OPEN ACCESS Freely available online

PLOS | GENETICS

An Excess of Gene Expression Divergence on the X Chromosome in *Drosophila* Embryos: Implications for the Faster-X Hypothesis

Melek A. Kayserili[1☯], Dave T. Gerrard[2☯], Pavel Tomancak[1]*, Alex T. Kalinka[1]*

1 Max Planck Institute for Molecular Cell Biology and Genetics, Dresden, Germany, **2** Faculty of Life Sciences, The University of Manchester, Manchester, United Kingdom

Abstract

The X chromosome is present as a single copy in the heterogametic sex, and this hemizygosity is expected to drive unusual patterns of evolution on the X relative to the autosomes. For example, the hemizgosity of the X may lead to a lower chromosomal effective population size compared to the autosomes, suggesting that the X might be more strongly affected by genetic drift. However, the X may also experience stronger positive selection than the autosomes, because recessive beneficial mutations will be more visible to selection on the X where they will spend less time being masked by the dominant, less beneficial allele—a proposal known as the faster-X hypothesis. Thus, empirical studies demonstrating increased genetic divergence on the X chromosome could be indicative of either adaptive or non-adaptive evolution. We measured gene expression in *Drosophila* species and in *D. melanogaster* inbred strains for both embryos and adults. In the embryos we found that expression divergence is on average more than 20% higher for genes on the X chromosome relative to the autosomes; but in contrast, in the inbred strains, gene expression variation is significantly lower on the X chromosome. Furthermore, expression divergence of genes on Muller's D element is significantly greater along the branch leading to the obscura sub-group, in which this element segregates as a neo-X chromosome. In the adults, divergence is greatest on the X chromosome for males, but not for females, yet in both sexes inbred strains harbour the lowest level of gene expression variation on the X chromosome. We consider different explanations for our results and conclude that they are most consistent within the framework of the faster-X hypothesis.

Citation: Kayserili MA, Gerrard DT, Tomancak P, Kalinka AT (2012) An Excess of Gene Expression Divergence on the X Chromosome in *Drosophila* Embryos: Implications for the Faster-X Hypothesis. PLoS Genet 8(12): e1003200. doi:10.1371/journal.pgen.1003200

Editor: Artyom Kopp, University of California Davis, United States of America

Received March 8, 2012; **Accepted** November 19, 2012; **Published** December 27, 2012

Funding: This work was supported by grant HFSP RGY0084. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: tomancak@mpi-cbg.de (PT); kalinka@mpi-cbg.de (ATK)

☯ These authors contributed equally to this work.

Introduction

It has long been suspected that the distinct properties of the X chromosome might in turn produce distinct patterns of evolution on the X relative to the autosomes [1,2]. In particular, the hemizygoisty of the X could be responsible for increased adaptive or non-adaptive evolution on this chromosome. Assuming an equal sex ratio and an equal variance in reproductive success in the two sexes, there will be three copies of the X in each mating pair versus four copies of each autosome thereby exposing the X to elevated levels of genetic drift [3]. If, however, we consider adaptive evolution, then the hemizygosity of the X is expected to facilitate the spread of recessive beneficial mutations, the selective benefit of which would otherwise be masked when in a heterozygous state on the autosomes [1,3–5]. Beneficial mutations with additive effects in heterozygotes are selectively equivalent on the X chromosome and on the autosomes, and would therefore be expected to evolve at similar rates across the chromosomes, whereas beneficial mutations that are dominant are expected to evolve faster on the autosomes [5]. A faster X may also be expected if mutations have sexually antagonistic effects, in which the sign of the selection coefficient is opposite in males and females [6]. In both adaptive and non-adaptive scenarios, it is the hemizygous context of the X chromosome in the heterogametic sex that is expected to drive more rapid evolution relative to the autosomes [7].

Determining the relative importance of different evolutionary forces in shaping the X chromosome is crucial for understanding several phenomena related to the X. For example, Haldane's rule, which is a classic generalization stating that in the hybrids of inter-species crosses the heterogametic sex is most often the inviable or sterile sex [8], could be explained by the fixation of recessive species-specific substitutions on the X chromosome which interact epistatically with autosomal loci [5]. Understanding how the X evolves could also help explain unusual distributions of genes across chromosomes [9], such as a disproportionate number of genes involved in cognitive function residing on the X in mammals [10] or an excess of sexually antagonistic genes on the X in *Drosophila* [11]. A fuller understanding of how selection acts differentially across autosomes and sex chromosomes could also shed light on the role of the X chromosome in the evolution of sexually-selected traits [12].

Empirical studies have sought to quantify the importance of adaptative processes in driving the evolution of the X. While many

Author Summary

There is a single copy of the X chromosome in males, yet two copies in females. This unique inheritance pattern has long been predicted to influence how the X chromosome evolves. In particular, the theory suggests that the single copy of the X in males could facilitate faster evolution of the X, although this faster evolution could be either adaptive or non-adaptive. We measured gene expression across the chromosomes in several different *Drosophila* species and also in several inbred strains of *D. melanogaster* for both embryos and adults. We found that gene expression is evolving significantly faster between species in the embryos, yet harbours significantly less variation within inbred strains. In adults, evolution between species appears to be much slower than in the embryos, yet they also harbour significantly lower levels of gene expression variation on the X chromosome in inbred strains. Overall, our results are consistent with there being an excess of adaptive evolution on the X chromosome in *Drosophila* embryos. Finally, we underscore the importance of biological context for understanding how chromosomes evolve in different species.

studies have found that the differences between species can often be attributed to X-linked loci of large effect [13–15], much of the recent work has found inconsistent evidence for an excess of positive selection of X-linked proteins. For example, studies of chimpanzee and human orthologs shows that X-linked loci have higher rates of adaptive protein evolution than autosomal loci [16–18], whereas in *Drosophila* species, whole-genome comparisons do not reveal any bias towards higher rates of protein evolution on the X chromosome [19–21]. Other *Drosophila* studies, which may use biased samples of genes [7], recover the faster-X effect found in mammals [22–25] including a study that demonstrated accelerated evolution of X-linked genes on the newly-formed X chromosome of *D. miranda* [26]. A recent study in aphids, an X0 sex determination system, found evidence for adaptive evolution of X-linked genes [27], and, interestingly, the same finding was reported for the Z chromosome (the equivalent of the X chromosome in the ZW sex determination system) in a comparison of chicken and zebra finch orthologs [28].

While the evidence for adaptive evolution of the X remains somewhat patchy, such discrepancies suggest that differences in the biology of different groups of species could strongly influence their chromosomal evolution. An important parameter in the faster-X theory is the presence or absence of dosage compensation in the heterogametic sex; that is, whether the presence of a single copy of a gene in the heterogametic sex is compensated, in terms of gene expression, to an extent that it is selectively equivalent to the two copies in the homogametic sex. Theory shows that beneficial mutations will evolve faster on the X compared to the autosomes, only if mutations are at least partially recessive [5]. Thus, to observe a global fast-X effect, most beneficial mutations must be at least partially recessive. In the absence of dosage compensation, however, theory suggests that beneficial mutations must be more recessive for the X to evolve faster provided that the weaker expression in males results in a correspondingly weaker beneficial selection coefficient [5] – this is because dosage compensation equalises the expression of genes expressed on the X in males and females, and is therefore assumed to also equalise their selection coefficients. Thus, fundamental differences in both the extent and mechanism of dosage compensation between different groups of species could have a dramatic effect on the rate of evolution of the X chromosome [5]. However, it is also possible that adaptive evolution of protein sequences accounts for a larger fraction of the evolutionary divergence between some groups of species relative to others. Therefore, while we may not see significantly higher adaptive protein evolution on the X in *Drosophila*, it is conceivable that adaptive differences in this group of species are most often seen in *cis*-regulatory, and therefore non-coding, regions of the genome [20,29].

We aimed to address evolution on the *Drosophila* X chromosome relative to the autosomes at the level of gene expression divergence. By focusing on gene expression, we relax the implicit assumption of previous studies that a majority of adaptive evolution occurs via changes in amino acid sequences. Additionally, by measuring divergence in terms of gene expression rather than coding sequences, we could compare expression divergence in embryos relative to adults and therefore ask whether gene expression is free to evolve independently in different stages of the animal's life-cycle. Our results show that mean gene expression divergence is higher for the X chromosome relative to autosomes and, more surprisingly, this effect is much stronger in the *Drosophila* embryos relative to the adults.

Results

Higher mean expression divergence on the X chromosome in *Drosophila* embryos

Evidence for accelerated evolution of the X in *Drosophila* has been sought in the adaptive evolution of protein sequences, but has so far produced mixed results [20–24]. We chose to focus on the evolution of gene expression with the advantage that we could detect the effects of divergence of non-coding regulatory sequences, and in addition we could directly compare evolution in different stages of the animal's life-cycle. To explore gene expression divergence across *Drosophila* chromosomes we used gene expression data from two distinct stages of the life-cycle – the embryo [30] and the adult [31]. In addition, we extracted RNA from the embryos of 17 inbred strains of *D. melanogaster* and hybridised the samples to whole-genome microarrays to provide insight into the maintenance of gene expression variation across chromosomes but within a single species. Similarly, for adult stages we used whole-genome microarray data from 40 adult inbred strains of *D. melanogaster* separated into males and females [32,33]. Table S1 summarises the chromosomal distributions of genes in each dataset.

In the between-species data for embryos, the X chromosome has the highest mean expression divergence ($P = 2.19 \times 10^{-7}$; Figure 1A) an effect that ranges from 18% up to 27% higher and in all cases is significant (see Table S2 for all chromosomal contrasts). In contrast, the X chromosome shows the lowest level of gene expression variation between the embryos of inbred *D. melanogaster* strains ($P = 1.16 \times 10^{-9}$; Figure 1B), ranging from 7% up to 10% lower (Table S3). Bootstrap resampling of the mean divergence across chromosomes confirms that it is significantly higher on the X between species (Figure 1C) and significantly lower on the X between strains (Figure 1D). In the between-species data, several specific branches in the phylogeny have significantly longer mean lengths judged by bootstrapping individual branches (Figure S1).

In the adults, mean divergence on the X is not higher than the autosomes in females ($P = 0.99$; Figure 2A; Table S4) yet gene expression variation is significantly lower on the X relative to the autosomes in female inbred strains ($P = 7.28 \times 10^{-6}$; Figure 2B; Table S5). In adult males, mean divergence is highest on the X, although it is not significant ($P = 0.35$; Figure 2E; Table S6), but

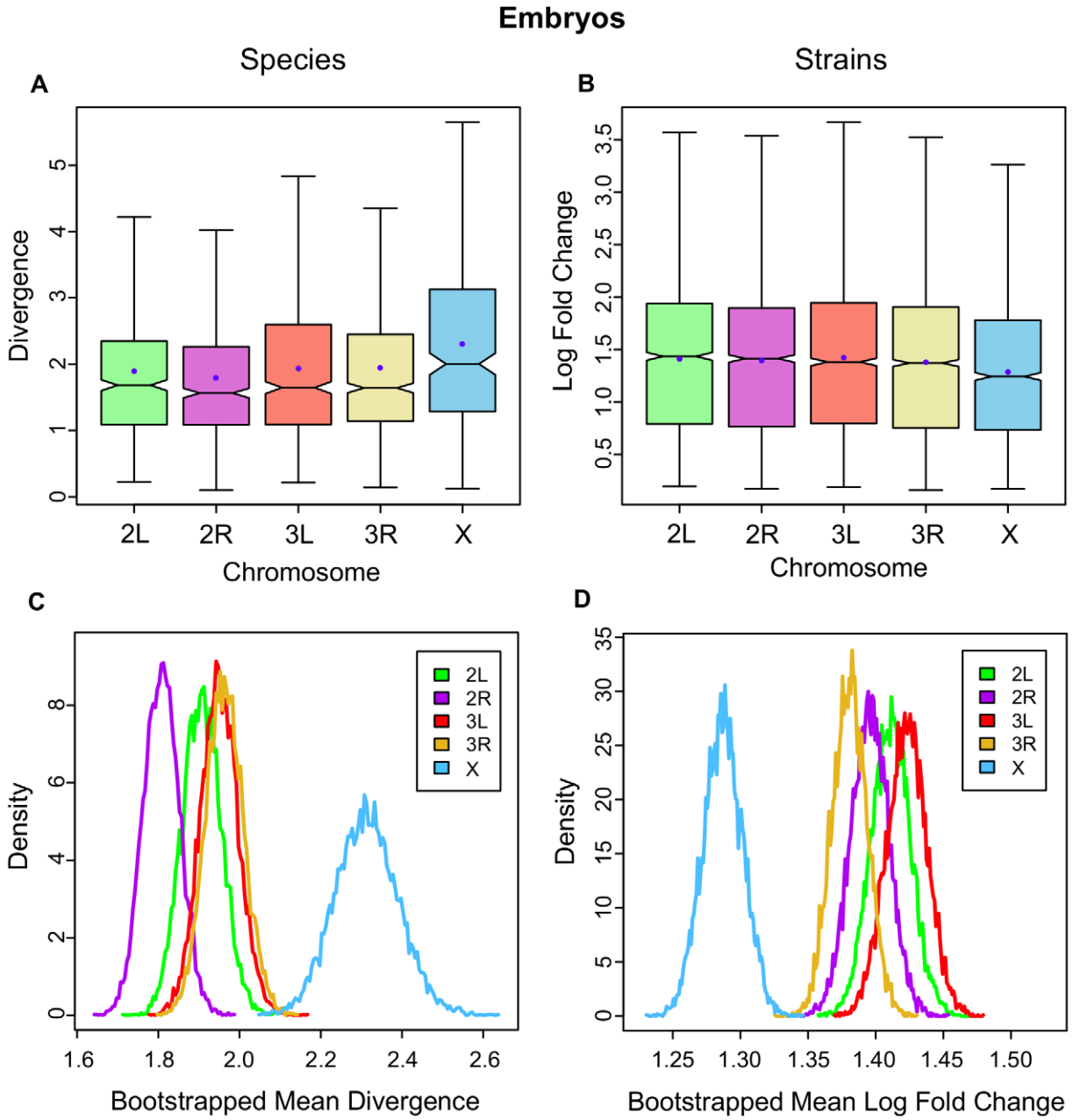

Figure 1. Gene expression divergence is higher on the X chromosome in *Drosophila* embryos and lower in *D. melanogaster* strains. The distributions of per gene expression divergence between Drosophila species separated onto each chromosome for A, embryos, and B, inbred strains of *D. melanogaster*. Divergence is measured per gene as the summed branch lengths for each gene tree for between-species data, and as mean log fold change for inbred strains as described in the Methods. Boxes show the upper and lower quartiles together with the median, error bars encompass data within 1.5 times the inter-quartile range, and blue circles indicate the means. Panels C and D show, for embryos and strains respectively, the distribution of 10,000 bootstrapped mean divergences for each chromosome using frequency polygons.
doi:10.1371/journal.pgen.1003200.g001

once again mean variation is significantly lower on the X in inbred strains ($P = 9.89 \times 10^{-11}$; Figure 2F; Table S7). Bootstrap resamples confirm that differences between the chromosomes are significant only in the strains (Figure 2C, 2D, 2G, 2H). When we reduce genes and species to a common set belonging to both the embryonic and adult between-species data, we find that the X remains more significantly divergent in the embryonic data (Tables S8, S9). In addition, we find that genes with sex-biased expression patterns also do not display an X effect in either sex confirming that the absence of any effect in adults is not caused by combining genes with different properties in the two sexes (see Methods; Figure S2).

We find that divergence on the X in embryos is not driven by a small subset of time points (Figure 3), nor can it be explained by artifacts caused by extreme expression levels (Figure S3) or by skews in the sex ratio (Figure S4; see Methods). Overall, these results indicate that there is a strong and significant excess of gene expression divergence on the X chromosome in *Drosophila* embryos together with a significant reduction of gene expression variation on the X within inbred strains of *D. melanogaster*. Divergence

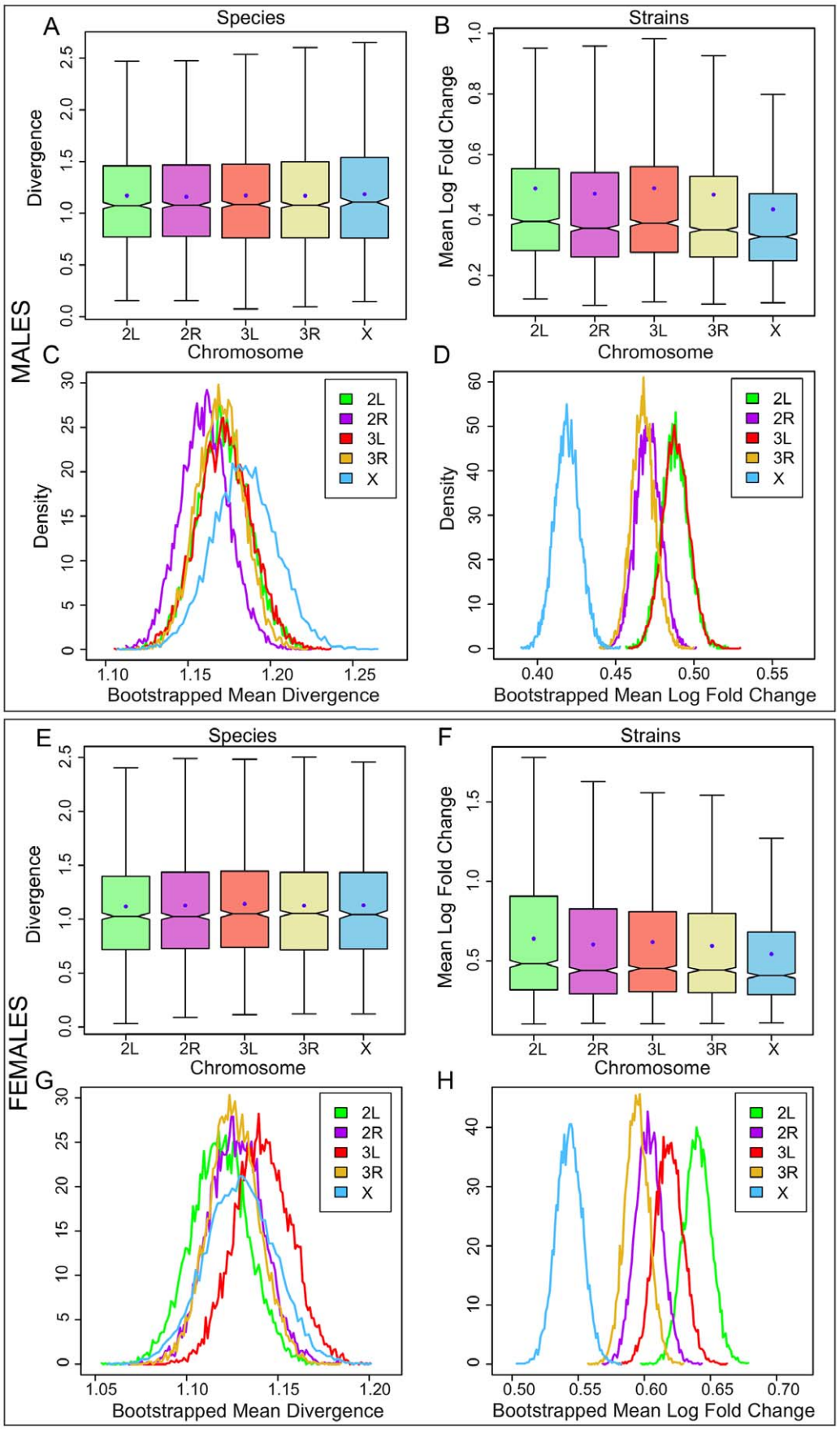
MALES
A
Species
Divergence
2L
2R
3L
3R
X
Chromosome
B
Strains
Mean Log Fold Change
Chromosome
C
Density
Bootstrapped Mean Divergence
D
Bootstrapped Mean Log Fold Change
FEMALES
E
F
G
H

Figure 2. Gene expression divergence is not higher on the X chromosome in *Drosophila* adults but is lower in *D. melanogaster* adult strains. The distributions of per gene expression divergence between Drosophila species separated onto each chromosome for A, adult males, B, inbred adult male strains of *D. melanogaster*, E, adult females, and F, inbred adult female strains of *D. melanogaster*. Divergence is measured per gene as the summed branch lengths for each gene tree for between-species data, and as mean log fold change for inbred strains as described in the Methods. Boxes show the upper and lower quartiles together with the median, error bars encompass data within 1.5 times the inter-quartile range, and blue circles indicate the means. Panels C, D, G, and H show, for adult males, inbred adult strains, adult females, and inbred adult female strains respectively, the distribution of 10,000 bootstrapped mean divergences for each chromosome using frequency polygons.
doi:10.1371/journal.pgen.1003200.g002

between species coupled with conservation within species is often viewed as a signature of adaptive evolution, and, at the least, is firm evidence against the observed divergence being driven by a relaxation of selective constraints.

Higher divergence on the ancestral branch of the neo-X in *Drosophila* embryos

In the obscura sub-group, Muller's element D (3L in *D. melanogaster*) has become X-linked and is referred to as a neo-X chromosome. If X-linkage were the cause of increased expression divergence, then we would expect to see accelerated evolution of gene expression on this chromosome relative to the remaining autosomes in this lineage [20]. As with the global X-effect, we see a small but significant increase in divergence on the ancestral branch of the obscura sub-group in the between-species embryonic dataset ($P = 0.0012$, Wilcoxon one-tailed test; Figure 4A). While the ancestral branch shows an excess of divergence (Figure 4A), the terminal branches do not (Figure S5). In the adult dataset, there is only one species in the obscura sub-group, and the branch leading to this species does not show an excess of divergence (Figure 4B).

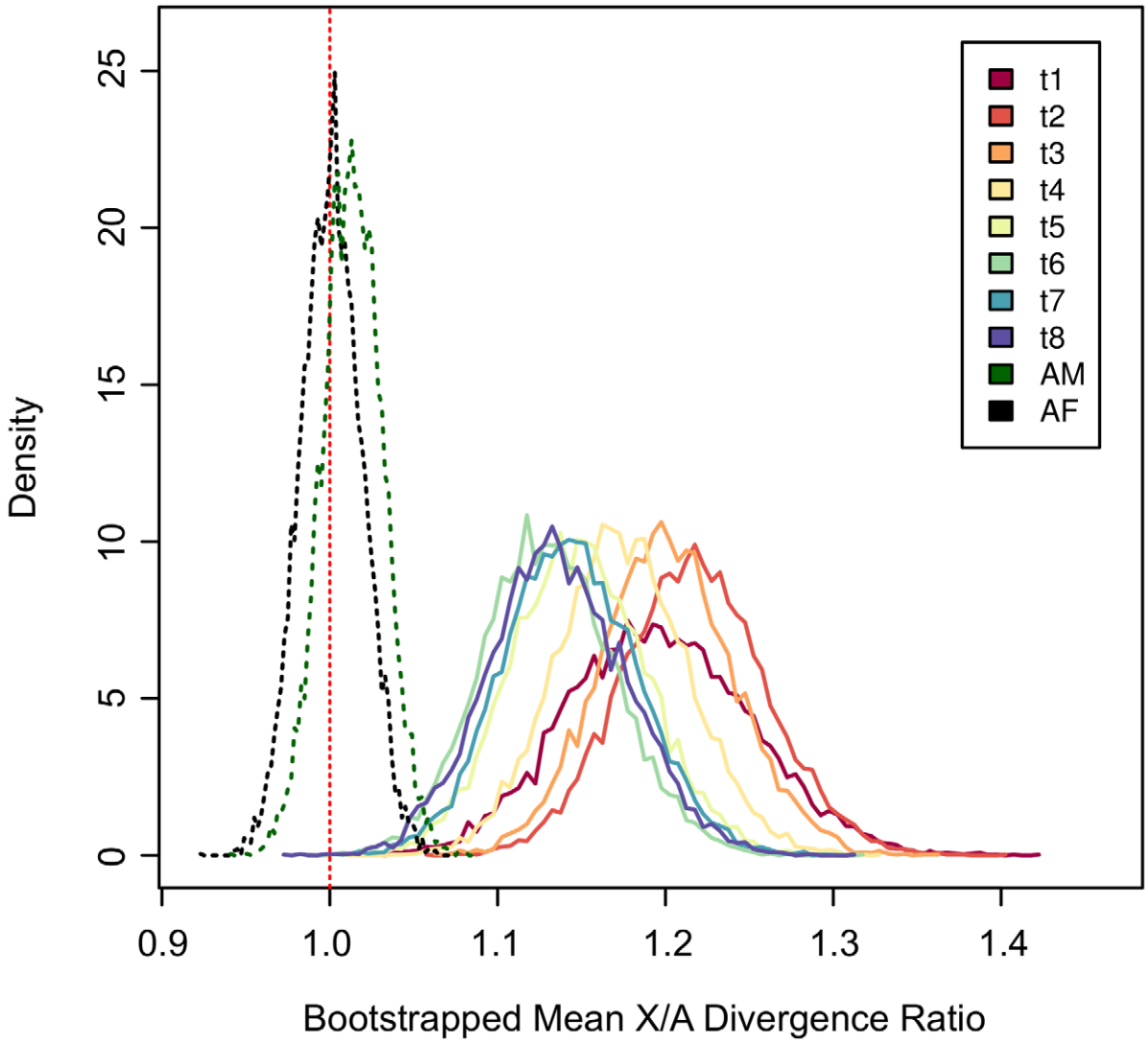

Figure 3. The X chromosome exhibits an excess of divergence throughout exmbryogenesis. Bootstrapped mean X/A divergence ratios for each time point throughout embryogenesis. Genes were resampled 10,000 times on each chromosome and the X/A ratio was scored for each time point separately. Bootstrapped distributions are shown as frequency polygons. Dashed green and black lines represent adult males (AM) and adult females (AF) respectively, and the vertical dashed red line marks an X/A ratio of 1.
doi:10.1371/journal.pgen.1003200.g003

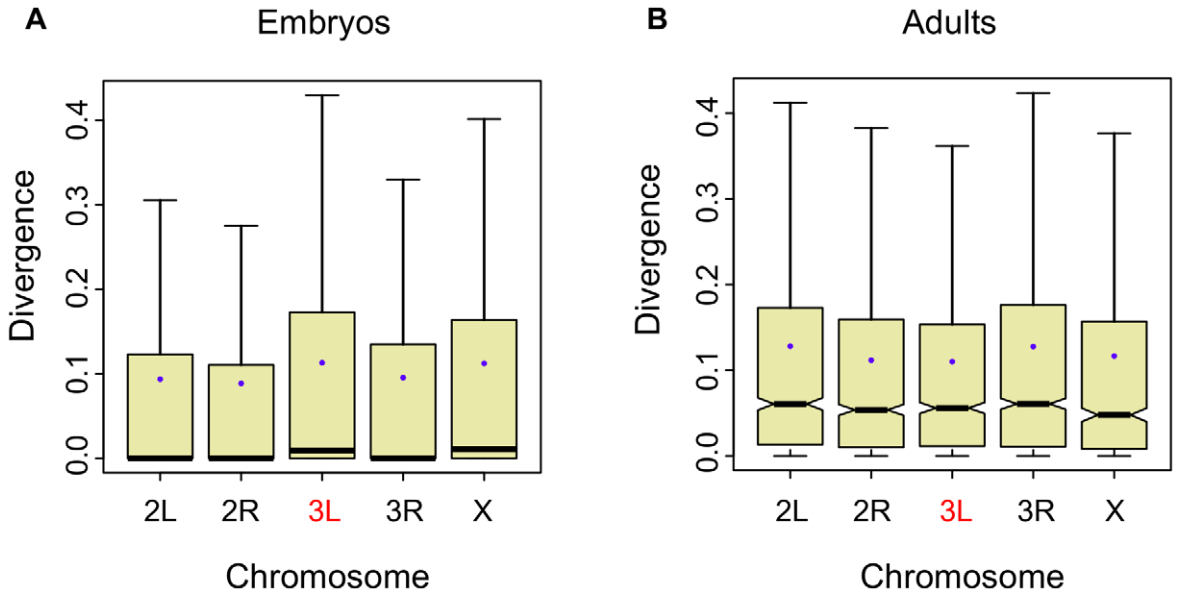

Figure 4. Expression divergence is higher for the ancestral branch of the neo-X (Muller element D). A, Per-gene, per-chromosome distributions of the length of the ancestral branch leading to the obscura sub-group (*D. persimilis* and *D. pseudoobscura*; see Figure S1) in the embryonic data divided by the sum of all branch lengths (3L is the neo-X chromosome in the obscura sub-group). B, Per-gene, per-chromosome distributions of the length of the branch leading to *D. pseudoobscura* in the adult data divided by the sum of all branch lengths.
doi:10.1371/journal.pgen.1003200.g004

An excess of gene expression divergence on the ancestral branch leading to the obscura sub-group for the neo-X suggests that evolution of this chromosome was accelerated more after its formation. More generally, this finding lends independent support to the notion that the X evolves more rapidly than the autosomes.

Lower mutational heritability on the *Drosophila* X

The discovery that *Drosophila* embryos have both an excess of divergence on the X chromosome between species (Figure 1A) and significantly lower levels of gene expression differentiation between strains of a single species (Figure 1B) is a pattern consistent with what we would expect to be driven by adaptive evolutionary processes. However, such a pattern could also be explained by random genetic drift since lower effective population sizes limit the amount of genetic variance a species can harbour [34] while simultaneously leading to the divergence of separate species through the accumulation of chance variations along separate lineages.

To determine whether it is likely that the X chromosome in *Drosophila* could accumulate mutations at a faster rate than the autosomes simply by virtue of being in a hemizygous state in males, we analysed data from mutation accumulation lines of *D. melanogaster* [35]. Twelve lines of *D. melanogaster* were allowed to accumulate mutations over a period of 200 generations. Since selection is relaxed in these lines, mutations are free to accumulate in the population and if the X has a biased accumulation of mutations due to its hemizygosity, we would expect an excess of gene expression variation between mutation accumulation lines for genes expressed on the X than for those on the autosomes. Gene expression was measured genome-wide at the late larval and puparium formation stages of the life-cycle. After fitting linear models to the data, the authors extracted the variance attributable to mutations and scaled it by the residual variance to give a measure of mutational heritability [35]. Mutational heritability is a dimensionless quantity, defined as the variance in a trait which is attributable to new mutations in each generation divided by the variance attributable to environmental variance (in an initially homozygous population) [36]. Thus, this measure captures the rate of increase in the heritability of a trait due to mutations. The trait of interest for us is gene expression, and this metric allows us to infer how quickly different mutation accumulation lines diverge from one another in terms of the accumulation of mutations affecting gene expression at individual genes.

The results show that, when we restrict the genes to those that have a measurable mutational heritability, the X has the lowest mutational heritability at both life-cycle stages ($P = 5.7 \times 10^{-8}$, Figure 5A; $P = 0.0143$, Figure 5B, Wilcoxon one-tailed tests). In addition, when we include those genes that do not have a measurable mutational heritability, we find that the X has both more genes with zero mutational heritability and less genes with a measurable mutational heritability than would be expected by chance (Figure 5C, 5D). These results suggest that, for these developmental stages at least, the fixation by random drift of mutations influencing gene expression is not biased on the X chromosome and hence is unlikely to be driving higher gene expression divergence on this chromosome. We note, however, that the mutation accumulation lines do not necessarily perfectly capture the conditions experienced by wild populations of *Drosophila* and so we believe it is important to conduct further studies designed to answer the question of whether the X fixes more mutations due to its hemizygosity.

A paucity of genes expressed in the cellular blastoderm on the *Drosophila* X

It was recently discovered that there is a paucity of adult tissue-specific gene expression on the *Drosophila* X chromosome [37]. This result suggests that the distribution of genes across chromosomes may influence observed differences in chromosomal rates of evolution. To test whether X chromosome genes have unusual embryonic tissue expression patterns, we used a controlled vocabulary of embryonic expression terms based on *in situ* expression data [38] to ask if there is under- or over-representation

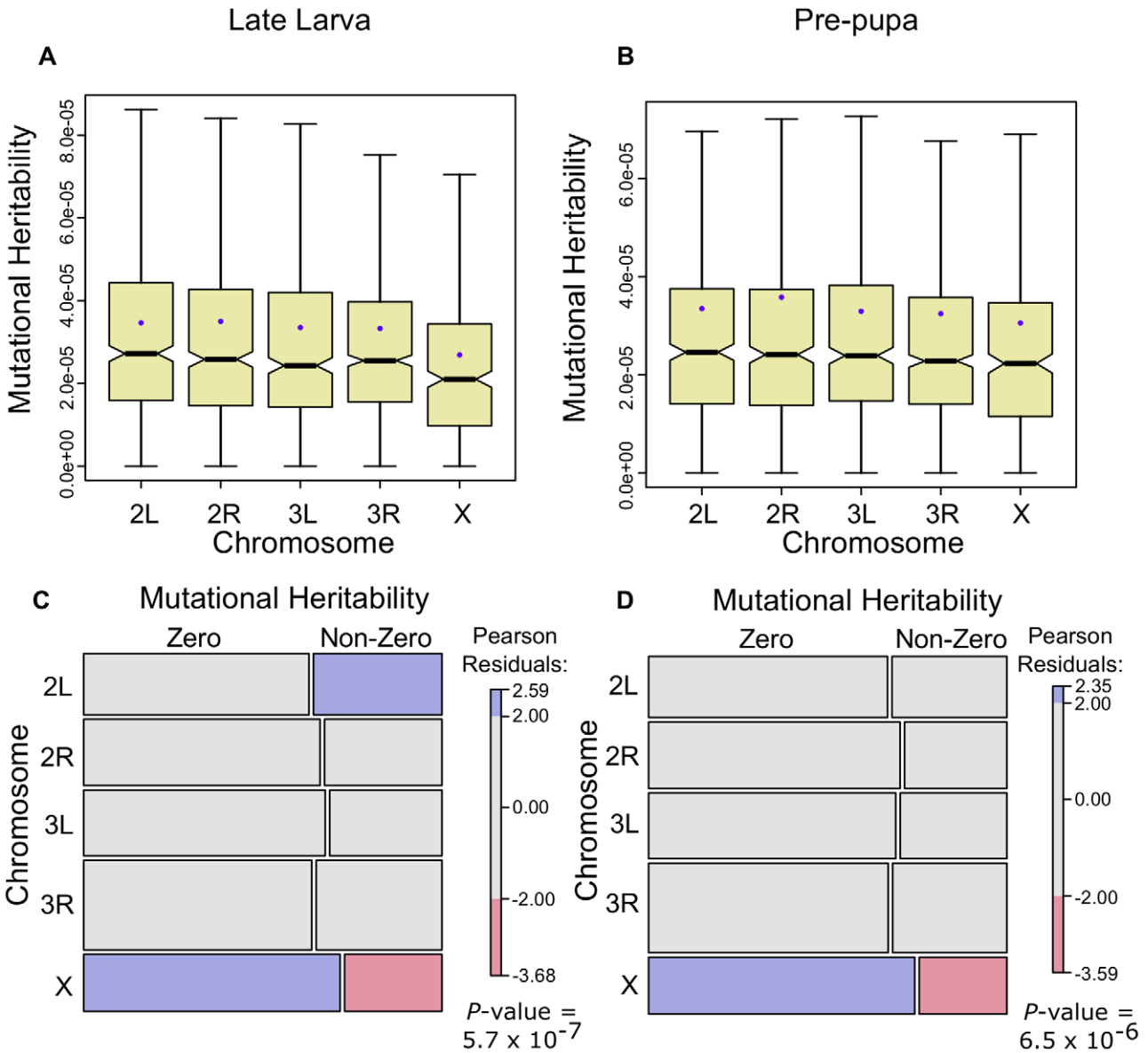

Figure 5. Gene expression mutational heritabilities are lower for the *Drosophila* X chromosome. Gene expression mutational heritabilities, estimated from mutation accumulation lines of *D. melanogaster* [35], separated onto chromosomes. Genes with measurable mutational heritabilities are shown for the late larva (A) and the pre-pupa (B). In C and D genes are categorized as displaying zero or non-zero mutational heritabilities for late larva and pre-pupa respectively and depicted using mosaic plots where the area in the rectangles is proportional to the number in that category combination. Pearson residual shading is used to depict deviations from null expectations – blue (excess) and red (paucity) colours indicate deviations from the expectation under the null hypothesis that the two variables, mutational heritability and chromosome, are independent [85]. *P*-values refer to the probability of independence (Chi-squared test).
doi:10.1371/journal.pgen.1003200.g005

of expression terms for genes on the X relative to the whole genome. After correcting for multiple testing, just one term showed a significant departure from its null expectation; genes expressed in the cellular blastoderm are significantly under-represented on the *Drosophila* X ($P_{adj} = 9.5 \times 10^{-5}$; Table S10).

This result makes sense when we consider that dosage compensation of X-expressed zygotic genes in male embryos via the MSL (Male-specific lethal) complex is not fully active until after the blastoderm stage [39,40]. The lag in activation of MSL-mediated dosage compensation may disfavour cellular blastoderm expressed genes from residing on the X, especially as they would need to evolve an alternative dosage compensation mechanism [40]. More generally, the absence of strong tissue-expression biases on the X chromosome suggests that an unusual chromosomal distribution of tissue-specific embryonic genes is unlikely to be driving the higher gene expression divergence that we find on the X chromosome.

The multi-locus faster-X effect with epistasis and linkage

Recent evidence suggests that epistatic interactions between genes constitutes a substantial fraction of the variation of quantitative traits in *Drosophila* [41]. Therefore, to determine the

relative benefits of chromosomal location and multi-locus co-evolution for beneficial alleles sweeping to fixation in a population, we analysed several diploid population genetic models of the faster-X effect. To compare evolution in equivalent genetic scenarios, we used the ratio of the selection gradient for X-linked versus autosomal cases (see Methods).

The results show that, although a faster-X effect exists in all the cases studied, by far the greatest advantage of X-linkage occurs when both epistatically interacting loci are linked on the same chromosome (Figure 6, blue circles; Table S11). When both loci are X-linked there will be no recombination in the heterogametic sex, and this will contribute to an increase in the rate of build-up of linkage disequilibrium between the loci. However, in species such as *D. melanogaster* there is also no recombination occurring between pairs of homologous autosomes in males, and therefore such an effect would contribute to increased evolution on the autosomes. To quantify the magnitude of this effect, we compared the X-linked case to a scenario in which there is no recombination between autosomally linked loci in males. The results show that the effect of a lack of recombination in males cannot account for the advantage enjoyed by X-linked loci, which when compared against the autosomal case in which there is male recombination shows that the advantage in this case is weak and dependent upon high-levels of genetic variance (Figure S6). Thus, the benefit of X-linkage in the multi-locus case accrues almost entirely from the increased efficacy of selection when acting on hemizygous males.

When positively-interacting alleles are located on separate chromosomes, it is extremely unlikely that they will sweep to fixation within a plausible time period because recombination will very effectively decay the linkage disequilibrium that is built up by selection in each generation [42]. When located on the same chromosome, interactions between loci could be considered to be either *cis-trans* or *cis-cis* interactions [42], thereby broadening the scope of possible genetic scenarios that are consistent with faster-X evolution. It remains possible, however, that beneficial *trans*-acting variants located on the autosomes, and interacting with fixed *cis* alleles on the X, are responsible for the excess of divergence that we find on the X. However, there are no reasons to suppose that such interactions ought to be biased in the direction of *trans*-autosomal to *cis*-X, since, due to symmetry, the opposite scenario of *trans*-X to *cis*-autosomal appears to be just as likely. Indeed, in a recent study of gene expression in hybrids of *D. yakuba* and *D. santomea*, hybrid male mis-expression was found to be greater for autosomal genes, most likely as a result of faster evolution of X-linked *trans*-acting factors [43]. Thus, the available evidence suggests that if there is a bias in positive species-specific interactions between the X and the autosomes, it is in the direction of *trans*-X to *cis*-autosomal. Overall, both theory and data support the notion that during adaptive evolution, X-linked alleles have a capacity to sweep to fixation faster than their autosomal equivalents, and this effect is greatly enhanced when there are beneficial interactions between two or more loci.

Higher co-ordination of gene expression in embryos relative to adults

In a recent study of gene expression evolution in mammals, evidence was reported for a faster-X effect [44] (although a separate study found no evidence for a faster-X effect for gene expression in two species of mice [45]). The authors correlated gene expression across homologous chromosomes in species pairs and used one minus Spearman's correlation coefficient as a measure of divergence. The same approach has also been used recently to find an excess of divergence on the X in adult males and females of *Drosophila* species [46]. Thus, we can ask why this correlation-based measure of divergence uncovers an X-effect in adults when our per-gene expression-level measure of divergence does not (at least not globally – see Figure S7).

To aid our search for an answer to this question, we first applied the correlation method to both embryos and adult males and females in the datasets that we have used. The results show that the X chromosome has a reduced cross-species correlation relative to the autosomes in the embryos (Figure 7A), just as it has in both adult males and females (Figure 8A,B; all pair-wise comparisons are shown in Figure S8) [46]. However, when we use an absolute distance metric to determine the per-chromosome differences between species, we find that, while the X consistently displays a greater distance between species in embryos (Figure 7B), in adults the X chromosome is largely equivalent to the autosomes (Figure 8C, 8D; Figure S9). Thus, the question arises as to why the X chromosome appears more divergent in terms of correlations but not in terms of distances?

The answer must be sought in the component of gene expression divergence that each measure is capturing. Spearman's rank correlation coefficient is a dimensionless number that in the context of gene expression in two species, determines the extent to which expression relationships between genes are retained across the two species, and the strength of the correlation is insensitive to absolute expression differences (Figure S10). Thus, this measure of divergence captures how co-ordinated expression is across a specific set of genes in two different species. In contrast, absolute distances, and per-gene expression changes, measure to what extent individual genes differ in expression level in two species, and these metrics are insensitive to how co-ordinated expression is between different genes. This suggests, therefore, that gene expression on the X chromosome in adults is weakly co-ordinated relative to expression on the autosomes even though absolute expression differences are not significantly greater on the X (Figure S10).

Furthermore, when we compare the chromosomal correlations in embryos and adults, we find that embryos have much higher correlations overall than the adults even when we reduce them both to a common set of genes and species (Figure S11). This suggests that gene expression is generally more highly co-ordinated in *Drosophila* embryos relative to adults.

Discussion

We have presented evidence that gene expression in *Drosophila* embryos evolves faster on the X chromosome between species, but slower on the X chromosome within species (Figure 1). The salience of this result is substantially strengthened by the discovery that the Muller D element has a significantly longer ancestral branch leading to the obscura sub-group in the embryonic data (Figure 4A). The Muller D element segregates as a neo-X chromosome in the obscura sub-group (*D. persimilis* and *D. pseudoobscura* in our data), and therefore provides a powerful, independent test for faster evolution of the X chromosome. In addition, we find that gene expression evolves faster on the X chromosome in embryos when we employ a more global measure of expression divergence (Figure 7A), a measure which we find can vary independently of per-gene expression level divergence (Figure 8, Figure S10). In what follows, we discuss different potential interpretations of these results.

Adaptive versus non-adaptive evolution

The excess of gene expression divergence that we find in the embryonic data could be driven by a relaxation of selective constraints acting on X-linked gene expression. We would predict

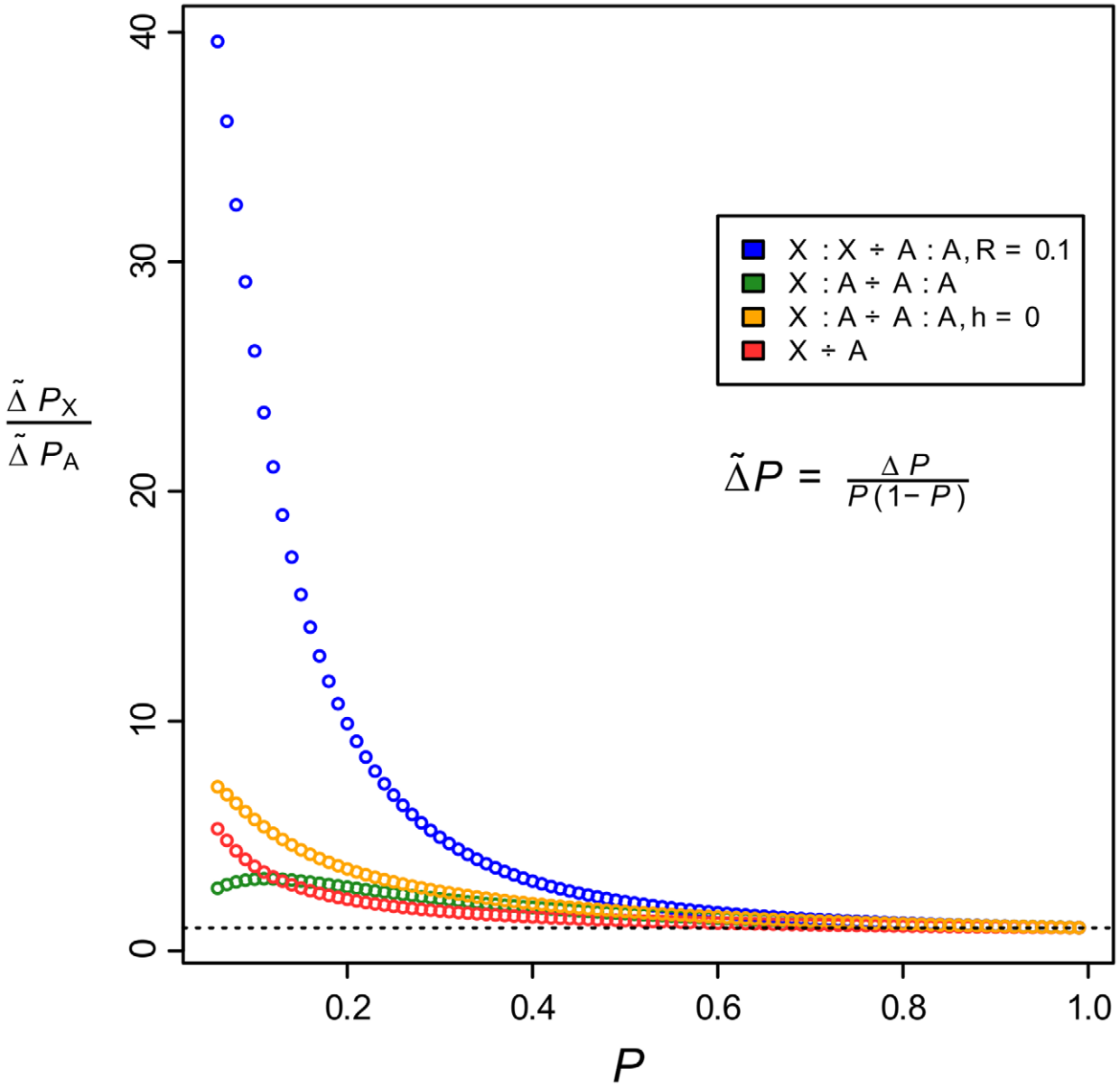

Figure 6. The faster-X effect is greatest when beneficially-interacting loci are linked on the same chromosome. The ratio of selection gradients for X-linked models versus their equivalent autosomal cases as a function of allele frequency. Blue points represent the case where both loci are linked on the same chromosome, orange and green points represent the case where the loci are on different chromosomes, and the red points are for the one-locus scenario. Unless otherwise stated in the legend, recombination rates, R, are equal to 0.5 (free recombination) and the dominance coefficient, h, is 0.01 ($h=0$ is close to identical to $h=0.01$ in the one-locus case and hence is not shown). The dashed line indicates a ratio of 1.
doi:10.1371/journal.pgen.1003200.g006

that relaxed selective constraints would lead to an elevation of within-species gene expression variation on the X, and, contrary to this prediction, we find that gene expression variation within inbred strains of *D. melanogaster* is significantly lower on the X relative to the autosomes (Figure 1B, 1D) suggesting that X-linked gene expression is not evolving under a relaxation of selective constraint. In support of this finding, we find a corresponding reduction in gene expression variation on the X in both adult males and females (Figure 2B, 2D, 2F, 2H) [46].

Nonetheless, it remains possible that elevated between-species variance coupled with diminished within-species variance is a consequence of random genetic drift, or demographic effects such as bottlenecks [3,47]. If the hemizygosity of the X chromosome in males, and the resulting potentially diminished effective population size of the X, were resposible for the lower within-species variance in X-linked gene expression, then we would expect to find an excess of fixation of X-linked gene expression mutations in separate mutation accumulation lines. However, we find the opposite pattern, that mutation accumulation lines display less gene expression variation for X-linked genes (Figure 5). Part of the reason for this could be due to the X chromosome presenting a smaller mutational target than the autosomes as a result of being in a hemizygous state in males, but this effect of hemizygosity will be present in wild populations of *Drosophila* as much as in lab-reared

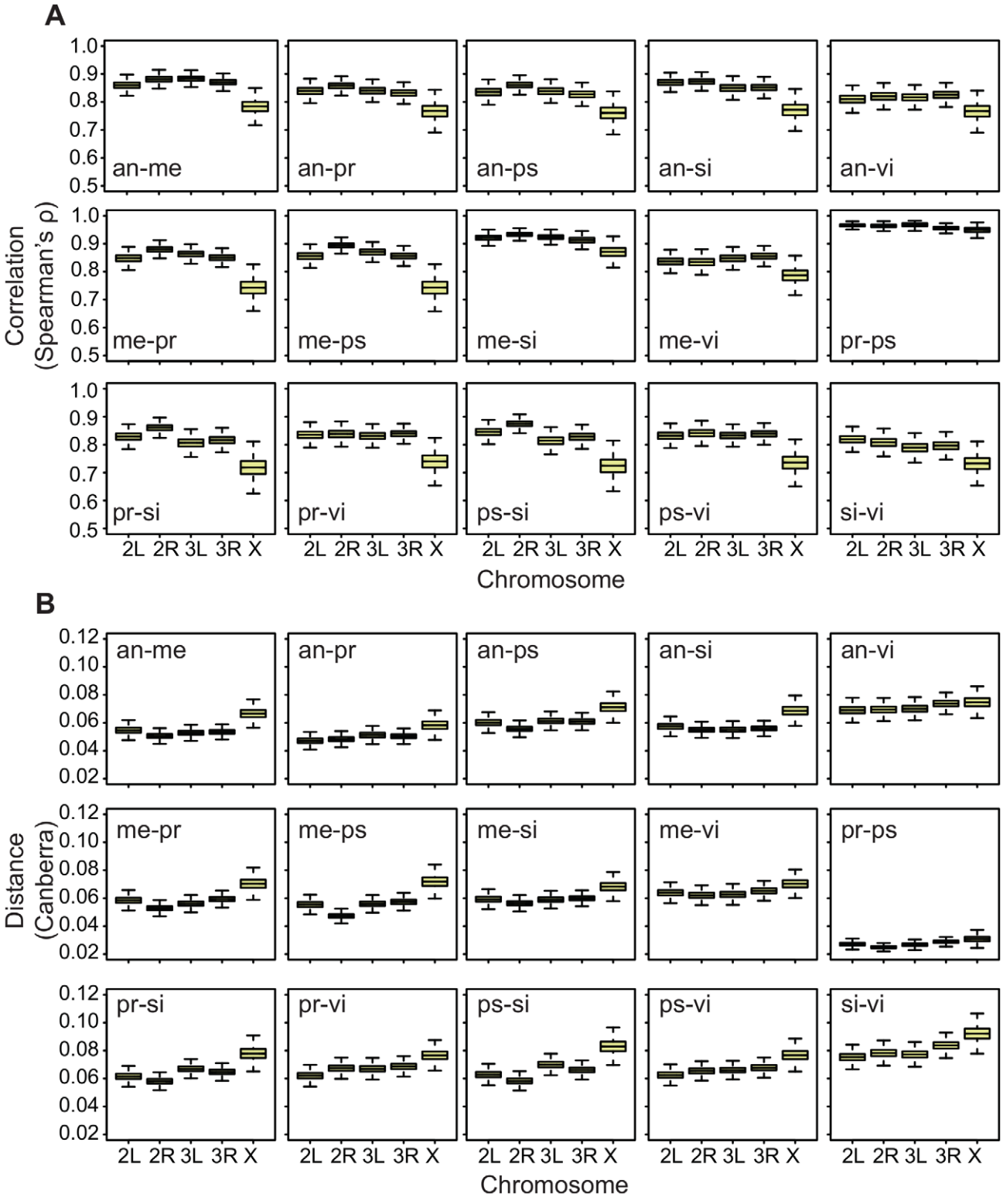

Figure 7. Divergence on the X in embryos is greater using both Spearman's ρ and the Canberra distance. Bootstrapped distributions of A, Spearman's ρ (divergence is $1-\rho$) and B, the mean Canberra distance across chromosomes in *Drosophila* embryos for all pair-wise species comparisons.
doi:10.1371/journal.pgen.1003200.g007

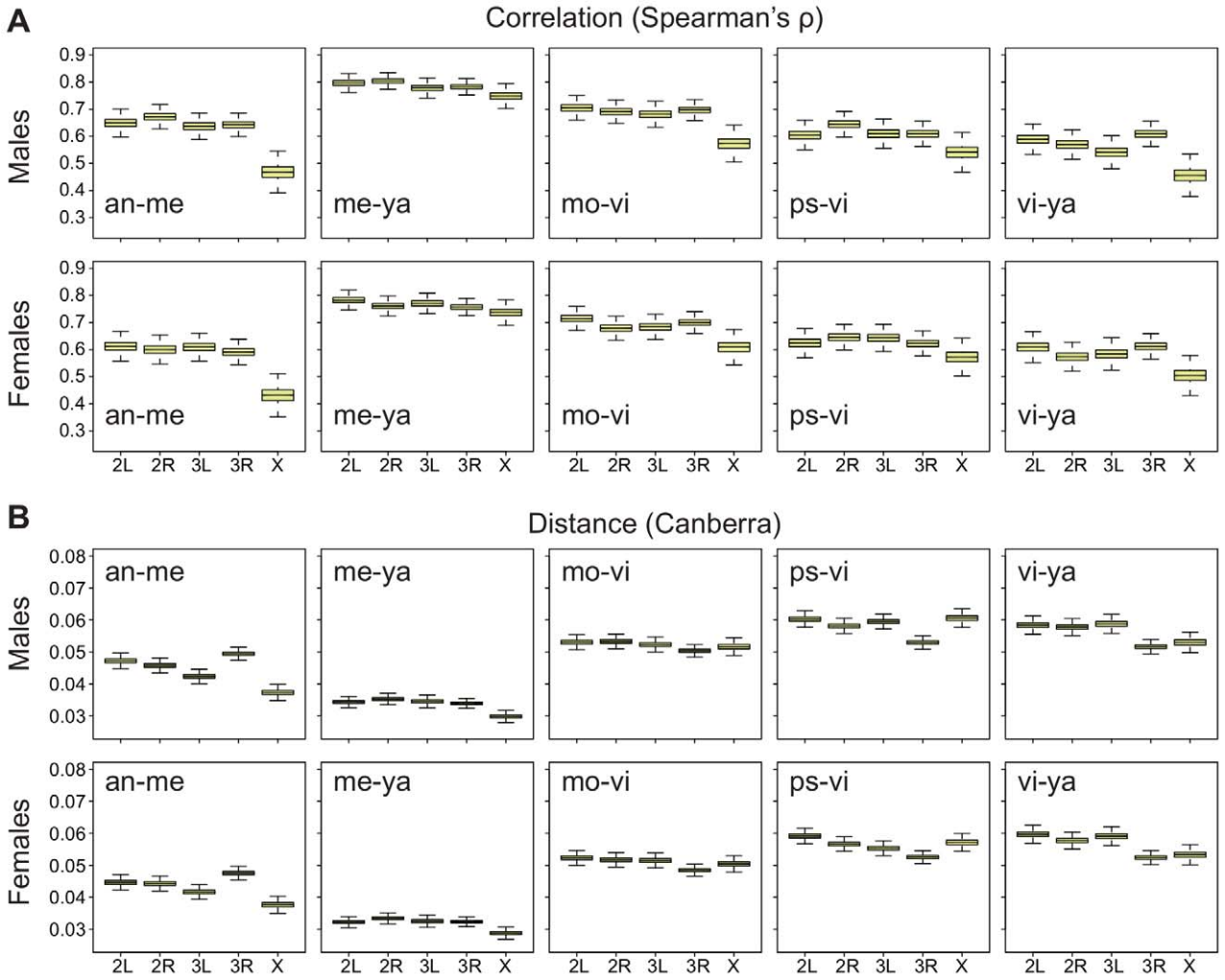

Figure 8. Divergence on the X in adults is greater using Spearman's ρ, but not the Canberra distance. Bootstrapped distributions of A, Spearman's ρ (divergence is $1-\rho$) and B, the mean Canberra distance across chromosomes in *Drosophila* males and females for a selection of pair-wise species comparisons (all pair-wise comparisons are shown in Figures S8, S9).
doi:10.1371/journal.pgen.1003200.g008

lines. It is also possible that, while the experimenters made every effort to neutralise the effects of mutations, selective effects remained in the accumulated mutations and that purifying selection is stronger on the X relative to the autosomes.

Prior studies have found that the X chromosome in *Drosophila* experiences more effective purifying selection against weakly deleterious and recessive mutations [48–51], and in non-recombining chromosomal regions, the X has been shown to experience the smallest reduction in the efficacy of selection [52]. In addition, studies of nucleotide diversity on the X in both coding and non-coding regions in *Drosophila* species suggest that adaptive processes best explain the observed variance on the X [29,47,53], including recent data showing that there is an absence of X-autosomal differences for putatively neutral sites [25]. Overall, our findings are consistent with there being an excess of adaptive evolution of X-linked gene expression, although this does not mean that drift or demographic effects are not involved in shaping gene expression evolution.

cis versus *trans* effects

Gene expression is influenced by both *cis*-acting regulatory sequences, and by *trans*-acting factors, such as transcription factors. Thus, while we observe an excess of X-linked divergence of gene expression, this could be the result of either *trans*-acting factors potentially located on other chromosomes, X-linked *cis*-acting variants, or a combination of both. Several studies have found evidence for both *cis* and *trans* effects influencing gene expression differences both within and between *Drosophila* species [54–59]. Thus far, however, the evidence suggests that there is an excess of *cis*-acting variants influencing divergence between species [54–56,60], and that *cis*-regulatory divergence increases with the divergence time between species [55,59]. One study reported an excess of *trans*-acting variation influencing gene expression in a comparison of *D. melanogaster* and *D. sechellia*, although as noted by the authors this could be related to the unusual demographic history and life-history evolution of *D. sechellia* [59].

It's possible that the excess of X chromosome divergence that we see is the result of a bias in the direction of autosomal *trans*-acting factors impacting the X chromosome more than the reverse situation of X-linked *trans*-acting factors affecting the autosomes. Current evidence suggests, however, that the opposite is the case – that there is a bias towards *trans*-acting factors on the X impacting autosomal *cis*-elements resulting in an excess of autosomal

mis-expression in *Drosophila* hybrids [43], including a study of mis-expression in hyrbid *D. simulans* males carrying an X-linked allele introgressed from *D. mauritiana* [61]. Therefore, if there are species-specific interactions between the X and the autosomes, it seems unlikely that they would be biased in such a way as to account for our results.

Theoretical considerations also do not favour the notion that *trans*-acting factors could be driving the majority of the divergence that we find, assuming that a substantial fraction of this divergence is adaptive. Mutations in *trans*-acting factors are more likely to be pleiotropic, and so should have less scope to influence adaptive evolution than the more modular effects of mutations in *cis*-regulatory regions [42,62–65]. Furthermore, population genetic models of the faster-X effect show that if there are two or more interacting loci with beneficial interactions between them, then X-linked loci enjoy a far greater benefit than autosomal loci (Figure 6). Whether adaptive changes occur in *cis* or in *trans* also has important consequences for the scope of mutations to have recessive or partially recessive effects on fitness, which in turn is of central importance for the faster-X phenomenon [5]. We address these issues towards the end of the Discussion.

Embryos versus adults

In the embryonic between-species data, we found evidence for faster evolution of gene expression on the X chromosome using two different measures of divergence (Figure 1A, Figure 7A). The first measure captures the change in expression levels on a per-gene basis (Figure 1A), and the second captures the extent to which gene expression relationships between genes have changed in pairs of species, and hence how co-ordinated expression is across a subset of genes (Figure 7A, Figure S10). In contrast, in the adults, we see evidence for higher divergence on the X chromosome using only the second measure of divergence (Figure 8A) and not the first (Figure 2A). This suggests that, while the X displays lower levels of co-ordinated expression in pairs of species in the adult, it does not exhibit significant differences in expression level on a per-gene basis. Then we must ask, why does the embryo diverge more on the X in terms of per-gene expression levels than the adults?

Embryogenesis is a highly dynamic process, driven by a cascade of gene expression unraveling through a highly co-ordinated developmental network leading to large batteries of genes being switched on and off at precise moments during development [66]. In contrast, in a fully developed adult, cells are largely fully differentiated, and gene expression is to a much lesser degree responding to a pre-determined developmental program, and is freer to respond to changes in the environment. Thus, it makes sense that we find gene expression to be overall much more highly co-ordinated in the embryo relative to the adults (Figure S11). But it is precisely because of the broad dynamic range of embryonic gene expression, with a large fraction of the zygotic genome being activated in a series of waves as embryogenesis proceeds (Figure S12), that even subtle shifts in timing could potentially produce large differences in expression levels. In a whole adult fly, however, genes are likely expressed in subsets of tissues and organs such that we will not find extremely low or high expression levels for most genes when we extract RNA from all of the tissues simultaneously, thereby diminishing the dynamic range of the data. Therefore, our results highlight the need to perform more precise organ-by-organ comparisons of gene expression in future between-species studies of adult flies. In addition, our analysis draws attention to the different components of divergence that are captured by different measures of gene expression divergence.

The faster-X hypothesis

Taking the above considerations and all of our results into account, we believe that the X effect we find in the embryos is best explained within the framework of the faster-X hypothesis. This does not mean that all of the divergence we see is driven by adaptive substitutions in *cis*-regulatory regions on the X chromosome, but rather that the excess of X chromosomal divergence that we find together with the reduction of expression variation in inbred strains of *D. melanogaster* is most consistent within an adaptive evolutionary scenario. In support of this interpretation, researchers found an excess of adaptive substitutions on the X chromosome in a long-term evolution experiment involving lines of *D. melanogaster* selected for increased rates of egg-to-adult development [67]. An interesting theoretical corollary of the fast-X interpretation is that it suggests that adaptive substitutions are more likely to occur via new mutations than from standing genetic variation [68].

If we adopt a faster-X interpretation of the data, then we must provide some explanation as to why beneficial *cis*-regulatory mutations have recessive or partially recessive effects on fitness, in keeping with the original model [1]. Current evidence in adult *Drosophila* species suggest the opposite, that *cis*-acting variants have largely additive effects relative to *trans*-acting factors, which show more deviations from additivity towards dominance and recessiveness [55,59]. However, these experiments determine the additivity of the phenotype of a *cis* variant (where the phenotype is its gene expression level), and not necessarily its effect on fitness. Theory suggests that mutations could have fitness consequences that are non-linear even if they have additive phenotypic effects [69]. Therefore, it is possible that phenotypic measures of *cis*-acting elements fail to capture their effects on fitness.

To understand the fitness effect of a mutation in an organismal context, we must focus on the biology of the organism, and not just on its genetics. One potential route towards non-additive intra-locus effects on fitness is canalisation. The canalisation of embryonic development, such that it is resistant to environmental or genetic perturbations, has long been recognized as a crucial element contributing to the evolution of robustness in developmental systems [70]. The evolution of dominance is a means by which the components of a network could become canalised [71–74]. While selection acting on modifiers of dominance will typically be weak (of the order of the mutation rate), it can be substantially stronger in non-equilibrium populations where genetic variation is maintained at high levels by processes such as migration and hybridisation [72,74]. The notion that the evolution of robustness (i.e., an attempt to prevent change of the phenotype) could lead to faster evolution of the X may seem counter-intuitive. However, the relationship between robustness and evolvability is well established, and suggests that the evolution of phenotypic robustness can often facilitate adaptive evolution [75–77]. We present this scenario partly to illustrate that the biological details of an individual species, such as species range and migratory pressures, might play a significant role in determining how its chromosomes evolve.

Outlook

We report evidence that gene expression evolves faster on the X chromosome in *Drosophila* embryos. While our results are consistent with adaptive evolutionary processes, more work is required to unravel the details underpinning this excess of divergence at the genetic, phenotypic, and fitness levels. We contend that variations in biological and life-history details, such as differences in dosage compensation menchanisms, can strongly impact how the chromosomes of different species evolve. We

therefore stress the importance of appreciating biological context when attempting to understand chromosomal evolution. Deciphering the relationship between species-specific biology and chromosomal patterns of evolution promises to provide fertile ground for future research.

Methods

Embryo collections and RNA isolation and labeling

We used inbred strains of *D. melanogaster*, originally collected from farmer's markets in North Carolina and provided as a resource by the Drosophila Genetic Reference Panel (DGRP; http://dgrp.gnets.ncsu.edu/) [33]. Seventeen strains were selected for the collection of 0–2 hour old embryos.

Populations of healthy adults from 3–7 days of age, were reared at 25°C and used for embryo collections. To synchronize the age of the embryos in each sample, we pre-laid the flies three times for 1 hour with a fresh apple juice plate with yeast paste before every collection. Another fresh plate with yeast was used to collect the embryos. After collection, embryos were rinsed with distilled water and then dechorionated in 100% bleach for 2 minutes before being washed in desalinated water. The embryos were then transferred into a 1.5-ml tube and snap-frozen in liquid nitrogen and stored at -80°C. Three biological replicates were collected for each strain.

To isolate RNA, embryos were thawed on ice and homogenized with a pellet pestle and a pellet pestle cordless motor (Kontes). RNA was isolated with the RNeasy Mini kit (Qiagen) and eluted with 30 ml of distilled water. The RNA concentration was measured with the NanoDrop spectrophotometer and RNA quality was assessed with Bioanalyser using the Agilent RNA 6000 Nano kit.

To prepare samples for hybridization to the chip, we followed the Agilent One-Colour Microarray-Based Gene Expression Analysis protocol version 6.5 (Low Input Quick Amp Labeling). The starting amount of RNA was normalized to 100 ng for all samples.

Gene expression data sets

Embryonic expression in *Drosophila* was taken from a species-specific microarray data set, in which eight time-points were sampled for the duration of embryogenesis of *D. melanogaster*, *D. simulans*, *D. ananassae*, *D. pseudoobscura*, *D. persimilis*, and *D. virilis* [30]. Adult *Drosophila* expression was collected from a microarray experiment that measured the gene expression of whole flies sorted into males and females and taken from *D. melanogaster*, *D. ananassae*, *D. mojavensis*, *D. pseudoobscura*, *D. simulans*, *D. virilis*, and *D. yakuba* [31]. Gene expression mutation accumulation data was taken from a microarray study of mutation accumulation lines of *D. melanogaster* [35]. Adult *D. melanogaster* strain data was taken from a whole-genome microarray study of gene expression in whole adult flies from 40 inbred strains separated into males and females [32].

Measures of chromosomal expression divergence and differentiation

To quantify gene expression divergence in a chromosomal context, we fitted the following linear model [78] to log_2 gene expression measures, y_{ijkl},

$$y_{ijkl} = \mu + S_j + C_k + GC_{i(k)} + SC_{jk} + GCS_{i(k)j} + e_{ijkl}$$

where S_j is the effect of the j'th species, C_k is the effect of the k'th chromosome, and $GC_{i(k)}$ is the effect of the i'th gene nested in the k'th chromosome. The interaction between the j'th species and the i'th gene nested in the k'th chromosome, $GCS_{i(k)j}$, provides information about species-specific chromosomal expression of a gene and is given by

$$GCS_{i(k)j} = \bar{y}_{ijk.} - \bar{y}_{i.k.} - \bar{y}_{.jk.} + \bar{y}_{..k.}$$

where values are averaged over missing subscripts indicated by dots. Thus, the effect of the i'th gene in the j'th species is the excess that cannot be explained by the expression of the i'th gene across species, the expression of the k'th chromosome in the j'th species, and the overall expression on the k'th chromosome. When there are multiple expression measures over a time-course, our measure of divergence is designed to detect translations up or down in expression level across the time course as a whole (see Figure S13).

Differentiation of gene expression between inbred strains was determined using the R package 'limma' [79]. Limma fits linear regression models to each gene separately. The differentiation of each gene was then scored as the mean log fold change of the gene across all pairwise strain comparisons.

Branch length analysis

Absolute pairwise species contrasts of the $GCS_{i(k)j}$ values were transformed into branch lengths using the Fitch-Margoliash least squares method (implemented in the PHYLIP program fitch) [80]. Negative branch lengths were set to zero, and for all genes the topology of the known phylogeny was used [81]. Per-gene expression divergence was then expressed as the sum of all of the branch lengths in each gene tree separately.

To test for acceleration on one lineage, for each gene we expressed the branch length of the focal lineage as a proportion of the total of all branch lengths. In the embryonic dataset we chose the ancestral branch leading to the common ancestor of *D. pseudoobscura* and *D. persimilis* but not including the terminal branches (Figure 4A). For the adult dataset, which does not have data for *D. persimilis*, we used the terminal branch leading to *D. pseudoobscura* (Figure 4B).

Resampling branch lengths

Mean summed branch lengths were bootstrapped by resampling the genes on each chromosome 10,000 times with replacement and in each bootstrap replicate calculating the mean summed branch lengths for the genes on each chromosome (Figure 1C, 1D). Individual branches in the embryonic and adult datasets were tested for an excess of divergence on the X chromosome using the number of bootstrap replicates in which mean autosomal branch lengths were greater than the mean on the X chromosome (Figure S1). All resampling was carried out using the R statistical programming environment [82].

In both of the *Drosophila* between-species data sets, the smallest sample of genes was on the X chromosome (Table S1). To determine whether the differences between the X and the autosomes could have been caused by a sampling bias on the X, we resampled the number of genes present on the X from the autosomes 10,000 times without replacement and each time recalculated the mean divergence. The distributions of these resampled means are shown in Figure S14.

Accounting for sex-biased expression in adults

Expression of genes in the adults can be biased towards one of the sexes [31], and it's possible that sex-biased genes might exhibit stronger differences in divergence across the chromosomes. We

focused on male and female-biased genes identified in [31] in each of the species. Genes that show a male-bias in at least one species show a significant excess of divergence in both males and females ($P_{male} = 1.57 \times 10^{-11}$; $P_{female} = 6.33 \times 10^{-6}$; Figures S15, S16) [83,84], and conversely female-biased genes are significantly more conserved in both males and females ($P_{male} = 6.27 \times 10^{-7}$; $P_{female} = 3.42 \times 10^{-10}$; Figures S15, S16). When we look at divergence across chromosomes, however, we find that sex-biased genes are not significantly more divergent on the X in either sex (Figure S2). Interestingly, when we restrict male-biased genes to those in *D. melanogaster* and *D. simulans* we do find a weak but significant excess of divergence on the X ($P = 0.0022$; Figure S7), which is absent for the same genes expressed in females ($P = 0.117$; Figure S7). The biological function of these genes is enriched for carbohydrate metabolism ($P_{adj} = 2.7 \times 10^{-6}$) and alcohol metabolism ($P_{adj} = 1.1 \times 10^{-6}$), which might suggest that these are genes that have evolved rapidly and relatively recently, thus preserving the signal of an excess of divergence on the X. Indeed, we find that these genes are significantly more divergent than average ($P = 1.0 \times 10^{-4}$; Figure S17).

The X-effect during embryogenesis

In the between-species embryonic data, our measure of divergence is designed to detect translations in expression up or down in different species across the embryonic time course as a whole (Figure S13). However, it remains possible that much of the difference that we detect between the X and the autosomes is driven by a subset of the time points. To test this, we extracted divergence measures from each time point separately. We then bootstrap resampled divergence measures for the X chromosome and the autosomes and in each bootstrap replicate calculated the ratio of mean X to mean autosomal divergence. The results show that at every time point the X chromosome displays an excess of divergence relative to the autosomes (X/A ratio >1; Figure 3). Furthermore, all of the resampled time point distributions heavily overlap with one another indicating that higher expression divergence on the X is not driven solely by one or a subset of time points.

Resampling according to gene expression level

Differences in gene expression divergence across chromosomes could be influenced by consistent differences in expression levels across chromosomes. In the between-species embryo data, the X chromosome has the weakest mean expression level (Figure S18), whereas in the adults, the X chromosome has the highest mean expression level (Figure S18). Higher expression in the adults could be a reflection of a paucity of adult tissue-specific expression on the X chromosome [37]. To elucidate the relationship between expression level and divergence in these data sets, we ranked genes by their expression level (lowest to highest), binned them into groups of 50 genes, and measured the deviation of each group's mean divergence from the global mean divergence.

The results show that for the embryos, the relationship is non-linear, with groups of the weakest expressed genes diverging less than the global average (Figure S19). Thus, although an increasing expression level does predict less divergence, divergence cannot be attributed simply to stochastic fluctuations of the weakest expressed genes. In the adults, the relationship is more linear, with the weakest expressed genes showing the highest divergence (Figure S19). Thus, higher expression on the X in adults may at least partly explain the lower levels of divergence relative to the embryos.

To clarify the relationship between expression level and chromosomal divergence, we bootstrap sampled genes from each chromosome while weighting their probability of being sampled according to their expression level. To sample genes according to expression level we weighted the probability of being sampled according to the cumulative distribution function of a normal distribution with a specified mean expression level and standard deviation. We defined the standard deviation as the standard deviation of the whole expression level distribution divided by the number of mean expression levels that were being sampled. Genes were then sampled with replacement 10,000 times for each mean expression level for each chromosome in both the embryonic and adult datasets. Fewer mean expression levels were taken for the adult data due to its lower expression level variance.

The results show that, in the embryo, divergence on the X is greater than the autosomes for intermediate gene expression levels, but not when expression is high or low (Figure S3A). In contrast to this result, in the adult data the X shows higher expression divergence when gene expression is low or high (Figure S3B). Thus, the higher expression divergence of the X in the embryos is not driven by expression levels at the extremes of the distribution.

Testing for sex ratio effects

While divergence on the X is not driven by particular periods during development, it is possible that there is a bias in the direction of expression differences between species. For example, if there was a persistent skew towards a male-biased sex ratio in one species relative to another and if dosage compensation in males was incomplete, then we would expect X-linked genes to show a skew towards lower expression in this species as the male-biased population would amplify the incomplete dosage compensation. To test this, we contrasted normalized expression in pairs of species and scored genes as up or down in one species relative to the other. We then asked if the X-chromosome showed significant skews in the number of genes scored as up or down in these species pairs relative to the autosomes. The results show this is not the case for any species pair (Figure S4), and this is shown in more detail for the *D. persimilis* versus *D. pseudoobscura* contrast (Figure S20), which is pertinent given that there is an excess of X chromosome divergence in this species comparison ($P = 0.0042$; Figures S1, S21). Therefore, there do not appear to be systematic biases in the direction of expression differences between species and hence this is unlikely to be a factor driving the higher divergence of the X chromosome.

Uncovering the relationship between expression evolution and excess chromosomal divergence

The discovery that different groups of genes exhibit differences in their chromosomal divergence in adults suggested that there may be a relationship between excess chromosomal divergence and the rate of gene expression evolution. To test this, we scored the ratio of mean divergence of genes belonging to each percentile of each chromosome's divergence distribution relative to the same percentile of the other chromosomes. The results show that in both the embryos and the adult males, excess divergence on the X chromosome increases as the genes become more divergent while such a pattern is not seen consistently on any of the other chromosomes (Figure S22). In addition we find that while in the embryos most of the genes on the X exhibit an excess of divergence relative to the autosomes, in adult males these genes are restricted to a subset of those on the X. The top enriched biological functions for these genes are primary sex determination, secondary metabolic process, and adult behavior (Table S12), all likely to be fast-evolving traits and processes. It is interesting to note that in both cases, the fastest evolving genes do not display an excess of divergence on the X. Overall, however, we find that fast-evolving genes tend to diverge more on the X in both embryos and adult males.

Correcting for non-expressed/weakly expressed genes

In the embryonic time course, an initially bimodal gene expression distribution gradually becomes unimodal as the zygotic genome is switched on during embryogenesis (Figure S12). If the X chromosome happened to be over-represented for genes in the lower mode of this bimodal distribution, then it is possible that much of the excess divergence we find on the X could be driven by spurious divergence between non-expressed genes. Therefore, to test for this we used the expectation-maximisation algorithm to determine a cutoff expression level (based on time point 1) below which a gene could be considered as non-expressed at any time point (log_2 expression of 8.513).

We then defined three gene sets based on increasingly more stringent criteria for being thrown out from the analysis. The first set (termed "Two") consists of genes that are not expressed in at least two species in at least one time point (1502 genes). The second set ("Six") consists of genes that are not expressed in at least six species in at least one time point (849 genes), and the final set ("Six-Eight") consists of genes that are not expressed in at least six species at every time point (536 genes). Expression distributions for these gene sets shows that they increasingly capture more weakly expressed genes as the criteria for exclusion becomes more stringent (Figure S23). When we compare gene expression divergence for the data set after removing these gene sets, we find that the excess of divergence on the X is not affected (Figure S24) showing that this effect is not driven by spurious divergence between non-expressed or weakly expressed genes.

Mutation accumulation analysis

To determine whether the lower effective population size of the X chromosome might increase the chance that it fixes weakly deleterious mutations, we used gene expression mutation accumulation data to assess potential chromosomal biases in the accumulation of gene expression differences. We used jack-knifed mutational variance estimates scaled by residual variances to provide estimates of the mutational heritability of gene expression changes between lines [35]. As a large fraction of the genes at both the late larval and puparium formation stages did not exhibit measurable mutational heritabilities, we separated the genes with measurable estimates (Figure 5A, 5B). In addition, we categorized genes as having measurable mutational heritabilities from those without and compared the ratios of these two categories across chromosomes using contingency tables. The results were visualized using residual-based shading with the R package 'vcd' [85] (Figure 5C, 5D).

Embryonic tissue expression enrichment analysis

A hierarchically-arranged controlled vocabulary (CV) of embryonic tissue expression terms based on an in situ expression data set [38] was used for assessing under- or over-representation of expression patterns for genes on the Drosophila X chromosome. Enrichment of terms was carried out in the R package 'topGO' [86] using custom-written code. The parent-child algorithm was employed to control for the inheritance bias between parent and child terms in the CV hierarchy [87] (Table S10). The resulting P-values were adjusted using the Benjamini-Hochberg correction in the R package 'multtest' [88].

Multi-locus population genetic models of the faster-X effect

In all of our models, we assume that selection coefficients are equal in the two sexes, which corresponds to the assumption of complete dosage compensation in [5], and, in the case of the two-locus models, that there is a beneficial epistatic interaction between one of the alleles at each locus. In addition, we assume that viability selection operates on the diploid zygotes, that mating is random, and that double heterozygotes experience half of the fitness benefit of single heterozygotes (Tables S11, S13).

We derived genotype frequency recurrence equations to describe the evolutionary dynamics in our models and then solved the equations numerically. To compare evolution in the equivalent X versus autosomal scenarios, we extracted the change in allele frequency of the *cis*-acting beneficial allele between generations, ΔP. We used the ratio of selection gradients in the equivalent models as a comparative statistic. The selection gradient describes the change in relative fitness as the allele frequency of the beneficial variant changes. Using the Robertson-Price identity [89,90] to describe the change in allele frequency, P, in terms of relative fitness, $\tilde{w}$,

$$\Delta P = Cov(\tilde{w},P),$$

and replacing with the regression coefficient, $Cov(\tilde{w},P)=\beta_{\tilde{w},P}\sigma_P^2$,

$$\Delta P = \beta_{\tilde{w},P}\sigma_P^2 = \frac{d\tilde{w}}{dP}P(1-P),$$

then the selection gradient, $\frac{d\tilde{w}}{dP}$, is equal to the change in allele frequency divided by its variance, $\tilde{\Delta}P = \frac{\Delta P}{P(1-P)}$. We plot the ratio of selection gradients in the X versus autosomal cases (Figure 6, Figure S6).

Correlation-based measures of divergence

Spearman's ρ was measured for pairs of chromosomes in pairs of species for both the embryonic and adult data. Correlation coefficients were bootstrapped by resampling the genes 10,000 times on each chromosome separately (Figure 7A, Figure 8A). For the embryos, we used expression averaged across time, and found that correlations derived from this measure agreed very well with correlations derived from expression within single time points in terms of a reduction of correlation on the X chromosome. In addition, we took the mean Canberra distance across chromosomes for pairs of species, averaging it by dividing by the number of genes on each chromosome separately (Figure 7B, Figure 8B).

The correlation approach captures the extent to which chromosomal subsets of genes tend to conserve their expression relationships in pairs of species. However, this approach fails to capture the level of conservation of gene expression in a chromosomal subset relative to a separate chromosomal subset across pairs of species. For example, we might wish to ask whether the expression relationship of genes on the X chromosome relative to the autosomal arm 2L shares a conserved pattern in a pair of species. To answer questions of this nature, we introduce a variant of Spearman's correlation coefficient which allows us to rank genes in a chromosomal subset relative to genes in a separate chromosomal subset for pairs of species. For the correlation of subset A relative to subset B in two species we have

$$\tilde{\rho}_{A:B} = \frac{\sum_i^n (x_{i_{A:B}} - \bar{x}_A)(y_{i_{A:B}} - \bar{y}_A)}{\sqrt{\sum_i^n (x_{i_{A:B}} - \bar{x}_A)^2 \sum_i^n (y_{i_{A:B}} - \bar{y}_A)^2}},$$

where $x_{i_{A:B}}$ and $y_{i_{A:B}}$ are the ranks of the i'th gene's expression level (from the n genes that belong to subset A) relative to gene

expression in subset B for species x and species y respectively. Thus, this relative measure captures whether expression in subset A is co-ordinated relative to subset B in pairs of species.

As it is established that correlation coefficients within subsets can vary, sometimes dramatically, from correlation at the level of aggregates (known as the Yule-Simpson effect [91–95]), we believe that it is necessary to account for possible discrepancies when measuring correlation within subsets drawn from a larger population (Figure S25). When we measure relativised correlations for chromosomal subsets in the embryonic and adult data, we find that the X chromosome displays a significantly higher correlation when correlating against an autosomal background in adult females (Figure S26). This suggests that in adult females the X is generally more co-ordinated in relation to the autosomes than in relation to itself ($P = 0.015$; Wilcoxon two-tailed test), a pattern that could be driven, in part, by gene interactions between the X and the autosomes. More generally, this result highlights the importance of considering cross-chromosome relationships when using correlation-based measures of divergence.

Supporting Information

Figure S1 Phylogenies of the species analyzed with the relative mean lengths of each branch for genes on the X vs genes on the autosomes depicted in blue and red respectively. Bold branches are significantly longer for genes on the X chromosome based on 10,000 bootstrap replicates at the 5% level.
(PDF)

Figure S2 Divergence of gene expression across chromosomes in both adult males and females for genes with sex-biased expression patterns.
(PDF)

Figure S3 Embryonic expression divergence on the X is not driven by extreme expression levels. Bootstrapped divergence measures generated by resampling genes according to their expression levels. Genes were resampled per chromosome using 10,000 bootstrap replicates for both embryos, A, and adults, B. There are more expression levels sampled for embryos because they have a broader gene expression level distribution than the adults.
(PDF)

Figure S4 Mosaic plots for all pair-wise species comparisons of normalized gene expression categorised as up or down relative to one of the species. Mosaic plots visualize categorical data (contingency table) using rectangles that are proportional to the number of counts in each row-column combination, and highlight in red variable combinations that have less than expected numbers and in blue those that have more than expected based on Pearson residuals [85]. *P*-values are based on Chi-squared tests, which test whether the two main variables, Expression and Chromosome, are independent.
(PDF)

Figure S5 The lengths of the summed terminal branches leading to *D. persimilis* and *D. pseudoobscura* as a fraction of the total branch length for *Drosophila* embryos.
(PDF)

Figure S6 Selection gradient ratios when there is no recombination between homologous pairs of male autosomes. The left panel shows the ratio when both loci are X-linked versus both loci being linked on the same autosome but with no male recombination. The right panel shows the ratio for autosomes when there is no recombination in males versus the case when there is. Parameter values: recombination rates, R, are equal to 0.5 (free recombination) and the dominance coefficient, h, is 0.01. The dashed line indicates a ratio of 1.
(PDF)

Figure S7 Divergence of gene expression across chromosomes in both adult males and females for 656 genes with male-biased expression in either *D. melanogaster* or *D. simulans*.
(PDF)

Figure S8 Bootstrapped (10,000 replicates) Spearman's ρ correlation coefficients for adult males and females for all pair-wise species comparisons.
(PDF)

Figure S9 Bootstrapped (10,000 replicates) Mean Canberra distances for adult males and females for all pair-wise species comparisons.
(PDF)

Figure S10 A schematic depicting gene expression in two genes showing why Spearman's ρ would produce a positive correlation despite large differences in expression level and a negative correlation when expression co-ordination between genes is diminished regardless of how much absolute gene expression levels have changed.
(PDF)

Figure S11 All bootstrapped Spearman's ρ correlations across all chromosomes for embryos and adult males and females.
(PDF)

Figure S12 The distribution of gene expression levels during embryogenesis of *D. melanogaster* showing that an initially bimodal distribution, where the lower mode represents unexpressed zygotic genes, becomes a unimodal distribution through time as the zygotic genome is activated.
(PDF)

Figure S13 Log_2 gene expression time course for the X-linked gene Vinculin (*Vinc*) for *D. ananassae* and *D. virilis* showing divergence across the whole time course.
(PDF)

Figure S14 The distributions of resampled mean divergences for each autosome with the mean of the X chromosome indicated by a dashed red line for embryos and adults. Autosomal genes were resampled so that they matched the number of genes on the X chromosome and in each of 10,000 resamples the mean divergence per chromosome was recorded.
(PDF)

Figure S15 Divergence of gene expression in adult males for genes that show unbiased, male-biased, and female-biased expression patterns.
(PDF)

Figure S16 Divergence of gene expression in adult females for genes that show unbiased, male-biased, and female-biased expression patterns.
(PDF)

Figure S17 Divergence of gene expression in adult males for 656 genes with male-biased expression in either *D. melanogaster* or *D. simulans* relative to all genes in the dataset.
(PDF)

Figure S18 Gene expression level by chromosome for embryos and adults in the Drosophila data sets. Expression level is shown as the deviation of each gene's mean $\log_2$ expression level from the global mean.
(PDF)

Figure S19 The relationship between expression level and divergence for embryos and adults in the Drosophila data sets. Genes are ranked by expression, from lowest to highest, binned into groups of 50, and their mean divergence deviation from the global mean (log divergence) is shown as a T-statistic, with significant values highlighted in red. A LOESS curve is fitted to the data.
(PDF)

Figure S20 Mosaic plots for the *D. persimilis-D. pseudoobscura* species comparison of normalized gene expression categorised as up or down relative to one of the species. Mosaic plots visualize categorical data (contingency table) using rectangles that are proportional to the number of counts in each row-column combination, and highlight in red variable combinations that have less than expected numbers and in blue those that have more than expected based on Pearson residuals [85]. *P*-values are based on Chi-squared tests, which test whether the two main variables, Expression and Chromosome, are independent.
(PDF)

Figure S21 Gene expression divergence per chromosome along the branches leading to *D. persimilis* and *D. pseudoobscura*.
(PDF)

Figure S22 Fast-evolving genes tend to diverge more on the X in embryos and adult males. The mean ratio of chromosomal divergence to divergence in the rest of the genome. Mean divergence is plotted for genes belonging to each percentile of a particular chromosome's divergence distribution (separately for 2L, 2R, etc) relative to genes in the same percentile of the divergence distribution of the the rest of the genome (all other chromosomes). The results show that, for the X chromosome, the excess of X/A divergence is higher for faster-evolving genes in both embryos and adult males. Lines are LOESS fits to the data and dashed lines indicate ratios of 1.
(PDF)

Figure S23 Log expression distributions for gene sets excluded for being non-expressed in at least two species in at least one time point ("Two"), in at least six species in at least one time point ("Six"), and in all species at all time points ("Six-Eight"). See Methods.
(PDF)

Figure S24 Gene expression divergence on the X chromosome relative to the autosomes for sets of genes with groups of non-expressed genes removed using various different criteria: non-expressed in at least two species in at least one time point ("Two"), non-expressed in at least six species in at least one time point ("Six"), and non-expressed in all species at all time points ("Six-Eight"). See Methods.
(PDF)

Figure S25 Simulated bivariate data illustrating the Yule-Simpson effect [91–95] when correlating subsets that belong to a larger aggregate. The red and blue points represent two subsets within the total population which display positive correlations when correlated as subsets (unbroken lines) yet a negative correlation when taken as a total population (dashed line). When we use a relativised Spearman's correlation (see Methods), however, we find that these subsets display negative correlations relative to each other thereby explaining why there is a negative correlation for the total population.
(PDF)

Figure S26 Distributions of pairwise species chromosome correlations for embryos, adult males, and adult females. In light blue are the distributions of a relativised Spearman's rank correlation coefficient (see Methods). The suffix "_r" indicates that these are the relative correlation coefficients for a particular chromosome in relation to the other chromosomes.
(PDF)

Table S1 The chromosomal distribution of genes in the expression datasets.
(PDF)

Table S2 Contrasts for *Drosophila* embryo species comparisons. Aut - all autosomes. W - Wilcoxon rank sum test statistic. P-values adjusted according to Benjamini-Hochberg correction.
(PDF)

Table S3 Contrasts for *D. melanogaster* embryo strain comparisons. Aut - all autosomes. W - Wilcoxon rank sum test statistic. P-values adjusted according to Benjamini-Hochberg correction.
(PDF)

Table S4 Contrasts for *Drosophila* adult female species comparisons. Aut - all autosomes. W - Wilcoxon rank sum test statistic. P-values adjusted according to Benjamini-Hochberg correction.
(PDF)

Table S5 Contrasts for *D. melanogaster* female adult strain comparisons. Aut - all autosomes. W - Wilcoxon rank sum test statistic. P-values adjusted according to Benjamini-Hochberg correction.
(PDF)

Table S6 Contrasts for *Drosophila* adult male species comparisons. Aut - all autosomes. W - Wilcoxon rank sum test statistic. P-values adjusted according to Benjamini-Hochberg correction.
(PDF)

Table S7 Contrasts for *D. melanogaster* male adult strain comparisons. Aut - all autosomes. W - Wilcoxon rank sum test statistic. P-values adjusted according to Benjamini-Hochberg correction.
(PDF)

Table S8 Contrasts for *Drosophila* embryos for a common set of 2072 genes and 5 species. W - Wilcoxon rank sum test statistic. P-values adjusted according to Benjamini-Hochberg correction.
(PDF)

Table S9 Contrasts for *Drosophila* adults for a common set of 2072 genes and 5 species. W - Wilcoxon rank sum test statistic. P-values adjusted according to Benjamini-Hochberg correction.
(PDF)

Table S10 Characterisation of the embryonic expression patterns of genes residing on the X chromosome in *Drosophila*. Enrichment is based on the 'parent-child' algorithm in the topGO R package and Fisher's exact test applied to 2228 genes that reside on the X chromosome in *Drosophila*, and enrichment is relative to the whole genome. Terms with uncorrected *P*-values below 0.05 are shown. # - total number of genes with this annotation in the dataset. Sig. - significant, Exp. - expected. P_{adj}-value - adjusted according to the Benjamini-Hochberg false discovery rate.
(PDF)

Table S11 Fitnesses in a diploid two-locus epistatic model with X-linkage. Fitnesses of different male-female gametic combinations when both the loci are located on the X chromosome. T/t - trans-acting gene; C/c - cis-acting locus; 00 - indicates a male gamete carrying a Y chromosome; *s* - selection coefficient; *h* - dominance coefficient.
(PDF)

Table S12 Characterisation of genes with a percentile X/A divergence ratio greater than 1.015 in adult males. Enrichment is based on the 'parent-child' algorithm in the topGO R package and Fisher's exact test applied to 352 genes that have an X/A percentile divergence ratio of >1.015 against the background of the genes in the dataset. # - total number of genes with this annotation in the dataset. Sig. - significant, Exp. - expected. P_{adj}-value - adjusted according to the Benjamini-Hochberg false discovery rate.
(PDF)

Table S13 Fitnesses in a diploid two-locus epistatic model. Fitnesses of different male-female gametic combinations when there is a beneficial partially recessive interaction between an autosomal allele and an X-linked allele (males are the heterogametic sex). T/t - trans-acting autosomal gene; C/c - cis-acting X-linked locus; 0 - indicates a male gamete carrying a Y chromosome; s - selection coefficient; h - dominance coefficient.
(PDF)

Acknowledgments

We gratefully acknowledge Julia Jarrells, Britta Jedamzik, and Nicola Gscheidel at the MPI-CBG Microarray Facility for their help with processing RNA samples for microarray hybridization. We thank Casey Bergman for helpful discussion and for proposing the analysis of differential rates of evolution on Muller's elements, and the analysis of mutation accumulation data. We also thank Nick Barton, Michael Hiller, and four anonymous reviewers for helpful comments on the manuscript, and Iva Kelava for preparing Figure 7, Figure 8, and Figures S8, S9, S10.

Author Contributions

Conceived and designed the experiments: MAK PT ATK. Performed the experiments: MAK. Analyzed the data: ATK. Contributed reagents/materials/analysis tools: PT. Wrote the paper: ATK. Discovered higher gene expression divergence on the X chromosome in Drosophila embryos and conceived the branch length analysis: DTG.

References

1. Haldane JBS (1924) A mathematical theory of natural and artificial selection. part i. Trans Camb Phil Soc 23: 19–41.
2. Muller HJ (1940) The New Systematics, Oxford University Press, chapter Bearings of the Drosophila work on systematics. pp. 185–268.
3. Avery PJ (1984) The population genetics of haplo-diploids and x-linked genes. Genet Res 44: 321–341.
4. Hartl DL (1971) Some aspects of natural selection in arrhenotokous populations. Am Zool 11: 309–325.
5. Charlesworth B, Coyne J, Barton NH (1987) The relative rates of evolution of sex chromosomes and autosomes. Am Nat 130: 113–146.
6. Rice WR (1984) Sex chromosomes and the evolution of sexual dimorphism. Evolution 38: 735–742.
7. Vicoso B, Charlesworth B (2006) Evolution on the x chromosome: unusual patterns and processes. Nat Rev Genet 7: 645–653.
8. Haldane JBS (1921) Sex-ratio and unisexual sterility in hybrid animals. J Genet 12: 101–109.
9. Gurbich TA, Bachtrog D (2008) Gene content evolution on the x chromosome. Curr Opin Genet Dev 18: 493–498.
10. Skuse DH (2005) X-linked genes and mental functioning. Hum Mol Genet 14 Spec No 1: R27–R32.
11. Innocenti P, Morrow EH (2010) The sexually antagonistic genes of drosophila melanogaster. PLoS Biol 8: e1000335. doi: 10.1371/journal.pbio.1000335
12. Zechner U, Wilda M, Kehrer-Sawatzki H, Vogel W, Fundele R, et al. (2001) A high density of x-linked genes for general cognitive ability: a run-away process shaping human evolution? Trends Genet 17: 697–701.
13. Dobzhansky T (1936) Studies on hybrid sterility. ii. localization of sterility factors in drosophila pseudoobscura hybrids. Genetics 21: 113–135.
14. Templeton AR (1977) Analysis of head shape differences between two interfertile species of hawaiian drosophila. Evolution 31: 630–641.
15. Coyne JA, Charlesworth B (1986) Location of an x-linked factor causing sterility in male hybrids of drosophila simulans and d. mauritiana. Heredity (Edinb) 57: 243–246.
16. Lu J, Wu CI (2005) Weak selection revealed by the whole-genome comparison of the x chromosome and autosomes of human and chimpanzee. Proc Natl Acad Sci U S A 102: 4063–4067.
17. Consortium CSA (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437: 69–87.
18. Hvilsom C, Qian Y, Bataillon T, Li Y, Mailund T, et al. (2012) Extensive x-linked adaptive evolution in central chimpanzees. Proc Natl Acad Sci U S A 109: 2054–2059.
19. Betancourt AJ, Presgraves DC, Swanson WJ (2002) A test for faster x evolution in drosophila. Mol Biol Evol 19: 1816–1819.
20. Thornton K, Bachtrog D, Andolfatto P (2006) X chromosomes and autosomes evolve at similar rates in drosophila: no evidence for faster-x protein evolution. Genome Res 16: 498–504.
21. Connallon T (2007) Adaptive protein evolution of x-linked and autosomal genes in drosophila: implications for faster-x hypotheses. Mol Biol Evol 24: 2566–2572.
22. Thornton K, Long M (2002) Rapid divergence of gene duplicates on the drosophila melanogaster x chromosome. Mol Biol Evol 19: 918–925.
23. Counterman BA, Ortz-Barrientos D, Noor MAF (2004) Using comparative genomic data to test for fast-x evolution. Evolution 58: 656–660.
24. Thornton K, Long M (2005) Excess of amino acid substitutions relative to polymorphism between x-linked duplications in drosophila melanogaster. Mol Biol Evol 22: 273–284.
25. Hu TT, Eisen MB, Thornton KR, Andolfatto P (2012) A second generation assembly of the drosophila simulans genome provides new insights into patterns of lineage-specific divergence. Genome Res.
26. Bachtrog D, Jensen JD, Zhang Z (2009) Accelerated adaptive evolution on a newly formed x chromosome. PLoS Biol 7: e1000082. doi:10.1371/journal.-pbio.1000082
27. Jaquiery J, Stoeckel S, Rispe C, Mieuzet L, Legeai F, et al. (2012) Accelerated evolution of sex chromosomes in aphids, an x0 system. Mol Biol Evol 29: 837–847.
28. Mank JE, Axelsson E, Ellegren H (2007) Fast-x on the z: rapid evolution of sex-linked genes in birds. Genome Res 17: 618–624.
29. Andolfatto P (2005) Adaptive evolution of non-coding dna in drosophila. Nature 437: 1149–1152.
30. Kalinka AT, Varga KM, Gerrard DT, Preibisch S, Corcoran DL, et al. (2010) Gene expression divergence recapitulates the developmental hourglass model. Nature 468: 811–814.
31. Zhang Y, Sturgill D, Parisi M, Kumar S, Oliver B (2007) Constraint and turnover in sex-biased gene expression in the genus drosophila. Nature 450: 233–237.
32. Ayroles JF, Carbone MA, Stone EA, Jordan KW, Lyman RF, et al. (2009) Systems genetics of complex traits in drosophila melanogaster. Nat Genet 41: 299–307.
33. Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, et al. (2012) The drosophila melanogaster genetic reference panel. Nature 482: 173–178.
34. Vicoso B, Charlesworth B (2009) Effective population size and the faster-x effect: an extended model. Evolution 63: 2413–2426.
35. Rifkin SA, Houle D, Kim J, White KP (2005) A mutation accumulation assay reveals a broad capacity for rapid evolution of gene expression. Nature 438: 220–223.
36. Houle D, Morikawa B, Lynch M (1996) Comparing mutational variabilities. Genetics 143: 1467–1483.
37. Mikhaylova LM, Nurminsky DI (2011) Lack of global meiotic sex chromosome inactivation, and paucity of tissue-specific gene expression on the drosophila x chromosome. BMC Biol 9: 29.
38. Tomancak P, Berman BP, Beaton A, Weiszmann R, Kwan E, et al. (2007) Global analysis of patterns of gene expression during drosophila embryogenesis. Genome Biol 8: R145.
39. Franke A, Dernburg A, Bashaw GJ, Baker BS (1996) Evidence that msl-mediated dosage compensation in drosophila begins at blastoderm. Development 122: 2751–2760.
40. Lott SE, Villalta JE, Schroth GP, Luo S, Tonkin LA, et al. (2011) Noncanonical compensation of zygotic x transcription in early drosophila melanogaster development revealed through single-embryo rna-seq. PLoS Biol 9: e1000590. doi:10.1371/journal.pbio.1000590
41. Huang W, Richards S, Carbone MA, Zhu D, Anholt RRH, et al. (2012) Epistasis dominates the genetic architecture of drosophila quantitative traits. Proc Natl Acad Sci U S A 109: 15553–15559.
42. Connallon T, Clark AG (2010) Sex linkage, sex-specific selection, and the role of recombination in the evolution of sexually dimorphic gene expression. Evolution 64: 3417–3442.
43. Llopart A (2012) The rapid evolution of x-linked male-biased gene expression and the large-x effect in drosophila yakuba, d. santomea and their hybrids. Mol Biol Evol 29: 3873–3886.
44. Brawand D, Soumillon M, Necsulea A, Julien P, Csrdi G, et al. (2011) The evolution of gene expression levels in mammalian organs. Nature 478: 343–348.

45. Good JM, Giger T, Dean MD, Nachman MW (2010) Widespread over-expression of the x chromosome in sterile fhybrid mice. PLoS Genet 6: e1001148. doi:10.1371/journal.pgen.1001148
46. Meisel RP, Malone JH, Clark AG (2012) Faster-x evolution of gene expression in drosophila. PLoS Genet 8: e1003013. doi:10.1371/journal.pgen.1003013
47. Singh ND, Macpherson JM, Jensen JD, Petrov DA (2007) Similar levels of x-linked and autosomal nucleotide variation in african and non-african populations of drosophila melanogaster. BMC Evol Biol 7: 202.
48. Singh ND, Davis JC, Petrov DA (2005) X-linked genes evolve higher codon bias in drosophila and caenorhabditis. Genetics 171: 145–155.
49. Singh ND, Larracuente AM, Clark AG (2008) Contrasting the efficacy of selection on the x and autosomes in drosophila. Mol Biol Evol 25: 454–467.
50. Vicoso B, Haddrill PR, Charlesworth B (2008) A multispecies approach for comparing sequence evolution of x-linked and autosomal sites in drosophila. Genet Res (Camb) 90: 421–431.
51. Takahashi KH, Tanaka K, Itoh M, Takano-Shimizu T (2009) Reduced x-linked rare polymorphism in males in comparison to females of drosophila melanogaster. J Hered 100: 97–105.
52. Campos JL, Charlesworth B, Haddrill PR (2012) Molecular evolution in nonrecombining regions of the drosophila melanogaster genome. Genome Biol Evol 4: 278–288.
53. Andolfatto P, Wong KM, Bachtrog D (2011) Effective population size and the efficacy of selection on the x chromosomes of two closely related drosophila species. Genome Biol Evol 3: 114–128.
54. Wittkopp PJ, Haerum BK, Clark AG (2004) Evolutionary changes in cis and trans gene regulation. Nature 430: 85–88.
55. Lemos B, Araripe LO, Fontanillas P, Hartl DL (2008) Dominance and the evolutionary accumulation of cis- and trans-effects on gene expression. Proc Natl Acad Sci U S A 105: 14471–14476.
56. Wittkopp PJ, Haerum BK, Clark AG (2008) Regulatory changes underlying expression differences within and between drosophila species. Nat Genet 40: 346–350.
57. Wang HY, Fu Y, McPeek MS, Lu X, Nuzhdin S, et al. (2008) Complex genetic interactions underlying expression differences between drosophila races: analysis of chromosome substitutions. Proc Natl Acad Sci U S A 105: 6362–6367.
58. Wittkopp PJ, Haerum BK, Clark AG (2008) Independent effects of cis- and trans-regulatory variation on gene expression in drosophila melanogaster. Genetics 178: 1831–1835.
59. McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, et al. (2010) Regulatory divergence in drosophila revealed by mrna-seq. Genome Res 20: 816–825.
60. Graze RM, McIntyre LM, Main BJ, Wayne ML, Nuzhdin SV (2009) Regulatory divergence in drosophila melanogaster and d. simulans, a genome-wide analysis of allele-specific expression. Genetics 183: 547–61, 1SI-21SI.
61. Lu X, Shapiro JA, Ting CT, Li Y, Li C, et al. (2010) Genome-wide misexpression of x-linked versus autosomal genes associated with hybrid male sterility. Genome Res 20: 1097–1102.
62. Stern DL (2000) Evolutionary developmental biology and the problem of variation. Evolution 54: 1079–1091.
63. Prud'homme B, Gompel N, Carroll SB (2007) Emerging principles of regulatory evolution. Proc Natl Acad Sci U S A 104 Suppl 1: 8605–8612.
64. Wray GA (2007) The evolutionary significance of cis-regulatory mutations. Nat Rev Genet 8: 206–216.
65. Rebeiz M, Pool JE, Kassner VA, Aquadro CF, Carroll SB (2009) Stepwise modification of a modular enhancer underlies adaptation in a drosophila population. Science 326: 1663–1667.
66. Peter IS, Davidson EH (2011) Evolution of gene regulatory networks controlling body plan development. Cell 144: 970–985.
67. Burke MK, Dunham JP, Shahrestani P, Thornton KR, Rose MR, et al. (2010) Genome-wide analysis of a long-term evolution experiment with drosophila. Nature 467: 587–590.
68. Orr HA, Betancourt AJ (2001) Haldane's sieve and adaptation from the standing genetic variation. Genetics 157: 875–884.
69. Sellis D, Callahan BJ, Petrov DA, Messer PW (2011) Heterozygote advantage as a natural consequence of adaptation in diploids. Proc Natl Acad Sci U S A 108: 20666–20671.
70. Waddington CH (1942) Canalization of development and the inheritance of acquired characters. Nature 150: 563–565.
71. Fisher RA (1928) The possible modification of the response of the wild type to recurrent mutations. Am Nat 62: 115–126.
72. Bourguet D (1999) The evolution of dominance. Heredity 83: 1–4.
73. Proulx SR, Phillips PC (2005) The opportunity for canalization and the evolution of genetic networks. Am Nat 165: 147–162.
74. Billiard S, Castric V (2011) Evidence for fisher's dominance theory: how many 'special cases'? Trends Genet 27: 441–445.
75. Wagner A (2008) Robustness and evolvability: a paradox resolved. Proc Biol Sci 275: 91–100.
76. Masel J, Trotter MV (2010) Robustness and evolvability. Trends Genet 26: 406–414.
77. Draghi JA, Parsons TL, Wagner GP, Plotkin JB (2010) Mutational robustness can facilitate adaptation. Nature 463: 353–355.
78. Kerr MK, Martin M, Churchill GA (2000) Analysis of variance for gene expression microarray data. J Comput Biol 7: 819–837.
79. Smyth GK (2005) Bioinformatics and Computational Biology Solutions using R and Bioconductor, Springer, chapter Limma: linear models for microarray data. pp. 397–420.
80. Felsenstein J (1989) Phylip - phylogeny inference package (version 3.2). Cladistics 5: 164–166.
81. Markow TA, O'Grady PM (2007) Drosophila biology in the genomic age. Genetics 177: 1269–1276.
82. R Development Core Team (2012) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. ISBN 3-900051-07-0.
83. Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL (2003) Sex-dependent gene expression and evolution of the drosophila transcriptome. Science 300: 1742–1745.
84. Meiklejohn CD, Parsch J, Ranz JM, Hartl DL (2003) Rapid evolution of male-biased gene expression in drosophila. Proc Natl Acad Sci U S A 100: 9894–9899.
85. Zeileis A, Meyer D, Hornik K (2007) Residual-based shadings for visualizing (conditional) independence. J Comp Graph Stat 16: 507–525.
86. Alexa A, Rahnenfhrer J, Lengauer T (2006) Improved scoring of functional groups from gene expression data by decorrelating go graph structure. Bioinformatics 22: 1600–1607.
87. Grossmann S, Bauer S, Robinson PN, Vingron M (2007) Improved detection of overrepresentation of gene-ontology annotations with parent child analysis. Bioinformatics 23: 3024–3031.
88. Pollard KS, Gilbert HN, Ge Y, S T, Dudoit S (2003) multtest: Resampling-based multiple hypothesis testing. Technical report, R package version 2.8.0.
89. Robertson A (1966) A mathematical model of the culling process in dairy cattle. Anim Prod 8: 95–108.
90. Price GR (1970) Selection and covariance. Nature 227: 520–521.
91. Yule GU (1903) Notes on the theory of association of attributes in statistics. Biometrika 2: 121–134.
92. Simpson EH (1951) The interpretation of interaction in contingency tables. J Roy Statist Soc B 13: 238–241.
93. Blyth CR (1972) On simpson's paradox and the sure-thing principle. J of the American Statistical Association 67: 364–366.
94. Wagner CH (1982) Simpson's paradox in real life. The American Statistician 36: 46–48
95. Wilcox AJ (2001) On the importance–and the unimportance–of birthweight. Int J Epidemiol 30: 1233–1241.