

LETTERS

Demasculinization of X chromosomes in the *Drosophila* genus

David Sturgill^{1*}, Yu Zhang^{1*}, Michael Parisi^{1†} & Brian Oliver¹

X chromosomes evolve differently from autosomes, but general governing principles have not emerged¹. For example, genes with male-biased expression are under-represented on the X chromosome of *D. melanogaster*², but are randomly distributed in the genome of *Anopheles gambiae*³. In direct global profiling experiments using species-specific microarrays, we find a nearly identical paucity of genes with male-biased expression on *D. melanogaster*, *D. simulans*, *D. yakuba*, *D. ananassae*, *D. virilis* and *D. mojavensis* X chromosomes. We observe the same under-representation on the neo-X of *D. pseudoobscura*. It has been suggested that precocious meiotic silencing of the X chromosome accounts for reduced X chromosome male-biased expression in nematodes⁴, mammals⁵ and *Drosophila*⁶. We show that X chromosome genes with male-biased expression are under-represented in somatic cells and in mitotic male germ cells. These data are incompatible with simple X chromosome inactivation models. Using expression profiling and comparative sequence analysis, we show that selective gene extinction on the X chromosome, creation of new genes on autosomes and changed genomic location of existing genes contribute to the unusual X chromosome gene content.

Several models have been advanced to explain the peculiar gene content on X chromosomes, which can be divided into those driven by gene-by-gene or chromosome-wide selective pressures^{1,7}. Antagonistic selection is a popular gene-by-gene model. Females and males are under different selective pressures and deploy the genome differently such that genes and expression states advantageous for one sex can be disadvantageous to the other. This is expected to have a profound influence on X chromosomes because hemizyosity and immediate selection of recessive alleles in males should be masculinizing, whereas the increased residency time of X chromosomes in females and immediate selection of dominant alleles should be a counteracting force. A popular chromosome-wide model suggests that X chromosome inactivation during spermatogenesis is responsible for the reduced number of X chromosome genes with male-biased expression. The X chromosome is precociously condensed, and thus silenced, in preparation for male meiosis, owing to the absence of a homologous pairing partner.

Before addressing particular models, we asked if X chromosome sex-biased expression patterns were consistent across the genus. We determined female:male expression ratios on species-specific microarrays⁸, normalized sex bias across species, and parsed the expression data by chromosome arm (Fig. 1a, and Supplementary Fig. 1). Homologous linkage groups in the *Drosophila* genus are referred to as 'Muller's elements' to standardize discordant species-specific chromosome nomenclature⁹. Muller A is part of the X chromosome in all the species examined. Muller A genes with male-biased expression were under-represented relative to autosomes in each of the seven species (30–43% less than expected,

$P < 10^{-4}$ by chi-squared test; Supplementary Table 1). No other chromosome arms showed a genus-wide significant departure from a random distribution, although we did observe a modest over-representation of genes with male-biased expression on Muller B ($P < 10^{-2}$ by chi-squared test) in *D. melanogaster*, *D. simulans*, *D. ananassae* and *D. mojavensis*. Therefore, a dearth of genes with male-biased expression on the X chromosome is characteristic of the genus.

Formally, reduced male-biased expression could be due to an historical accident on the Muller A element rather than a property of the X chromosome itself. This is directly testable owing to the chromosome arrangement in *D. pseudoobscura*, in which the ancestral autosome, Muller D, fused to the X chromosome ~8–12 million years ago¹⁰. The neo-X chromosome (Fig. 1a) also showed a strong under-representation of male-biased expression (37% of the expected random distribution, $P < 10^{-4}$ by chi-squared test). The consistent under-representation of genes with male-biased expression on the *D. pseudoobscura* neo-X and on the ancestral X chromosomes, suggests that the effect of linkage on male-biased expression is a property of X chromosomes and that it takes less than 12 million years to reach a stereotypical depleted status.

Important predictions of the X chromosome inactivation model for reduced male-biased expression are readily testable: Expression of all X chromosome genes should be reduced, and X chromosome deficits in male-biased expression should be restricted to late spermatocytes. There was no global reduction in X chromosome gene expression in males (Fig. 1b), suggesting that global inactivation in spermatocytes does not disrupt the over-all balance of gene expression at the organismal level. These data are also consistent with genus-wide high-fidelity X chromosome dosage compensation in both the male soma and germline as observed in *D. melanogaster*¹¹. The paucity of X chromosome genes with male-biased expression seems due to reduced numbers of genes with overt male-biased expression, not a chromosome-wide effect. Although these data do not support the X chromosome inactivation model, they are also based on whole-animal expression profiles that greatly dilute late spermatocyte expression.

If male-biased expression from X chromosomes is reduced owing to a germline-specific event, then there should be no similar reduction in the soma. Most orthologues show sex-biased expression in all tested species⁸. Therefore, we estimated somatic sex-biased expression by computationally removing orthologues with testis-biased expression in *D. melanogaster* (1,569 genes). We observed significant under-representation ($P < 10^{-4}$ by chi-squared test) of X chromosome genes with putative somatic male-biased expression in five of the seven species (Fig. 1c). This included the *D. pseudoobscura* neo-X chromosome. We also re-analysed non-gonadal *D. melanogaster* expression data (Fig. 2a)². Whereas the number of genes

¹Laboratory of Cellular and Developmental Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Department of Health and Human Services, Bethesda Maryland 20892, USA. †Present address: Department of Biology, University of Pennsylvania, Pennsylvania 19104, USA.

*These authors contributed equally to this work.

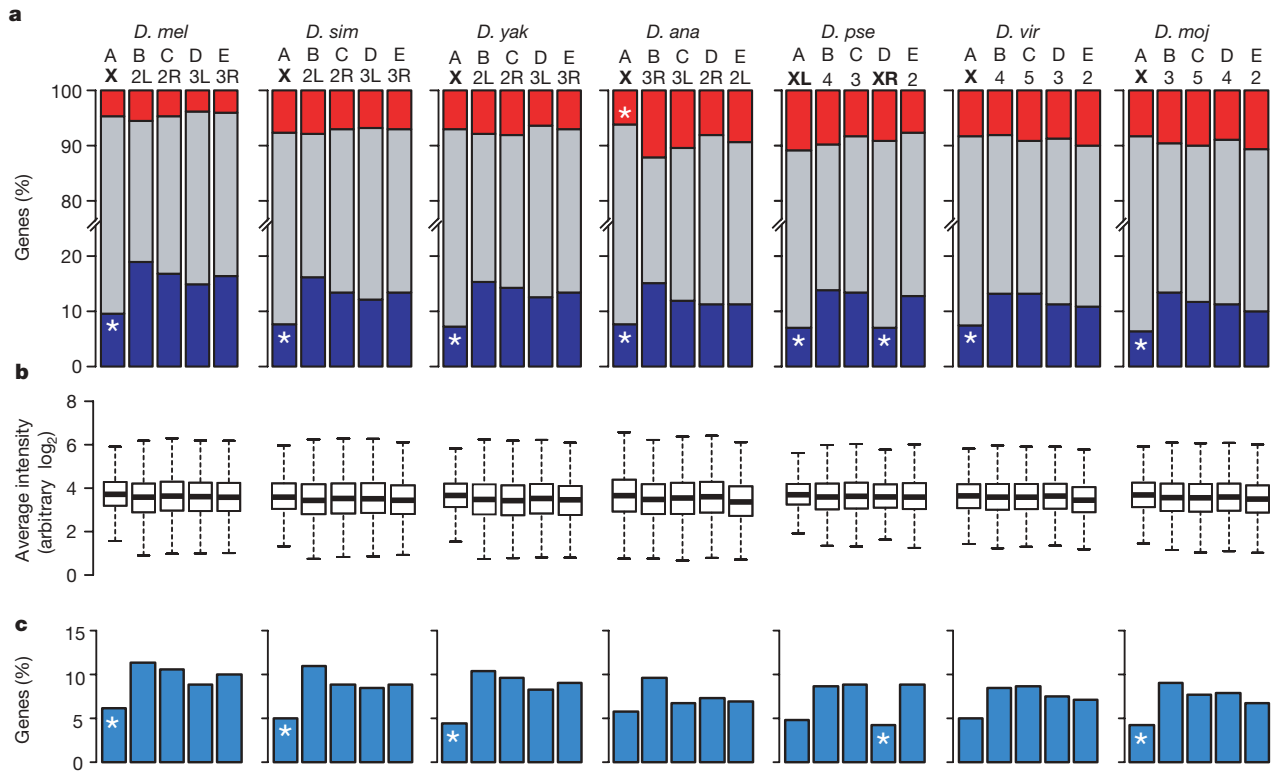


Figure 1 | Genes that show sex-biased expression, on chromosome arms. **a**, Percentage of genes with female-biased (red), non-biased (grey) and male-biased (blue) expression on chromosome arms. Muller's elements and arms are indicated (X chromosomes in bold). Significant under-representation from a random distribution of genes in an expression class (chi-squared test) is noted ($*P < 10^{-4}$). The top 20% of differentially expressed genes (by ranked P values) are assigned a sex-biased expression class. **b**, Box plots of

average hybridization intensities for males by chromosome arm. Twenty-fifth to seventy-fifth percentiles (boxes), medians (lines in boxes) and ranges (whiskers) are indicated. **c**, As in **a**, but with genes with predicted testis-biased expression removed. *D. melanogaster*, *D. mel*; *D. simulans*, *D. sim*; *D. yakuba*, *D. yak*; *D. ananassae*, *D. ana*; *D. pseudoobscura*, *D. pse*; *D. virilis*, *D. vir*; *D. mojavensis*, *D. moj*.

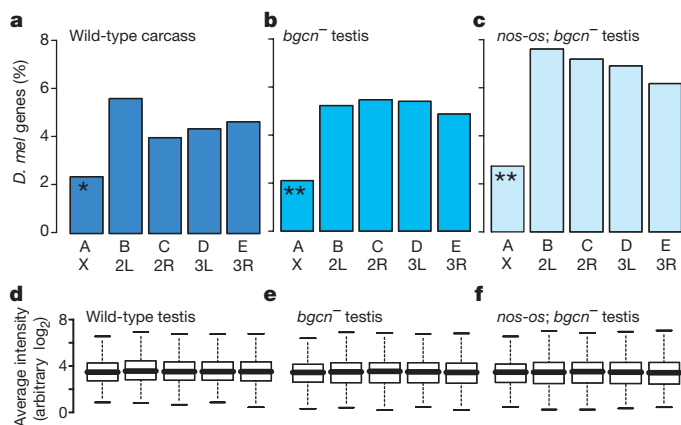


Figure 2 | Male-biased expression not subject to X chromosome inactivation. **a–c**, Percentage of *D. melanogaster* genes with male-biased expression. Significant under-representation as in Fig. 1 is noted ($*P < 10^{-2}$; $**P < 10^{-4}$). **a**, Non-gonadal soma (gonadectomized carcasses)². **b, c**, Testis from mutants with mitotically active germline tumours¹². Male-biased expression was determined by present calls in wild-type and mutant testis and absent calls in wild-type ovary expression profiles¹³ from FlyMine²⁰. Testis from *bgn*⁻ males (primary spermatocyte-biased; **b**) accumulates mitotically active cysts of interconnected germ cells, whereas testis from males overexpressing *os* (*nos-os*) in a *bgn*⁻ background (male germline stem-cell-biased; **c**) accumulates mitotically active germ cells that undergo complete cytokinesis¹². **d–f**, Box plots of average intensities (arbitrary scale) across hybridizations for wild-type (**d**), *bgn*⁻ (**e**), or *nos-os*; *bgn*⁻ (**f**) testis separated by chromosome arm. Twenty-fifth to seventy-fifth percentiles (boxes), medians (lines in boxes) and ranges (whiskers) are indicated.

with male-biased expression is much lower in the non-gonadal soma, we observed clear under-representation of X chromosome genes with male-biased expression ($P < 10^{-2}$), as previously observed².

To test the X chromosome inactivation model more directly, we used testis expression data from *D. melanogaster* mutants blocked during primary spermatocyte amplification¹². Testes from these mutants accumulate vast numbers of mitotic germ cells. We used independently profiled ovary samples as a reference¹³. X chromosome genes with testis-biased expression were under-represented in these mutant profiles ($P < 10^{-4}$ by chi-squared test, Fig. 2b, c). Additionally, there was no evidence for global reduction in the expression of all X chromosome genes in either wild-type or mutant testis (Fig. 2d–f). These data suggest that the X chromosome is a poor location for germline male-biased expression before X chromosome inactivation, which occurs in late post-mitotic spermatocytes⁶. X chromosome inactivation may be too late to affect most X chromosome transcript levels. We suggest that the paucity of X chromosome genes with male-biased expression in *Drosophila* is due to selection at the gene level, not global chromatin status.

Regardless of the thematic model, there are a limited number of physical mechanisms for achieving the observed under-representation of X chromosome genes with male-biased expression^{2,14}. We explored whether X chromosome genes with male-biased expression convert expression class, preferentially move to autosomes, are preferentially lost, or fail to arise on X chromosomes, using the *D. pseudoobscura* neo-X chromosome as a well-controlled natural experiment. We inferred the expression pattern of the ancestral Muller D of *D. pseudoobscura* before translocation to the X chromosome by using expression data from six phylogenetically flanking species. The extant expression pattern was directly

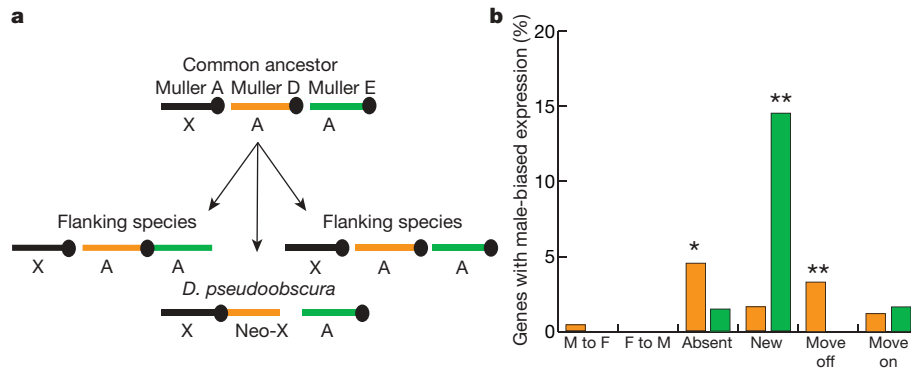


Figure 3 | The neo-X chromosome. **a**, *D. pseudoobscura* neo-X chromosome formation cartoon. Muller A, D and E are attached to different centromeres (filled circles) in *D. pseudoobscura* relative to flanking species. **b**, Bar plot showing changes to *D. pseudoobscura* Muller D neo-X chromosome (orange) and Muller E autosomal arm (green), based on the predicted expression pattern of the ancestral autosomes. Categories are: switches in sex-biased expression (F, female-biased; M, male-biased), not found

determined. In the other species, Muller D and Muller E are arms of the same chromosome or are individual chromosomes. Muller E provided a reference for the analysis (Fig. 3a, Supplementary Tables 3, 4, and Supplementary Fig. 2).

Of the 242 Muller D genes with ancestral or extant male-biased expression (Fig. 3b) that can be unambiguously assigned linkage and expression patterns, 216 (89%) remain in place on the neo-X chromosome. Only one gene on the neo-X chromosome switched from male-biased expression to female-biased expression. Although this particular gene *monkey king* (*mkg-p*) has undergone an interesting functional radiation, and sex-biased expression switches in the melanogaster subgroup¹⁵, sex-bias change on the neo-X chromosome is not significantly different from Muller E (no switches). Thus, there is no overt evidence that expression changes are responsible for under-representation of genes with male-biased expression on X chromosomes.

We found that 5% of genes with male-biased expression on the ancestral autosome had been lost from the genome of *D. pseudoobscura*, whereas only 1% of ancestral genes with male-biased expression were lost from Muller E ($P < 0.05$ by chi-squared test). Only 2% of genes with male-biased expression on the neo-X chromosome were unique to *D. pseudoobscura*, whereas 15% of the Muller E genes with male-biased expression in *D. pseudoobscura* are unique. The relative excess of new genes with male-biased expression on the autosome is highly significant ($P < 10^{-4}$ by chi-squared test). Determining if a gene is absent or new rather than highly diverged is difficult. However, exploration of genus-wide gene content using relaxed alignment instead of the consensus gene models indicates that many of these genes are likely to be completely absent rather than highly diverged (Supplementary Fig. 3). We also found clear evidence that non-synonymous substitutions (K_A) relative to synonymous substitution (K_S) rates are higher for X chromosome genes with male-biased expression than for those on autosomes (Supplementary Fig. 4). These data suggest that there is net loss of genes with male-biased expression from the X chromosome (as well as more subtle changes to sequence) and that functions lost from the X chromosome may be replaced by new gene formation on autosomes.

Analysis of gene transposition showed a net movement of genes with male-biased expression off the neo-X chromosome, with 3% of ancestral genes with male-biased expression moving to autosomes and no genes with male-biased expression moving from Muller E ($P < 10^{-4}$ by chi-squared test). Male-biased expression was always maintained in the new location. There was no significant difference between the neo-X and Muller E with respect to newly arriving genes, because 1% of ancestral genes with male-biased expression moved to

(absent), only found in *D. pseudoobscura* (new), and genes moving from (move off) and to (move on) the indicated arm. Significant differences (chi-squared test) between Muller D and Muller E are noted (** $P < 10^{-4}$; * $P < 0.05$). The ancestral condition of the neo-X chromosome was inferred by examining chromosome linkage and expression on homologous arms in species that diverged from the *Drosophila* lineage before and after *D. pseudoobscura*.

the neo-X chromosome and 2% moved to Muller E. These data clearly support the idea that movement of genes with male-biased expression to autosomes promotes long-term survival of those genes. Gene loss, gain and movement unambiguously accounted for 13% of the loss of male-biased expression from the neo-X chromosome and 73% of the gain of male-biased expression on Muller E (Supplementary Table 5). If these unambiguous results reflect the changes that we are unable to trace, then gene loss, gain and movement are the dominant mechanisms for depleting male-biased expression on the X chromosome relative to autosomes.

Our data strongly support the idea that the X chromosome has an unusual distribution of genes with male-biased expression^{4,5,16–18}. Although there is compelling evidence that precocious X chromosome inactivation occurs in at least some species, and that this may contribute to the reduced density of genes with male-biased expression on X chromosomes^{4,5}, our data indicate that this is not a major contributor to the pattern in the *Drosophila* genus. We suggest that the dominant mechanisms for achieving this under-representation are preferential extinction of X chromosome genes and formation of new autosomal genes with male-biased expression², along with movement of genes with male-biased expression from the X chromosome¹⁴. Somewhat surprisingly, altered gene expression does not seem to be a major contributor.

METHODS SUMMARY

Array data sources. Expression data for sex-sorted whole adults of the seven *Drosophila* species⁸ are from GEO¹⁹ (GSE6640). Data for *D. melanogaster* gonadectomized male and female carcass on the FlyGEM platform were published previously², GEO (GSE442). Affymetrix data for *bgn*⁻ and UAS-*os*, *bgn*⁻/*nos-Gal4-VP16*, *bgn*⁻ mutant testes expression¹² were obtained from GEO (GSE4188, GDS2228) and wild-type ovary data were obtained from FlyMine²⁰ release v.8.0.

Sequence analysis. *D. melanogaster* annotation v.4.3 (ref. 21) and Comparative Analysis Freeze 1 (CAF1) assemblies²² were used throughout. Orthology relationships between the seven species including orthologue, paralogue and no_ homologue were assigned to consensus gene predictions²². Gene content changes were also determined by comparing *D. melanogaster* amino acid sequence against six-frame translated genomic DNA of each species by BLAST tblastn²³.

For gene loss and translocation, the ancestral state of the lost/moved gene was inferred by the consensus expression class and chromosome linkage for each gene. For the *D. pseudoobscura* neo-X chromosome (Muller D) and autosomal Muller E, gene gain/loss/movement was manually counted gene-by-gene. The ancestral gene content of this arm was inferred by phylogeny using genes present in species that diverged before (*D. virilis* and *D. mojavensis*) and after (*D. ananassae* and the melanogaster subgroup) the *melanogaster/obscura* group split. The ancestral state refers to the chromosome linkage and expression class of genes at the origin of this node in a rooted phylogenetic tree.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 September; accepted 28 September 2007.

- Vicoso, B. & Charlesworth, B. Evolution on the X chromosome: unusual patterns and processes. *Nature Rev. Genet.* **7**, 645–653 (2006).
- Parisi, M. *et al.* Paucity of genes on the *Drosophila* X chromosome showing male-biased expression. *Science* **299**, 697–700 (2003).
- Hahn, M. W. & Lanzaro, G. C. Female-biased gene expression in the malaria mosquito *Anopheles gambiae*. *Curr. Biol.* **15**, R192–R193 (2005).
- Bean, C. J., Schaner, C. E. & Kelly, W. G. Meiotic pairing and imprinted X chromatin assembly in *Caenorhabditis elegans*. *Nature Genet.* **36**, 100–105 (2004).
- Khil, P. P., Smirnova, N. A., Romanienko, P. J. & Camerini-Otero, R. D. The mouse X chromosome is enriched for sex-biased genes not subject to selection by meiotic sex chromosome inactivation. *Nature Genet.* **36**, 642–646 (2004).
- Wu, C. I. & Xu, E. Y. Sexual antagonism and X inactivation—the SAXI hypothesis. *Trends Genet.* **19**, 243–247 (2003).
- Ellegren, H. & Parsch, J. The evolution of sex-biased genes and sex-biased gene expression. *Nature Rev. Genet.* **8**, 689–698 (2007).
- Zhang, Y. *et al.* Constraint and turnover in sex-biased gene expression in the genus *Drosophila*. doi:10.1038/nature06323 (this issue).
- Ashburner, M., Golic, K. G. & Hawley, R. S. *Drosophila: A Laboratory Handbook* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY, 2004).
- Richards, S. *et al.* Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution. *Genome Res.* **15**, 1–18 (2005).
- Gupta, V. *et al.* Global analysis of X-chromosome dosage compensation. *J. Biol.* **5**, 3 (2006).
- Terry, N. A., Tulina, N., Matunis, E. & DiNardo, S. Novel regulators revealed by profiling *Drosophila* testis stem cells within their niche. *Dev. Biol.* **294**, 246–257 (2006).
- Chintapalli, V. R., Wang, J. & Dow, J. A. Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nature Genet.* **39**, 715–720 (2007).
- Betran, E., Thornton, K. & Long, M. Retroposed new genes out of the X in *Drosophila*. *Genome Res.* **12**, 1854–1859 (2002).
- Wang, W., Yu, H. & Long, M. Duplication–degeneration as a mechanism of gene fission and the origin of new genes in *Drosophila* species. *Nature Genet.* **36**, 523–527 (2004).
- Ranz, J. M., Castillo-Davis, C. I., Meiklejohn, C. D. & Hartl, D. L. Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science* **300**, 1742–1745 (2003).
- Reinke, V. *et al.* A global profile of germline gene expression in *C. elegans*. *Mol. Cell* **6**, 605–616 (2000).
- Wang, P. J., McCarrey, J. R., Yang, F. & Page, D. C. An abundance of X-linked genes expressed in spermatogonia. *Nature Genet.* **27**, 422–426 (2001).
- Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).
- Lyne, R. *et al.* FlyMine: An integrated database for *Drosophila* and *Anopheles* genomics. *Genome Biol.* **8**, R129 (2007).
- Grumbling, G. & Strelets, V. FlyBase: anatomical data, images and queries. *Nucleic Acids Res.* **34**, D484–D488 (2006).
- Drosophila* 12 Genomes Consortium. Evolution of genes and genomes of the *Drosophila* phylogeny. doi:10.1038/nature06341 (this issue).
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the *Drosophila* 12 Genomes Consortium for access to the assembly, alignment and annotation of the 12 sequenced *Drosophila* genomes, S. Davis for valuable technical advice, and C. Vinson, A. Clark and members of the B. Oliver laboratory for helpful discussion and comments on the manuscript. We are supported by the Intramural Research Program of the NIH, NIDDK.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.S. (davidsturgill@nidk.nih.gov) and B.O. (oliver@helix.nih.gov).

METHODS

Expression data for whole adults of the seven *Drosophila* species were generated with custom-designed oligonucleotide arrays (NimbleGen Systems). Array design, sample preparation and labelling, and data handling were as described⁶. Data for *D. melanogaster* gonadectomized male and female carcass² were re-analysed from raw intensity; these data were re-normalized by within-slide print-tip loess normalization, followed by between-slide quantile normalization using Bioconductor²⁴. Differential expression was then called in the same manner as the custom Nimblegen arrays, using Mann–Whitney tests with false-discovery rate correction. For Affymetrix data, we used ‘present’ and ‘absent’ calls to determine sex-bias. Intensity data were used directly from GEO without additional normalization, and averaged over replicates. Data were remapped to match DrosGenome2 to DrosGenome1 identifiers by Affymetrix.

‘Gained genes’ were defined as genes present in only one species (Supplementary Fig. 2). ‘Lost genes’ are genes absent from only one. ‘Moved genes’ are those that are present in all seven species, but with different chromosome linkage in at least one species. For gene loss and translocation, the ancestral state of the lost/moved gene is inferred by the consensus expression class and chromosome linkage for each gene. Genes with different arm linkage, owing to a large pericentric inversion in the *D. yakuba* lineage²¹, were not counted as translocations.

We also calculated content changes independent of gene prediction in the newly sequenced species. We performed a comparison of *D. melanogaster* amino acid sequence against six-frame translated genomic DNA of each species via BLAST tblastn²³. BLAST results were parsed and hits tiled together using BioPerl²⁵. *D. melanogaster* genes were called ‘absent’ in the other species if there was no hit below an *E*-value cut-off (10^{-3} and 10^{-17} are shown in Supplementary Fig. 3).

For the *D. pseudoobscura* neo-X chromosome (Muller D) and autosomal Muller E, gene gain/loss/movement was manually counted gene-by-gene. The ancestral gene content of this arm was inferred by phylogeny, using genes present in species that diverged before (*D. virilis* and *D. mojavensis*) and after (*D. ananassae* and the *melanogaster/obscura* group split). The ancestral state refers to the chromosome linkage and expression class of genes at the origin of this node in a rooted phylogenetic tree^{21,22}. For both Muller D and E, we filtered genes that are linked to these arms in any species. These were then filtered to include those that are Muller D- or Muller E- linked in at least one species downstream and one species upstream, or in *D. pseudoobscura* itself. Then genes with male-biased expression in at least one species are filtered. Each remaining gene was then manually assigned a putative ancestral expression class and arm location and post neo-X chromosome fate that considered the entire *Drosophila* lineage. There are cases involving species on the ‘edge’ of the tree that could not be resolved by this approach. For example, a gene that is Muller B-linked in *D. mojavensis/D. virilis* (on the edge of the rooted tree), but Muller D-linked in every other species would not be counted, because it is not possible to discriminate the direction of translocation. Our approach includes only unambiguous cases (Supplementary Tables 2–4).

Multiple sequence alignments of orthologues were imported using the seqinR package²⁶. K_A/K_S estimates adjusted for differences in transition and transversion rates were calculated from these alignments²⁷.

All computation was performed in the R/Bioconductor environment²⁴.

24. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
25. Stajich, J. E. *et al.* The Bioperl toolkit: Perl modules for the life sciences. *Genome Res.* **12**, 1611–1618 (2002).
26. Charif, D., Thioulouse, J., Lobry, J. R. & Perriere, G. Online synonymous codon usage analyses with the ade4 and seqinR packages. *Bioinformatics* **21**, 545–547 (2005).
27. Li, W. H. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J. Mol. Evol.* **36**, 96–99 (1993).