# Constraint and turnover in sex-biased gene expression in the genus *Drosophila*

Yu Zhang[1]*, David Sturgill[1]*, Michael Parisi[1]†, Sudhir Kumar[2] & Brian Oliver[1]

Both genome content and deployment contribute to phenotypic differences between species[1–5]. Sex is the most important difference between individuals in a species and has long been posited to be rapidly evolving. Indeed, in the *Drosophila* genus, traits such as sperm length, genitalia, and gonad size are the most obvious differences between species[6]. Comparative analysis of sex-biased expression should deepen our understanding of the relationship between genome content and deployment during evolution. Using existing[7,8] and newly assembled genomes[9], we designed species-specific microarrays to examine sex-biased expression of orthologues and species-restricted genes in *D. melanogaster, D. simulans, D. yakuba, D. ananassae, D. pseudoobscura, D. virilis* and *D. mojavensis*. We show that averaged sex-biased expression changes accumulate monotonically over time within the genus. However, different genes contribute to expression variance within species groups compared to between groups. We observed greater turnover of species-restricted genes with male-biased expression, indicating that gene formation and extinction may play a significant part in species differences. Genes with male-biased expression also show the greatest expression and DNA sequence divergence. This higher divergence and turnover of genes with male-biased expression may be due to high transcription rates in the male germline, greater functional pleiotropy of genes expressed in females, and/or sexual competition.

There are numerous case studies demonstrating that orthologues with sex-biased function diverge more rapidly than genes with non-biased function[10]. To determine systematically the relative contributions of gene content and expression divergence to sexual differences, we sampled sex-biased expression within the *Drosophila* genus using species-specific microarrays designed for the closely related *D. melanogaster, D. simulans* and *D. yakuba* group (common ancestor, 10–13 million years ago), and for the more distantly related *D. ananassae, D. pseudoobscura, D. virilis* and *D. mojavensis* (common ancestor, 40–65 million years ago) (Supplementary Table 1). The species-specific platform eliminated confounding effects of sequence divergence on hybridization and allowed us to assay the expression of lineage-restricted genes.

Previous work has demonstrated that sex-biased expression in *D. melanogaster* adults is substantial, primarily owing to gametogenesis[10]. This seems to be characteristic for the genus (Fig. 1, and Supplementary Fig. 1). Generally, we observed greater male-biased expression (~7–14% of the transcriptome) relative to female-biased expression (~3–9% of the transcriptome), at a significance value of
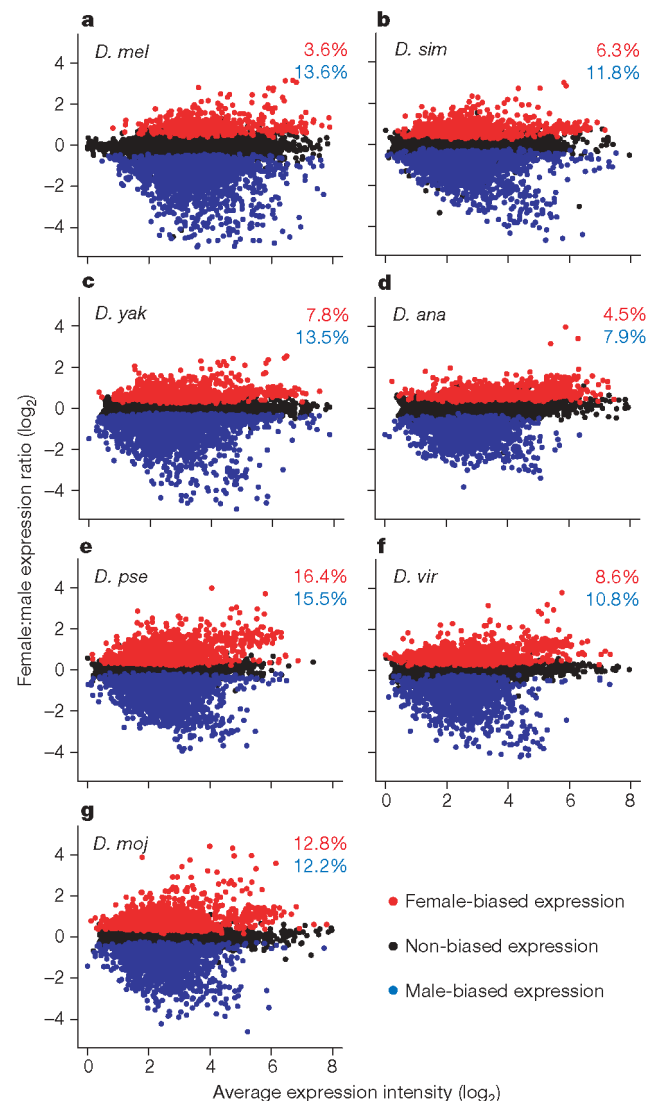
Figure 1 | Sex-biased expression in *Drosophila* species. a–g, Sex-biased female:male expression ratio (log$_2$) versus average expression intensity (log$_2$) plots for each *Drosophila* species. Expression intensities are arbitrary, where zero represents the minimum value. Values for genes with significant ($P \le 0.01$, false-discovery-rate-corrected Mann–Whitney test) female-biased, male-biased and non-biased expression are shown. The per cent of genes with female-biased or male-biased expression is inset in each panel. *D. melanogaster, D. mel; D. simulans, D. sim; D. yakuba, D. yak; D. ananassae, D. ana; D. pseudoobscura, D. pse; D. virilis, D. vir;* and *D. mojavensis, D. moj.*

[1]Laboratory 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. [2]Center for Evolutionary Functional Genomics, Biodesign Institute, Arizona State University, Tempe, Arizona 85287, USA. †Present address: Department of Biology, University of Pennsylvania, Pennsylvania 19104, USA.
*These authors contributed equally to this work.

$P \leq 0.01$ (Mann–Whitney, false-discovery-rate-corrected). The exceptions were *D. pseudoobscura* (~16% female- and male-biased expression) and *D. mojavensis* (~12% female- and male-biased expression). Additionally, the magnitude of male-biased expression was generally greater than the female-biased expression—the average $\log_2$ female:male expression ratio was -1.2 for genes with male-biased expression and 0.8 for genes with female-biased expression. This indicates that there were more genes approaching male-specific expression than female-specific expression. The genes that showed sex-biased expression in each species are listed in Supplementary Information (Supplementary Tables 3–16).

To examine expression divergence over time, we parsed the genes with orthologues in every species and constructed a pairwise matrix of $\log_2$ female:male expression ratios. We compared expression within species (two strains of *D. simulans*), between species within the closely related melanogaster subgroup, and between all seven species (Fig. 2a–c). Similar pairwise matrices for quadruplicate replicates within each species were also plotted as a baseline measurement of technical noise and biological variability (Supplementary Fig. 2). All expression ratio plots were linear and showed increasing expression divergence with inferred genetic distance.

There was an especially clear relationship between sequence and expression divergence. Neighbour-joining trees of expression divergence (from the pairwise expression ratios between each species; $1 -$ Pearson's r; Supplementary Fig. 3), or by sequence divergence[9,11] have the same topology (Fig. 2d). Expression divergence tightly correlated with time (Fig. 2e, $r^2 = 0.96$), which may provide a useful tool in molecular phylogenetics.

Although the whole-genome trends in expression divergence were both obvious and clear, at the gene level, the magnitude of expression

divergence was modest. Only 384 orthologue pairs (0.3%) showed significant female-biased expression in one species and significant male-biased expression in another. Switches between highly female-biased expression and highly male-biased expression were never observed (Fig. 2c). Extensive (20%) categorical changes in sex-bias class, especially for genes with male-biased expression, were previously reported between *D. melanogaster* and *D. simulans*[12,13]. We observed a categorical change in sex-biased expression in 12% of the orthologues between these two species, but the changes were dominated by low magnitude changes between modest sex-biased expression and non-sex-biased categories. These values are highly sensitive to arbitrary significance-level cut-offs; however, it was clear in exploratory plots of expression ratios that genes with male-biased expression showed greater expression divergence (Fig. 2b, c). Plots of expression ratio standard deviations against average expression ratio (Fig. 3a) also showed a clear excess of variable expression among orthologues with male-biased expression ($P < 10^{-8}$, chi-squared test). Thus, male-biased expression contributes heavily to overall expression divergence.

To determine if particular types of genes show greater or lesser expression divergence we analysed Gene Ontology[14] (GO) terms. Unsurprisingly, genes annotated as 'unknown function' are significantly over-represented ($P < 10^{-8}$, Fisher's exact test) among genes with variable expression. Genes with 'transcriptional regulation' annotations were under-represented in the same gene set ($P < 10^{-4}$, Fisher's exact test), suggesting that genes involved in transcription regulation are under constraint. Similar constrained expression of transcriptional regulators was observed in a study of metamorphosis in the melanogaster subgroup[5].

Just as changes in DNA sequence can have consequences ranging from deleterious to neutral to advantageous[15], changes in gene
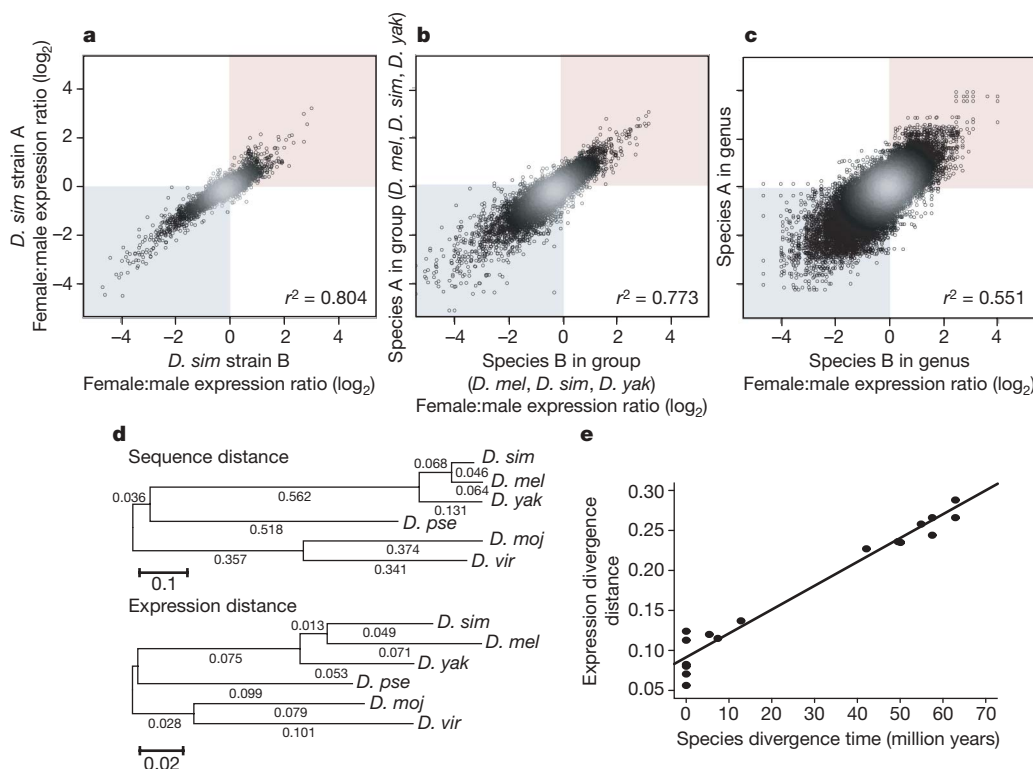
Figure 2 | Expression divergence among common orthologues.
Female:male expression ratios ($\log_2$) for orthologue pairs, plotted against each other: **a**, two different *D. simulans* strains; **b**, the melanogaster subgroup (*D.melanogaster*, *D. simulans* and *D. yakuba*); **c**, all seven *Drosophila* species. All the density (grey for high, black for low) scatter plots include every 1:1 pair of common orthologues for which both have an expression value. In **b** and **c**, the species A and B designation is arbitrary, but A is assigned to the species in the pair most closely related to

D. melanogaster. **d**, Neighbour-joining trees with branch lengths inferred using sequence distance (genomic mutation distance[11]) and the expression distance ($1 -$ Pearson's r) for all pairs of species except the pairs between the *D. ananassae* outlier and other species (Supplementary Fig. 3). **e**, Expression distance values plotted against estimated divergence time[11] for all possible species pairs and replicates within species. Quadruplicate replicates within each species were used at a time of 0 million years.

expression should have variable effects, owing to underlying mutations in transcription factors, *cis*-regulatory sites and post-transcriptional regulators, and the resulting variance will be subject to drift and selection[2,3,5,13,16–18]. We were able to distinguish expression differences between species well enough to show a linear relationship with time at the full-transcriptome level, but does this apply to individual genes?

To determine if there is a common set of orthologues that can tolerate variable expression (that can be thought of as the thematic equivalent of a synonomous codon substitution), we asked if expression divergence between orthologues within the melanogaster subgroup correlates with the expression divergence between more distantly related species. We found no significant correlation between orthologue expression divergence between groups of species (r^2 = 0.08, Fig. 3b). Genes with greater expression divergence in the melanogaster subgroup and the remaining species are different. Thus, although overall expression divergence shows a clock-like behaviour (reflecting mutation accumulation in a neutral model, or an adaptive speed limit in a selection model), different individual genes contribute to this global expression divergence in different amounts. This suggests that there is not a common set of genes that tolerate large drifts in sex-biased expression ratios.

To analyse further the orthologues with the most divergent expression, we selected orthologues with the greatest expression divergence (s.d. > 0.5) and subjected them to cluster analysis with species-order fixed (Fig. 3c). Strikingly, even those genes with the most variable expression were organized into well-defined clusters. Each of the clusters was subsequently analysed to look for patterns of change. We observed three distinct cluster types revealing expression divergence between lineages, aberrant expression in a single species, and unpatterned variability (Fig. 3c, d). For example, cluster 'A' shows higher male-biased expression in just the melanogaster

subgroup (*D. melanogaster*, *D. simulans* and *D. yakuba*); cluster 'B' shows increased male-biased expression in *D. pseudoobscura* only; and cluster 'C' shows no evidence for a phylogenetic trend. Briefly, among the 5% of common orthologues with the most variable expression, 52% exhibited lineage-specific, 22% species-specific and only 25% unpatterned expression variability.

Having only a few sequenced genomes seriously hinders the study of genes that are species- or lineage-specific (species-restricted). We took advantage of the species-specific array design to determine the contribution of common orthologues and species-restricted genes to overall sex-biased expression patterns (Fig. 4a, b). Female-biased expression was over-represented ($P < 10^{-2}$, chi-squared test) among common orthologues in four of the seven species, whereas male-biased expression was always under-represented. The pattern was reversed among the species-restricted genes. Female-biased expression of species-restricted genes was less prevalent in all species except *D. virilis*, and male-biased expression was more prevalent in each of the species examined. Female-biased expression was also under-represented among paralogues (Supplementary Fig. 4). Similar results were obtained using TBLASTN methods to detect genes that had diverged to obscure orthology (Supplementary Fig. 5). These suggest that genes with male-biased expression have higher effective birth and extinction rates.

We also asked if sex-bias and expression divergence correlate with sequence divergence among orthologues. If similar selective pressure acts on both protein-coding capacity and expression at a given locus, then they should correlate. However, protein-coding capacity and expression divergence need not be tightly coupled. For example, high expression divergence can result from changes in upstream transcription factors or the *cis*-regulatory sites that they bind[19].
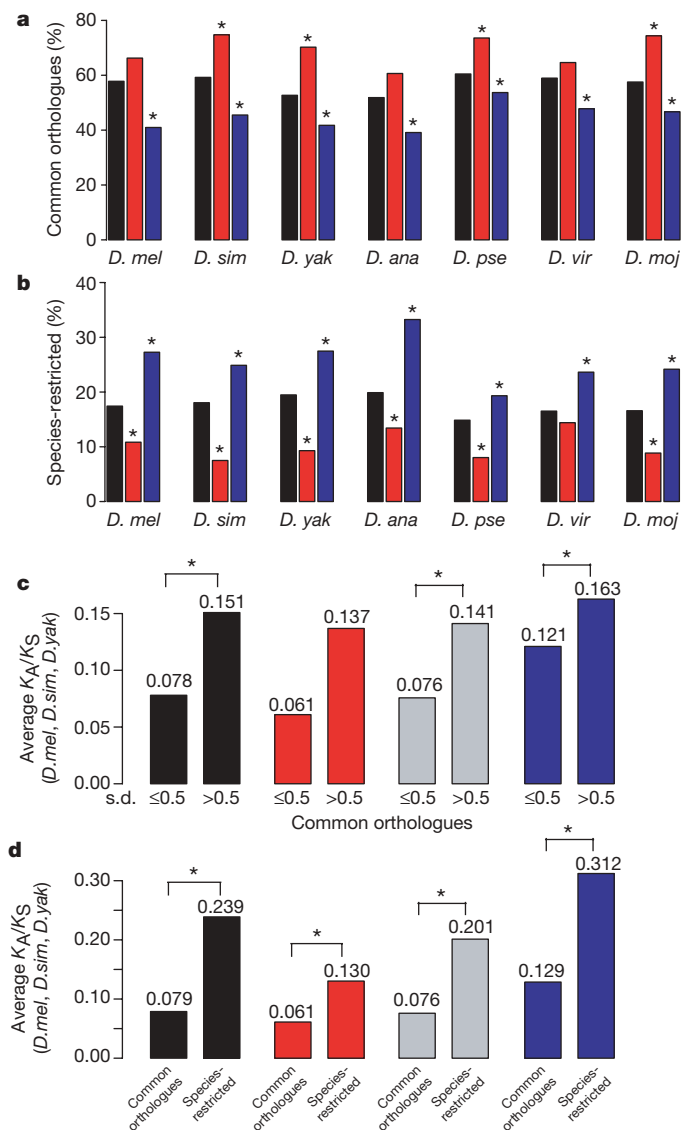
Synonymous (K_S) and non-synonymous substitution rates (K_A) in protein-coding genes were used to examine sequence divergence[20].
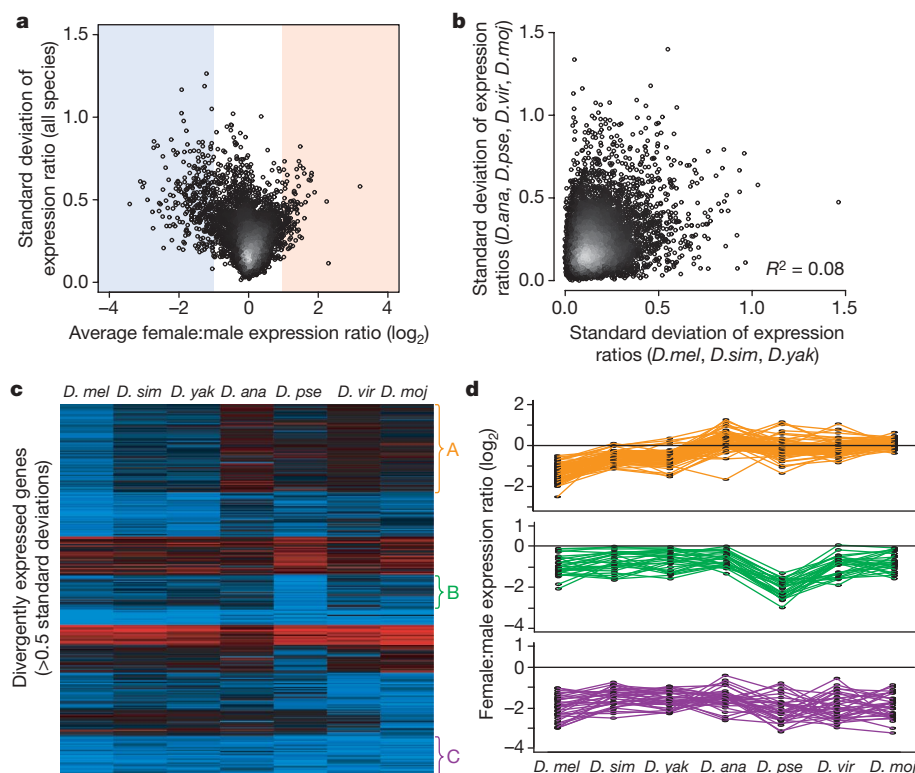
Figure 3 | Expression divergence within and between species and groups. **a**, Average female:male expression ratios for common orthologues plotted against expression divergence (expression ratio standard deviations between 7 species) for the same orthologues. **b**, Expression ratio standard deviations among members of the melanogaster subgroup (*D. melanogaster*, *D. simulans* and *D. yakuba*) plotted against standard deviations among the other four species (*D. ananassae*, *D. pseudoobscura*, *D. virilis* and *D. mojavensis*). **c**, *K*-means clustering (*K* = 10, species-order fixed) of expression ratios where s.d. > 0.5. Female-biased (red), male-biased (blue) and non-biased (black) expression is indicated. **d**, Examples of gene clusters that are indicated on the Eisengram (**c**). Species (*x* axis) and log_2 female:male expression ratio (*y* axis) of common orthologues are shown.

Multiple substitutions occur at a given site between distantly related species (for example, *D. melanogaster* and *D. mojavensis*) making K_A/K_S ratios much less reliable, and therefore K_A/K_S ratios were used only within the melanogaster subgroup (Fig. 4c, d). Genes with male-biased expression were expected to show higher K_A/K_S ratios[10]. Indeed, common orthologues with male-biased expression had K_A/K_S values within the melanogaster subgroup (0.129), more than two times those of common orthologues with female-biased expression (0.061). Interestingly, common orthologues with non-biased expression showed intermediate K_A/K_S values. We observed a strong correlation between expression and sequence divergence among the genes showing the greatest expression divergence (Fig. 4c), as

has also been seen in mammals[21]. Additionally, species-restricted genes had higher sequence-divergence than common orthologues for all expression categories (Fig. 4d), as has been seen in vertebrates[22]. Perhaps expression divergence, gene turn-over, sex-bias and sequence divergence of individual genes are often coupled to the same selective forces.

The contrasting divergence and turnover patterns of genes with male-biased expression relative to those with female-biased expression is somewhat surprising. Reproduction is the function of a couple, not an individual; therefore co-evolution of reproductive traits is expected to occur. For example, selection for sperm tail length in *Drosophila* males is coupled to selection for length of the seminal receptacle in females[23]. There are a number of possible explanations. There may be greater *de novo* generation of genes with male-biased expression as a result of simple sequence requirements for core promoter generation[24] and extremely high levels of RNA polymerase in spermatocytes[25]. This combination might result in excessive transcription of intragenic regions[26]. A few of these new genes with male-biased expression might be functional, but most of these '*de novo*' genes would be expected to rapidly degenerate. Alternatively, genes required for oogenesis may be more constrained because of pleiotropy or the under-representation of paralogues with partially overlapping functions. Many *D. melanogaster* genes required for female fertility are also required for organismal viability[27], and genes with clear multiple functions, such as those encoding ribosomal proteins, are overexpressed in ovaries relative to testes[28]. Finally, male–male competition might be particularly strong[29]. The addition of more sequenced genomes will provide ample opportunities to explore these questions further.

METHODS SUMMARY

Flies. Species were grown on standard media (Tucson *Drosophila* Stock Center). We isolated messenger RNA from adult females and males grown at 22 °C (5–7 days post eclosion), and labelled and hybridized using standard methods.

Arrays. Oligonucleotide arrays of 50-mers (NimbleGen Systems) were designed against draft assemblies and *ab initio* annotations we contributed to the gene model reconciliation[9] (Supplementary Table 2). *D. melanogaster* 60-mer expression array (NimbleGen Design ID 2005-10-17_Dmel4_60mer_exp) designed on the basis of Flybase annotation V4.2 was used for *D. melanogaster* hybridizations. For this report, we remapped all array elements to current consensus gene models[9]. Our expression results and conclusions were similar using the original models. Because low-magnitude expression divergence is difficult to distinguish from noise, we performed at least quadruplicate replicates for each species and only channels passing a stringent quality control regimen were used in the final analysis (72 channels total). Full platform descriptions and data are available at the GEO under accession GSE6640.

For each gene, $\log_2$ intensities for female and male expression were compared by non-parametric two-sample Mann–Whitney tests to generate the significance of sex-biased expression ($P \leq 0.01$) and ratios of each gene were calculated as the average probeset intensity in female channels divided by the intensity in male channels. Common orthologues are present in all 7 species and species-restricted genes are present in at least one species, but absent in ≥1 species.

Figure 4 | Relationship between sex-biased expression, gene content and sequence divergence. Gene content and expression of common orthologues (**a**) and species-restricted (**b**) genes. The percentages of all genes (black) and with female (red) or male (blue) -biased expression are shown. Significant differences ($P < 10^{-2}$, chi-squared test) between sex-biased classes and total genes are indicated (asterisks). See Supplementary Fig. 5 for paralogues. Average K_A/K_S ratios within the melanogaster subgroup for common orthologues with high or low expression-ratio s.d. (**c**) and for all common orthologues or species-restricted genes (**d**). Bars are colour-coded as in **a**, with the addition of grey bars, which represent non-biased expression. Significant differences ($P < 10^{-2}$, Mann–Whitney test) between common orthologues with constrained expression and variable expression, or between common orthologues and species-restricted genes are indicated (asterisks).

1. Denver, D. R. *et al.* The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans. Nature Genet.* **37**, 544–548 (2005).
2. Jordan, I. K., Marino-Ramirez, L. & Koonin, E. V. Evolutionary significance of gene expression divergence. *Gene* **345**, 119–126 (2005).
3. Khaitovich, P., Enard, W., Lachmann, M. & Paabo, S. Evolution of primate gene expression. *Nature Rev. Genet.* **7**, 693–702 (2006).
4. Rifkin, S. A., Houle, D., Kim, J. & White, K. P. A mutation accumulation assay reveals a broad capacity for rapid evolution of gene expression. *Nature* **438**, 220–223 (2005).
5. Rifkin, S. A., Kim, J. & White, K. P. Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nature Genet.* **33**, 138–144 (2003).
6. Powell, J. R. *Progress and Prospects in Evolutionary Biology: The Drosophila Model* (Oxford Univ. Press, Oxford, 1997).
7. Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster. Science* **287**, 2185–2195 (2000).

8. Richards, S. *et al.* Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and *cis*-element evolution. *Genome Res.* **15**, 1–18 (2005).

9. *Drosophila* 12 Genomes Consortium. Evolution of genes and genomes in the *Drosophila* phylogeny. *Nature* doi:10.1038/nature06341 (this issue).

10. Ellegren, H. & Parsch, J. The evolution of sex-biased genes and sex-biased gene expression. *Nature Rev. Genet.* **8**, 689–698 (2007).

11. Tamura, K., Subramanian, S. & Kumar, S. Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. *Mol. Biol. Evol.* **21**, 36–44 (2004).

12. Nuzhdin, S. V., Wayne, M. L., Harmon, K. L. & McIntyre, L. M. Common pattern of evolution of gene expression level and protein sequence in *Drosophila. Mol. Biol. Evol.* **21**, 1308–1317 (2004).

13. 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).

14. Harris, M. A. *et al.* The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32**, D258–D261 (2004).

15. Ohta, T. & Gillespie, J. H. Development of neutral and nearly neutral theories. *Theor. Popul. Biol.* **49**, 128–142 (1996).

16. Hsieh, W. P., Chu, T. M., Wolfinger, R. D. & Gibson, G. Mixed-model reanalysis of primate data suggests tissue and species biases in oligonucleotide-based gene expression profiles. *Genetics* **165**, 747–757 (2003).

17. Khaitovich, P., Paabo, S. & Weiss, G. Toward a neutral evolutionary model of gene expression. *Genetics* **170**, 929–939 (2005).

18. Lemos, B., Meiklejohn, C. D., Caceres, M. & Hartl, D. L. Rates of divergence in gene expression profiles of primates, mice, and flies: stabilizing selection and variability among functional categories. *Evolution Int. J. Org. Evolution* **59**, 126–137 (2005).

19. Wray, G. A. The evolutionary significance of *cis*-regulatory mutations. *Nature Rev. Genet.* **8**, 206–216 (2007).

20. Li, W. H. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J. Mol. Evol.* **36**, 96–99 (1993).

21. Khaitovich, P. *et al.* Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science* **309**, 1850–1854 (2005).

22. Subramanian, S. & Kumar, S. Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome. *Genetics* **168**, 373–381 (2004).

23. Miller, G. T. & Pitnick, S. Sperm-female coevolution in *Drosophila. Science* **298**, 1230–1233 (2002).

24. FitzGerald, P. C., Sturgill, D., Shyakhtenko, A., Oliver, B. & Vinson, C. Comparative genomics of *Drosophila* and human core promoters. *Genome Biol.* **7**, R53 (2006).

25. Schmidt, E. E. & Schibler, U. High accumulation of components of the RNA polymerase II transcription machinery in rodent spermatids. *Development* **121**, 2373–2383 (1995).

26. Struhl, K. Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nature Struct. Mol. Biol.* **14**, 103–105 (2007).

27. Perrimon, N., Mohler, D., Engstrom, L. & Mahowald, A. P. X-linked female-sterile loci in *Drosophila melanogaster. Genetics* **113**, 695–712 (1986).

28. Parisi, M. *et al.* A survey of ovary-, testis-, and soma-biased gene expression in *Drosophila melanogaster* adults. *Genome Biol.* **5**, R40 (2004).

29. Singh, S. R., Singh, B. N. & Hoenigsberg, H. F. Female remating, sperm competition and sexual selection in *Drosophila. Genet. Mol. Res.* **1**, 178–215 (2002).

METHODS

Flies. Stocks (Tucson Drosophila Stock Center) of *D. simulans* (14021-0251.198 and 14021-0251.011), *D. yakuba* (14021-0261.01), *D. ananassae* (14024-0371.13), *D. pseudoobscura* (14011-0121.94), *D. mojavensis* (15081-1352.22) and *D. virilis* (15010-1051.87) were grown on standard cornmeal media (Tucson Drosophila Stock Center), with the exception of *D. pseudoobscura*, *D. mojavensis* and *D. virilis*, which were grown on Banana-Opuntia media (Tucson Drosophila Stock Center).

Arrays. Oligonucleotide 50-mer arrays (NimbleGen Systems) were designed on the basis of draft or versioned genomic assemblies of six *Drosophila* species (*D. simulans* assembly: PCAP assembly for white501, GSC, Wash U, 01 December 04; *D. yakuba* assembly: GSC (WashU), 07 April 2004. *D. ananassae* assembly: Arachne Assembler, Agencourt, 06 December 2004; *D. pseudoobscura* assembly: v. 1.03 from FlyBase, December 2004; *D. mojavensis* assembly: Arachne Assembler, Agencourt, 06 December 2004; *D. virilis* assembly: Arachne Assembler, Agencourt, 29 October 2004). An average of 10 array probes were selected without bias with respect to position within each of our OLIV gene models. The OLIV set includes both high- and low-confidence models, which included non-overlapping draft EIS gene models based on *D.melanogaster* orthology (M. Eisen laboratory, v1.0, Feb 2005), *ab initio* GeneID[30] predictions using the *D. melanogaster* training set, FlyBase[31] genes and expressed sequence tag sequence from GenBank[32]. Array probes were remapped to the final genome assemblies and gene predictions GLEANR[9] by BLAT V25x1 (ref. 33). Only probes uniquely and perfectly matched to both annotation and assembly were used for final analysis.

Hybridization was according to the manufacturer's instructions (NimbleGen Systems), except that hybridization was done in custom-made chambers. Arrays were scanned on an Axon GenePix 4000B (Molecular Devices Corporation) and data were captured using NimbleScan 2.1 (NimbleGen Systems). For each species, at least four hybridizations, including technical (dye-flipped) replicates for each of two discrete samples (biological replicates) were performed. Extra hybridizations were performed for a different *D. simulans* stain (14021-0251.011).

Data handling. We used a multi-step quality control pipeline. Hybridization channels were retained when experimental intensities were >1 s.d. above mean on-spot background (from non-*Drosophila* control elements) and the inter-quartile ranges of $\log_2$ intensities were >1. Passed channels were normalized using variance stabilization normalization[34]. Signal variability between replicate channels was then tested by calculating the inter-quartile range of the relative log expression values for each channel against a virtual reference (the median value in all replicate channels for each array element). The channels with inter-quartile ranges of the relative log expression values greater than one were rejected. Passed channels were re-normalized by variance stabilization normalization from the raw data. This approach does not over-normalize the data while assuring that hybridization intensity is consistent between replicate channels.

For each gene, $\log_2$ intensities from female-sample single channels and male-sample single channels were compared by non-parametric two-sample Mann–Whitney tests to generate a significance measure. Ratios were then calculated as the average probeset intensity in female channels divided by the average probeset intensity in male channels for each gene. *P* values were false-discovery-rate-corrected[35]. The cut-off for sex-biased expression we used was

false-discovery-rate-adjusted $P \leq 0.01$. Expression was called 'below background' when the average probeset intensity was less than the average intensity of negative controls (probes targeting four *Arabidopsis* genes and two yeast genes) in both sexes.

Orthology calls are as described[9]. Common orthologues represent orthologues present in all 7 species and species-restricted genes are present in one species, but absent in ≥ 1 species. Paralogues are excluded from most of the data analysis. Multiple sequence alignments of orthologues were imported using the seqinR package[36]. K_A/K_S estimates adjusted for differences in transition and transversion rates were calculated from these alignments[20]. Average K_A/K_S of common orthologues were calculated from all possible K_A/K_S values between the melanogaster subgroup, then median K_A/K_S values were calculated for each category of genes with different sex-biased expression and genes with different expression divergence.

DNA sequence divergence and expression divergence ($1 -$ Pearson's *r* between the sex-biased expression ratio of two species) between each species pair was calculated. DNA sequence divergence was presented using genomic mutation distance by the method described previously[11]. Neighbour-joining trees were then inferred using DNA sequence divergence and expression distance from six species separately in MEGA4 (ref. 37). The common orthologues with most variable expression among species (s.d. > 0.5) were *K*-means clustered with 10 nodes using the euclidean similarity metric in Cluster 3.0/Tree-View[38].

D. melanogaster v.4.3 and *D. pseudoobscura* v.2.0 sequence and annotation from FlyBase were used as queries against the final genome assemblies[9] of all other six species by TBLASTN[39].

Unless otherwise noted all data handling was performed in BioConductor[40].

30. Parra, G., Blanco, E. & Guigo, R. GeneID in *Drosophila*. *Genome Res.* 10, 511–515 (2000).
31. Grumbling, G. & Strelets, V. FlyBase: anatomical data, images and queries. *Nucleic Acids Res.* 34, D484–D488 (2006).
32. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Wheeler, D. L. GenBank. *Nucleic Acids Res.* 34, D16–D20 (2006).
33. Kent, W. J. BLAT–the BLAST-like alignment tool. *Genome Res.* 12, 656–664 (2002).
34. Huber, W., von Heydebreck, A., Sultmann, H., Poustka, A. & Vingron, M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* 18 (Suppl 1), S96–S104 (2002).
35. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B. Methodological* 57, 289–300 (1995).
36. Charif, D. & Lobry, J. R. in *Structural approaches to sequence evolution: Molecules, networks, populations* (eds Bastolla, U. M.P., Roman, H. E. and Vendruscolo, M.) (2006).
37. Tamura, K., Dudley, J., Nei, M. & Kumar, S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Mol. Biol. Evol.* 24, 1596–1599 (2007).
38. de Hoon, M. J., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* 20, 1453–1454 (2004).
39. 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).
40. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5, R80 (2004).