hmel2.5. A. LOD scores for morphoscore as a function of genotype. **B.** Heat map for LOD_f scores (upper right triangle) and LOD_{int} scores (lower left triangle) between pairwise combinations of markers across the genome. Blues indicate low scores, reds indicate high scores. Max LOD $_{int}$ is left of the colour ramp, and max LOD_f is to the right of the colour ramp.