

Note

Pervasive and Largely Lineage-Specific Adaptive Protein Evolution in the Dosage Compensation Complex of *Drosophila melanogaster*

Mia T. Levine,¹ Alisha K. Holloway, Umbreen Arshad and David J. Begun

Section of Evolution and Ecology, University of California, Davis, California 95616

Manuscript received July 25, 2007

Accepted for publication September 18, 2007

ABSTRACT

Dosage compensation refers to the equalization of X-linked gene transcription among heterogametic and homogametic sexes. In *Drosophila*, the dosage compensation complex (DCC) mediates the twofold hypertranscription of the single male X chromosome. Loss-of-function mutations at any DCC protein-coding gene are male lethal. Here we report a population genetic analysis suggesting that four of the five core DCC proteins—MSL1, MSL2, MSL3, and MOF—are evolving under positive selection in *D. melanogaster*. Within these four proteins, several domains that range in function from X chromosome localization to protein–protein interactions have elevated, *D. melanogaster*-specific, amino acid divergence.

THE *Drosophila* core dosage compensation complex (DCC) is composed of five proteins and two noncoding RNA transcripts that complex to mediate the twofold hypertranscription of the single male X chromosome, largely through chromatin remodeling (LUCCHESI *et al.* 2005). Loss-of-function mutations in *mSl1*, *mSl2*, *mSl3*, *mof*, or *mle* result in male-specific lethality (BELOTE and LUCCHESI 1980; SKRIPSKY and LUCCHESI 1982; HILFIKER *et al.* 1997).

These five proteins are evolutionarily unrelated and functionally distinct. MOF acetylates lysine 16 of histone 4 (BONE *et al.* 1994; HILFIKER *et al.* 1997), a histone mark known to stimulate transcription (IKEDA *et al.* 1999). Two other components, MSL1 and MSL2, are thought to scaffold the complex to the X chromosome (LYMAN *et al.* 1997; SCOTT *et al.* 2000), while MSL3 bound to MSL1 stimulates histone acetylation induced by MOF (MORALES *et al.* 2004). Finally, MLE interacts with at least one of the two noncoding RNAs associated with the DCC complex—roX1 and roX2—likely guiding these RNAs to the DCC and reinforcing their integration into the complex (COPPS *et al.* 1998). Other proteins such as JIL-1 (JIN *et al.* 2000) and the NURF complex components (BAI *et al.* 2007) interact with the DCC but also are known to engage in several other biological processes. We therefore refer to MSL1, MSL2, MSL3, MOF, and MLE as the “core” DCC proteins. We note that these five

proteins are also considered core DCC components in an alternative dosage compensation model posited by BIRCHLER *et al.* (2003).

The DCC components are coded by essential genes that are conserved across *Drosophila* and at least four other drosophilid genera (MARIN *et al.* 1996). Nevertheless, a genome scan of *D. melanogaster*–*D. simulans* divergence revealed several candidate Gene Ontology terms, including “dosage compensation,” that were enriched for genes with unusual patterns of divergence (BEGUN *et al.* 2007). However, those data were insufficient for inferring the population genetic mechanism, which motivated the detailed *D. melanogaster* polymorphism survey presented here. Our population genetic analysis supports the hypothesis that four of the five core DCC proteins are evolving adaptively, and surprisingly, this pattern is largely restricted to the *D. melanogaster* lineage.

We found a consistent and significant excess of amino acid fixations at *mSl1*, *mSl2*, *mSl3*, and *mof* along the *D. melanogaster* branch (Table 1), suggesting that recurrent directional selection is an important force driving rapid evolution at these four loci in *D. melanogaster*. We found no evidence of recurrent adaptive protein evolution for *mle*; however, limited statistical power due to low levels of polymorphism at this locus makes our inability to reject neutrality difficult to interpret. Reduced polymorphism at *mle* may be explained by its location near a region of reduced crossing over at the 2R centromere. In *D. simulans*, only *mof* exhibited evidence of adaptive protein evolution (Table 1). These data suggest that adaptive protein divergence of DCC components is

¹Corresponding author: Section of Evolution and Ecology, University of California, 1 Shields Ave., Davis, CA 95616.
E-mail: mialevine@ucdavis.edu

TABLE 1

Polymorphism and divergence along the *D. melanogaster* branch, divergence along the *D. simulans* branch, and results of *D. melanogaster* McDonald–Kreitman tests for each of the five DCC genes

| Gene | π (N/S) ^a | n_{sim} | mel/sim ^a | | <i>D. melanogaster</i> branch | | | | | <i>D. simulans</i> branch | | | | |
|--------------|--------------------------|-----------|----------------------|-------------|-------------------------------|------------------|-------------------|------------------|------------------|---------------------------|------------------|-------------------|------------------|------------------|
| | | | d_N | d_S | N _{poly} | N _{fix} | S _{poly} | S _{fix} | FET ^b | N _{poly} | N _{fix} | S _{poly} | S _{fix} | FET ^b |
| <i>mle</i> | 0.000/0.000 | NA | 0.010/0.010 | 0.047/0.054 | 1 | 18 | 3 | 33 | 1.00 | NA | NA | NA | NA | NA |
| <i>mssl1</i> | 0.001/0.006 | 4 | 0.024/0.015 | 0.039/0.060 | 3 | 34 | 8 | 19 | 0.042 | 11 | 13 | 21 | 24 | 1.00 |
| <i>mssl2</i> | 0.003/0.021 | 4 | 0.026/0.020 | 0.067/0.093 | 8 | 24 | 21 | 17 | 0.010 | 23 | 11 | 34 | 16 | 1.00 |
| <i>mssl3</i> | 0.002/0.027 | 13 | 0.018/0.004 | 0.083/0.060 | 1 | 10 | 22 | 12 | 0.002 | 5 | 2 | 31 | 4 | 0.25 |
| <i>mof</i> | 0.000/0.005 | 13 | 0.013/0.009 | 0.086/0.071 | 0 | 17 | 8 | 30 | 0.048 | 5 | 10 | 32 | 5 | 0.03 |

π , nucleotide diversity; N, nonsynonymous; S, synonymous; n_{sim} , *D. simulans* sample size (for *D. melanogaster*, $n = 10$ for all); d , divergence; poly, polymorphism; fix, fixations; FET, Fisher's exact test *P*-value).

^aTen inbred lines of *D. melanogaster* (North Carolina, T. Mackay) were sequenced directly to obtain polymorphism data from each of the five core protein-coding loci in the DCC. For *D. simulans*, sequences were obtained from a population genomic data set, which consists of light-shotgun sequencing (one to two times coverage) of six lines syntenically aligned to *D. melanogaster* v4 assembly, and were used for population data for all five DCC genes (BEGUN *et al.* 2007). These *D. simulans* data were subjected to extensive quality control (<http://www.dpgp.org>). Levels of nucleotide diversity (π) were estimated as in NEI (1987) and WEIR (1990). The numbers of silent and replacement sites were estimated using the method of NEI and GOJOBORI (1986). The pathway between two codons was calculated as the average number of silent and replacement changes from all possible paths between the pair. Lineage-specific divergence was estimated by maximum likelihood using PAML v3.14 (YANG 1997). PAML was run in batch mode using a BioPerl wrapper (STAJICH *et al.* 2002) for codeml with codon frequencies estimated from the data. Sequence data for this article have been submitted to GenBank under accession nos. EU167087–EU167150.

^bTo test for adaptive protein evolution, we used the McDonald–Kreitman test (MCDONALD and KREITMAN 1991), which determines whether the number of fixations relative to polymorphisms for synonymous and nonsynonymous sites deviates from neutral expectations. We used parsimony to infer whether a fixation occurred on the *D. melanogaster* or the *D. simulans* branch, restricting our attention to codons that varied at a single position among the three species. The syntenic alignment of *D. yakuba* was used to polarize changes to the *D. melanogaster* or *D. simulans* branch (BEGUN *et al.* 2007). For *mssl3* and *mof*, which had borderline significant McDonald–Kreitman tests for the *D. simulans* lineage, 10 additional inbred lines were sequenced (Winters, CA, S. Nuzhdin).

considerably more common in *D. melanogaster* than in *D. simulans*. An attempt to use PAML to understand long-term patterns of molecular evolution was thwarted by unreliable alignments among *Drosophila* species outside the *D. melanogaster* subgroup, which raises the possibility of rapid evolution on other distantly related lineages.

We used finer-scale analysis to gain biological insight into particular protein functions that might be under selection (Table 2). No known protein domains of MOF or MSL2 exhibited evidence of rapid evolution; in fact, the nonsynonymous divergence for these domains was relatively low compared to genomic averages (BEGUN *et al.* 2007). However, the MSL1 glycine-rich domain, which is thought to be involved in MSL1–MSL1 interactions (LI *et al.* 2005), as well as the MSL1 X chromosome-binding domain, have unusually high d_N . The MSL1 leucine zipper/apolar/coiled-coil domain, essential for MSL2 binding as well as X chromosome localization (LI *et al.* 2005), and the polar region of MSL3, which is thought to cooperate with the chromatin-binding domain of MSL3 (BUSCAINO *et al.* 2006), also had elevated d_N .

The four proteins at which adaptive evolution was inferred in *D. melanogaster* interact directly with at least one other member of the complex, which suggests the possibility of a history of coevolutionary interactions within this single complex; recurrent directional selec-

tion of one or more members of the complex may drive compensatory evolution in other members. Protein domains known to be involved in protein self-association, localization to the X chromosome-bound complex, intra-protein interactions, and interprotein interactions were found to be evolving particularly rapidly. These findings are consistent with the hypothesis of within-complex and even within-protein coevolutionary interactions and suggest protein domains where adaptive evolution is concentrated. These domains are excellent candidates for investigating DCC functional divergence between *D. melanogaster* and *D. simulans*.

The nature of DCC functional divergence between these sister species depends on the biological or molecular function targeted by directional selection. Adaptive protein evolution at *mssl1*, *mssl2*, *mssl3*, and *mof* may result from selection of DCC function itself or of a dosage-independent function carried out by either a DCC protein or a non-DCC protein undergoing adaptive evolution that interacts with a dosage protein. For example, the nuclear pore protein, NUP153, interacts with MSL3 and MOF (MENDJAN *et al.* 2006) and exhibits patterns of adaptive protein evolution along both the *D. melanogaster* and *D. simulans* lineages (PRESGRAVES and STEPHAN 2007). However, given the observation that NUP153 has been under directional selection in both *D. melanogaster* and *D. simulans*, it is not clear why the putative coevolutionary interactions among the core

TABLE 2

Polymorphism and divergence along the *D. melanogaster* and *D. simulans* branches by previously described domains for the four DCC genes exhibiting adaptive protein evolution

| Gene | Domain | Codon position | <i>D. melanogaster</i> branch | | | | | <i>D. simulans</i> branch | | |
|-------------|-----------------------|----------------|-------------------------------|-----------|-------|-------|-----------|---------------------------|-------|-----------|
| | | | π (N) | π (S) | d_N | d_S | d_N/d_S | d_N | d_S | d_N/d_S |
| <i>mle</i> | RNA binding | 2–70 | 0.0000 | 0.0000 | 0.019 | 0.040 | 0.48 | 0.000 | 0.083 | 0.00 |
| | Helicase/ATP binding | 394–560 | 0.0000 | 0.0020 | 0.011 | 0.014 | 0.79 | 0.008 | 0.069 | 0.12 |
| | C-terminal helicase | 639–813 | 0.0000 | 0.0000 | 0.005 | 0.290 | 0.19 | 0.003 | 0.022 | 0.12 |
| <i>mof</i> | Extradomain | | 0.0002 | 0.0006 | 0.001 | 0.046 | 0.21 | 0.016 | 0.059 | 0.28 |
| | Chromodomain | 395–450 | 0.0000 | 0.0100 | 0.000 | 0.180 | 0.00 | 0.010 | 0.019 | 0.51 |
| | HAT-Zn | 575–760 | 0.0000 | 0.0000 | 0.002 | 0.090 | 0.03 | 0.002 | 0.110 | 0.02 |
| <i>msl1</i> | Extradomain | | 0.0040 | 0.0060 | 0.027 | 0.072 | 0.26 | 0.012 | 0.057 | 0.21 |
| | X-binding | 1–26 | NA | NA | 0.080 | 0.065 | 1.3 | 0.000 | 0.000 | 0.00 |
| | Glycine rich | 26–74 | 0.0000 | 0.0000 | 0.070 | 0.050 | 1.5 | 0.019 | 0.012 | 0.26 |
| <i>msl2</i> | LZ/apolar/coiled coil | 113–159 | 0.0000 | 0.0180 | 0.070 | 0.100 | 0.67 | 0.000 | 0.190 | 0.00 |
| | MOF interaction site | 766–939 | 0.0000 | 0.0030 | 0.018 | 0.019 | 0.01 | 0.000 | 0.000 | 0.00 |
| | MSL3 interaction site | 973–1039 | 0.0000 | 0.0060 | 0.015 | 0.017 | 0.08 | 0.000 | 0.000 | 0.00 |
| | Extradomain | | 0.0020 | 0.0060 | 0.028 | 0.071 | 0.39 | 0.021 | 0.053 | 0.16 |
| | Cysteine rich | 521–561 | 0.0000 | 0.0210 | 0.000 | 0.000 | 0.00 | 0.000 | 0.065 | 0.00 |
| <i>msl3</i> | Extradomain | | 0.0040 | 0.0220 | 0.035 | 0.120 | 0.29 | 0.011 | 0.060 | 0.17 |
| | CBD | 35–72 | 0.0000 | 0.0070 | 0.000 | 0.150 | 0.00 | 0.000 | 0.031 | 0.00 |
| | Polar region | 72–151 | 0.0020 | 0.0140 | 0.012 | 0.006 | 2.00 | 0.005 | 0.100 | 0.05 |
| | MRG | 200–480 | 0.0020 | 0.0300 | 0.009 | 0.073 | 0.12 | 0.004 | 0.000 | NA |
| | Extradomain | | 0.0020 | 0.0400 | 0.040 | 0.120 | 0.34 | 0.003 | 0.110 | 0.03 |

d , divergence; N, nonsynonymous; S, synonymous; extradomain, region outside currently annotated domains.

DCC proteins would be restricted to the *D. melanogaster* lineage. Indeed, lineage specificity must characterize whatever biological or molecular process is driving adaptive evolution at the DCC.

An alternative hypothesis for explaining *D. melanogaster*-specific adaptive protein evolution at several core DCC loci may be a host–pathogen interaction. The male-killing bacteria, *Spiroplasma*, has been found in several natural populations of *D. melanogaster* that occur near the equator (MONTENEGRO *et al.* 2000; POOL *et al.* 2006). VENETI *et al.* (2005) showed that *Spiroplasma*-infected *D. melanogaster* males carrying a loss-of-function mutation in any one of the five core DCC components survived to the third larval instar stage, while similarly infected males that carried no DCC mutations died as embryos. This result suggests that the deleterious effect of *Spiroplasma* on male fitness may be mediated through interactions with the dosage compensation complex and that a host–pathogen arms race localized to the DCC proteins may drive recurrent adaptive protein evolution. Interestingly, *Spiroplasma* has yet to be identified in *D. simulans*. Evidence that DCC adaptive protein evolution is more prevalent in *Drosophila* species with histories of *Spiroplasma* infections in nature would support this host–pathogen hypothesis.

The authors thank T. Mackay for providing the inbred *D. melanogaster* lines, S. Nuzhdin for providing the inbred *D. simulans* lines, and M. Kuroda for helpful comments on an earlier draft of the manuscript. M.T.L. was supported by a National Science Foundation (NSF) Graduate Research Fellowship. A.K.H. was supported by an NSF

Postdoctoral Fellowship in Biological Informatics. This work was funded by National Institutes of Health grant GM071926 to D.J.B.

LITERATURE CITED

- BAI, X., E. LARSCHAN, S. Y. KWON, P. BADENHORST and M. I. KURODA, 2007 Regional control of chromatin organization by noncoding roX RNAs and the NURF remodeling complex in *Drosophila melanogaster*. *Genetics* **176**: 1491–1499.
- BEGUN, D. J., A. K. HOLLOWAY, K. S. STEVENS, L. W. HILLIER, Y. P. POH *et al.*, 2007 Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol.* **5**(11): e310.
- BELOTE, J. M., and J. C. LUCCHESI, 1980 Male-specific lethal mutations of *Drosophila melanogaster*. *Genetics* **96**: 165–186.
- BIRCHLER, J. A., M. PAL-BHADRA and U. BHADRA, 2003 Dosage dependent gene regulation and the compensation of the X chromosome in *Drosophila* males. *Genetica* **117**: 179–190.
- BONE, J. R., J. LAVENDER, R. RICHMAN, M. J. PALMER, B. M. TURNER *et al.*, 1994 Acetylated histone H4 on the male X chromosome is associated with dosage compensation in *Drosophila*. *Genes Dev.* **8**: 96–104.
- BUSCAINO, A., G. LEGUBE and A. AKHTAR, 2006 X-chromosome targeting and dosage compensation are mediated by distinct domains in MSL-3. *EMBO Rep.* **7**: 531–538.
- COPPS, K., R. RICHMAN, L. M. LYMAN, K. A. CHANG, J. RAMPERSAD-AMMONS *et al.*, 1998 Complex formation by the *Drosophila* MSL proteins: role of the MSL2 RING finger in protein complex assembly. *EMBO J.* **17**: 5409–5417.
- HILFIKER, A., D. HILFIKER-KLEINER, A. PANNUTI and J. C. LUCCHESI, 1997 *mof*, a putative acetyl transferase gene related to the Tip60 and MOZ human genes and to the SAS genes of yeast, is required for dosage compensation in *Drosophila*. *EMBO J.* **16**: 2054–2060.
- IKEDA, K., D. J. STEGER, A. EBERHARTER and J. L. WORKMAN, 1999 Activation domain-specific and general transcription stimulation by native histone acetyltransferase complexes. *Mol. Cell Biol.* **19**: 855–863.

- JIN, Y., Y. WANG, J. JOHANSEN and K. M. JOHANSEN, 2000 JIL-1, a chromosomal kinase implicated in regulation of chromatin structure, associates with the male specific lethal (MSL) dosage compensation complex. *J. Cell. Biol.* **149**: 1005–1010.
- LI, F., D. A. PARRY and M. J. SCOTT, 2005 The amino-terminal region of *Drosophila* MSL1 contains basic, glycine-rich, and leucine zipper-like motifs that promote X chromosome binding, self-association, and MSL2 binding, respectively. *Mol. Cell. Biol.* **25**: 8913–8924.
- LUCCHESI, J. C., W. G. KELLY and B. PANNING, 2005 Chromatin remodeling in dosage compensation. *Annu. Rev. Genet.* **39**: 615–651.
- LYMAN, L. M., K. COPPS, L. RASTELLI, R. L. KELLEY and M. I. KURODA, 1997 *Drosophila* male-specific lethal-2 protein: structure/function analysis and dependence on MSL-1 for chromosome association. *Genetics* **147**: 1743–1753.
- MARIN, I., A. FRANKE, G. J. BASHAW and B. S. BAKER, 1996 The dosage compensation system of *Drosophila* is co-opted by newly evolved X chromosomes. *Nature* **383**: 160–163.
- MCDONALD, J. H., and M. KREITMAN, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**: 652–654.
- MENDJAN, S., M. TAIPALE, J. KIND, H. HOLZ, P. GEBHARDT *et al.*, 2006 Nuclear pore components are involved in the transcriptional regulation of dosage compensation in *Drosophila*. *Mol. Cell* **21**: 811–823.
- MONTENEGRO, H., W. N. SOUZA, D. DA SILVA LEITE and L. B. KLACZKO, 2000 Male-killing selfish cytoplasmic element causes sex-ratio distortion in *Drosophila melanogaster*. *Heredity* **85**(Pt. 5): 465–470.
- MORALES, V., T. STRAUB, M. F. NEUMANN, G. MENGUS, A. AKHTAR *et al.*, 2004 Functional integration of the histone acetyltransferase MOF into the dosage compensation complex. *EMBO J.* **23**: 2258–2268.
- NEI, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- NEI, M., and T. GOJOBORI, 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**: 418–426.
- POOL, J. E., A. WONG and C. F. AQUADRO, 2006 Finding of male-killing *Spiroplasma* infecting *Drosophila melanogaster* in Africa implies transatlantic migration of this endosymbiont. *Heredity* **97**: 27–32.
- PRESGRAVES, D. C., and W. STEPHAN, 2007 Pervasive adaptive evolution among interactors of the *Drosophila* hybrid inviability gene, *Nup96*. *Mol. Biol. Evol.* **24**: 306–314.
- SCOTT, M. J., L. L. PAN, S. B. CLELAND, A. L. KNOX and J. HEINRICH, 2000 MSL1 plays a central role in assembly of the MSL complex, essential for dosage compensation in *Drosophila*. *EMBO J.* **19**: 144–155.
- SKRIPSKY, T., and J. C. LUCCHESI, 1982 Intersexuality resulting from the interaction of sex-specific lethal mutations in *Drosophila melanogaster*. *Dev. Biol.* **94**: 153–162.
- STAJICH, J. E., D. BLOCK, K. BOULEZ, S. E. BRENNER, S. A. CHERVITZ *et al.*, 2002 The Bioperl toolkit: Perl modules for the life sciences. *Genome Res.* **12**: 1611–1618.
- VENETI, Z., J. K. BENTLEY, T. KOANA, H. R. BRAIG and G. D. HURST, 2005 A functional dosage compensation complex required for male killing in *Drosophila*. *Science* **307**: 1461–1463.
- WEIR, B., 1990 *Genetic Data Analysis*. Sinauer Associates, Sunderland, MA.
- YANG, Z., 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**: 555–556.

Communicating editor: D. M. RAND