Neighboring genes for DNA-binding proteins rescue male sterility in *Drosophila* hybrids

Marjorie A. Liénard[a,b,1], Luciana O. Araripe[a,2], and Daniel L. Hartl[a,1]

[a]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138; and [b]Department of Biology, Lund University, 22362 Lund, Sweden

Contributed by Daniel L. Hartl, May 25, 2016 (sent for review February 22, 2016; reviewed by Norman A. Johnson, Mohamed A. F. Noor, and Giuseppe Saccone)

Crosses between closely related animal species often result in male hybrids that are sterile, and the molecular and functional basis of genetic factors for hybrid male sterility is of great interest. Here, we report a molecular and functional analysis of *HMS1*, a region of 9.2 kb in chromosome 3 of *Drosophila mauritiana*, which results in virtually complete hybrid male sterility when homozygous in the genetic background of sibling species *Drosophila simulans*. The *HMS1* region contains two strong candidate genes for the genetic incompatibility, *agt* and *Taf1*. Both encode unrelated DNA-binding proteins, *agt* for an alkyl-cysteine-*S*-alkyltransferase and *Taf1* for a subunit of transcription factor TFIID that serves as a multifunctional transcriptional regulator. The contribution of each gene to hybrid male sterility was assessed by means of germ-line transformation, with constructs containing complete *agt* and *Taf1* genomic sequences as well as various chimeric constructs. Both *agt* and *Taf1* contribute about equally to *HMS1* hybrid male sterility. Transgenes containing either locus rescue sterility in about one-half of the males, and among fertile males the number of offspring is in the normal range. This finding suggests compensatory proliferation of the rescued, nondysfunctional germ cells. Results with chimeric transgenes imply that the hybrid incompatibilities result from interactions among nucleotide differences residing along both *agt* and *Taf1*. Our results challenge a number of preliminary generalizations about the molecular and functional basis of hybrid male sterility, and strongly reinforce the role of DNA-binding proteins as a class of genes contributing to the maintenance of postzygotic reproductive isolation.

postzygotic reproductive isolation | hybrid male sterility | gene conflict | transcription factor

Sterility, lethality, or other abnormalities observed among the offspring of crosses between species are known as hybrid incompatibilities. Usually ascribed to the dysfunction of parental coadapted gene complexes, hybrid incompatibilities have intrigued geneticists since almost the beginning of modern genetics (1). Hybrid incompatibilities are important evolutionarily because they act as reproductive barriers that can both promote speciation in sympatric populations and maintain the integrity of the species following allopatric divergence. They are also paradoxical because they cannot usually arise as a direct result of natural selection in the diverging parental lineages; rather, they are incidentally acquired in the ordinary course of evolutionary divergence and only manifest in the unique hybrid genomic background (2–4).

In the century since hybrid incompatibilities were first called out as an intriguing and important issue in genetics and evolutionary biology, the identification of the causal genes and their molecular mechanisms has been hampered by hybrid incompatibility itself, because any sterile or lethal hybrid individual constitutes a virtual dead end for genetic analysis. Additionally, even when only one sex is sterile or lethal, the number of genes contributing to hybrid incompatibility is usually large and their interactions complex (5–8). There is also the inherent difficulty of distinguishing when during the course of speciation incompatibilities may have arisen, especially because postspeciation incompatibilities may accumulate exponentially in time (9).

Recent reviews detail virtually all analyzed cases of hybrid incompatibility in a variety of organisms (flowering plants, yeast, copepods, fruit flies, fish, and mouse), which altogether still contribute fewer than two dozen genetic factors (3, 10). Considerable discussion has focused on whether particular classes of genes are overrepresented in this limited number of examples, with a high proportion of genes involved in internal genomic conflicts, especially those involving genes encoding proteins that bind to DNA or chromatin (3, 10). Among the genetic factors identified to date in *Drosophila*, the latter category includes the homeobox-containing gene *OdsH* (11–13), two heterochromatin proteins *Lhr* (14) and *Hmr* (15), and the coding gene *Ovd* (16).

Most empirical evidence about hybrid incompatibility comes from *Drosophila* owing to the special advantages and resources available for genetic analysis in this organism (5, 17). Among the well-studied drosophilid species are *Drosophila simulans* and its island-endemic sibling species *Drosophila mauritiana*, which diverged ∼250,000 y ago (18). Among the hybrids, males are sterile but females are fertile, in accordance with Haldane's rule that the heterogametic sex manifests hybrid incompatibilities sooner than the homogametic sex (9, 19, 20), and makes genetic studies by backcrossing feasible. Short genomic regions introgressed from *D. mauritiana* in an otherwise isogenic *D. simulans* genetic background identify numerous regions associated with hybrid incompatibilities (21–24). Many of these

Significance

Hybrid sterility is a frequent outcome of crosses between closely related plant and animal species because of incompatibilities that have evolved in the parental genomes. Here, we show that a small region associated with hybrid male sterility between two closely related species of *Drosophila* contains two genes, both encoding DNA-binding proteins, each of which contributes to the hybrid male sterility. These results emphasize that hybrid incompatibility between well-established species is the result of numerous genetic factors, each contributing quantitatively to the incompatibility. Among these factors, DNA-binding proteins are disproportionately represented. Each incompatibility is complex, resulting from interactions between nucleotide sites in different regions of the gene, and is likely to have evolved long after the initial establishment of reproductive isolation.

Author contributions: M.A.L., L.O.A., and D.L.H. designed research; M.A.L. and L.O.A. performed research; M.A.L. and D.L.H. analyzed data; and M.A.L. and D.L.H. wrote the paper.

Reviewers: N.A.J., University of Massachusetts; M.A.F.N., Duke University; and G.S., Università degli Studi di Napoli "Federico II," Naples.

The authors declare no conflict of interest.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. KX225407–KX225410).

[1]To whom correspondence may be addressed. Email: mlienard@fas.harvard.edu or dhartl@oeb.harvard.edu.

[2]Present address: Laboratório de Biologia Molecular de Insetos, Fundação Oswaldo Cruz, Instituto Oswaldo Cruz, CEP 21040-360, Rio de Janeiro, RJ, Brazil.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1608337113/-/DCSupplemental.

regions show complex nonadditive epistatic interactions modulating male fertility (25). Although genomic conflicts over sex chromosome transmission contribute significantly to the evolution of reproductive isolation (23, 24, 26–28), the autosomes alone contain ∼40 genetic regions that contribute significantly to hybrid male sterility (29).

Most of the hybrid incompatibility regions are relatively large and may contain more than one contributing genetic factor. Here, we report the genetic analysis of a 9.2-kb region in chromosome 3 of *D. mauritiana*, denoted hybrid male sterility one (*HMS1*), which is associated with hybrid male sterility when introgressed into the genome of *D. simulans* (30). By means of germ-line transformation with a number of constructs of sequences within or near *HMS1*, we show that *HMS1* contains two neighboring genes for unrelated DNA-binding proteins, each of which contributes quantitatively to hybrid male sterility. One gene, *Taf1*, encodes a component of transcription factor TFIID with a testes-enriched alternatively spliced isoform. The other gene, *agt*, encodes a DNA alkyltransferase. In both cases, the incompatibility is not due to a single-nucleotide replacement in the gene but to interactions among multiple sites along the gene. These results emphasize the quantitative, multifactorial basis of hybrid male sterility and strongly support DNA-binding proteins as a prominent class of genes contributing to genetic conflicts resulting in reproductive isolation.

Results

Identification of *HMS1*. *HMS1* refers to a region in the right arm of chromosome 3 of *Drosophila mauritiana* that is associated with hybrid male sterility between *D. mauritiana* and its sibling species *D. simulans*. When the *HMS1* region from *D. mauritiana* (*HMS1[mau]*) is homozygous in an otherwise *D. simulans* genetic background, the males show a dramatic reduction in fertility. *HMS1* is one of about 20 *HMS* factors in chromosome 3 originally mapped as quantitative-trait loci affecting male fertility (29). Most *HMS* factors have relatively small effects so that generally at least two or three different factors are required to produce complete sterility (29). In contrast, males homozygous for *HMS1* alone are usually quasisterile (29), although the effect of *HMS1* is strongly dependent on genetic background (30).

Originally localized to a region of 1.26 Mb between the genes *Regena* (*Rga*) and *Antennapedia* (*Antp*) in chromosome 3R within the overlap between *P*-element introgressions *P*32.110 and *P*45.6 (Fig. 1*A*) (29), the *HMS1* interval was significantly narrowed by four successive rounds of recombination to a region of 9.2 kb (Fig. 1*B*) corresponding to band 84A1 in the polytene salivary-gland chromosomes (30). Within this region are four genes, two of which exhibit significant coding and noncoding differences between the species, making them candidate genes for hybrid male sterility (30). The candidate genes are *Taf1* (CG17603) and *agt* (CG1303) (Fig. 1*B*), both of which encode DNA-binding proteins The genetic mapping and initial analysis left it unclear, however, which one of these genes, or possibly both, contributed to the male-sterility phenotype.

HMS1 male sterility is manifested in a *D. simulans* genetic background (strain SimB) into which the *HMS1* region of *D. mauritiana* (strain Mau12) has been introgressed. For simplicity, we will refer to the genotype of the SimB introgression genotype as *HMS1[mau]/HMS1[mau]* (see refs. 29 and 30 for details of the SimB and Mau12 strains and the introgression methods). Homozygous *HMS1[mau]/HMS1[mau]* males are almost completely sterile: 97.9% of tested males are sterile (n = 109), and among those with progeny the average number is 1.5 ± 0.71 (Fig. 1*C* and Table S1). The sterility is restored in heterozygous *HMS1[mau]/HMS1[sim]* males, among which all are fertile (n = 183) with an average offspring number of 82.6 ± 52.1 (Fig. 1*C* and Table S1).

To ascertain the genetic basis of *HMS1* sterility, we carried out germ-line transformation using various constructs inserted into the *piggyBac* vector (31). The strategy was to create *HMS1[mau]/HMS1[mau]* male-sterile genotypes carrying an extra, transgenic, *piggyBac* copy of all or part of the *HMS1* region from SimB (*HMS1[sim]*) to look for evidence of complementation, which would be manifested as partial or complete recovery of male fertility. One problem with this strategy is that *HMS1[mau]/HMS1[mau]* is male sterile and so cannot be transformed directly. Instead, we transformed SimB with *piggyBac* constructs containing all or part of *HMS1[sim]* (symbolized *pB::HMS1[sim]*) (Fig. S1) to generate transgenic *HMS1[sim]/HMS1[sim]; pB::HMS1[sim]* animals. In each case, the location

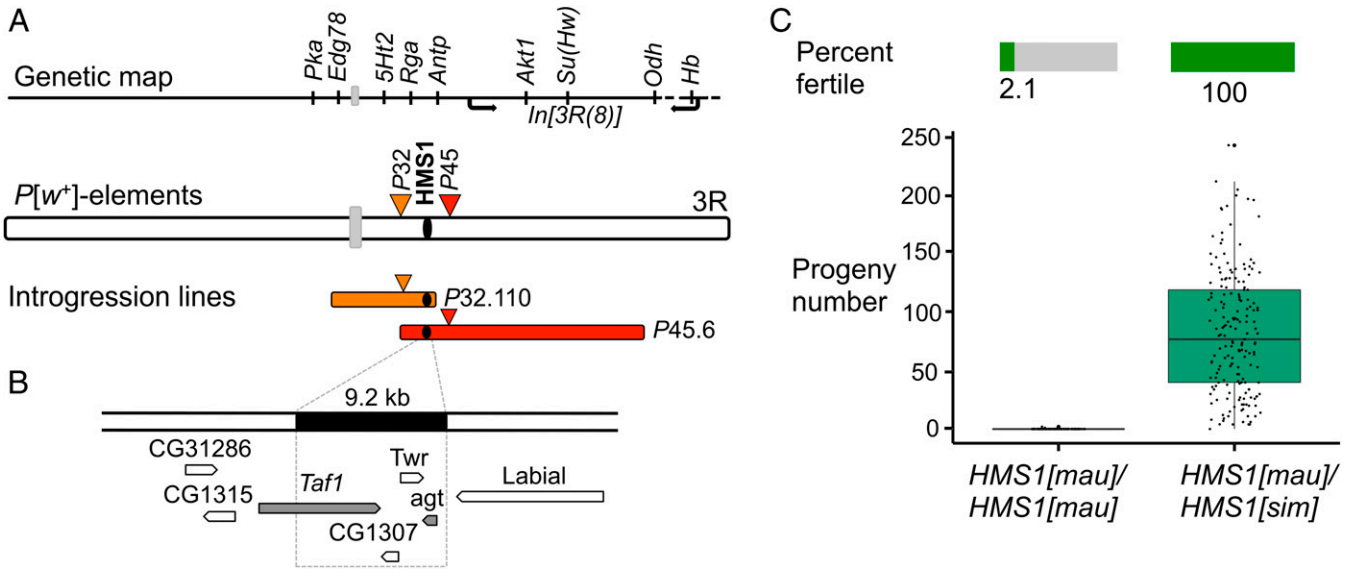

Fig. 1. *HMS1* structure and phenotype. (*A*) Genetic map of *HMS1* region showing location of *P*32 and *P*45 *P*-element insertions as well as the extent of *D. mauritiana* Mau12 genome present in the *P*32.110 and *P*45.6 introgressions obtained by repeated backcrossing. *P*-elements *P*45 and *P*32 localize to polytene bands 83A and 84E8, respectively (76). (*B*) Expanded map of *HMS1* region showing the genes in and near the region. (*C*) The phenotype of the *HMS1[mau]/HMS1[mau]* homozygous males showing that the great majority of males are completely sterile, whereas only a few are quasisterile; the phenotype of the *HMS1[mau]/HMS1[sim]* heterozygous males is complete fertility rescue with high numbers of progeny among most of the fertile males.

of the transgene in the genome was determined by the inverse PCR (32). The *pB-HMS1[sim]* transgene was then transmitted among genotypes by a stepwise crossing design to create the required *HMS1[mau]/HMS1[mau]; pB::HMS1[sim]* genotypes to assay for fertility (Fig. S2). The genetic markers for the *HMS1[mau]* introgression and also the *pB::HMS1[sim]* constructs all affect eye color; however, the copy number effects on phenotype usually allow the genotypes to be distinguished (Fig. S3). All assayed males, both fertile and sterile, were unambiguously genotyped postmating by means of PCR, and in the case of fertile males the reliability of the molecular genotyping was further verified by observing segregation of the eye color phenotypes among the progeny.

Both *agt* and *Taf1* Rescue *HMS1* Sterility, But Neither Wholly. Among the *piggyBac* constructs created to dissect the genetic basis of *HMS1* were *agt[sim]* and *Taf1[sim]* (Fig. S4). The former includes the entirety of the *agt* gene, and the latter, the entirety of *Taf1*, both derived from the SimB strain of *D. simulans*. We tested three independent insertions of *pB::agt[sim]* and two of *pB::Taf1[sim]* in a genetic background of *HMS1[mau]/HMS1[mau]* to detect possible position effects of the insertions. Although introgressed *HMS1[mau]/HMS1[mau]* males are virtually all sterile (Fig. 1*C*), all three *pB::agt[sim]* and both *pB::Taf1[sim]* transgenes restore fertility to some extent (Fig. 2). Among transgenic males that are fertile, the level of fertility is high and not significantly different overall among all *pB::agt[sim]* and *pB::Taf1[sim]* transformants (*t* test on progeny number: $t = 0.369$, df = 152, $P = 0.713$). The fertility rescue is only partial, however, as indicated by the relatively high proportions of sterile males among all transgenic lines (Fig. 2). The percentage of sterile males averages 42 ± 4% among *pB::agt[sim]* transformants and 54 ± 5% among *pB::Taf1[sim]* transformants (Fig. 2). For *agt[sim]*, the number of progeny does not differ significantly for the three genomic transgene insertions (one-way ANOVA: $F = 2.65$, df = 2, $P = 0.075$) and the fractions fertile are also homogeneous [$P = 0.172$ ($X^2 = 3.523$, df = 2, $n = 185$)]. For *Taf1[sim]*, the number of progeny differs somewhat from one insertion site to the next (*t* test on progeny number: $t = -3.030$, df = 43, $P = 0.004$) as indicated by the asterisk in Fig. 2, but importantly the fraction fertile remains homogenous [χ^2 test on number fertile: $P = 0.063$ ($X^2 = 3.453$, df = 1, $n = 94$)], which attests to the limited impact of positional effects at the sites of insertion studied. From the standpoint of rescuing complete sterility, *agt[sim]* is thus not significantly different from *Taf1[sim]* [$P = 0.079$ ($X^2 = 3.076$, df = 1, $n = 279$)]. In both cases, the partial rescue is specific to the *agt[sim]* or *Taf1[sim]* transgene, because control males carrying an empty *piggyBac* vector with no insert fail completely to rescue *HMS1* sterility (Table S2).

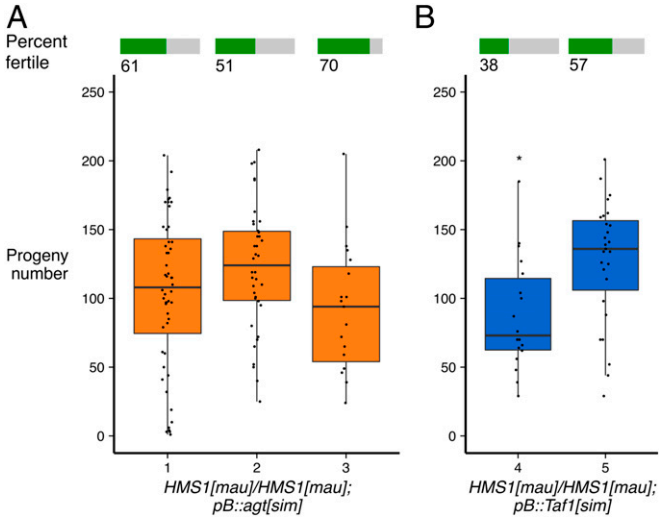

Fig. 2. Phenotypes of *HMS1/HMS1* sterile males with *agt[sim]* or *Taf1[sim]* transgenes. (*A*) Males carrying a transgene of *piggyBac* containing the complete *agt[sim]* gene show partial rescue of male sterility with nearly normal numbers of progeny compared with males that are fertile in Fig. 1*C*. (*B*) Males carrying a transgene of *piggyBac* containing the complete *Taf1[sim]* gene show a similar phenotype, with partial rescue of sterility and nearly normal number of progeny yielded by fertile males except for insertion 4 that has a reduced fecundity (*t* test on progeny number: $t = -3.030$, df = 43, $P = 0.004$), as indicated by the asterisk (*B*). All of these *piggyBac* insertions are in introns or intergenic regions.

To further investigate the mechanism by which fertility is restored in the transgenic lines, we assayed whether decreased expression levels would account for the sterility at either candidate locus. We first quantified mRNA levels in male reproductive tissue of both *agt* and *Taf1* loci in fertile (*HMS1[mau]/HMS1[sim]*) and sterile (*HMS1[mau]/HMS1[mau]*) introgression males, which are isogenic except for a small region including *HMS1* (*SI Materials and Methods* and Fig. S5*A*). These males yield baseline levels of *agt* and *Taf1* expression for males bearing either one or two *sim* alleles of each gene. In each of two genetic backgrounds, SimB and w501, our quantitative PCR results indicate no significant difference in mRNA level of either *Taf1* or *agt* between fertile (*HMS1[mau]/HMS1[sim]*) and sterile (*HMS1[mau]/HMS1[mau]*) males (Fig. S5*B* and Table S3). This finding supports the inference that the hybrid male sterility is not mediated by differences in gene expression of either *agt* or *Taf1*.

Impact of a Candidate Amino Acid Replacement in *agt*. A previous analysis had suggested that an amino acid replacement in *agt* might contribute to hybrid male sterility (30). In *D. simulans*, amino acid position 121 is occupied by aspartic acid, whereas in *D. mauritiana* position 121 is an asparagine. This D121N difference is the only fixed amino acid difference then known between the species (30), hence making it a strong candidate for an incompatibility. To test this hypothesis, we created an *agt[simD121N]* sequence (Fig. S4) that is identical to *agt[sim]* except for a G-to-A nucleotide change in codon 121 resulting in D121N. A transgenic *agt[simD121N] piggyBac* transformant was obtained, crossed into an *agt[mau]/agt[mau]* genetic background (Fig. S2), and tested for male fertility. The results are shown in Fig. 3. If D121N were solely responsible for the *agt[mau]* incompatibility, then transgenic *pB::agt[simD121N]* males should be sterile. However, 27% of the males are fertile (Fig. 3), and their average number of offspring does not differ significantly from that observed for *pB::agt[sim]* ($t = 0.184$, df =121, $P = 0.89$) (Fig. 2). Nevertheless the fertility rescue of *pB::agt[simD121N]* is significantly less than that of *pB::agt[sim]*, with 73 ± 7% sterile in the former and 42 ± 4% in the latter [$P = 0.014$ ($X^2 = 6.09$, df = 1, $n = 225$)]. These results apply to a single genomic insertion; however, the observation that position effects for degree of rescue are not observed for distinct genomic insertion sites for agt[*sim*] (Fig. 2), together with the absence of transcriptional difference in *agt* between *HMS1[mau]/HMS1[sim]* and *HMS1[mau]/HMS1[mau]* males, suggest that the reduced fraction of fertile males in *pB::agt[simD121N]* may be heavily attributable to the coding mutation.

Follow-Up Analysis of *agt[mau]* Incompatibility. The findings with *agt[simD121N]* suggest that the *agt[mau]* incompatibility is due only in part to D121N and in part to other nucleotide differences in *agt* between the SimB and Mau12 strains. A total of 16 nucleotide differences are positioned across the gene (Table 1), or 132 differences if those in the intergenic regions are also considered; hence there is considerable opportunity for interactions among sites. To test for such interactions, we created an *agt[simP_mauCDS]* construct that includes the promoter region from *D. simulans* and the ORF from *D. mauritiana* (*agt* contains no introns), and the reciprocal *agt[mauP_simCDS]* construct (Fig. S4). Two independent transgenic *piggyBac* lines were obtained and tested from each construct (Table S1), yielding the grouped results shown in Fig. 3.

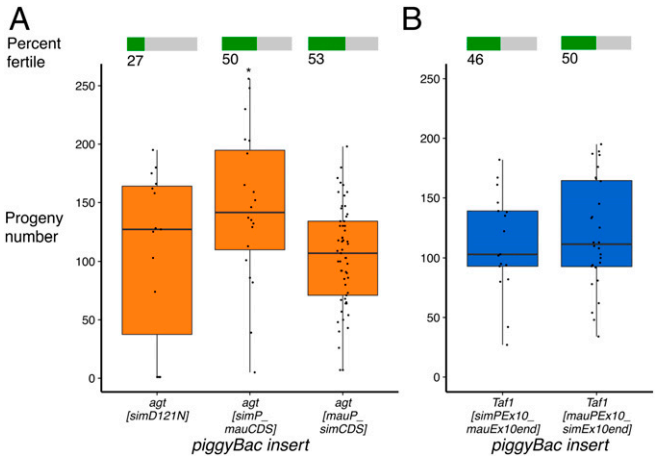

Fig. 3. (*A* and *B*) Phenotypes of *HMS1/HMS1* sterile males with *agt* or *Taf1* chimeric genes. The *agt[simD121N]* transgene contains the *agt[sim]* sequence from SimB with a D121N amino acid replacement, which is a fixed difference between *D. simulans* and *D. mauritiana*; this transgene rescues sterility significantly less than does *agt[sim]*; however, progeny production of the fertile males is in the normal range. The *agt[simP_mauCDS]* and *agt[mauP_simCDS]* are reciprocal chimeras carrying the *agt* promoter from one species and the *agt* coding sequence from the other. Both rescue fertility to the same extent as *agt[sim]*, whereas *agt[simP_mauCDS]* exhibits higher fecundity than *agt[sim]* (one-way ANOVA: $F = 2.998$, df = 5, $P = 0.012$, Tukey post hoc test), as indicated by the asterisk (*A*). The *Taf1[simPEx10_mauEx10end]* and *Taf1[mauPEx10_simEx10end]* are reciprocal chimeric constructs containing *Taf1* of one species from the promoter through an AvrII site in exon 10 fused to *Taf1* of the other species from the AvrII site in exon 10 through to the end of the gene. Both chimeras complement sterility and productivity to the same extent as does *Taf1[sim]*. All of these *piggyBac* insertions are in introns or intergenic regions.

Both types of construct show levels of complete fertility rescue comparable to those of *agt[sim]* [$P = 0.257$ ($X^2 = 5.309$, df = 4, $n = 335$)] and, among the fertile males, only *agt[simP_mauCDS]* sire more progeny (one-way ANOVA: $F = 2.998$, df = 5, $P = 0.012$, Tukey post hoc test). These results support a model in which the incompatibility in *agt[mau]* results from epistatic interactions of nucleotide sites within or near the *agt* gene, interactions that the *agt[simP_mauCDS]* and *agt[mauP_simCDS]* constructs disrupt.

Complex Nature of *Taf1[mau]* Incompatibility. Differences between SimB and Mau12 in the *Taf1* region give even more scope for complex interactions than in *agt*. *Taf1* is a much larger gene than *agt* (9,412 vs. 758 nt). Across the *Taf1* gene region, there are 127 nucleotide differences between the SimB and Mau12 strains (Table 1), and 164 differences if the 3′-untranslated and intergenic regions are also taken into account. To test for interactions among these sites, we created two chimeric constructs designated *Taf1[simPEx10_mauEx10end]* and *Taf1[mauPEx10_simEx10end]* (Fig. S4). The first construct joins a DNA fragment containing the SimB *Taf1* promoter through an AvrII restriction site in exon 10 with another DNA fragment containing Mau12 *Taf1* coding sequence from the AvrII restriction site through to the 3′ end of the gene. The AvrII site was chosen as the joining site because it is near where recombination analysis had suggested the approximate 5′ upstream boundary of *HMS1* was located (30). The second construct is the reciprocal, joining the 5′ end of Mau12 *Taf1* with the 3′ end of simB *Taf1*, again with the joining site at the AvrII site in exon 10.

Results with the reciprocal *Taf1* transformants are shown in Fig. 3. The two constructs do not differ significantly in either the proportion of sterile males or the fecundity of fertile males, nor do they differ significantly from the *Taf1[sim]* insertions in Fig. 2 (χ^2 on number fertile: $P = 0.913$, $X^2 = 0.182$, df = 2, $n = 189$) or fecundity from *Taf1[sim]* insertion 5 (one-way ANOVA on fecundity: $F = 0.612$, df = 2, $P = 0.545$). Considering the limited contribution of position effects at the sites of insertion studied in terms of degree of rescue (although not globally for fecundity; i.e., lower progeny numbers in *Taf1[sim]* insertion 4), these results suggest that the *Taf1[mau]* incompatibility is due to interactions between sites in or near the gene. In the case of *Taf1*, between the strains SimB and Mau12, there are 5 nonsynonymous differences and 44 synonymous differences in the coding sequence preceding the AvrII site, and 9 nonsynonymous differences and 23 synonymous differences following the AvrII site; hence, opportunities for interactions among sites are plentiful.

Tests for Positive Selection. To look for signals of positive selection, we sequenced the ORF of 32 alleles of *agt* and exons 10–14 of 15 alleles of *Taf1* from genomic DNA extracted from individual flies from 17 diverse *D. simulans* populations from South, Central, and North Africa, Europe, Japan, Australia, and North America as well as 16 independent acquisitions of *D. mauritiana* (an island endemic), from Mauritius (Table S4). Statistical tests for selection included comparing the mean number of pairwise differences with the number of segregating sites [Tajima's *D* (33)], as well as comparing old and recent mutations according to where they occur in the gene genealogies using mutation rates either estimated from nucleotide diversity [Fu and Li's *D* (34)] or else estimated from nucleotide polymorphism [Fu and Li's *F* (34)]. We also analyzed nonsynonymous versus synonymous polymorphism and divergence [McDonald–Kreitman test (35)], as well as site-specific models of selection using phylogenetic analysis by maximum likelihood (PAML) (36). None of these tests reached statistical significance, even when uncorrected for multiple testing (Table S4).

Discussion

Genetic Basis of *HMS1*. Our findings demonstrate the genetic complexity of a small region of 9.2 kb designated *HMS1* that is associated with hybrid male sterility when introgressed from *D. mauritiana* strain Mau12 into *D. simulans* strain SimB. The region contains two likely candidate genes for hybrid male sterility, *agt* and *Taf1*, and we find that both genes contribute to the incompatibility. In both cases, transgenes carrying a compatible allele rescue the fertility of otherwise sterile *HMS1* males. The rescue phenotype has the unusual property that, although only about one-half of the transgenic males regain fertility, those that are fertile produce on average as many progeny as *HMS1[mau]/HMS1[sim]*

Table 1. Nucleotide differences across *agt* and *Taf1* between SimB and Mau12

	Promoter region		Coding sequence		Other	
Gene	Intergenic	5′-UTR	Nonsynonymous	Synonymous	Intron	3′-UTR/intergenic
agt, length (bp)	598*	84	579		N/A	95
agt, no. differences	116	4	6	6	N/A	0
Taf1, length (bp)	1,174	100	6,393		2,440	427
Taf1, no. differences	26	1	14	67	45	11

*Portion of the intergenic region included in the *piggyBac* construct; the total intergenic region between *agt* and *Lab* is 1,421 bp and includes 147 differences. N/A, not applicable.

heterozygous males (Figs. 1 and 2). We note that position effects cannot be fully excluded from having contributed to some of the results, even if the few replicates noted do not show statistically significant differences after multiple comparisons. In *agt*, a transgene carrying an amino acid replacement that had previously been identified as a candidate incompatibility factor (30) rescued *HMS1* sterility less than controls; however, the full effect of *agt* appears to involve an interaction among two or more nucleotide sites along the gene (Fig. 3). Likewise, the *Taf1* incompatibility seems to result from interactions among nucleotide sites, as chimeric constructs do ameliorate the *HMS1* sterile phenotype (Fig. 3). Sequence analysis of allelic diversity among naturally occurring *agt* and *Taf1* alleles gives no evidence of positive selection.

Reproductive Isolation as a Polygenic Quantitative Trait. Our findings with *HMS1* challenge a number of hypotheses regarding hybrid incompatibilities while they support others. The first is that hybrid incompatibilities are due to single genes of large effect. Although a number of large-effect genes are known that result in partial rescue of hybrid sterility or lethality (13, 16, 37, 38), *HMS1* does not represent a single incompatibility gene but a nearby pair of incomplete incompatibilities that each afford only partial rescue. The partial incompatibility due to each gene is itself attributable to epistatic interactions among sites in or near the gene, as expected of ongoing evolution through slightly deleterious mutations compensated by other mutations elsewhere in the gene (39). The number of interacting sites could be as few as two, but experimental evidence implies that higher-order interactions are more likely (40, 41). Considering the within-gene interactions and the effects of environment and genetic background (30), hybrid male sterility in this case behaves like a classical quantitative trait affected by multiple genes with relatively small effects (42) as well as environmental factors. Although a number of genes associated with hybrid incompatibility show evidence of positive selection (3, 10, 43, 44), this is not the case for either *agt* or *Taf1*; however, our findings do not rule out strong selection in the ancient past or ongoing weak selection because these processes would not necessarily leave statistically detectable signals of selection on contemporary polymorphisms.

Fig. 4 illustrates a threshold model of *HMS1* sterility treated as a quantitative trait subject to environmental effects. The horizontal axis represents a hypothetical fertility potential, and the vertical axis is proportional to the fraction of males in a genetically homogeneous population. Any male with a fertility potential less than some threshold (dashed vertical line) is completely sterile. The *mau*/*mau* curve on the *Left* corresponds to the recessive hybrid male-sterile *HMS1[mau]*/*HMS1[mau]* carrying two copies of the *HMS1* introgression, whereas the *sim*/*mau* curve depicts the fertility potential of SimB male individuals carrying one copy of the *HMS1* introgression (*HMS1[sim]*/*HMS1[mau]*). Because the incompatibility is recessive, *sim* represents the dominant allele conferring male fertility. Note that a small proportion of *mau*/*mau* males are fertile, and likewise a small proportion of *sim*/*mau* males are sterile. The *sim* + *mau*/*mau* curve in the *Center* represents the effects of the rescue transgenes *pB::agt[sim]* and *pB::Taf1[sim]* in Fig. 2 and the chimeric genes in Fig. 3. One copy of the dominant fertility allele is sufficient to restore fecundity in the normal range, on average, in 45% of the transgenic males. This kind of fecundity rescue would be expected if, in the transgenic males, any germ cells that survive the incompatibility proliferate more than they otherwise would, thereby compensating for germ cells that succumb to the incompatibility and restoring male fecundity to high levels. It is worth noting that a substantial proportion of males who fail to rescue fertility produce some motile but dysfunctional sperm (Table S1). The proportion of sterile individuals with motile sperm is relatively constant (50%) regardless of their transgenic genotype.

Support for DNA-Binding Proteins. One hypothesis that the genetic analysis of *HMS1* strongly supports is the prominence of internal

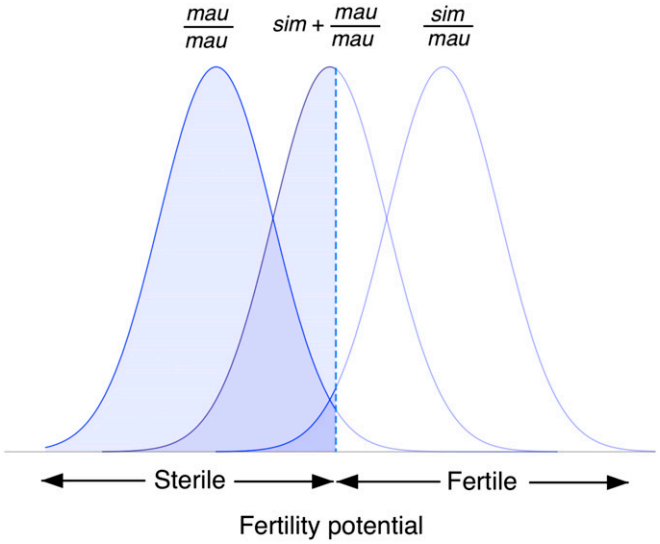

Fig. 4. Threshold model of *HMS1* sterility treated as a quantitative trait subject to environmental effects. The horizontal axis represents the magnitude of a hypothetical fertility potential, and the vertical axis is proportional to the fraction of males in a genetically homogeneous population. Males homozygous for a Mau12 introgression containing *HMS1* (*HMS1[mau]*/*HMS1[mau]*) are virtually sterile, whereas the heterozygous SimB lines *HMS1[sim]*/*HMS1[mau]* are fertile.

genomic conflicts in hybrid incompatibilities, especially those conflicts involving genes for DNA- or chromatin-binding proteins (3, 10). In *Drosophila*, spermatogenesis proceeds through a mitotic phase of stem cell renewal and differentiation followed by meiosis and posttranscriptional spermatid maturation into fully motile, individualized sperm (45). Germ cell development is influenced by genotypic variability as well as microenvironmental perturbations that are ultimately reflected in quantitative differences in the function of mature spermatocytes and spermatozoa within and among individual males.

Several lines of evidence suggest that alternatively spliced isoforms of *Taf1* transcripts regulate spermatogenesis in a tissue-specific manner. *Taf1* encodes a TAF (TATA-box–binding associated factor) constituting one subunit of transcription factor TFIID broadly required for transcription by RNA polymerase II (46, 47). The product of *Taf1* serves as a multifunctional transcriptional regulator operating in the normal cell cycle, the assembly of other TAFs and TBPs (TATA-binding proteins), promoter activities, histone acetylation, and other chromatin modifications that can modulate chromatin structure enabling it to access transcriptionally repressed chromatin (48). In mice, the product of the *Taf1* homolog *Bdrt* associates with hyperacetylated histone H4 and functions in chromatin remodeling following histone hyperacetylation as postmeiotic germ cells mature into fully differentiated sperm (49). In *Drosophila*, the primary transcript of *Taf1* is alternatively spliced into at least four isoforms whose protein products contain AT hooks that directly bind with DNA (50). One of the isoforms (Taf1-2) is enriched in *Drosophila melanogaster* testes where it accounts for 45% of *Taf1* mRNA versus 10% of *Taf1* mRNA in adult male whole flies (50). Among other possible functions, the Taf1-2 product can bind promoters of testes-specific genes including sperm-specific dynein intermediate chain (*Sdic*), heat shock protein 70bc (*hsp70*), β2-tubulin (β2*t*), and don juan (*dj*) (51), and the role of Taf1-2 as transcriptional activator in the testes is likely conserved across *Drosophila*. At 10 kb, *Taf1* is a relatively large gene for *Drosophila*, and there are about 1.5% nucleotide sequence differences in *Taf1* between SimB and Mau12 (Table 1). Several fixed nucleotide differences are located in the gene region around to exon 12a, which is a spliced exon

required for isoform Taf1-2. The sequence divergence might therefore affect alternative splicing efficiency (52), promoter-binding affinity, chromatin modification, or other processes that could contribute to hybrid male sterility as a quantitative trait (53, 54). The role of *Taf1* in hybrid male sterility is perhaps not surprising in light of the rapid evolution and functional diversification of testes-specific TAFs in *Drosophila* (55).

The mechanism by which *agt* contributes to *HMS1* sterility is more obscure. The gene is induced by genotoxic stress, and the protein is an alkyl-cysteine *S*-alkyltransferase that removes alkyl groups from DNA, notably from O^6-methylguanine (56), a repair process observed in eukaryotes, notably in postmeiotic male germ cells in *Drosophila* (57). A much smaller gene than *Taf1*, with an ORF of only 192 codons and no introns, *agt* is 2.4% different between SimB and Mau12 (Table 1). It is also widely divergent in the species subgroup, with nearly every species having a different amino acid at position 121 (30). The D121N difference between SimB and Mau12 clearly contributes to the hybrid male sterility (Figs. 2 and 3); however, interactions among sites in *agt* are also implicated (Fig. 3). Although the role of *agt* in hybrid male sterility in *Drosophila* is uncertain, in the mouse the DNA repair gene *Ercc1* is essential for functional integrity of germ cell DNA and normal spermatogenesis (58), and a histone methyltransferase encoded by *Prdm9* is necessary for meiosis (59). This emphasizes the intriguing relationships between gametogenesis and the variety of DNA repair enzymes functioning in parallel pathways in spermatogenesis. Also, it suggests that in eukaryotes dysfunction of genes involved in any DNA repair mechanism in male germ cell stages may influence the degree of fertility (60, 61).

Temporal Origins of Reproductive Incompatibility. It is unlikely that either *Taf1* or *agt* was involved in the origin of reproductive isolation between *D. simulans* and *D. mauritiana*. Recent experimental evidence (62) and theory (63–67) imply that certain allele combinations causing partial reproductive incompatibility can be found segregating in natural populations, which suggests that, on an evolutionary timescale, reproductive isolation can evolve very rapidly from standing genetic variation for deleterious allelic mutations causing partial reproductive isolation, even in sympatric populations (68, 69).

A quick, heuristic calculation also suggests that the incompatibilities of *Taf1* and *agt* are both likely to have evolved more recently than the species divergence time of 250,000 y. Between *D. simulans* and *D. mauritiana*, there are an estimated 40 autosomal genes that contribute to hybrid male sterility (29); however, the X chromosome, which evolves incompatibilities faster than the autosomes, has about 100 such factors (24). Each factor reduces male fertility by about 10%, and, taken together, they constitute 15 "HMS equivalents," where each HMS equivalent includes a sufficient number of incompatibilities to result in hybrid male sterility (24). In effect, hybrid males of *D. simulans* and *D. mauritiana* are sterile 15 times over. The factors do not act independently, however: epistasis among incompatibility genes is pervasive (7, 29, 70).

If hybrid incompatibility factors evolve at a rate that is linear in time, and those in the X chromosome evolve at a rate 2.5 times that of those in the autosomes (24), then each HMS equivalent would entail, on average, 9.3 total incompatibility factors of which 2.7 would be autosomal. Because 40 autosomal factors have evolved in 250,000 y, then with a linear increase in number through time, 2.7 autosomal factors would be expected to have evolved in 17,000 y. This number is less than 7% of the total time since species divergence, so any incompatibility factor between *D. simulans* and *D. mauritiana* is very likely to have evolved since their divergence. On the other hand, population genetics theory suggests that the accumulation of incompatibility factors might be exponential rather than linear (9). In the exponential case, the situation for early evolution is a bit more optimistic, but not by much. With exponential increase from 0 to 40 autosomal factors in 250,000 y, 2.7 factors would be expected to accumulate in 27,000 y, or about 10% of the total time since divergence. Both estimates imply that most hybrid incompatibility factors identified in well-established, reproductively isolated species are likely to have arisen since speciation. Unveiling the genetic origins of species, "that mystery of mysteries" (71), requires an untangling of complex combinations of the evolutionary forces that create incompatibilities and the reproductive isolating mechanisms that result.

Materials and Methods

Fly Lines. All flies were reared on cornmeal–molasses–agar medium sprinkled with yeast grains (*SI Materials and Methods*). The *D. simulans* stocks used in this study are *w*501 (University of California, San Diego, line 14021-0251.011), *w;e* (*w;II:*e), and simB (*w;nt;III*), where *II* and *III* represents isogenic second and third chromosomes, respectively, from *D. simulans* line (13*w* × JJ) (24). The *D. mauritiana* × *D. simulans* heterozygous introgression lines used in this study are designated *P*45.6 and *P*32.110 (72); the creation of these lines has been described in detail earlier (72, 73). *P* represents the immobile *P*-transposon element *P*[w^+] marking the portion of *D. mauritiana* [Mau12, a white-eyed inbred laboratory stock (14021 0241.60)] on the right arm of the third chromosome. *P*[w^+]-elements are semidominant markers sensitive to the location and copy number of the miniwhite w^+ gene (Fig. S3 and *SI Materials and Methods*). All crosses were performed at room temperature (20–22 °C), except the cross that generated *HMS1[mau]/HMS1[mau]; pB::HMS1[sim]* ("3*P*") progeny, which was performed at 18 °C. Absence of recombination in *Drosophila* males allows maintaining the original introgression lines by backcrossing each generation of *P*[w^+]-males with virgin SimB females.

Cloning and Germ-Line Transformation. Germ-line injections were carried out by BestGene. *D. simulans* w^- embryos were injected with the purified MW-FPNS *piggyBac* (*pB*) plasmid carrying either *agt[sim], Taf1[sim], agt[simD121N], agt[simP_mauCDS], agt[mauP_simCDS], Taf1[simPEx10_mauEx10end], Taf1 [mauPEx10_ simEx10end]*, or a control MW-FPNS plasmid with no insert. Full details of cloning procedures are provided in *SI Materials and Methods*. Injected embryos were subsequently raised in the laboratory, and emerging individual virgin adult w^- flies were backcrossed to SimB. Progeny were screened via presence or absence of the w^+ eye marker (74), and those containing the *pB*[w^+] transposed DNA insert were selected to establish stable male transgenic lines. Altogether, the progeny of 429 (*agt*) and 160 (*Taf1*) injected adult flies was scored, all constructs considered, and yielded similar transformation efficiencies of 2.79% and 3.1%. Flanking genomic insertions were determined via inverse PCR (Table S1 and *SI Materials and Methods*). Stable male transgenic lines representing eight unique *agt* insertions, four unique *Taf1* insertions, and two control lines were used to assess the role of each candidate gene in fertility rescue crosses (Fig. S1).

Phenotyping of Rescue Hybrid Males. Rescue hybrid males bearing a copy of the SimB allele for either *agt* or *Taf1* (*HMS1[mau]/HMS1[mau]; pB::HMS1[sim]*), or a chimeric *agt* or *Taf1* construct, were assayed for their ability to sire progeny in single-mating fertility assays with three virgin *w;e* females before scoring of sperm motility and genotyping of all *P*[w^+]-elements. Full details of these procedures are provided in *SI Materials and Methods*.

Molecular Analysis. The genomic region comprising *agt* or *Taf1* and their respective regulatory regions were amplified for *D. mauritiana* Mau12 and *D. simulans* SimB using overlapping oligonucleotide primers, sequenced, and analyzed in Geneious, version 9.0.5 (75). The sequences have been deposited in the GenBank database under accession nos. KX225407–KX225410.

Tests of Positive Selection. Patterns of DNA divergence were calculated to detect departure from neutral models of molecular evolution among sequenced *agt* and *Taf1* alleles from 17 *D. simulans* populations spanning all continents, as well as 16 independent isogenic lines of *D. mauritiana* collected on Mauritius island (see *Results, Tests for Positive Selection*). Various specific site models from the PAML, v4.8, package (36) were used to test for interspecific site-specific positive selection. Likelihood ratio tests were used to evaluate whether the model pairs allowing positive selection provided a significantly better fit to the data. Results from theses comparisons were consistent, all supporting neutral evolution in and around the *HMS1* region.

Models for the Accumulation of Hybrid Male-Sterility Factors. The linear model used for the accumulation of 40 hybrid male-sterility factors in 250,000 y was $f(t) = 40 \times (t/250{,}000)$; the exponential model used was $f(t) = 40 \times [\mathrm{Exp}(t/250{,}000) - 1]/[\mathrm{Exp}(1) - 1]$.

ACKNOWLEDGMENTS. We thank BestGene, Inc., for carrying embryo microinjections; Dr. Norbert Perrimon's laboratory for discussion on quantitative PCR normalization; Dr. Russ Corbett-Detig and Dr. Jean-Marc Lassance for valuable discussions; Kalsang Namgyal for technical assistance; and the reviewers and coreviewer Katharine Korunes for their constructive comments on the manuscript. This work was supported by National Institutes of Health Grant AI106734 (to D.L.H.), and postdoctoral fellowships from the Swedish Research Council and the European Molecular Biology Organization (to M.A.L.).

1. Bateson W (1909) Heredity and variation in modern lights. *Darwin and Modern Science*, ed Seward AC (Cambridge Univ Press, Cambridge, UK), pp 85–101.
2. Mallet J (2006) What does *Drosophila* genetics tell us about speciation? *Trends Ecol Evol* 21(7):386–393.
3. Johnson NA (2010) Hybrid incompatibility genes: Remnants of a genomic battlefield? *Trends Genet* 26(7):317–325.
4. Presgraves DC (2010) The molecular evolutionary basis of species formation. *Nat Rev Genet* 11(3):175–180.
5. Coyne JA, Orr HA (2004) *Speciation* (Sinauer, Sunderland, MA).
6. Moyle LC, Nakazato T (2009) Complex epistasis for Dobzhansky-Muller hybrid incompatibility in solanum. *Genetics* 181(1):347–351.
7. Chang AS, Bennett SM, Noor MA (2010) Epistasis among *Drosophila persimilis* factors conferring hybrid male sterility with *D. pseudoobscura bogotana*. *PLoS One* 5(10): e15377.
8. Turner LM, White MA, Tautz D, Payseur BA (2014) Genomic networks of hybrid sterility. *PLoS Genet* 10(2):e1004162.
9. Orr HA, Turelli M (2001) The evolution of postzygotic isolation: Accumulating Dobzhansky-Muller incompatibilities. *Evolution* 55(6):1085–1094.
10. Maheshwari S, Barbash DA (2011) The genetics of hybrid incompatibilities. *Annu Rev Genet* 45:331–355.
11. Ting CT, Tsaur SC, Wu ML, Wu CI (1998) A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* 282(5393):1501–1504.
12. Ting C-T, et al. (2004) Gene duplication and speciation in *Drosophila*: Evidence from the *Odysseus* locus. *Proc Natl Acad Sci USA* 101(33):12232–12235.
13. Sun S, Ting C-T, Wu C-I (2004) The normal function of a speciation gene, *Odysseus*, and its hybrid sterility effect. *Science* 305(5680):81–83.
14. Brideau NJ, et al. (2006) Two Dobzhansky-Muller genes interact to cause hybrid lethality in *Drosophila*. *Science* 314(5803):1292–1295.
15. Barbash DA, Ashburner M (2003) A novel system of fertility rescue in *Drosophila* hybrids reveals a link between hybrid lethality and female sterility. *Genetics* 163(1): 217–226.
16. Phadnis N, Orr HA (2009) A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. *Science* 323(5912):376–379.
17. Barbash DA (2010) Ninety years of *Drosophila melanogaster* hybrids. *Genetics* 186(1): 1–8.
18. McDermott SR, Kliman RM (2008) Estimation of isolation times of the island species in the *Drosophila simulans* complex from multilocus DNA sequence data. *PLoS One* 3(6): e2442.
19. Haldane JBS (1922) Sex ratio and unisexual sterility in animal hybrids. *J Genet* 12(2): 101–109.
20. Turelli M, Orr HA (1995) The dominance theory of Haldane's rule. *Genetics* 140(1): 389–402.
21. True JR, Mercer JM, Laurie CC (1996) Differences in crossover frequency and distribution among three sibling species of *Drosophila*. *Genetics* 142(2):507–523.
22. Hollocher H, Wu C-I (1996) The genetics of reproductive isolation in the *Drosophila simulans* clade: X vs. autosomal effects and male vs. female effects. *Genetics* 143(3): 1243–1255.
23. Tao Y, Hartl DL, Laurie CC (2001) Sex-ratio segregation distortion associated with reproductive isolation in *Drosophila*. *Proc Natl Acad Sci USA* 98(23):13183–13188.
24. Tao Y, Chen S, Hartl DL, Laurie CC (2003) Genetic dissection of hybrid incompatibilities between *Drosophila simulans* and *D. mauritiana*. I. Differential accumulation of hybrid male sterility effects on the X and autosomes. *Genetics* 164(4):1383–1397.
25. Palopoli MF, Wu C-I (1994) Genetics of hybrid male sterility between *Drosophila* sibling species: A complex web of epistasis is revealed in interspecific studies. *Genetics* 138(2):329–341.
26. Masly JP, Presgraves DC (2007) High-resolution genome-wide dissection of the two rules of speciation in *Drosophila*. *PLoS Biol* 5(9):e243.
27. Meiklejohn CD, Tao Y (2010) Genetic conflict and sex chromosome evolution. *Trends Ecol Evol* 25(4):215–223.
28. Garrigan D, Kingan SB, Geneva AJ, Vedanayagam JP, Presgraves DC (2014) Genome diversity and divergence in *Drosophila mauritiana*: Multiple signatures of faster X evolution. *Genome Biol Evol* 6(9):2444–2458.
29. Tao Y, Zeng Z-B, Li J, Hartl DL, Laurie CC (2003) Genetic dissection of hybrid incompatibilities between *Drosophila simulans* and *D. mauritiana*. II. Mapping hybrid male sterility loci on the third chromosome. *Genetics* 164(4):1399–1418.
30. Araripe LO, Montenegro H, Lemos B, Hartl DL (2010) Fine-scale genetic mapping of a hybrid sterility factor between *Drosophila simulans* and *D. mauritiana*: The varied and elusive functions of "speciation genes". *BMC Evol Biol* 10:385.
31. Handler AM, Harrell RA, 2nd (1999) Germline transformation of *Drosophila melanogaster* with the *piggyBac* transposon vector. *Insect Mol Biol* 8(4):449–457.
32. Rehm EJ (2015) Inverse PCR and cycle sequencing of *P* element insertions for STS generation. Available at www.fruitfly.org/about/methods/inverse.pcr.html. Accessed May 6, 2016.
33. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123(3):585–595.
34. Fu Y-X, Li W-H (1993) Statistical tests of neutrality of mutations. *Genetics* 133(3): 693–709.
35. McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351(6328):652–654.
36. Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8):1586–1591.
37. Satyaki PRV, et al. (2014) The Hmr and Lhr hybrid incompatibility genes suppress a broad range of heterochromatic repeats. *PLoS Genet* 10(3):e1004240.
38. Phadnis N, et al. (2015) An essential cell cycle regulation gene causes hybrid inviability in *Drosophila*. *Science* 350(6267):1552–1555.
39. DePristo MA, Weinreich DM, Hartl DL (2005) Missense meanderings in sequence space: A biophysical view of protein evolution. *Nat Rev Genet* 6(9):678–687.
40. Stam LF, Laurie CC (1996) Molecular dissection of a major gene effect on a quantitative trait: The level of alcohol dehydrogenase expression in *Drosophila melanogaster*. *Genetics* 144(4):1559–1564.
41. Weinreich DM, Lan Y, Wylie CS, Heckendorn RB (2013) Should evolutionary geneticists worry about higher-order epistasis? *Curr Opin Genet Dev* 23(6):700–707.
42. Naveira HF (1992) Location of X-linked polygenic effects causing sterility in male hybrids of *Drosophila simulans* and *D. mauritiana*. *Heredity (Edinb)* 68(Pt 3):211–217.
43. Orr HA, Presgraves DC (2000) Speciation by postzygotic isolation: Forces, genes and molecules. *BioEssays* 22(12):1085–1094.
44. Orr HA, Masly JP, Presgraves DC (2004) Speciation genes. *Curr Opin Genet Dev* 14(6): 675–679.
45. White-Cooper H, Doggett K, Ellis RE (2009) The evolution of spermatogenesis. *Sperm Biology: An Evolutionary Perspective*, eds Birkhead TR, Hosken DJ, Pitnick S (Elsevier, Burlington, MA), pp 151–183.
46. Metcalf CE, Wassarman DA (2007) Nucleolar colocalization of TAF1 and testis-specific TAFs during *Drosophila* spermatogenesis. *Dev Dyn* 236(10):2836–2843.
47. Feller C, Forné I, Imhof A, Becker PB (2015) Global and specific responses of the histone acetylome to systematic perturbation. *Mol Cell* 57(3):559–571.
48. Dynlacht BD, Hoey T, Tjian R (1991) Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. *Cell* 66(3):563–576.
49. Morinière J, et al. (2009) Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. *Nature* 461(7264):664–668.
50. Katzenberger RJ, Marengo MS, Wassarman DA (2006) ATM and ATR pathways signal alternative splicing of *Drosophila* TAF1 pre-mRNA in response to DNA damage. *Mol Cell Biol* 26(24):9256–9267.
51. Metcalf CE, Wassarman DA (2006) DNA binding properties of TAF1 isoforms with two AT-hooks. *J Biol Chem* 281(40):30015–30023.
52. Wang GS, Cooper TA (2007) Splicing in disease: Disruption of the splicing code and the decoding machinery. *Nat Rev Genet* 8(10):749–761.
53. Johnson NA, Porter AH (2000) Rapid speciation via parallel, directional selection on regulatory genetic pathways. *J Theor Biol* 205(4):527–542.
54. Porter AH, Johnson NA (2002) Speciation despite gene flow when developmental pathways evolve. *Evolution* 56(11):2103–2111.
55. Li VC, et al. (2009) Molecular evolution of the testis TAFs of *Drosophila*. *Mol Biol Evol* 26(5):1103–1116.
56. Kooistra R, et al. (1999) Identification and characterisation of the *Drosophila melanogaster* O^6-alkylguanine-DNA alkyltransferase cDNA. *Nucleic Acids Res* 27(8):1795–1801.
57. Nivard MJ, Pastink A, Vogel EW (1992) Molecular analysis of mutations induced in the vermilion gene of *Drosophila melanogaster* by methyl methanesulfonate. *Genetics* 131(3):673–682.
58. Hsia KT, et al. (2003) DNA repair gene Ercc1 is essential for normal spermatogenesis and oogenesis and for functional integrity of germ cell DNA in the mouse. *Development* 130(2):369–378.
59. Mihola O, Trachtulec Z, Vlcek C, Schimenti JC, Forejt J (2009) A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* 323(5912):373–375.
60. Baarends WM, van der Laan R, Grootegoed JA (2001) DNA repair mechanisms and gametogenesis. *Reproduction* 121(1):31–39.
61. Cooke HJ, Saunders PTK (2002) Mouse models of male infertility. *Nat Rev Genet* 3(10): 790–801.
62. Corbett-Detig RB, Zhou J, Clark AG, Hartl DL, Ayroles JF (2013) Genetic incompatibilities are widespread within species. *Nature* 504(7478):135–137.
63. Tulchinsky AY, Johnson NA, Porter AH (2014) Hybrid incompatibility despite pleiotropic constraint in a sequence-based bioenergetic model of transcription factor binding. *Genetics* 198(4):1645–1654.
64. Phillips PC, Johnson NA (1998) The population genetics of synthetic lethals. *Genetics* 150(1):449–458.
65. Lachance J, Johnson NA, True JR (2011) The population genetics of X-autosome synthetic lethals and steriles. *Genetics* 189(3):1011–1027.
66. Khatri BS, Goldstein RA (2015) A coarse-grained biophysical model of sequence evolution and the population size dependence of the speciation rate. *J Theor Biol* 378:56–64.
67. Tulchinsky AY, Johnson NA, Watt WB, Porter AH (2014) Hybrid incompatibility arises in a sequence-based bioenergetic model of transcription factor binding. *Genetics* 198(3):1155–1166.
68. Turelli M, Lipkowitz JR, Brandvain Y (2014) On the Coyne and Orr-igin of species: Effects of intrinsic postzygotic isolation, ecological differentiation, X chromosome size, and sympatry on *Drosophila* speciation. *Evolution* 68(4):1176–1187.

69. Landguth EL, Johnson NA, Cushman SA (2015) Clusters of incompatible genotypes evolve with limited dispersal. *Front Genet* 6:151.
70. Chang AS, Noor MAF (2010) Epistasis modifies the dominance of loci causing hybrid male sterility in the *Drosophila pseudoobscura* species group. *Evolution* 64(1): 253–260.
71. Darwin C (1859) *The Origin of Species* (New American Library, New York).
72. Tao Y, Hartl DL (2003) Genetic dissection of hybrid incompatibilities between *Drosophila simulans* and *D. mauritiana*. III. Heterogeneous accumulation of hybrid incompatibilities, degree of dominance, and implications for Haldane's rule. *Evolution* 57(11):2580–2598.
73. True JR, Weir BS, Laurie CC (1996) A genome-wide survey of hybrid incompatibility factors by the introgression of marked segments of *Drosophila mauritiana* chromosomes into *Drosophila simulans*. *Genetics* 142(3):819–837.
74. Roote J, Prokop A (2013) How to design a genetic mating scheme: A basic training package for *Drosophila* genetics. *G3 (Bethesda)* 3(2):353–358.
75. Kearse M, et al. (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647–1649.
76. Araripe L, Eckstrand N, Hartl D, Tao Y (2006) Flanking regions of *P*-elements inserted in the 3rd chromosome of *Drosophila mauritiana*. *Dros Inf Ser* 89:54.

Supporting Information

Liénard et al. 10.1073/pnas.1608337113

SI Materials and Methods

Cornmeal–Molasses–Agar Medium. The fly food consists of agar (9.29 g; Affymetrix), torula yeast (32.35 g; Affymetrix), cornmeal (61.17 g; MP Biomedicals), dextrose (64.70 g; Affymetrix), molasses (47.05 mL; Genesee Scientific), 10% (vol/vol) tegosept solution in ethyl alcohol (9.41 mL; Genesee Scientific), propionic acid (5.88 mL; VWR), and distilled water (to 1 L).

Plasmid Constructs for *agt* and *Taf1*. For *agt[sim]*, a 1.36-kb genomic region (Fig. S4*A*) was amplified using SimB genomic DNA as template, which comprises 681 bp upstream flanking DNA sequence including: (*i*) the predicted transcription start at position −270, (*ii*) the 579-bp intronless *agt[sim]* ORF, and (*iii*) 95 bp of the downstream intergenic region between *agt* and *spase*. This PCR fragment was inserted into a MW-FPNS *piggyBac* vector [pBac(3xP3-EGFPafm)::MCS::(pW8 miniwhite)] obtained from Dave Miller in the laboratory of Thomas Kauffman, Indiana University, Bloomington, IN. The intergenic flanking region between *agt* and *spase* totals 145 bp and is identical in sequence between Mau12 and SimB. A PCR QuikChange site-targeted mutagenesis (Stratagene) and modified primers were used to generate the *agt[simD121N]* point-mutation construct, following the manufacturer's procedure. The chimeric constructs *agt[simP_mauCDS]* and *agt[mauP_simCDS]* were generated by blunt-end ligation between the *agt[SimB]* coding and UTR region and the *agt[Mau12]* promoter region, and conversely, by fusing the *agt[Mau12]* coding and UTR regions with the *agt[SimB]* promoter region (Fig. S4*A*). Specifically, oligonucleotide primers encompassing AvrII and StuI or StuI and FseI restriction sites were used to generate PCR fragments corresponding to the promoter or coding sequence (+95 bp downstream sequence) using SimB or Mau12 genomic DNA as template, respectively, followed by ligation in PCR TOPO 2.1 (Invitrogen) and transformation in *Escherichia coli* TOPO 10 competent cells. The internal StuI restriction site (AGGGCCT) was engineered in the oligonucleotide primer via a single-nucleotide change from the original SimB/Mau *agt* DNA sequence that is located 5 bp upstream from the *agt* 5′-UTR region. Individual plasmid DNA constructs were verified by double-restriction digestion (AvrII/StuI or StuI/FseI) and sequenced on both strands with ABI 3730xl automated capillary sequencing instruments and ABI PRISM BigDye chemistry to ensure that no mutations were introduced, before subcloning into the AvrII and FseI cloning sites of the linearized *pB* vector. Internal fusion between coding sequence and promoter regions was achieved in this ligation step by blunt-end ligation at the StuI restriction site. All final constructions encompassed the same gene region, only with variable portions of the SimB or Mau12 alleles of *agt*.

For *Taf1[sim]*, oligonucleotide primers were designed encompassing the FseI and Sbf1 restriction sites (Fig. S4*B*) to amplify a 10.5-kb fragment of the SimB allele corresponding to: (*i*) 1,273 bp of upstream intergenic region comprising the 100-bp-long 5′-UTR and 1,173 bp of the promoter region including the predicted transcription start at position −1,098, (*ii*) the 8,883-bp region corresponding to the total *Taf1[SimB]* (16 exons and 15 introns), and (*iii*) 431 bp downstream sequence comprising the 3′-UTR and the entire intergenic region between *Taf1* and gene *CG1307*. For the *Taf1* chimeras, *Taf1[simPEx10_mauEx10end]* and *Taf1[mauPEx10_simEx10end]* (Fig. S4*B*), DNA fragments corresponding to each part were amplified using TaKaRa LA Taq (Clontech) from SimB or Mau12 genomic DNA templates, respectively, and religated at an internal AvrII site adjacent to a previously mapped molecular marker marking the *HMS1* boundary (Fig. S4*B*) in a single subcloning reaction in the linearized *pB* vector. First, PCR fragments were ligated in TOPO XL PCR (Invitrogen) and transformed in *E. coli* TOPO 10 cells (Invitrogen), before plasmid purification and sequencing. Upon plasmid restriction digestion, the 5.9-kb *Taf1[mau]* or *Taf1[sim]* DNA fragment (including the promoter region and partial coding sequence) and delimited by FseI and AvrII was respectively fused at AvrII to the 4.6-kb *Taf1[SimB]* or *Taf1[Mau12]* fragment encompassing the coding sequence end and 3′-UTR. Each *Taf1* chimera was religated in a single directional cloning step at the FseI and Sbf1 restriction sites of the linearized *pB* vector. The empty MW-FPNS *piggyBac* vector [pBac(3xP3-EGFPafm)::MCS::(pW8 miniwhite)] alone was used for germ-line transformation generating control lines. The MW-FPNS *piggyBac* vector, and final *agt* and *Taf1* constructs were sequenced on both strands with ABI 3730xl automated capillary sequencing instruments and ABI PRISM BigDye chemistry to ensure that no mutations had been introduced before germ-line transformation.

Localization of *pB* Chromosomal Insertions. An inverse PCR protocol adapted from the Berkeley *Drosophila* genome project (32) was used to determine the 5′- and 3′-end sequences flanking *pB* insertions. High-quality genomic DNA was extracted from five w^+ males of each line and digested using *Sau3A* (5′ end) or *HinP1* (3′ end), before overnight self-ligation using T4 DNA ligase (New England Biolabs). Ligation products were subjected to two successive rounds of PCR using the following cycling conditions: 94 °C for 5 min, 35 cycles of 94 °C for 30 s, 58 °C for 1 min, 72 °C for 2 min 30 s, and a final step at 72 °C for 10 min using primers designed based on the MW-FPNS plasmid on the boundaries of the left and right insertion sites (all primer sequences used in this study are available upon request). Strong unique bands were detected by agarose gel electrophoresis in all PCRs, and the PCR products were purified using Exo and SAP enzymes (Fermentas) before sequencing. DNA sequences were analyzed in Geneious, version 7.1.7, and searched against the *D. simulans* and *D. melanogaster* reference genomes using the flybase BLAST tool, altogether confirming that each transgenic strain has a single-copy gene inserted in one genomic location; locations are provided in Table S1.

Phenotyping of P[w$^+$]-Elements. Each *P*[w^+]-element is a semidominant marker bearing a copy of the miniwhite w^+ gene. Eye color is sensitive to the copy number and position of the *P*[w^+]-insert, which allows to distinguish 1*P* or 2*P* heterozygotes from 2*P* and 3*P* homozygotes phenotypes. *P*45.6 males (*HMS1[mau]*/*HMS1[sim]*) typically have dark red eyes, whereas *P*32.110 males (*HMS1[mau]*/*HMS1[sim]*) have bright orange eyes. Corresponding virgin females have lighter eye coloration, orange and yellow, respectively; however, the eye colors darken with an individual's age. Both *P*45.6 and *P*32.110 heterozygous lines carry a *Drosophila mauritiana* introgression that covers *HMS1* and confers full sterility to males when homozygous. The homozygous 2*P* (*P*45.6/*P*32.110) flies, with two copies of *P*[w^+]-inserts have a clearly different, bright red eye color. 3*P* rescue homozygotes are not distinguishable from homozygotes 2*P*, requiring progeny phenotyping and molecular genotyping of all *P*[w^+]-elements.

Genotyping. Genomic DNA was extracted in 50 μL of fresh squishing buffer (10 mM Tris, pH 8.2, 1 mM EDTA, 25 mM NaCl, and 200 μg/mL Proteinase K), and the extract was incubated for 1 h at 37 °C and a 2-min inactivation at 95 °C. Single genotyping PCRs were run using 1 μL of male DNA as template and sets of specific primers targeting *pB*, *P*45, and *P*32 *P*-elements under the following

cycling conditions: 94 °C for 4 min, 35 cycles of 94 °C for 30 s, 60 °C for 30 s, 72 °C for 1 min 30 s, and 72 °C for 10 min, before visualization of amplicon presence/absence on 1% agarose gels. *HMS1[mau]/HMS1[mau]; pB::HMS1[sim]* or 3*P* males were also genotyped to ensure no recombination at *P*45.6 or *P*32.110 had taken place. Whether *pB::HMS1[sim]* alleles were inserted on the second or third chromosome, males bearing the *pB* and *P*32.110 elements were crossed to *P*45 females. Accordingly, 3*P* fertile males were genotyped to ensure no recombination at *P*45. 3*P* fertile male progeny bearing *pB* on chromosome 3 were genotyped at *P*32.110 to ensure the absence of recombination in intermediate 2*P* females (*P*32.110 + *pB* or *HMS1[mau]/HMS1[sim]; pB::HMS1[sim]*).

Fertility Assays. For fertility assays in single-male mating tests, bright red-eye male progeny corresponding to *HMS1[mau]/HMS1[mau]; pB::HMS1[sim]* and also homozygous 2*P* genotypes obtained for crosses on chromosome 2 (*HMS1[mau]/HMS1[mau]*) were selected and crossed at 18 °C with three 2- to 3-d-old virgin females of the *D. simulans* tester *w;e* line, where the *e* (ebony) recessive marker is used to detect potential non virgins. Similarly, *HMS1[mau]/HMS1[sim]* (*P*45.6) dark red-eyed male progeny were crossed to record the progeny distribution of "fertile" control heterozygous. On the 10th day, females were cleared, and dental cotton was inserted in the medium to maximize pupation. Only sterile vials for which the male and at least one female were still alive at the time of collection were included in the analysis to prevent inclusion of males who might have died before mating. A test male was considered sterile if he produced zero progeny, and fertile if he produced one progeny or more. Progeny were counted on the 20th and 25th day. The male parent genotype was determined based both on the segregation of *P*-element eye coloration in the progeny, and based on PCR genotyping at all three *P*-elements. The latter is particularly critical to distinguish homozygous 2*P* from 3*P* parent males because both bear several genetic w^+ elements in addition to the dominant red *P*45.6, and therefore display similar bright vermillion eye coloration. Results for the tested transgenic lines carrying *pB::agt[simP_mauCDS]* (insertions 7 and 8) and *pB::agt[mauP_simCDS]* (insertions 9 + 10) were grouped for clarity as presented in Fig. 3 after confirming that the degree and level of fertility rescue were similar between lines for each transgene, respectively [$P = 0.236$ ($X^2 = 1.406$, df = 1, $n = 36$); $P = 0.889$ ($X^2 = 0.019$, $df = 1$, $n = 114$)].

Sperm Motility. The reproductive organs of hybrid males that did not sire offspring on the 10th day were dissected in Ringer's isotonic solution (111 mM NaCl, 5.6 mM KCl, 2.2 mM $CaCl_2$, and 2.4 mM $NaHCO_3$), and gently squashed preparations were visualized under phase contrast microscopy to assess the presence of elongated spermatids, individualized spermatocytes, and motile sperm during a 60-s observation period. Because sperm motility is very difficult to measure quantitatively, we did not define a sperm motility index based on the number of motile sperm, and characterized males as bearing motile sperm whenever one or more sperm were observed moving. Reproductive organs from an arbitrary subset of fertile males were observed and always showed evidence of abundant motile sperm. None of the examined sterile individuals showed evidence for smaller, absent, or abnormal testes (e.g., hybrid dysgenesis in the form of gonadal abnormalities), and elongated spermatocytes were observed in all cases. All *pB* lines considered, sperm motility was assessed in a subset of 35 2*P* sterile homozygous males and 256 sterile 3*P* males.

Monitoring mRNA Relative Expression Level in 2*P* Males. 2*P* males were generated by crossing virgin 2-d-old *P*32.110 males (Sim B *w;nt;III/HMS1[mau]*) or *P*32.75 males (Sim B *w;nt;III/HMS1[sim]*) with 4-d-old virgin females from the *P*45.6 tester line (SimB *w;nt;III/HMS1[mau]*). In the *w;nt;III/HMS1[sim]* introgression line, the *D. mauritiana* portion of the chromosome does not span the *HMS1* region (30). Two genetic backgrounds were used for the experiment, in which *HMS1* was introgressed in the *SimB* or *w*501 strains.

Individual 2*P* male fertility phenotypes were assessed by crossing each male to three 4-d-old virgin *w;e* females following the procedure described under the fertility assay section. On the seventh day, 2*P* males were anesthetized and testes were dissected in sterile Ringer's solution, immediately transferred in 20 μL of TRIzol and snap-frozen in TRIzol. Testes tissues in TRIzol were subsequently disrupted on dry ice using a sterile pestle rotor before storage at –80 °C. The corresponding individual male carcasses were used for DNA extraction and genotyping at *P*45.6 to ensure absence of recombination in females, before tissue pooling and RNA extraction.

For each 2*P* genotype category, RNA was extracted using the TRIzol procedure (Invitrogen) before overnight precipitation at –20 °C and with NaAc/ethanol purification and treatment with DNase I (New England Biolabs) from biological replicates representing a population of 10 testes samples for which males displayed fertile or sterile phenotypes. cDNAs were synthesized using the SuperScript II Reverse Transcriptase (Invitrogen), an Oligo d(T) 23 VN (New England Biolabs), and the RNase Out (Invitrogen) following procedures provided by the manufacturer.

The 25-μL quantitative PCRs (qPCRs) were run on an Applied Biosystems 7900HT using SYBR Green PC master mix (Applied Biosystems) and 20 ng of cDNA, 200 nM GSPs, and 50 nM Rox dye with the following cycling conditions: 50 °C for 2 min, 95 °C for 2 min, followed by 40 cycles of 95 °C for 30 s, 60 °C for 1 min, and 72 °C for 1 min. The dissociation curve analysis was as follows: 95 °C for 1 min, 60 °C for 30 s (and a gradual heating gradient to 95 °C at 0.01 °C/s. Target genes primer sets were designed in AlleleID software (PREMIER Biosoft International) as follows: *Taf1*: 5′-CAACGACGGCAAGGAATA-3′ and 5′-TGCGAACTGCTTGATGAA-3′ (amplicon size: 112 bp); and *agt*: 5′-CCAATTGCGACTTGGTCTTT-3′ and 3′-CCAGAACACCCAAGTTCGTT-3′ (amplicon size: 126 bp). Quantification reactions for each cDNA were run in duplicates in two separate qPCRs ($n = 4$ for genotypes with one biological replicate of 10 testes samples, or $n = 8$ for genotypes with two biological replicates of 10 testes samples). Relative expression was calculated using Δ*Ct* values after robust normalization against three stable endogenous control genes (*RpL32*, m*RpL15*, and *Alpha-tubulin84D*) targeting the following gene regions: *RpL32* (*Dsim*/GD17388): 5′-ATCGGTTACGGATCGAACAA-3′ and 5′-GACGATCTCCTTGCGCTTCT-3′ (amplicon size: 165 bp); *mRpL15* (*Dsim*/GD12186): 5′-GGTCATAGCGGCCATAGAGA-5′ and 5′-CAGCATGCGGCTAGGAATAG-3′ (amplicon size: 130 bp); and *Alpha-tubulin84D* (*Dsim*/GD19940): 5′-TGTCGCGTGTGAAACACTTC-3′ and 5′-AGCAGGCGTTTCCAATCTG-3′ (amplicon size: 96 bp). All primer pair reaction efficiencies were initially validated with standard curves calculated from running preexperimental qPCRs with serial dilutions of template cDNA. We initially tested several primer pairs and additional reference genes including that encoding actin, and the stability of the reference gene expression patterns chosen for our analysis was assessed postexperimentally by statistical analyses of variation using the Bestkeeper software program for reference gene selection (www.gene-quantification.com/bestkeeper.html).

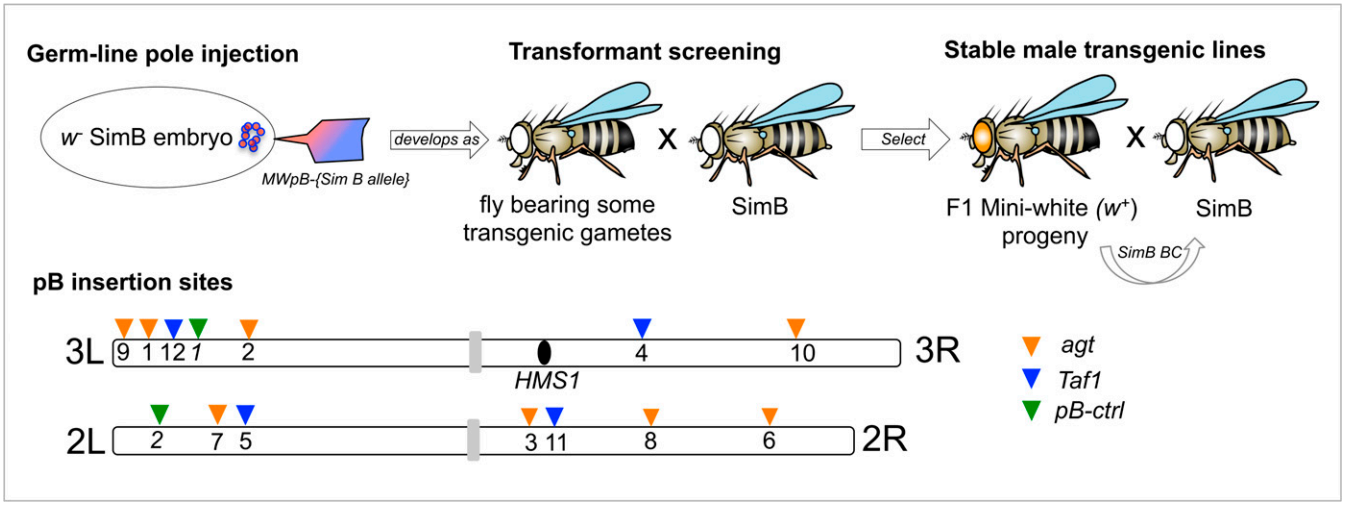

Fig. S1. Germ-line transformation and *piggyBac* insertion sites. SimB *white-eye* fruit fly embryos were transformed with the donor plasmid MW-FPNS alone (*pB-ctrl*) or containing sequences of *agt* or *Taf1*, alongside a helper recombinase plasmid. The *pB* carries a miniwhite (w^+) gene for screening F_1 offspring. Stable male rescue lines are maintained via backcross. Insertion sites are indicated by upside-down triangles, and the numbers inside the chromosomes refer to the insertion nucleotide site for each line as listed in Table S1. The centromere of each chromosome is represented by a gray rectangle. The position of *HMS1* is indicated by a black oval. Fly images were designed using the Genotype Builder (74).

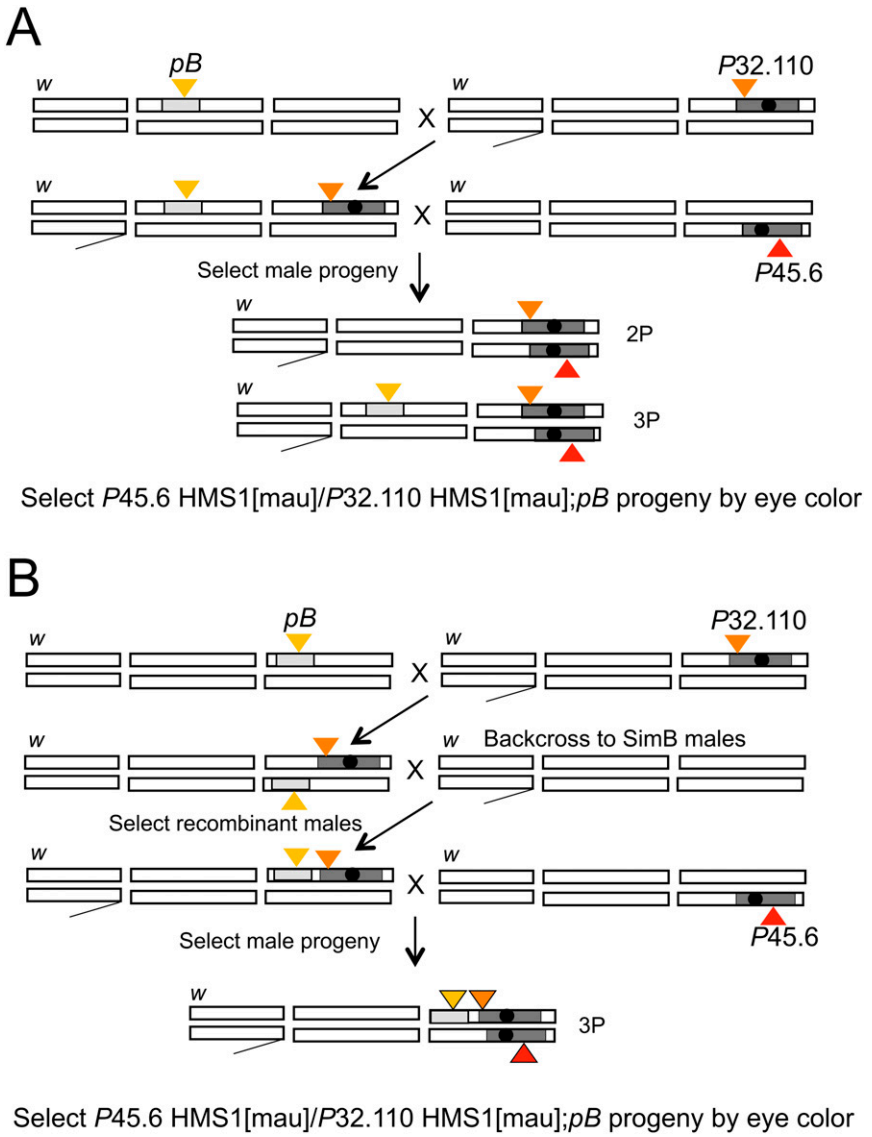

Fig. S2. Mating scheme for introducing *piggyBac* (*pB*) constructs into a homozygous *HMS1[mau]/HMS1[mau]* male-sterile introgression background. (*A*) When *pB* construct is in chromosome 2. Homozygous *HMS1[mau]/HMS1[mau]* males (2*P*) served as controls. (*B*) When *pB* construct is in chromosome 3.

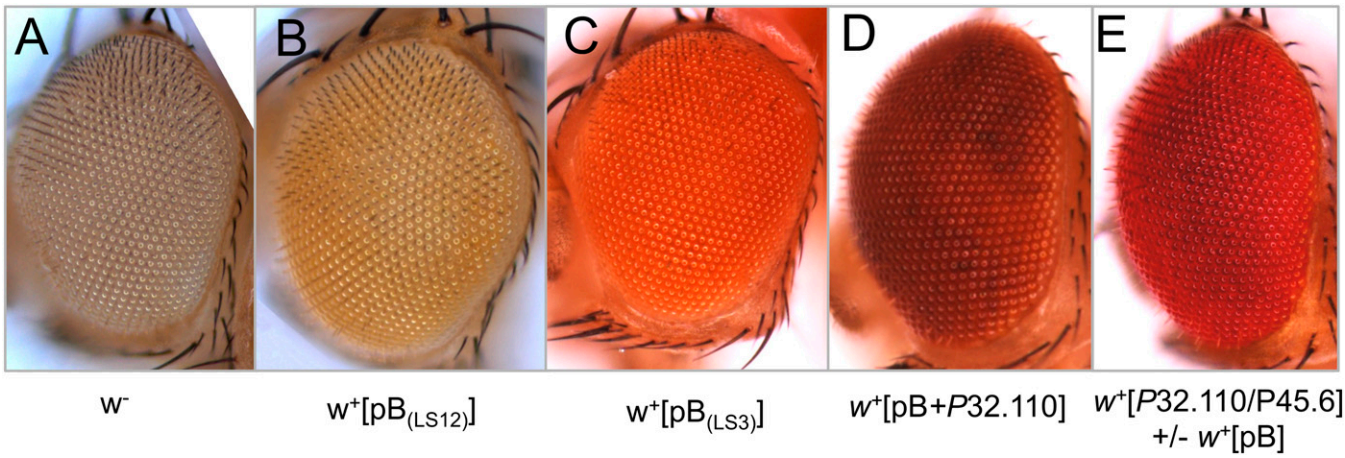

Fig. S3. Positional and copy number effect of the miniwhite w^+ eye marker phenotype in *D. simulans* males. (*A*) SimB [w^-] parental line. (*B* and *C*) Examples of *piggyBac*-[w^+] insertions on chr3 (*B*) and chr2 (*C*), respectively. Variable eye color phenotypes are observed depending on the marker genomic location, typically ranging from pale yellow to bright orange. (*D*) Male eye coloration from genotype *HMS1[mau]/HMS1[sim(P32.110)]; pB::HMS1[sim]*). (*E*) Typical bright vermillion red eye phenotype obtained in progenies with genotype *HMS1[mau(P45.6)]/HMS1[mau(P32.110)]; pB::HMS1[sim])* or *HMS1[mau(P45.6)]/HMS1[mau(P32.110)]*. All images were acquired from 1- to 2-d-old individuals except for a 7-d-old individual in *D* on an AxioZoom.V16 stereo zoom microscope equipped with the AxioCam HRm using the *z*-stack module with bright-field illumination in the Zen acquisition suite (Zeiss). No editing was done except cropping because the *z*-stack imaging mode allows to collect focused pixels from a collection of real-time images.

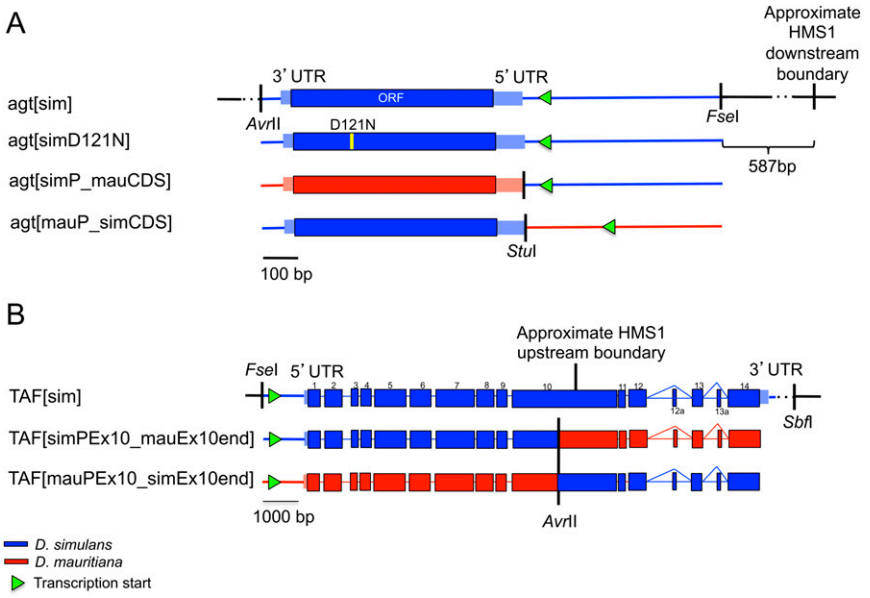

Fig. S4. Constructs inserted into *piggyBac* for testing possible male-sterility complementation by *agt* and *Taf1*. Blue and red represent DNA from SimB or Mau12, respectively. (*A*) *agt* constructs. (*B*) *Taf1* constructs.

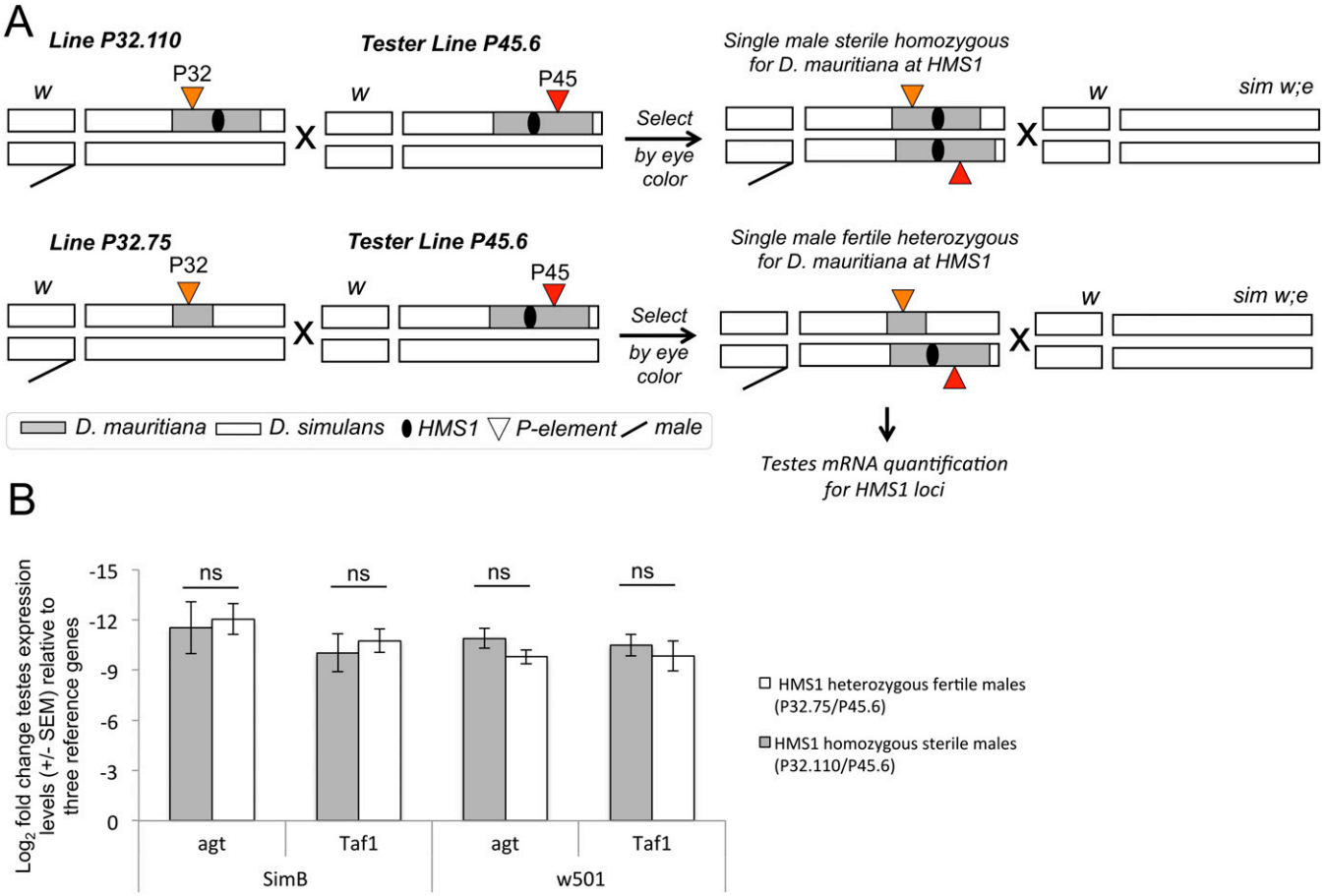

Fig. S5. *Taf1* and *agt* testes cDNA expression patterns monitored by qPCR. (*A*) Crossing scheme used to generate individual males heterozygous or homozygous for the *D. mauritiana HMS1* region. Individual tester virgin females bearing the *D. mauritiana P*45.6 introgression were crossed to males bearing a distinct heterozygous introgression flanking the *P*-element *P*32 (covering or not covering *HMS1*) generating a proportion of male offspring with both *P*-elements (2*P* males) that could be distinguished by their bright-red eyes. Under this mating scheme, *HMS1* is made homozygous from two independently maintained *P*[w^+] stocks, which controls for male sterility arising via the accumulation of slightly deleterious spontaneous mutations. Each 2*P* male was subsequently crossed to three virgin *w;e* females to confirm its expected phenotype (sterile when homozygous for *HMS1*, fertile when heterozygous for *HMS1*), before testes dissection and RNA extraction, and further genotyping to ensure the absence of recombination in *P*45.6 females. A similar crossing scheme was used for the *P*32 and *P*45.6 introgressions in two *D. simulans* genetic backgrounds (SimB and w501), generating four distinct 2*P* male categories. (*B*) Testes cDNA expression profiles of *Taf1* and *agt* monitored by qPCR. The mean relative $\log_2$ fold change expression scores were calculated from raw cycle threshold values (±SEM, n = 4–8) relative to three reference genes (*SI Materials and Methods*). Expression levels are not significantly (ns) different between heterozygous and homozygous lines in either genetic background at either locus (independent t tests: SimB *agt*: $t = -0.311$, df = 9, $P = 0.763$; SimB *Taf1*: $t = -0.568$, df = 10, $P = 0.582$; w501 *agt*: $t = 0.102$, df = 9, $P = 0.315$; w501 *Taf1*: $t = 0.580$, df = 10, $P = 0.575$). In other words, the presence of one or two *mau* copies at *HMS1* is not associated with a gene expression imbalance for either of the candidate loci *agt* or *Taf1*. This finding supports the inference that the incompatibility does not derive from variation in expression level of the *sim* and *mau* alleles, but rather from regulatory and functional allelic variation in or near the genes themselves.

Table S1. Hybrid fertility rescue and sperm motility in introgression males

Insertion	Insertion site	*pB* nucleotide insertion (bp)	Percent fertile males*	Mean progeny (±SD)	No. of sterile males	Proportion of sterile males with motile sperm (%)
HMS1[mau]/HMS1[sim]	N/A	N/A	100 (n = 183)	82.61 (± 52.1)	0	0
HMS1[mau]/HMS1[mau]†	N/A	N/A	2.1 (n = 109)	1.5 (± 0.71)	107	34.6‡
pB::agt[sim]	1	3L:2,110,233 (2.1 Mb)	61 (n = 85)	104 (± 55.04)	33	39.8
	2	3L: 9,809,050 (9.8 Mb)	51 (n = 74)	122.2 (± 45.2)	36	55.5
	3	2R:4.463,970 (4.46 Mb)	70 (n = 27)	91.79 (± 46.6)	10	ND
pB:: agt[simD121N]	6	2R:15,215,710 (15.2 Mb)	27 (n = 40)	106.47 (± 72.7)	25	ND
pB::agt[simP_mauCDS]	7+8	7 = 3L:172,185 (1.72 Mb); 8 = 3R:22,151,753 (22.15 Mb)	53 (n = 36)	145.7 (± 65.6)	16	ND
pB::agt[mauP_simCDS]	9+10	9 = 2L:7,793,729 (7.79 Mb); 10 = 2R:5,961,808 (5.96 Mb)	50 (n = 114)	104.8 (± 43.8)	56	49.7
	All *agt* insertions		52.0% fertility			49.8
pB::TAF[sim]	4	3R:8,564,338 (8.56 Mb)	38 (n = 47)	87.7 (± 40.9)	29	57.1
	5	2L:7,872,345 (7.87 Mb)	57.5 (n = 47)	127.26 (± 44.1)	20	38.9
pB::TAF[simPEx10_mauEx10end]	11	2R:5,362,355 (5.36 Mb)	45.9 (n = 37)	112.23 (±42.37)	20	43.5
pB::TAF[mauPEx10_simEx10end]	12	3L:3,155,806 (3.15 Mb)	50 (n = 56)	119.5 (± 45.95)	28	42.9
	All *TAF* insertions		47.85% fertility			45.6

*n = number of males individually assayed in a fertility cross with three *w;e* virgin females.
†*HMS1[mau]/HMS1[mau]* progeny derived from all rescue crosses.
‡Thirty-five males were examined: 66.4% had no motile sperm, and 34.6% had very few (<10) motile sperm. ND, nondetermined.

Table S2. Progeny distribution of *pB::control* lines

pB insertion site*	Male†	Genotype‡	Phenotype	Progeny
pB::ctrl-1	1_1	3*P*	S	0
	1_2	3*P*	S	0
	1_3	3*P*	S	0
	1_10	3*P*	S	0
	1_12	3*P*	S	0
pB::ctrl-2	2_1	3*P*	S	0
	2_2	3*P*	S	0
	2_4	3*P*	S	0
HMS1[mau]/HMS1[sim]	1_6	*P*45	F	101
	2_3	*P*45	F	62
	2_5	*P*45	F	142
	2_6	*P*45	F	141
	2_7	*P*45	F	140

F, fertile; S, sterile.
**pB* nucleotide insertion: *pB::ctrl-1* at 3L:6,000,733 (6 Mb); *pB::ctrl-2* at 2L:2,972,661 (2.97 Mb).
†Each male was individually assayed in a fertility cross with three *w;e* virgin females.
‡3*P* = *HMS1[mau]/HMS1[mau]; pB::ctrl*; *P*45 = *HMS1[mau]/HMS1[sim]*.

Table S3. Quantitative expression data from male testes in *D. simulans* SimB and w501 lines bearing *HMS1* from *D. mauritiana* when heterozygous (*P*32.75) or homozygous (*P*32.110)

Genetic background	Male testes cDNAs (*HMS1* in heterozygosity, *sim*/*mau* = fertile, versus *HMS1* in homozygosity, *mau*/*mau* = sterile)*	*Ct* values housekeeping genes				*Ct* values target genes		Log fold-expression relative to three reference genes		Genetic background	Locus	Mean log fold expression relative to three reference genes			
		$Ct_{(mRPL15)}$	$Ct_{(\alpha\ tub\ 84D)}$	$Ct_{(RPL32)}$	Mean *Ct* for three reference genes	$Ct_{(TAF)}$	$Ct_{(agt)}$	Delta $Ct_{(TAF)}$	Delta $Ct_{(agt)}$			HMS1 heterozygous fertile males (*P*32.75/*P*45.6)	HMS1 homozygous sterile males (*P*32.110/*P*45.6)	SEM fertile	SEM sterile
Exp1_w501	w501_P32.75/P45.6_fertile	31.36	25.64	24.51	27.40	36.38	37.85	−8.98	−10.45	w501	Taf1	−9.85	−10.49	0.90	0.63
		32.62	25.65	24.61		36.44	37.3	−9.04	−9.90		agt	−9.79	−10.90	0.41	0.60
	w501_P32.110/P45.6_sterile1	32.25	25.78	24.58	27.54	37.9	40.82	−10.36	−13.28	SimB	Taf1	−10.76	−10.04	0.69	1.15
		32.75	26.71	23.19		38.96	38.11	−11.42	−10.57		agt	−12.05	−11.53	0.91	1.56
	w501_P32.110/P45.6_sterile2	37.47	31.73	28.15	31.74	41.11	39.88	−9.37	−8.14						
		36.6	29.38	27.11		40.21	41.41	−8.47	−9.67						
Exp2_w501	w501_P32.75/P45.6_fertile	32.34	26.28	25.83	27.77	40.31	—	−12.54	—						
		32.1	25.44	24.64		36.61	36.8	−8.84	−9.03						
	w501_P32.110/P45.6_sterile1	33.67	26.53	24.81	28.38	40.34	40.14	−11.97	−11.77						
		33.38	27.12	24.74		42.18	40.25	−13.81	−11.88						
	w501_P32.110/P45.6_sterile2	34.93	26.87	25.52	28.98	38.51	41.26	−9.53	−12.28						
		34.35	27.61	24.58		37.96	38.59	−8.98	−9.61						
Exp1_SimB	SimB_P32.75/P45.6_fertile1	30.54	25.4	24.72	27.10	37.36	—	−10.26	—						
		31.46	25.46	25.04		36.97	39.44	−9.87	−12.34						
	SimB_P32.75/P45.6_fertile2	31.47	25.3	24.54	27.06	36.72	36.91	−9.66	−9.85						
		31.71	24.76	24.59		36.11	38.35	−9.05	−11.29						
	SimB_P32.110/P45.6_sterile	31.96	25.4	25.98	27.81	36.88	38.11	−9.07	−10.30						
		31.82	25.73	25.97		35.15	35.68	−7.34	−7.87						
Exp2_SimB	SimB_P32.75/P45.6_fertile1	32.35	26.22	25.69	28.20	41.95	44.5	−13.75	−16.30						
		33.01	26.29	25.62		38.37	39.38	−10.17	−11.18						
	SimB_P32.75/P45.6_fertile2	33.1	25.98	24.43	27.91	41.87	41.88	−13.96	−13.97						
		33.22	26.76	23.95		37.24	37.34	−9.33	−9.43						
	SimB_P32.110/P45.6_sterile	32.6	24.65	24.37	27.15	39.69	40.1	−12.54	−12.95						
		32.22	25.11	23.94		38.35	42.16	−11.20	−15.01						

*cDNA samples were obtained from testes of individuals isogenic for the *D. mauritiana P*45.6 introgression on chromosome 3 (that contains *Hms1*), with on their second chromosome 3, either the *D. mauritiana P*32.75 introgression (that does not cover *HMS1*), or the *D. mauritiana P*32.110 introgression (that covers *HMS1*). The absence or presence of a second copy of the *HMS1* region confers male fertility or male sterility, respectively, due to expression of the recessive incompatibility phenotype.

Table S4. Population genetic analysis of *agt* and *Taf1*

Gene	*agt*	*Taf1*
No. alleles	32	15
Length (bp)	576	2178
No. polymorphic sites	13	21
Nonsynonymous polymorphisms	15	10
Synonymous polymorphisms	10	21
Nonsynonymous fixed differences	1	4
Synonymous fixed differences	4	8
No. haplotypes	8	7
Haplotype diversity	0.895	1.00
Nucleotide diversity (π)	0.007	0.004
Nucleotide polymorphism (θ)	0.007	0.004
Tajima's *D*	−0.11 NS	0.43 NS
Fu and Li's *F*	NS	NS
Fu and Li's *D*	NS	NS
McDonald–Kreitman	0.16 NS	0.99 NS
PAML	0.10–0.15 NS	0.27–0.28 NS

Note: *P* values uncorrected for multiple tests.