Species-specific positive selection of the male-specific lethal complex that participates in dosage compensation in *Drosophila*

Monica A. Rodriguez[†‡], Danielle Vermaak[†], Joshua J. Bayes[†§], and Harmit S. Malik[†¶]

[†]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109; and [§]Molecular and Cellular Biology Program, University of Washington, Seattle, WA 98195

Communicated by Robert N. Eisenman, Fred Hutchinson Cancer Research Center, Seattle, WA, August 7, 2007 (received for review June 27, 2007)

In many taxa, males and females have unequal ratios of sex chromosomes to autosomes, which has resulted in the invention of diverse mechanisms to equilibrate gene expression between the sexes (dosage compensation). Failure to compensate for sex chromosome dosage results in male lethality in *Drosophila*. In *Drosophila*, a male-specific lethal (MSL) complex of proteins and noncoding RNAs binds to hundreds of sites on the single male X chromosome and up-regulates gene expression. Here we use population genetics of two closely related *Drosophila* species to show that adaptive evolution has occurred in all five protein-coding genes of the MSL complex. This positive selection is asymmetric between closely related species, with a very strong signature apparent in *Drosophila melanogaster* but not in *Drosophila simulans*. In particular, the MSL1 and MSL2 proteins have undergone dramatic positive selection in *D. melanogaster*, in domains previously shown to be responsible for their specific targeting to the X chromosome. This signature of positive selection at an essential protein–DNA interface of the complex is unexpected and suggests that X chromosomal MSL-binding DNA segments may themselves be changing rapidly. This highly asymmetric, rapid evolution of the MSL genes further suggests that misregulated dosage compensation may represent one of the underlying causes of male hybrid inviability in *Drosophila*, wherein the fate of hybrid males depends on which species' X chromosome is inherited.

genetic conflict | McDonald–Kreitman test | X chromosome | spiroplasmal | retrotransposons

Chromosomal aneuploidy is highly deleterious; deletions larger than 3% of the genome and duplications larger than 10% are not tolerated in *Drosophila* (1), presumably because an imbalance of expression levels of many genes is hard to accommodate in stoichiometric complexes involving many different proteins (2). In organisms with highly diverged sex chromosomes, there is frequently a difference in number of sex chromosomes versus autosomes in the heterogametic sex (XY or ZW). This difference requires "dosage compensation" strategies to equilibrate expression levels in both sexes. Recent evidence suggests that these strategies operate at two levels. A primary mechanism is to increase gene expression of the single X chromosome by 2-fold in the heterogametic sex, a strategy that appears to be universally conserved in animals (3, 4). However, different animal lineages have adopted diverse, secondary strategies to equilibrate gene expression in the two sexes (5). In mammals, this secondary modification involves the inactivation of one of the two female X chromosomes, whereas in *Caenorhabditis elegans*, it is achieved by 2-fold lower transcriptional output from both X chromosomes in hermaphrodites. Flies adopt a different strategy; they double the transcriptional output of the single male X chromosome in somatic cells (6, 7), which requires the targeting of a male-specific lethal complex (MSL) to the X chromosome but not to autosomes in *Drosophila* males (5).

In *Drosophila melanogaster*, the MSL complex consists of proteins encoded by five genes: *male-specific lethal* genes, *msl1*, *msl2*, and *msl3*, *maleless* (*mle*) and *males absent on the first* (*mof*) (Fig. 1*A*), as well as two noncoding RNAs (roX1 and roX2). MSL1 and MSL2 play a central role in the assembly of the MSL complex and targeting to the X chromosome (Fig. 1*B*). It is believed that this targeting of the MSL complex enables MOF to specifically acetylate lysine-16 on histone H4 tails, a histone modification correlated with active transcription (8–10). High-resolution mapping of MSL-binding sites has revealed a strong bias for the middle and 3′ ends of coding sequences, suggesting that any transcriptional up-regulation may involve increased elongation efficiency by RNA polymerase (11–14). Indeed, experiments in *Saccharomyces cerevisiae* have shown that MOF recruitment (using a Gal4 DNA-binding domain fusion) results in transcriptional up-regulation (5, 10, 13). However, it has been suggested that the absence of the rest of the MSL complex in these experiments complicates the exact implication of this result regarding X chromosome up-regulation in *Drosophila* (15).

Which DNA sequences target the MSL complex specifically to the X chromosome? There are ≈35–40 high-affinity sites on the X chromosome that are bound by this complex (16, 17). In total, there are estimated to be 700 separable regions where the MSL complex is bound as shown by chromatin immunoprecipitation experiments. These regions cover roughly 25% of the X chromosome and, presumably, include sites ranging in affinity (11, 12, 18). These regions occur mostly in coding sequences of genes (transposable elements were not included in the arrays used in the analyses) and may be enriched in GAGA motifs (11, 12, 19). There is some debate about the role of active transcription in attracting or maintaining the MSL complex (for review, see ref. 13), but active transcription alone cannot explain the strong bias for binding to X chromosomal DNA. Computational sequence analysis can identify some "higher-order" features on X chromosomal sequences that distinguish them from autosomal DNA, yet extensive efforts at identifying common sequence predictors of MSL-binding sites have yielded limited prediction power at best (12, 20). These findings have led to the suggestion that degenerate and multiple weak signals may contribute to targeting (12, 19, 21). Arguably, comparative genomics methodology has limited applicability to this problem because MSL-binding sites have not been mapped in divergent *Drosophila* species or

Author contributions: M.A.R., D.V., and J.J.B. contributed equally to this work; D.V. and H.S.M. designed research M.A.R., D.V., J.J.B., and H.S.M. performed research; M.A.R., D.V., J.J.B., and H.S.M. analyzed data; and M.A.R., D.V., J.J.B., and H.S.M. wrote the paper.

The authors declare no conflict of interest.

Abbreviations: D–M model, Dobzhansky–Muller model; HKA test, Hudson–Kreitman–Aguade test; MK test, McDonald–Kreitman test; MLE, maleless; MOF, males absent on the first; MSL, male-specific lethal.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. EF630368–EF630425, EF634160–EF634184, EF653847–EF653867, and EF682039–EF682060).

[‡]Present address: Department of Genetics, Stanford University, Stanford, CA 94305.

[¶]To whom correspondence should be addressed. E-mail: hsmalik@fhcrc.org.

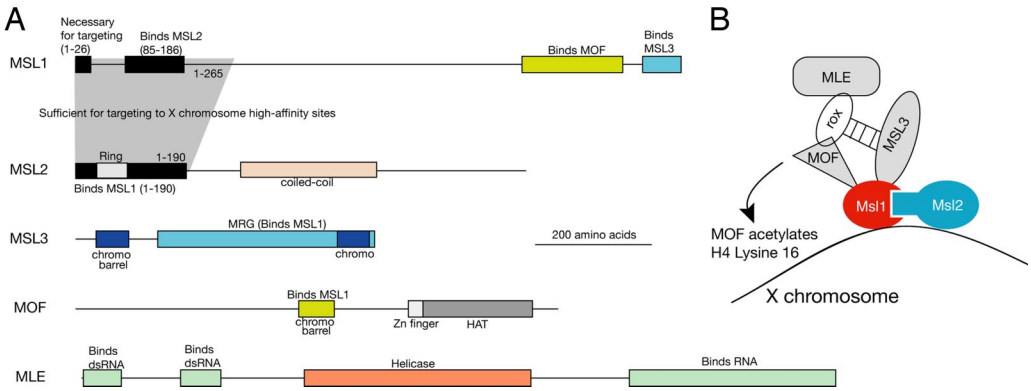

Fig. 1. MSL1 and MSL2 play a key role in assembly and targeting of the dosage compensation complex to the male X chromosome. (*A*) Five known protein components and two known RNAs comprising the MSL complex. The five MSL proteins are drawn to scale with known domains highlighted. MSL1 serves as a scaffold for the entire MSL complex. MSL1 binds to MSL2, and together they bind to the X chromosome. Amino acids 85–186 of MSL1 are necessary and sufficient for binding to amino acids 1–190 of MSL2. Together, amino acids 1–265 of MSL1 and amino acids 1–190 of MSL2 are sufficient for targeting to high-affinity binding sites on the male X chromosome (30, 33, 34). Targeting of MSL1 is abolished by deletion of amino acids 1–26 (32). MSL1 also binds to other components of the MSL complex, including MSL3 [which contains a chromobarrel domain (61) that binds RNA (62)] and MOF [which contains a chromobarrel domain (61), a zinc finger, and an acetyltransferase domain with specific activity for histone H4 (10, 63)]. MLE encodes ATPase and RNA/DNA helicase activities (64). (*B*) Schematic model of the assembly of the MSL complex onto the male X chromosome that highlights the central scaffolding role of MSL1 and MSL2.

even methodically in different *D. melanogaster* strains. In addition, genetic experiments have determined that any substantial segment of the X chromosome sequence appears to possess the ability to attract the MSL complex autonomously (22, 23). Although autosomal genes inserted onto the X chromosome will also frequently undergo dosage compensation, X chromosomal sequences are superior in their ability to recruit the MSL complex compared with autosomal sequences despite varying in their affinity for MSL recruitment (24). The questions remain: What is recruiting the MSL complex specifically to the male X, and why are these sequences so difficult to define?

We hypothesized that a MSL-binding site consensus is hard to define because these motifs might be evolutionarily labile. A selective pressure that prevented the stable coevolution of MSL proteins and DNA might have prevented the fixation of an optimal DNA sequence that could recruit the MSL complex. Such a scenario would preclude the identification of a consensus MSL-binding signature in the DNA. We explored this possibility by investigating the selective pressures shaping genes encoding MSL proteins as a "surrogate" to studying the MSL-binding sites themselves directly. We found strong evidence of positive selection acting on all five genes encoding protein components of the MSL complex. This finding is highly unexpected because MSL function is essential for male viability. We further found that the signature for rapid evolution is strikingly asymmetric, affecting *D. melanogaster* but (largely) not *Drosophila simulans*. Some of the strongest signatures of positive selection can be localized to the MSL domains responsible for X chromosomal targeting, suggesting that MSL-recruiting DNA segments may also have evolved rapidly in the *D. melanogaster* lineage, where all of the mapping studies have been done. Together, these findings also suggest that incompatibilities between MSL proteins and MSL-recruiting DNA elements on the *D. melanogaster* X chromosome may represent one of the underlying causes of male hybrid inviability in *Drosophila*.

Results

We sequenced all five protein-coding MSL complex genes from multiple strains of *D. melanogaster* and *D. simulans*, two species that diverged 2.5 million years ago. Summary statistics for polymorphisms seen in these genes are presented in Table 1. From these statistics, there is no evidence for a pattern of rare,

Table 1. Summary statistics for polymorphisms in MSL complex genes

Gene	Location	Codon usage (ENC)	Nucleotide diversity	Tajima's D	Fu and Li *F**	No. of strains	No. of bp
D. melanogaster							
msl1	2L (36F11–37A1)	56.8	0.00279	−0.262	−0.819	14	3,180
msl2	2L (23F3)	51.8	0.00961	−0.215	−0.426†	11	2,374
msl3	3L (65E4)	52	0.00879	−0.0993	0.435	10	1,890
mof	X (5C5)	48	0.00468	−0.114	0.033	13	2,484
mle	2R (42A6)	53.1	0.00206	−1.55	−2.58‡	14	5,447
D. simulans							
msl1		56.3	0.00691	−0.227	−0.658	14	3,195
msl2		50.7	0.00939	−0.0745	−0.340†	15	2,350§
msl3		52.2	0.01091	−0.0734	0.168	12	1,881
mof		47.1	0.00669	−0.549	0.0958	19	2,484
mle		53.5	0.00896	0.236	−0.0576	7	5,164

The Fu and Li *F* statistic was calculated by using either *D. simulans* or *D. melanogaster* as an outgroup species. ‡, $P < 0.05$.
†Seventy-five nucleotides were excluded from analysis because of one region between *D. melanogaster* and *D. simulans* with ambiguity in the alignment.
§Eighteen codons were excluded from analysis due to two polymorphic indels of 12 and 42 nucleotides within *D. simulans*.

Table 2. HKA tests on the MSL complex genes in *D. melanogaster*

Gene	Intraspecific: No. of segregating sites	Intraspecific: Total no. of sites	Intraspecific: Sample size†	Interspecies, total no. of differences*	*msl2*	*msl3*	*mle*	*mof*
msl1	29	3,168	14	162.63	0.117	**0.029**	0.742	0.117
msl2	61	2,245	11	150.51		0.560	**0.034**	0.982
msl3	48	1,880	10	86.14			**0.003**	0.545
mle	46	4,743	16	293.65				**0.020**
mof	37	2,484	13	112.36				

P values for all pairwise comparisons between the *D. melanogaster* MSL complex genes are shown with significant deviations from neutral expectations highlighted in bold.
**D. simulans* was used for the interspecies comparison.
†No. of *D. melanogaster* strains sequenced.

singleton polymorphisms that might suggest recovery of polymorphisms after a recent adaptive sweep in most of the MSL genes. However, we see a Tajima's D value of −1.55 and a Fu and Li value of −2.58 ($P < 0.05$) in the *mle* gene of *D. melanogaster*, strongly supporting the possibility that a recent sweep has affected the polymorphism spectrum (Table 1) (25, 26). To investigate this possibility further, we compared all of the MSL genes in *D. melanogaster* by using a Hudson–Kreitman–Aguade (HKA) test, which examines whether interspecies divergence and intraspecies polymorphisms are correlated, as would be predicted under neutrality (27). We find a significant discordance in the polymorphism patterns between *mle* and three other MSL genes, *msl2*, *msl3*, and *mof* (Table 2). Thus, both the Fu and Li F^* statistic and the HKA test results strongly implicate a recent adaptive sweep in the *mle* gene in *D. melanogaster*. None of the MSL genes shows a significantly discordant polymorphism pattern by the HKA test in *D. simulans*.

We also tested for positive selection by using the McDonald–Kreitman (MK) test (Table 3). This test evaluates whether an excess number of replacement (amino acid altering) changes versus synonymous changes had been fixed between the two species compared with replacement and synonymous polymorphisms within each species (28). Under this test, we find that four of five MSL genes, *msl1*, *msl2*, *msl3*, *mof*, but not *mle*, show robust signatures of positive selection when compared across the whole gene. This finding is highly unusual for essential genes because a high degree of evolutionary constraint is expected to act to preserve function. Using *Drosophila yakuba* as an outgroup species, we can ascertain which lineage has been affected by positive selection by assigning the fixed changes to either the *D.*

Table 3. MK test for positive selection on MSL complex genes

MSL complex protein-encoding gene	Observed Sp	Observed Rp	Observed Sf	Observed Rf	*G* value	*P* value
msl1						
Pooled	54	41	56	78	5.025	**0.025**
D. melanogaster	18	10	24	53	9.097	**0.0026**
D. simulans	36	32	23	22	0.036	0.85
*msl2**						
Pooled	83	40	51	63	12.47	**0.00041**
D. melanogaster	36	22	24	41	7.743	**0.0054**
D. simulans	47	18	22	18	3.201	0.074
msl3						
Pooled	49	10	22	19	9.913	**0.00164**
D. melanogaster	24	3	11	15	13.23	**0.00028**
D. simulans	24	6	7	3	0.379	0.54
mof						
Pooled	84	19	48	37	13.97	**0.00019**
D. melanogaster	30	7	25	23	7.86	**0.0051**
D. simulans	55	12	19	13	5.531	**0.012**
mle						
Pooled	50	35	75	48	0.096	0.76
D. melanogaster	14	16	37	24	1.562	0.21
D. simulans	36	19	29	21	0.608	0.44

We compared the ratio of replacement and synonymous changes that were polymorphic within the species (Rp:Sp) with the ratio of replacement and synonymous changes that were found fixed between the species (Rf:Sf). If no alteration in selective regimes occurred during evolution, we expect these ratios to be statistically indistinguishable. However, an excess of fixed replacement changes is a clear indication of positive selection. Pooled changes refer to all the polymorphism and fixed changes using both lineages. However, by using an outgroup species (*D. yakuba*), we can also make the same comparison, specific either to the *D. melanogaster* or *D. simulans* lineage.
*A region of 75 nucleotides was excluded from all *msl2* analysis because of ambiguity in the alignment between *D. simulans* and *D. melanogaster* (the region corresponds to nucleotides 904–978 relative to *D. melanogaster*).

melanogaster or *D. simulans* lineages. We find that there is robust evidence for positive selection acting on the *D. melanogaster* lineage for four of five MSL genes (all except *mle*) by using the MK test. Interestingly, *mle* is the only gene that appears to have undergone a recent sweep in the *D. melanogaster* species (Tables 1 and 2) and has likely reduced our ability to detect positive selection by using the MK test on this gene. In contrast, we find that only one of five genes (*mof*) has been subjected to positive selection along the *D. simulans* lineage; even in this case, the intensity of positive selection is weaker than in the *D. melanogaster* lineage. Also, of the five MSL genes analyzed here, *mof* is the only one that is also expressed robustly in females (8), although the functional significance of MOF function in females is still unclear.

The *msl1* and *msl2* genes are key players in targeting the MSL complex to the male X chromosome. Mutational analyses of each MSL gene have shown that MSL1 and MSL2 are capable of targeting "high-affinity" sites, independent of other known MSL components (29–31). Targeting requires an interaction between the N-terminal domains of MSL1 and MSL2 (Fig. 1) and is abolished by deletion of the first 26 amino acids of MSL1 (30, 32–34). Because these targeting domains have been roughly mapped, we next addressed whether positive selection had shaped these regions in particular, focusing our analyses only on the *D. melanogaster* lineage. If we parse the fixed or polymorphic changes that have taken place in MSL1 in the *D. melanogaster* lineage, we find that the N-terminal domain (amino acids 1–265), which is necessary and sufficient for both X chromosomal targeting and for interactions with MSL2 (30, 32–34), bears all of the hallmarks of positive selection (Rf:Sf::Rp:Sp = 30:6::5:8, G value 8.406, $P < 0.005$). In contrast, the remainder of the MSL1 protein [amino acids 266-1039, which includes interaction interfaces with both MOF and MSL3 (33)] shows no evidence for positive selection (Rf:Sf::Rp:Sp = 23:18::5:10, G value 2.229, $P > 0.15$).

Similarly, when we parse our fixed and polymorphic changes for the MSL2 protein, we find that the domain required for binding MSL1 and thereby, targeting to DNA (amino acids 1–190) (34) evolves under positive selection in *D. melanogaster* (Rf:Sf::Rp:Sp = 9:9::1:10, $P < 0.025$). In contrast to MSL1, however, even the remainder of the MSL2 protein, which consists of a central coiled-coil domain and a C-terminus of as yet-undefined function (Fig. 1), shows a robust signature of positive selection (Rf:Sf::Rp:Sp = 32:15::21:26, G value 5.203, $P < 0.025$).

Thus, the N-terminal domains of both MSL1 and MSL2 are hot spots for positive selection, despite the fact that these domains are essential for the MSL1–MSL2 interaction both with each other and to binding sites on the X chromosome. Our findings support the idea that the rules that guide X chromosomal sequence-directed DNA binding have been evolutionarily labile, at least in the *D. melanogaster* lineage.

Discussion

Rapid evolution of the MSL complex is at odds with the expectation that proteins so essential for male viability ought to be highly constrained and under purifying selection. Moreover, interactions between MSL proteins and their cognate DNA-binding sites should be especially well constrained because any mutations in one MSL component would have to be accommodated in the other MSL proteins and in the DNA target sites to retain the essential function of the complex. Indeed, it is reasonable to speculate that the selective force that drove the rapid evolution must have imposed a stringent selective cost, which would drive changes in the whole MSL complex.

Male-killing bacteria provide an example of just such a selective cost. For instance, *Spiroplasma poulsonii* specifically kill male *D. melanogaster* flies, as they are transmitted exclusively through females. Recent studies have directly implicated the presence of a functional MSL complex as a requirement for this male-specific killing by *S. poulsonii* (35). Under such a "genetic conflict" scenario, one could imagine bacterial proteins evolving to "detect" MSL components through direct binding, whereas MSL components could be under strong selective pressure to evolve away from this recognition. This "arms race" would result in changes in one or all of the MSL components because fixation of slightly deleterious mutations in the MSL complex would be preferred over bacteria-induced male lethality.

A second possible source driving positive selection of the MSL complex could be genetic conflict with retrotransposable elements. It has been suggested previously that LINE1 non-LTR retrotransposons may provide "landing sites" for dosage compensation (X inactivation) in mammalian X chromosomes (36). Under this second possibility, MSL binding to retrotransposons may be an important defense against them (37). Repeated specialization of the MSL complex to recognize retroelements may also result in some of these elements becoming preferred landing sites for the MSL complex, effectively altering the landscape of MSL binding to the X chromosome. It is important to note, however, that retrotransposons primarily mobilize in the germ line, whereas the MSL proteins discussed here are acting predominantly in somatic tissues.

Both of these conflict scenarios fit well with our finding of highly asymmetric positive selection because either the male-killing bacteria or retrotransposons may provide a lineage-specific selective pressure, not affecting even closely related species. It has been suggested (38) that because the relative stoichiometries of regulator proteins are so intricately linked to each other, rapid evolution of any one component driven by genetic conflict under any model [by male-killing bacteria or retrotransposons or even by "centromere drive" (39)] could inevitably trigger a "ripple effect of adaptation" in other MSL genes. Each such alteration would trigger a coevolutionary episode in which other target genes and regulator proteins would adjust to the changed landscape to ensure optimal function (38). Although it is unlikely that a single ripple event can explain the pervasive positive selection we have seen in multiple domains of all MSL proteins, selection on one member of a complex might bring along changes in other members at any number of domains if that produces an eventual fitness advantage by restoring optimal function.

One possible consequence of such rapid evolution is that MSL components may quickly become incompatible in different species. Such incompatibilities are thought to occur under a Dobzhansky–Muller (D–M) model wherein independently occurring allelic changes in different interacting components could manifest as negative epistasis in resulting hybrids (the simplest two-locus form is schematized in Fig. 2*A*). The possibility of this negative epistasis is greatly increased with accelerated change; genes responsible for postzygotic isolation are frequently subject to positive selection (40–42). Rapid evolution can also result in asymmetric postzygotic isolation (Fig. 2*B*). There are two formal possibilities for D–M incompatibilities arising from the positive selection of the genes encoding MSL1 and MSL2. The first possibility is that the two components could represent the MSL1 and MSL2 proteins themselves (Fig. 2*C*) because it is the protein–protein interaction surface between these two proteins that is a hot spot for positive selection. However, recognizing that these domains also determine the exclusive targeting to the X chromosome, the D–M incompatibility could exist between the MSL1–MSL2 proteins, and MSL-targeting sites on the X chromosome (Fig. 2*D*). Either of these incompatibilities would lead to compromised MSL function and thereby male inviability in interspecies hybrids.

Intriguingly, male hybrids have different outcomes in a cross between *D. melanogaster* and *D. simulans*, depending on which

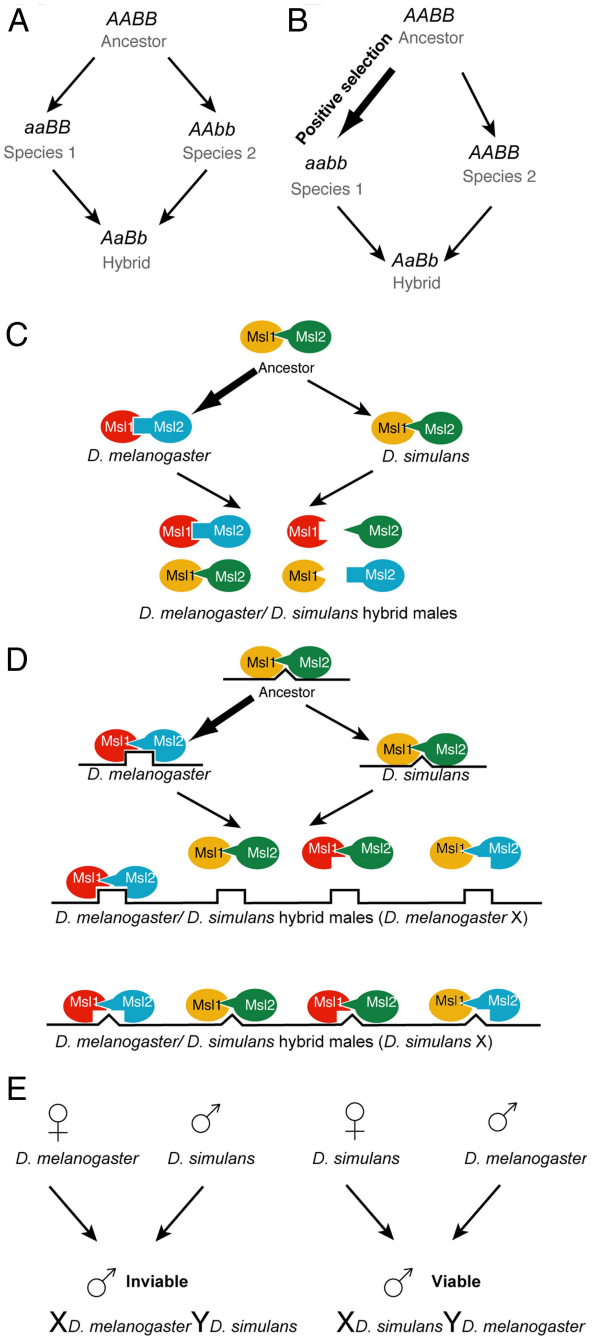

Fig. 2. Positive selection of the MSL complex and X chromosomal MSL-binding sites might result in hybrid incompatibility. (*A*) Two-locus D–M model for hybrid incompatibility between closely related species. Loci A and B interact in the ancestral species. During (reproductive or recombinational) isolation, there is a neutral fixation of the *a* and *b* alleles in the two populations, which is tolerated because the new alleles (*a* and *b*) are still compatible with the old alleles (*B* and *A*, respectively). However, this fixation results in hybrid incompatibility because of negative epistatic interactions between the new *a* and *b* alleles. This model can explain the onset of incompatibilities even under neutral evolution (56). (*B*) In the case of positive selection (bold arrows) driving the interaction of the *A* and *B* loci, only one lineage may evolve to the new *a* and *b* alleles, resulting in incompatibility with the other lineage, which still preserves the ancestral *A* and *B* alleles. (*C*) MSL1 and MSL2 could represent the *A* and *B* loci in the D–M model, with the positive selection (bold arrows) at their interaction interface resulting in hybrid inviability. Under this model, male hybrids containing either the *D. melanogaster* or *D. simulans* X chromosomes would be inviable because the protein composition is expected to be the same in both cases. (*D*) Model for hybrid incompatibility with MSL1–MSL2 and the X chromosomal MSL-binding sites, as *A* and *B* loci, respectively. Positive selection (bold arrows) in *D. melanogaster* has resulted in rapid evolution of the MSL1–MSL2 genes and (we infer) the X chromosomal MSL-binding sites. In male hybrids, *D. simulans* MSL1 and MSL2 are unable to recognize ''newly evolved'' MSL-binding sites on the *D. melanogaster* X chromosome resulting in mislocalization of the MSL complex in hybrids with a *D. melanogaster* X chromosome (44). However, hybrids with a *D. simulans* X chromosome localize the MSL complex normally (45) because the *D. melanogaster* MSL1 and MSL2 proteins retain an ancestral DNA-binding ability. (*E*) Known male hybrid incompatibility in *D. melanogaster* crosses to *D. simulans*. Male inviability occurs when a *D. melanogaster* X chromosome is combined with a hybrid autosomal background (43).

X chromosome is inherited in the hybrid males (Fig. 2*E*). Hybrid males that inherit the *D. simulans* X chromosome are viable. However, hybrid males that inherit the *D. melanogaster* X chromosome suffer larval lethality, dying at a developmental stage similar to that of pure-species *D. melanogaster* males that have mutated MSL components (43, 44). Because the *msl1* and *msl2* genes reside on autosomes (Table 1), both hybrids should acquire both *D. simulans* and *D. melanogaster* versions of both genes. The viability of the *D. simulans* X-bearing hybrid males implies that an incompatibility between MSL1 and MSL2 proteins (Fig. 2*B*) is not likely to be causal for hybrid male inviability. Instead, our findings suggest that although the MSL1–MSL2 interaction is not severely affected in hybrids, hybrid inviability may result from negative epistasis between the *D. melanogaster* X chromosome and *D. simulans* MSL components (Fig. 2*C*). Because most of the MSL positive selection has occurred along the *D. melanogaster* lineage, *D. simulans* MSL proteins may not have the requisite changes for the correct targeting to the *D. melanogaster* X. In support of this idea, a recent study has found that MSL components do not correctly target in *D. melanogaster* X-bearing hybrid males (44). In contrast, MSL localization and function are known to be normal in *D. simulans* X-bearing hybrid males (44, 45).

The genetic dissection of the determinants of postzygotic isolation in *Drosophila* has been greatly aided by the discovery of hybrid rescue genes, so called because mutations in these genes restore hybrid viability. (MSL components are not expected to be hybrid rescue genes because compromised MSL function would be strongly deleterious.) Hybrid males that inherit a *D. melanogaster* X chromosome and are otherwise inviable, can be rescued by a naturally occurring mutation in the *Lhr* (*lethal hybrid rescue*) gene (43). A recent study firmly establishes that the *Lhr* gene has a heterochromatic localization (41). This study is especially noteworthy because several heterochromatin proteins and remodeling factors have been directly implicated in transcriptional regulation of the male X chromosome (46–49). Indeed, it is possible that other hybrid incompatibility factors mapped in this cross may arise from defects in dosage compensation. For instance, the intriguing finding that nuclear pore complex proteins cause hybrid male lethality (40) could be viewed in light of recent findings connecting them to the MSL proteins (50). It is likely that a balance of "negative" dosage regulators and "positive" MSL proteins is required to achieve the correct level of transcription in *Drosophila* hybrid males (51). Such nonadditive expression phenotypes have been observed in hybrids of *D. melanogaster* and *D. simulans* (52) with an apparent overabundance of misregulated genes on the X chromosome (53).

The generality of Haldane's rule, wherein it is more likely that the heterogametic sex will be inviable, has led to several attractive theories about how hybrid inviability could represent a breakdown in dosage compensation in hybrids (54, 55). Indeed, it has been clear for quite some time that the X chromosome plays a disproportionate role in hybrid incompatibilities, referred to as the "large-X" effect (56). Until recently, it has not been clear that D–M incompatibilities could arise in such systems because they are so essential for function and thereby predicted

to evolve under a high degree of constraint. However, our present analysis on MSL complex genes suggests that such genes can and do evolve rapidly, which implies that even genes that participate in essential chromatin functions such as dosage compensation (as described here), chromosome segregation (57), and defining origins of DNA replication (58) are not immune from being called to participate in genetic conflict and adaptation. Indeed, D–M incompatibilities arising because of rapid evolution of these essential protein–DNA interactions are more likely to result in hybrid inviability and sterility rather than incompatibilities between two proteins that carry out a nonessential role in either species.

Methods

All *Drosophila* strains were obtained from the Species Stock Center (Tucson, AZ) except for the African isofemale lines that were a gift from Daven Presgraves (University of Rochester, Rochester, NY). Genomic DNA was prepared as described previously (57). Genes were amplified by using PCR Supermix High Fidelity (Invitrogen, Carlsbad, CA) and primers based on *D. melanogaster* genomic sequence. Most PCR products were sequenced directly except in the case of *mle* for some *D. simulans* strains. When direct sequencing of PCR products was not possible because of a low yield of PCR products, these products were cloned by using Topo-TA vectors (Invitrogen), and sequencing was done on at least three separate colonies. ClustalX (59) was used to obtain multiple alignments, which were subsequently hand-edited with the amino acid sequence as a guide. The DNASP software package (60) was used to perform several tests for positive selection, including the Tajima's D (25) and Fu and Li tests (26), as well as the MK (28) and the HKA (27) tests.

We thank the *Drosophila* Species Center and Daven Presgraves for the various *Drosophila* strains used in this work, and Jim Birchler, Nels Elde, Julie Kerns, Eric Smith and an anonymous reviewer for comments on the manuscript. This work was supported by National Institutes of Health (NIH) Grant GM074108 (to H.S.M.) and a Searle Scholar Award (to H.S.M.). J.J.B. was supported by NIH Training Grant PHS NRSA T32 GM07270.

1. Lindsley DL, Sandler L, Baker BS, Carpenter AT, Denell RE, Hall JC, Jacobs PA, Miklos GL, Davis BK, Gethmann RC, *et al*. (1972) *Genetics* 71:157–184.
2. Birchler JA, Yao H, Chudalayandi S (2007) *Biochim Biophys Acta* 1769:422–428.
3. Gupta V, Parisi M, Sturgill D, Nuttall R, Doctolero M, Dudko OK, Malley JD, Eastman PS, Oliver B (2006) *J Biol* 5:3.
4. Nguyen DK, Disteche CM (2006) *Nat Genet* 38:47–53.
5. Straub T, Becker PB (2007) *Nat Rev Genet* 8:47–57.
6. Hamada FN, Park PJ, Gordadze PR, Kuroda MI (2005) *Genes Dev* 19:2289–2294.
7. Straub T, Gilfillan GD, Maier VK, Becker PB (2005) *Genes Dev* 19:2284–2288.
8. Hilfiker A, Hilfiker-Kleiner D, Pannuti A, Lucchesi JC (1997) *EMBO J* 16:2054–2060.
9. Bone JR, Lavender J, Richman R, Palmer MJ, Turner BM, Kuroda MI (1994) *Genes Dev* 8:96–104.
10. Akhtar A, Becker PB (2000) *Mol Cell* 5:367–375.
11. Alekseyenko AA, Larschan E, Lai WR, Park PJ, Kuroda MI (2006) *Genes Dev* 20:848–857.
12. Gilfillan GD, Straub T, de Wit E, Greil F, Lamm R, van Steensel B, Becker PB (2006) *Genes Dev* 20:858–870.
13. Schubeler D (2006) *Genes Dev* 20:749–753.
14. Mito Y, Henikoff JG, Henikoff S (2005) *Nat Genet* 37:1090–1097.
15. Pal-Bhadra P, Bhadra U, Kundu J, Birchler JA (2005) *Genetics* 169:2061–2074.
16. Kelley RL, Meller VH, Gordadze PR, Roman G, Davis RL, Kuroda MI (1999) *Cell* 98:513–522.
17. Meller VH, Gordadze PR, Park Y, Chu X, Stuckenholz C, Kelley RL, Kuroda MI (2000) *Curr Biol* 10:136–143.
18. Legube G, McWeeney SK, Lercher MJ, Akhtar A (2006) *Genes Dev* 20:871–883.
19. Dahlsveen IK, Gilfillan GD, Shelest VI, Lamm R, Becker PB (2006) *PLoS Genet* 2:e5.
20. Stenberg P, Pettersson F, Saura AO, Berglund A, Larsson J (2005) *BMC Bioinformatics* 6:158.
21. Gilfillan GD, König C, Dahlsveen IK, Prakoura N, Straub T, Lamm R, Fauth T, Becker PB (2007) *Nucleic Acids Res* 35:3561–3572
22. Fagegaltier D, Baker BS (2004) *PLoS Biol* 2:e341.
23. Oh H, Bone JR, Kuroda MI (2004) *Curr Biol* 14:481–487.
24. Demakova OV, Kotlikova IV, Gordadze PR, Alekseyenko AA, Kuroda MI, Zhimulev IF (2003) *Chromosoma* 112:103–115.
25. Tajima F (1989) *Genetics* 123:585–595.
26. Fu YX, Li WH (1993) *Genetics* 133:693–709.
27. Hudson RR, Kreitman M, Aguade M (1987) *Genetics* 116:153–159.
28. McDonald JH, Kreitman M (1991) *Nature* 351:652–654.
29. Gu W, Szauter P, Lucchesi JC (1998) *Dev Genet* 22:56–64.
30. Lyman LM, Copps K, Rastelli L, Kelley RL, Kuroda MI (1997) *Genetics* 147:1743–1753.
31. Palmer MJ, Richman R, Richter L, Kuroda MI (1994) *Genes Dev* 8:698–706.
32. Li F, Parry DA, Scott MJ (2005) *Mol Cell Biol* 25:8913–8924.
33. Scott MJ, Pan LL, Cleland SB, Knox AL, Heinrich J (2000) *EMBO J* 19:144–155.
34. Copps K, Richman R, Lyman LM, Chang KA, Rampersad-Ammons J, Kuroda MI (1998) *EMBO J* 17:5409–5417.
35. Veneti Z, Bentley JK, Koana T, Braig HR, Hurst GD (2005) *Science* 307:1461–1463.
36. Lyon MF (2000) *Proc Natl Acad Sci USA* 97:6248–6249.
37. McDonald JF, Matzke MA, Matzke AJ (2005) *Cytogenet Genome Res* 110:242–249.
38. Birchler JA, Veitia RA (2007) *Plant Cell* 19:395–402.
39. Malik HS, Bayes JJ (2006) *Biochem Soc Trans* 34:569–573.
40. Presgraves DC, Balagopalan L, Abmayr SM, Orr HA (2003) *Nature* 423:715–719.
41. Brideau NJ, Flores HA, Wang J, Maheshwari S, Wang X, Barbash DA (2006) *Science* 314:1292–1295.
42. Barbash DA, Siino DF, Tarone AM, Roote J (2003) *Proc Natl Acad Sci USA* 100:5302–5307.
43. Sawamura K, Watanabe TK, Yamamoto MT (1993) *Genetica* 88:175–185.
44. Pal-Bhadra M, Bhadra U, Birchler JA (2006) *Genetics* 174:1151–1159.
45. Lakhotia SC, Mishra A, Sinha P (1981) *Chromosoma* 82:229–236.
46. Spierer A, Seum C, Delattre M, Spierer P (2005) *J Cell Sci* 118:5047–5057.
47. de Wit E, Greil F, van Steensel B (2005) *Genome Res* 15:1265–1273.
48. Corona DF, Clapier CR, Becker PB, Tamkun JW (2002) *EMBO Rep* 3:242–247.
49. Liu LP, Ni JQ, Shi YD, Oakeley EJ, Sun FL (2005) *Nat Genet* 37:1361–1366.
50. Mendjan S, Taipale M, Kind J, Holz H, Gebhardt P, Schelder M, Vermeulen M, Buscaino A, Duncan K, Mueller J, *et al*. (2006) *Mol Cell* 21:811–823.
51. Birchler JA, Riddle NC, Auger DL, Veitia RA (2005) *Trends Genet* 21:219–226.
52. Ranz JM, Namgyal K, Gibson G, Hartl DL (2004) *Genome Res* 14:373–379.
53. Barbash DA, Lorigan JG (2007) *J Exp Zool* 308:74–84.
54. Forsdyke DR (1995) *J Theor Biol* 172:335–345.
55. Orr HA (1989) *Genetics* 122:891–894.
56. Coyne JA, Orr HA (2004) *Speciation* (Sinauer Associates, Sunderland, MA).
57. Malik HS, Henikoff S (2001) *Genetics* 157:1293–1298.
58. Wiggins BL, Malik HS (2007) *Fly* 1:155–163.
59. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) *Nucleic Acids Res* 25:4876–4882.
60. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) *Bioinformatics* 19:2496–2497.
61. Nielsen PR, Nietlispach D, Buscaino A, Warner RJ, Akhtar A, Murzin AG, Murzina NV, Laue ED (2005) *J Biol Chem* 280:32326–32331.
62. Morales V, Regnard C, Izzo A, Vetter I, Becker PB (2005) *Mol Cell Biol* 25:5947–5954.
63. Smith ER, Pannuti A, Gu W, Steurnagel A, Cook RG, Allis CD, Lucchesi JC (2000) *Mol Cell Biol* 20:312–318.
64. Lee CG, Chang KA, Kuroda MI, Hurwitz J (1997) *EMBO J* 16:2671–2681.