Research

Shared evolutionary trajectories of three independent neo-sex chromosomes in *Drosophila*

Masafumi Nozawa,[1,2] Yohei Minakuchi,[3] Kazuhiro Satomura,[1,5] Shu Kondo,[4,6] Atsushi Toyoda,[3] and Koichiro Tamura[1,2]

[1]Department of Biological Sciences, Tokyo Metropolitan University, Hachioji, Tokyo 192-0397, Japan; [2]Research Center for Genomics and Bioinformatics, Tokyo Metropolitan University, Hachioji, Tokyo 192-0397, Japan; [3]Comparative Genomics Laboratory, Department of Genomics and Evolutionary Biology, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan; [4]Invertebrate Genetics Laboratory, Department of Chromosome Science, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan

Dosage compensation (DC) on the X Chromosome counteracts the deleterious effects of gene loss on the Y Chromosome. However, DC is not efficient if the X Chromosome also degenerates. This indeed occurs in *Drosophila miranda*, in which both the neo-Y and the neo-X are under accelerated pseudogenization. To examine the generality of this pattern, we investigated the evolution of two additional neo-sex chromosomes that emerged independently in *D. albomicans* and *D. americana* and re-analyzed neo-sex chromosome evolution in *D. miranda*. Comparative genomic and transcriptomic analyses revealed that the pseudogenization rate on the neo-X is also accelerated in *D. albomicans* and *D. americana* although to a lesser extent than in *D. miranda*. In males, neo-X-linked genes whose neo-Y-linked homologs are pseudogenized tended to be up-regulated more than those whose neo-Y-linked homologs remain functional. Moreover, genes under strong functional constraint and genes highly expressed in the testis tended to remain functional on the neo-X and neo-Y, respectively. Focusing on the *D. miranda* and *D. albomicans* neo-sex chromosomes that emerged independently from the same autosome, we further found that the same genes tend to become pseudogenized in parallel on the neo-Y. These genes include *Idgf6* and *Jhl-26*, which may be unnecessary or even harmful in males. Our results indicate that neo-sex chromosomes in *Drosophila* share a common evolutionary trajectory after their emergence, which may prevent sex chromosomes from being an evolutionary dead end.

[Supplemental material is available for this article.]

Sex chromosomes are widely present in diverse organisms (Bachtrog et al. 2014). Species with sex chromosomes can maintain a stable sex ratio (i.e., approximately one), irrespective of their surrounding environment, which may be advantageous in many species with panmixia. However, once sex chromosomes emerge from a pair of autosomes by acquiring sex-determining and sexually antagonistic genes, the proto-X and proto-Y Chromosomes mostly stop meiotic recombination to retain stable sex determination, resulting in Y Chromosome (hereafter, Y) degeneration (Charlesworth et al. 2005). Consequently, there are by far fewer genes on the Y than on the X Chromosome (hereafter, X) in many species (Koerich et al. 2008; Cortez et al. 2014; Zhou et al. 2014; Dupim et al. 2018; Bracewell and Bachtrog 2020), and the copy number of many X-linked genes is imbalanced (i.e., two in females and one in males). Thus, Y degeneration could potentially be deleterious for males and eventually for entire species with sex chromosomes.

Dosage compensation (DC), originally proposed by Muller (1932) and refined by Ohno (1967; see Gartler 2014), counteracts the dosage imbalance of X-linked genes between sexes. Global or chromosome-wide DC has been reported in many organisms (Disteche 2012; Graves 2016; for ineffective DC in some organisms, see also Ellegren et al. 2007; Zha et al. 2009; Vicoso and Bachtrog 2011). In *Drosophila*, male-specific lethal (MSL) binds to chromatin entry sites (CESs) and trigger lysine acetylation at the 16th residue in histone H4 (H4K16ac), which in turn stimulates male-X up-regulation (i.e., global DC) (Alekseyenko et al. 2012; McElroy et al. 2014; Valsecchi et al. 2021). Global DC has also been detected in neo-sex chromosomes (hereafter, neo-X and neo-Y, respectively) in *Drosophila miranda* (Fig. 1), as well as in *Drosophila melanica* and *Drosophila robusta,* in which transposable elements are co-opted into the pathway of global DC and function as CESs (Ellison and Bachtrog 2013; Zhou et al. 2013; Ellison and Bachtrog 2019).

In addition to global DC, Nozawa et al. (2018) reported that individual neo-X-linked genes are likely to have been up-regulated in response to the pseudogenization of neo-Y-linked homologs in *D. miranda* (i.e., gene-by-gene or localized DC) (see also Nozawa et al. 2014). Gene-by-gene DC seems particularly important at the initial stage of sex chromosome evolution, when global DC has not yet been established, to counteract the deleterious effects of the sequential losses of individual neo-Y-linked genes. However, the pseudogenization rate is accelerated not only on the neo-Y but also on the neo-X in *D. miranda* (Nozawa et al. 2016), implying that the acquisition of sex chromosomes is more deleterious than previously recognized. Yet, it is not clear
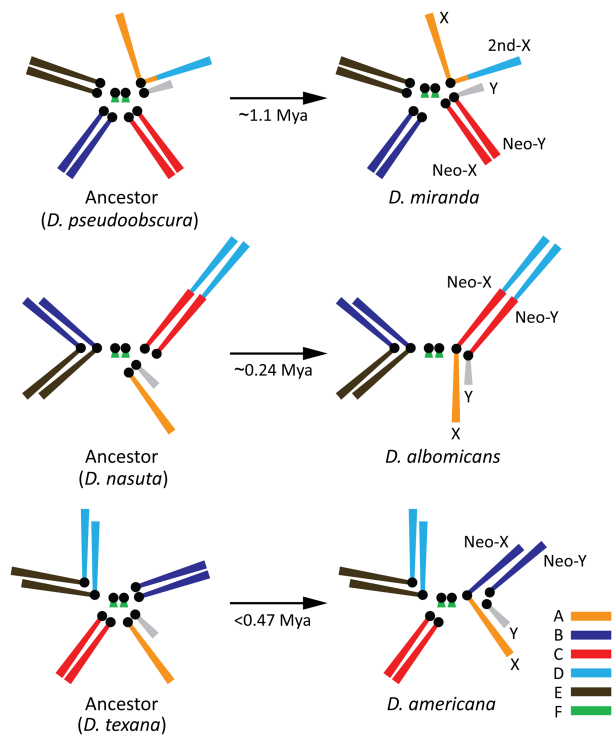
Figure 1. Male karyotypes of three *Drosophila* species with neo-sex chromosomes and their close relatives (ancestors). Colors correspond to the Muller elements shown on the *bottom right*.

whether the observations in *D. miranda* apply to other species and whether sex chromosomes show a common evolutionary trajectory. To address these questions, we need to investigate additional young sex chromosomes with independent origins.

In this study, we therefore aimed to investigate two neo-sex chromosomes that originated independently in *Drosophila* lineages by comparing with their orthologous autosomes in closely related species. One of such neo-sex chromosome emerged in *Drosophila albomicans* ~0.24 myr ago (Mya), after the divergence from *Drosophila nasuta* (Satomura and Tamura 2016; Wei and Bachtrog 2019; Mai et al. 2020). The neo-sex chromosome in this species emerged via the fusion of Chromosome 3 (Muller elements C and D) to the canonical X and Y (Fig. 1). The other neo-sex chromosome in *Drosophila americana* likely emerged <0.47 Mya, after the split from its sibling species, *Drosophila texana* (Vieira et al. 2003). One of the Chromosome 4 (Muller element B) was fused to the X to generate the neo-X, and the other unfused Chromosome 4 became the neo-Y (Fig. 1). We also aimed to reinvestigate the neo-sex chromosome (Muller element C) in *D. miranda* using a substantially updated genome sequence (Mahajan et al. 2018; Bachtrog et al. 2019). Because Muller element C became neo-sex chromosomes independently in *D. miranda* and *D. albomicans* (Fig. 1) and the genes on the Muller element are largely orthologous, we also addressed whether parallel gene loss occurred in these two neo-sex chromosomes.

Results

Genomes and transcriptomes of nine *Drosophila* species

Using long-read (PacBio RS II and Sequel, Pacific Biosciences) and short-read (HiSeq 2500, Illumina) sequencing platforms, we sequenced and assembled the genomes of *D. albomicans* and *D. amer-*

icana with neo-sex chromosomes (Supplemental Table S1). As closely related species of the former and the latter, the genomes of *D. nasuta* and *D. texana,* without neo-sex chromosomes, were also sequenced. In addition, we sequenced the genomes of *Drosophila kohkoa* and *Drosophila novamexicana* as outgroup species to estimate the ancestral states of *D. albomicans* and *D. nasuta* and of *D. americana* and *D. texana* and polarize the changes at the molecular level. The PacBio-based genome sequence of *D. albomicans* (strain no. 15112-1751.03 from Nankung, Taiwan) has already been published (Mai et al. 2020), but we de novo sequenced the genome of another strain (NG3 strain from Okinawa, Japan). To obtain the neo-Y assemblies for *D. albomicans* and *D. americana*, male and female short-reads were separately mapped onto the genome assembly based on female DNA, and the neo-X sequence was replaced with male-specific single-nucleotide polymorphisms (SNPs) and insertions/deletions (indels) to obtain the neo-Y sequence (for details, see "Genome assembly" in Supplemental Methods). The number of sites replaced with male-specific SNPs/indels on the neo-X was 769,441 (~1.4%) and 553,667 (~1.9%) in *D. albomicans* and *D. americana*, respectively. If only exon regions were considered, the proportion of differences between neo-X and neo-Y was ~0.9% for *D. albomicans* and ~1.1% for *D. americana*, namely, on average two or three SNPs in one sequencing fragment. Among 24,754 exons on the neo-X in *D. albomicans*, 9462 (38.2%) exons were identical between neo-X and neo-Y, but a majority of such exons (5253/9462 = 55.5%) were <200 bp in length, that is, shorter than the read length from each fragment. Similarly, among 11,072 exons on the neo-X in *D. americana*, 2984 (27.0%) exons were identical between neo-X and neo-Y, but a majority of such exons (1795/2984 = 60.2%) were <200 bp. Moreover, 92.5% and 92.4% of the neo-X-linked genes had at least one SNP/indel that was distinguishable with the neo-Y-linked homologs in *D. albomicans* and *D. americana*, respectively. Therefore, the majority of neo-X and neo-Y-linked genes were distinguishable in the two species.

For *D. albomicans* and *D. nasuta*, the contig N50 values were ~22.0 Mb and ~17.9 Mb, respectively (Supplemental Table S2), roughly equivalent to the average length of euchromatic regions of a Muller element (Celniker et al. 2002). For other species, the contig N50 was shorter (~1.6 to ~3.6 Mb); however, we obtained fewer than 1000 contigs for all sequenced genomes. To evaluate these genome assemblies, we used benchmarking universal single-copy orthologs (BUSCO) ver. 4.0.4 (Seppey et al. 2019). For all nine genome assemblies, including those for *D. miranda*, *Drosophila pseudoobscura*, and *Drosophila obscura*, >98% of the 3285 single-copy genes in dipteran species were identified as complete open reading frames (ORFs) (Supplemental Table S3).

For transcriptome analyses, we conducted RNA sequencing (RNA-seq) of whole bodies of females and males at the larval, pupal, and adult stages with two biological replicates (Supplemental Table S4). The number of functional protein-coding genes ranged from approximately 9700 to approximately 13,900, depending on the species (Supplemental Table S5). We also conducted RNA-seq of the ovary and testis with two biological replicates (Supplemental Table S4). The chromosomal location, expression level, and functionality of each gene and transcript are listed in Supplemental Tables S6 and S7, respectively.

Proportion of pseudogenes on neo-sex chromosomes

Bachtrog et al. (2019) reported a substantial increase in genes on the neo-Y after the formation of the neo-sex chromosome in

D. miranda. They suggested that testis-specific and dosage-sensitive genes were disproportionately amplified on the neo-Y to increase male fitness. We also confirmed that the number of genes on the Y/neo-Y was approximately 3.4 times greater than that on the neo-X (Fig. 2A). However, ~80% of genes were classified as pseudogenes owing to ORF disruption and/or gene silencing. The number of functional genes was slightly smaller on the neo-Y than on the neo-X (1862 and 2082, respectively). When all inparalogs for each category (i.e., functional, silenced, disrupted, silenced-disrupted, and unclassified genes) were combined, the number of functional gene groups on the neo-Y became much smaller than that on the neo-X (1439 and 2031 groups, respectively) (Supplemental Fig. S1). Therefore, although the total number of genes on the Y/neo-Y was substantially greater than that on the neo-X, the Y/neo-Y had considerable degeneration in *D. miranda* and a narrower functional repertoire than that on the neo-X.

In contrast, the numbers of functional, silenced, and disrupted genes on the neo-X and neo-Y were similar in *D. albomicans* and *D. americana*, reflecting their more recent origins compared with *D. miranda* (Fig. 2B,C). In addition, the number of orthologous gene groups was similar to the number of genes (Supplemental Fig. S1), indicating that gene duplication was not common on the neo-sex chromosomes after their emergence in these two species. Therefore, degeneration and specialization on neo-sex chromosomes were less conspicuous in these two species than in *D. miranda* (see also Wei and Bachtrog 2019).

DC on neo-Xs

DC is often evaluated by comparing the expression levels of X-linked and autosomal genes in males (e.g., Deng et al. 2011; Vicoso and Bachtrog 2011). However, because the X-linked and autosomal genes are not orthologous, this comparison cannot be ap-
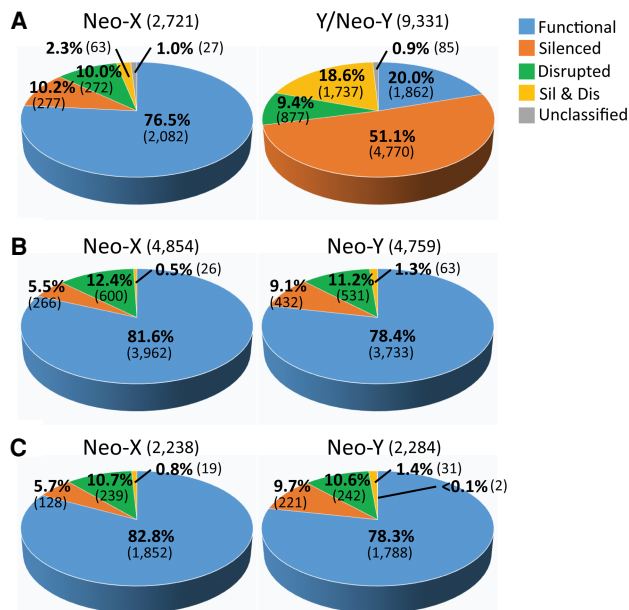
plied to evaluate gene-by-gene DC. Instead, the expression level of an X-linked gene in a target species is compared with that of an autosomal ortholog in a different species based on the R_{Lin} index (for details, see "Estimation of the extent of dosage compensation" in the Methods) (Lin et al. 2012; see also Nozawa et al. 2014). R_{Lin} is expected to be 1.0 if DC is complete, whereas a value of 0.5 indicates that DC is absent. Using the R_{Lin} index, Nozawa et al. (2018) detected gene-by-gene DC on the *D. miranda* neo-X. In *D. miranda*, the R_{Lin} value for neo-X-linked genes with pseudogenized neo-Y-linked homologs (X_F–Y_P, where the subscripts F and P indicate functional genes and pseudogenes, respectively) was significantly greater than that for neo-X-linked genes with functional neo-Y-linked homologs (X_F–Y_F). The updated data set for *D. miranda* confirmed this pattern in all tissues examined, although the difference in R_{Lin} was not significant in testis in which meiotic sex chromosome inactivation may operate (Fig. 3A) (e.g., Vibranovski 2014; Argyridou and Parsch 2018). In *D. albomicans*, the trend was the same as that in *D. miranda*; however, the difference in R_{Lin} values between the X_F–Y_F and X_F–Y_P groups was significant only in larvae and adults (Fig. 3B). In *D. americana*, the difference was significant in adults but not at other stages, although the trend was similar to that in the two species (Fig. 3C). Therefore, gene-by-gene DC is present on neo-Xs in the two species but is less conspicuous than that in *D. miranda*. It should be mentioned that adult male whole body containing testis showed a significant level of gene-by-gene DC in the three species, possibly because other tissues such as head and thorax are under relatively stringent gene-by-gene DC as shown in *D. miranda* (Nozawa et al. 2018).

The neo-Xs in these species also differed in R_{Lin} for the X_F–Y_F group. As already mentioned, the MSL recognizes the CES, which triggers histone acetylation to initiate global DC on the male neo-X in *D. miranda* (Ellison and Bachtrog 2013). Therefore, in *D. miranda*, in addition to X_F–Y_P genes, X_F–Y_F genes showed R_{Lin} values that were significantly >0.5, the expected value under no DC (Fig. 3A; see also Nozawa et al. 2018). In contrast, X_F–Y_F gene up-regulation was not obvious in *D. albomicans* and *D. americana*. In particular, R_{Lin} values for X_F–Y_F genes were ~0.5 in all samples examined in *D. albomicans* (Fig. 3B). These results suggest that global DC has not yet been established in *D. albomicans* or *D. americana*. When R_{Lin} was computed with corrected transcripts per million mapped reads (cTPM; TPM normalized by the median) (Supplemental Fig. S2) instead of corrected fragments per kilobase of exon per million mapped reads (cFPKM) (Fig. 3) for the estimation of gene expression levels, the results were qualitatively the same.

It should be mentioned that we used multiply mapped fragments in addition to uniquely mapped fragments for estimating gene expression level. When we restricted our analysis to uniquely mapped fragments, the general pattern of gene-by-gene DC remained unchanged, although statistical significance changed in some cases (Supplemental Fig. S3). However, the R_{Lin} values were considerably underestimated in this case because cFPKM values in species with neo-sex chromosomes decrease by the removal of multiply mapped fragments, whereas those in the close relatives remain almost unchanged.


Figure 2. Classification of neo-X- and neo-Y-linked genes in *Drosophila miranda* (*A*), *D. albomicans* (*B*), and *D. americana* (*C*). Numbers in parentheses indicate the numbers of genes in each category. Colors on each pie chart correspond to the categories shown on the *top right* corner. Silenced, disrupted, and silenced-and-disrupted genes are pseudogenes.

Accelerated pseudogenization on neo-Ys and neo-Xs

We next confirmed the previously reported accelerated pseudogenization on both the neo-X and neo-Y in *D. miranda* (Nozawa et al. 2016) with updated data sets (Fig. 4A,B). The pseudogenization rate on the neo-X lineage (branch *a*) was significantly greater
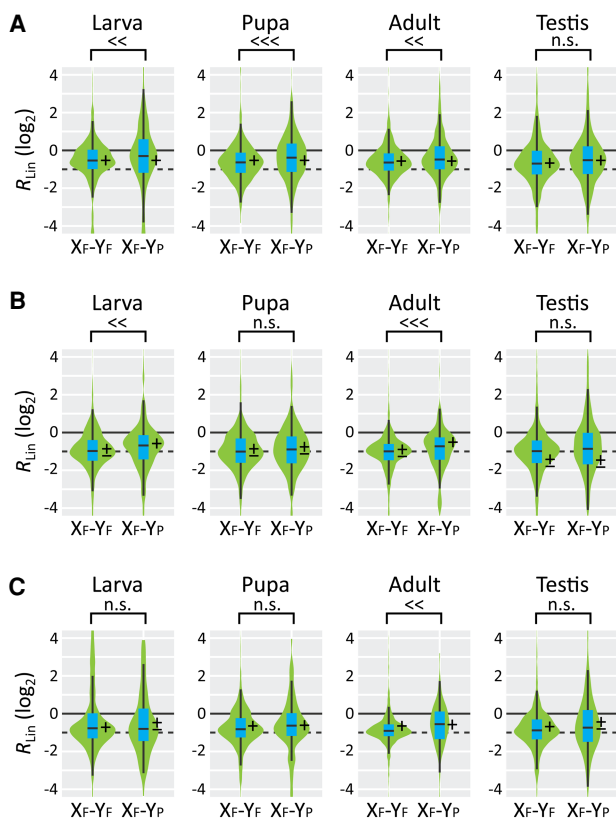
Figure 3. Relationship between the functionality of neo-Y-linked genes and the DC of neo-X-linked homologs in the larva, pupa, adult, and testis of *Drosophila miranda* (*A*), *D. albomicans* (*B*), and *D. americana* (*C*). To compute R_{Lin}, corrected fragments per kilobase of exon per million mapped reads (cFPKM; FPKM normalized by median) was used. (X_F-Y_F) A group of genes with functional neo-X-linked and neo-Y-linked homologs; (X_F-Y_P) a group of genes with functional neo-X-linked homologs and pseudogenized neo-Y-linked homologs. In the X_F-Y_F and X_F-Y_P groups, 664 and 435 genes, respectively, were analyzed for *D. miranda*, 2449 and 218 for *D. albomicans*, and 1204 and 93 for *D. americana*. A box plot is also shown on each violin plot. Differences of R_{Lin} values between groups were evaluated by Mann–Whitney U test with correction for multiple testing by the Benjamini–Hochberg method (Benjamini and Hochberg 1995) under the null hypothesis R_{Lin} (X_F-Y_F) = R_{Lin} (X_F-Y_P): (<<<) $Q < 0.001$; (<<) $Q < 0.01$; (n.s.) $Q \geq 0.05$. The solid line indicates an R_{Lin} value of 1.0 (0 in $\log_2$), indicating perfect DC, whereas the broken line corresponds to a value of 0.5 (−1 in $\log_2$), indicating no DC. (+, −, and ±) Along each plot, they indicate that the median R_{Lin} value is >0.5, <0.5, or 0.5, respectively, at the 5% significance level based on a bootstrap test with 10,000 replications with correction for multiple testing.

than that on the orthologous autosome in *D. pseudoobscura* (branch *e*) ($Q = 2.5 \times 10^{-6}$ by a χ^2 test with correction for multiple testing). Similarly, the pseudogenization rate on the neo-X lineage (branch *a*) was significantly greater than that on the *D. miranda* lineage, before the emergence of neo-sex chromosome (branch *c*; $Q = 1.2 \times 10^{-9}$). The sum of pseudogenization events on branches *a* and *c* was also significantly greater than the sum of those on branches *d* and *e* ($Q = 3.7 \times 10^{-3}$). One may think that inversion polymorphism on Chromosome 3 in *D. pseudoobscura* (Olvera et al. 1979) affected the analysis. Such inversion polymorphism may certainly increase the rate of pseudogenization on the Chromosome 3 compared with other autosomes in *D. pseudoobscura* owing to inefficacy of natural selection. Indeed, the rate of pseudogenization on the Chromosome 3 (i.e., Muller element C)

in the *D. pseudoobscura* lineage was a bit higher than that on other autosomes (Supplemental Fig. S4A). Then, the difference in pseudogenization rate between the neo-X and the Chromosome 3 must be smaller than the difference under the condition of no inversion polymorphism. In other words, our analysis must be conservative to our conclusion. Therefore, the trend of accelerated pseudogenization on the neo-X in *D. miranda* is robust.

We then evaluated whether a similar process occurred on the neo-sex chromosomes in *D. albomicans* and *D. americana*. Similar patterns were also observed in the *D. albomicans* lineage; that is, differences in the pseudogenization rates on branches *a* and *e*, branches *a* and *c*, and branches *a* + *c* and *d* + *e* were all statistically significant (Fig. 4C,D). In contrast, the difference was not significant in the *D. americana* lineage (96 + 31 = 127 for *a* + *c* and 109 for *d* + *e*, $Q = 0.24$), although the trend was qualitatively similar (Fig. 4E,F). The difference in the pseudogenization rates on the neo-X and neo-Y was also less clear ($Q = 0.028$), and the extent of accelerated pseudogenization on the neo-Y compared with the orthologous autosome in *D. texana* was less substantial (i.e., *b* + *c* vs. *d* + *e*, $Q = 2.1 \times 10^{-3}$). Although the time of neo-sex chromosome emergence in *D. americana* is not clear (Vieira et al. 2003), this observation might reflect the more recent origin than those of the neo-sex chromosomes in *D. miranda* and *D. albomicans*. It should be noted that the pseudogenization rate for most of the other chromosomes did not differ significantly between species with neo-sex chromosomes and closely related species without neo-sex chromosomes (Supplemental Fig. S4). Therefore, the mutation rate and general functional constraints were similar in the former and latter species. These results indicate that the three species with neo-sex chromosomes share an evolutionary trajectory in which pseudogenization is accelerated on the neo-X as well as the neo-Y, although the degree of acceleration varies among species, possibly owing to the age difference among the neo-sex chromosomes.

Pseudogenization pattern on neo-sex chromosomes

We next examined relationships between functional constraints on genes before being linked to neo-sex chromosomes, estimated as the ratio of nonsynonymous to synonymous nucleotide substitutions per site (d_N/d_S), and pseudogenization (Fig. 5). In the *D. miranda* lineage, pseudogenes on the neo-X (i.e., X_P-Y_F and X_P-Y_P genes) tended to have less functional constraints in the ancestral lineage before the emergence of the neo-X and neo-Y (branch *c* in Fig. 5A). A similar but less clear pattern was observed in the *D. albomicans* and *D. americana* lineages (branch *c* in Fig. 5B,C). X_P-Y_F genes tended to evolve faster than X_F-Y_F and X_F-Y_P genes on the neo-X lineages in all three species, as expected (Fig. 5A–C, branch *a*). However, X_P-Y_F genes tended to evolve at a faster rate than X_P-Y_P genes, particularly in the *D. miranda* lineage (Fig. 5A, middle panel). X_F-Y_P genes, in contrast, did not evolve faster than other gene categories in the ancestral lineage (branch *c*), although they evolved faster on the neo-Y lineages (branch *b*), as expected (Fig. 5A–C, right panel). Therefore, neo-X-specific pseudogenization events primarily occurred in genes under less functional constraints before becoming neo-sex-linked, whereas neo-Y-specific pseudogenization events are unlikely to be associated with functional constraint in the ancestor.

We detected enrichment for many Gene Ontology (GO) terms in functional genes and pseudogenes on the neo-X and neo-Y (Supplemental Tables S8–S19). However, the enriched terms were largely shared between neo-X and neo-Y Chromosomes in
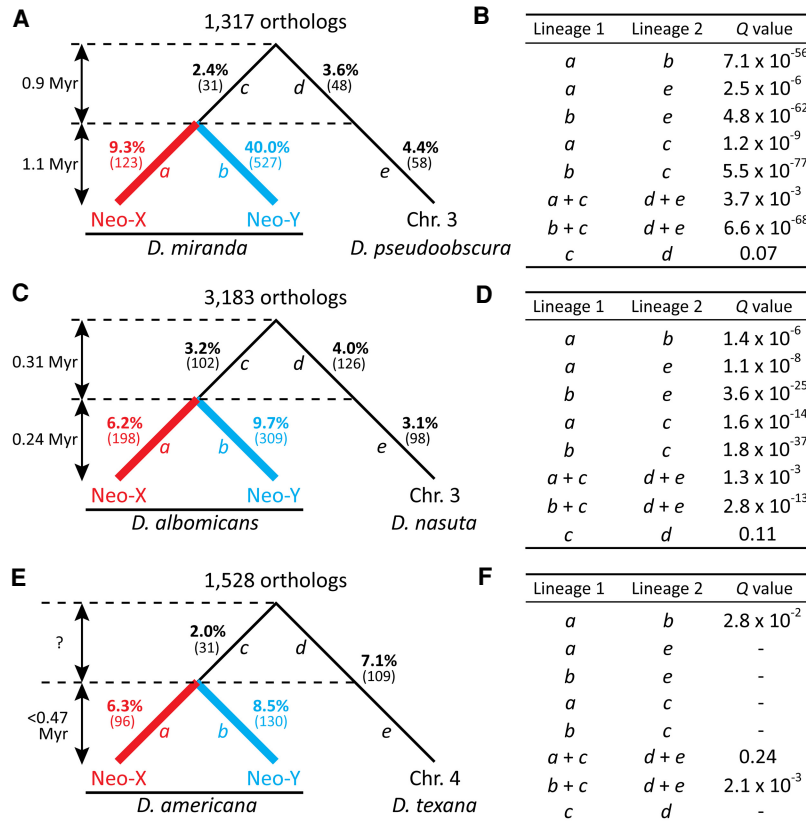
Figure 4. Pseudogenization events before and after neo-sex chromosome emergence. (*A,C,E*) Proportions of genes that were pseudogenized in *Drosophila miranda* (*A*), *D. albomicans* (*C*), and *D. americana* (*E*) lineages compared with those in the corresponding closely related species. Numbers in parentheses indicate the number of pseudogenized genes in each lineage. Orthologs regarded as functional in the outgroup species (*D. obscura, D. kohkoa,* and *D. novamexicana,* respectively) and located on the same Muller element without any inparalogs were used for this analysis. The number of such orthologs is shown *above* each tree. (*B,D,F*) Analysis of differences in the pseudogenization rate between lineages of *D. miranda* (*B*), *D. albomicans* (*D*), and *D. americana* (*F*) by a χ^2 test with correction for multiple testing under the null hypothesis of equal pseudogenization rates between lineages. Lineages *a* to *e* correspond to the branches in *A, C,* and *E*.

the three species. The shared evolutionary history, until recently, can explain the shared enrichment for GO terms between the neo-X and neo-Y but cannot explain shared enrichment among species. We speculate that the results reflect general features of all chromosomes rather than the neo-sex chromosomes alone. Indeed, the detection of chemical stimuli was a commonly enriched GO term in pseudogenes, consistent with the dynamic turn-over and pseudogenization of chemosensory receptor genes in many animals (Niimura and Nei 2007; Nozawa and Nei 2007; Nei et al. 2008; Hayakawa et al. 2014).

We next examined whether pseudogenization events are associated with spatiotemporal gene expression patterns. For this analysis, genes were classified based on the tissue and stage in which they were most highly expressed. Because it is not possible to obtain RNA from ancestral species, we used gene expression levels in closely related species without neo-sex chromosomes as a proxy for gene expression levels in the ancestors (i.e., *D. pseudoobscura, D. nasuta,* and *D. texana* for *D. miranda, D. albomicans,* and *D. americana,* respectively).

On the canonical X, genes with the highest expression in the ovary tended to maintain their functions, although the trend was

not significant after the correction of multiple testing (Fig. 6A–C, second panel from the left). A similar but nonsignificant pattern was also observed in the *D. miranda* and *D. americana* neo-Xs (Fig. 6A,C, see also Nozawa et al. 2016). Therefore, the neo-X could be in a transitional state toward the canonical X with respect to gene content.

On the neo-Y, the proportion of pseudogenes was significantly lower for genes showing the highest expression in the testis than for genes showing the highest expression in somatic tissues for the *D. miranda* and *D. albomicans* lineages (Fig. 6A,B, rightmost panels), consistent with previous results for *D. miranda* (Kaiser et al. 2011; Nozawa et al. 2018). The same trend was also observed in the *D. americana* lineage (Fig. 6C, rightmost panel). Genes expressed most highly in the ovary also tended to maintain their functions, even after becoming neo-sex-linked in all three lineages, consistent with the findings in *D. miranda* (Nozawa et al. 2018). It should be noted that there was no clear-cut difference in the frequency of pseudogenization among tissues on autosomes (Fig. 6A–C, leftmost panels), probably owing to biparental inheritance with equal contributions from males and females. Notably, the same general pattern was obtained when TPM was used instead of cFPKM to estimate gene expression levels particularly for the neo-Y (Supplemental Fig. S5). Therefore, the similar pseudogenization pattern on the neo-Y in three species would be primarily owing to the linkage of the genes to the Y.

Parallel pseudogenization on the neo-sex chromosomes in *D. miranda* and *D. albomicans*

Because Muller element C independently became the neo-sex chromosome in *D. miranda* and *D. albomicans,* we next examined whether orthologs were pseudogenized in parallel in the two lineages (Fig. 7). Three orthologs were pseudogenized in parallel on the neo-X lineages, which did not differ from the random expectation for pseudogenization on each lineage (Fig. 7A). In contrast, there were 35 parallel pseudogenization events on the neo-Y lineages, which was significantly greater than the expectation under random pseudogenization (Fig. 7B). We did not detect parallel pseudogenization on the Muller element C in the ancestral lineage before the emergence of the neo-sex chromosomes (Fig. 7C) or on other Muller elements (Supplemental Fig. S6). It should be noted that parallel pseudogenization was identified on a couple of Muller elements in the comparisons of *D. miranda* and *D. americana* (Supplemental Fig. S7), as well as *D. albomicans* and *D. americana* (Supplemental Fig. S8); however, the extent was by far the greatest on the neo-Ys of *D. miranda* and *D. albomicans.* Therefore, parallel pseudogenization on the two neo-Ys in
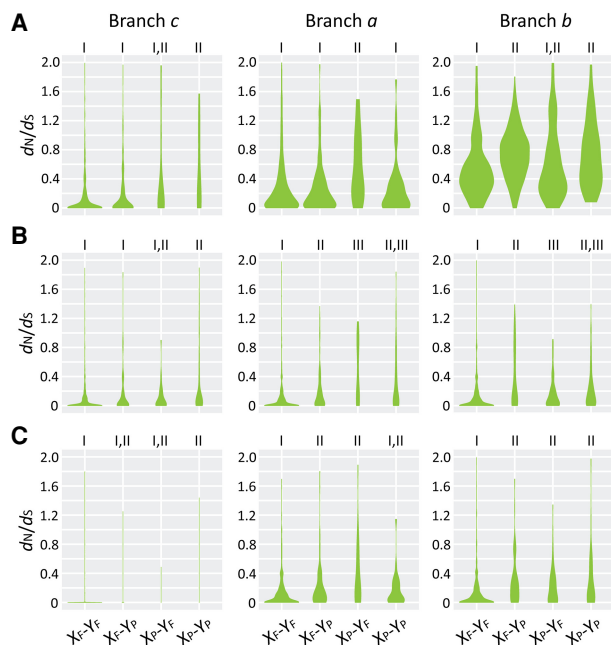
Figure 5. Relationship between the functionality of neo-sex-linked genes and constraint, estimated as the ratio of nonsynonymous to synonymous nucleotide substitutions per site (d_N/d_S) in *Drosophila miranda* (*A*), *D. albomicans* (*B*), and *D. americana* (*C*) lineages. Branches *c* (*left*), *a* (*center*), and *b* (*right*) correspond to those in Figure 4. Violin plots were used to show the distribution of d_N/d_S. Orthologs regarded to be functional in the outgroup species (*D. obscura, D. kohkoa,* and *D. novamexicana,* respectively) and located on the same Muller element without any inparalogs were used for this analysis. In total, 681 genes were analyzed for X_F-Y_F, 474 genes for X_F-Y_P, 70 genes for X_P-Y_F, and 92 genes for X_P-Y_P in the *D. miranda* lineage; 2551 genes were analyzed for X_F-Y_F, 277 genes for X_F-Y_P, 166 genes for X_P-Y_F, and 189 genes for X_P-Y_P in the *D. albomicans* lineage; and 1267 genes were analyzed for X_F-Y_F, 119 genes for X_F-Y_P, 85 genes for X_P-Y_F, and 57 genes for X_P-Y_P in the *D. americana* lineage. Differences of d_N/d_S ratios between groups were evaluated by a Mann–Whitney *U* test with correction for multiple testing under the null hypothesis of an equal d_N/d_S ratio. The same roman numerals indicate $Q \geq 0.05$, whereas different numerals indicate $Q < 0.05$ between categories.

D. miranda and *D. albomicans* likely reflects a common evolutionary trajectory owing to the linkage to the Y.

We further searched for unique features of the genes that experienced parallel pseudogenization. The d_N/d_S ratios for the genes that experienced parallel pseudogenization and lineage-specific pseudogenes were similar, although the ratios for the pseudogenes tended to be higher than those for functional genes on the neo-Y lineages (Supplemental Fig. S9, branch *b*). However, the orthologs of the neo-Y pseudogenes in the closely related species (*D. pseudoobscura* for *D. miranda* and *D. nasuta* for *D. albomicans,* respectively) showed a greater degree of tissue-specific expression, as indicated by τ (Yanai et al. 2005), than that of functional genes, although there was no significant difference in τ between orthologs of lineage-specific and shared pseudogenes (Fig. 8A). The number of tissues in which genes were expressed also tended to be smaller for the orthologs of shared pseudogenes than those of functional genes and lineage-specific pseudogenes on the neo-Y (Fig. 8B). Moreover, the orthologs of shared neo-Y-linked pseudogenes in *D. pseudoobscura,* closely related to *D. miranda,* tended to show more female-biased expression compared with the expression of orthologs of functional genes and lineage-specific pseudogenes at the larval and pupal stages (Fig. 8C), although the pattern was

less clear in *D. nasuta*, closely related to *D. albomicans* (Fig. 8D). Therefore, female-biased genes expressed only in a small number of tissues tend to be pseudogenized in parallel on the neo-Y.

Discussion

We investigated whether a common evolutionary trajectory prevents neo-sex chromosomes from being evolutionary dead ends using the three *Drosophila* species that independently acquired neo-sex chromosomes and closely related species without neo-sex chromosomes. We observed the following shared features among neo-sex chromosomes of the three species: (1) Gene-by-gene or localized DC operated on neo-X-linked genes; (2) the rate of pseudogenization was accelerated not only on the neo-Y but also on the neo-X; (3) neo-Y-linked genes with the highest expression in the testis or ovary tended to remain functional; and (4) the same orthologs were frequently pseudogenized in parallel on the neo-Y in *D. miranda* and *D. albomicans*.

DC on the neo-X in *Drosophila*

When individual Y-linked genes are pseudogenized, the immediate up-regulation of X-linked homologs is critical to counteract the deleterious effects. We found the gene-by-gene up-regulation of neo-X-linked homologs in three species with independent origins of neo-sex chromosomes; this up-regulation likely offset the negative effect of losing neo-Y genes. However, gene-by-gene DC was weaker in *D. albomicans* and *D. americana*, with relatively young neo-sex chromosomes, than in *D. miranda,* with the older neo-sex chromosome. These results imply that there was insufficient time for the evolution of the gene-by-gene DC mechanism in the former two species. It should be mentioned that this interpretation is not necessarily consistent with previous results showing that some neo-X-linked genes already acquired the potential for up-regulation before the corresponding neo-Y homologs are lost in *D. miranda* (Nozawa et al. 2018). Yet, this previous analysis was inconclusive owing to large sampling errors in the pseudogenization time for each neo-Y-linked gene owing to the low sequence divergence between the neo-sex chromosomes in *D. miranda*. Hence, we did not conduct the same analysis for *D. albomicans* and *D. americana*, in which the sequence divergence between neo-X and neo-Y is even smaller. Therefore, the time frame for the establishment of the mechanism underlying gene-by-gene DC after neo-Y-linked gene loss remains unclear.

We also detected both global and gene-by-gene DC on the neo-X in *D. miranda,* as reported previously (Ellison and Bachtrog 2013; Nozawa et al. 2018), whereas in *D. albomicans* and *D. americana*, global DC is unlikely to exist. These results support a previous proposal (Nozawa et al. 2018) that gene-by-gene DC develops in the early stage of sex chromosome evolution and global DC develops later, gradually taking over the role of gene-by-gene DC as Y degeneration proceeds. Under this scenario, the neo-Xs in *D. albomicans* and *D. americana* are too young to establish global DC, whereas the neo-X in *D. miranda* is in a transient stage from gene-by-gene to global DC. Consistent with this scenario, the neo-Xs that arose independently in *D. melanica* and *D. robusta* about 4–15 Mya (Flores et al. 2008) have already acquired the MSL binding sites, that is, the global DC mechanism (Ellison and Bachtrog 2019). In contrast, the neo-X with an age of ~0.85 Mya in *Drosophila busckii* is unlikely to have acquired global DC (Zhou and Bachtrog 2015). These observations imply that a period of ~1 Myr is necessary for the development of a partial global DC
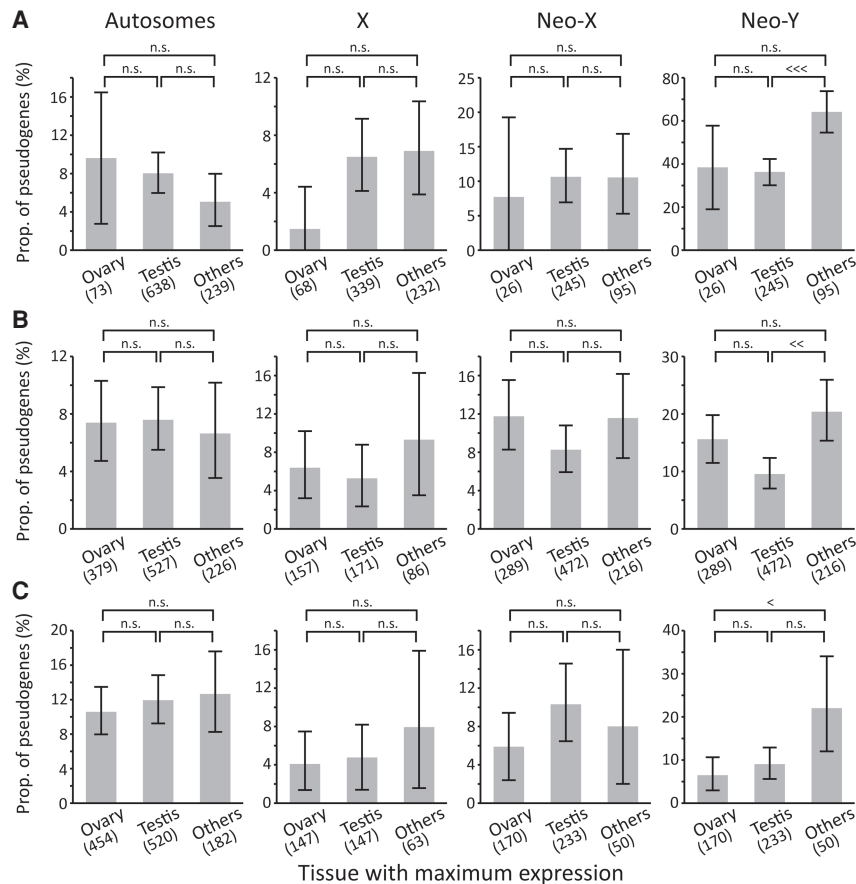
Figure 6. Relationship between spatiotemporal gene expression pattern and pseudogenization. All genes were classified into three groups based on the tissue with the highest expression in *Drosophila pseudoobscura* (A), *D. nasuta* (B), and *D. texana* (C), closely related species of *D. miranda*, *D. albomicans*, and *D. americana*, respectively, based on cFPKM values. To remove genes with similar cFPKM in multiple tissues, only genes with at a difference in cFPKM of at least twofold between a particular tissue and other tissues examined were included. Orthologs regarded to be functional in the close relatives (*D. pseudoobscura* and *D. obscura* for *D. miranda*, *D. nasuta* and *D. kohkoa* for *D. albomicans*, and *D. texana* and *D. novamexicana* for *D. americana*, respectively) and located on the same Muller element without any inparalogs were used for this analysis. Number of genes in each category is shown in parenthesis. Error bars show the 95% confidence interval based on a bootstrap resampling with 10,000 replicates. Statistical significance between groups was tested by Fisher's exact test with correction for multiple testing: (<<<) $Q < 0.001$; (<<) $Q < 0.01$; (<) $Q < 0.05$; (n.s.) $Q \geq 0.05$.

mechanism on the neo-X in *Drosophila* species. It should be mentioned that neo-Y degeneration was not conspicuous in *D. albomicans* and *D. americana*. Therefore, the effects of gene loss from the neo-Y on fitness may be limited, even without strong gene-by-gene DC.

We do not have enough data to discuss the mechanism underlying gene-by-gene DC, but we speculate that it involves auto-regulation or buffering at the mRNA and/or protein level. X inactivation (i.e., inactivation of one allele of X-linked genes) in humans also seems to be primarily driven by the loss of Y-linked homologs to balance gene dosage between females and males (Wilson Sayres and Makova 2013), which may also be accomplished by a buffering mechanism. In addition, Xia et al. (2021) reported that when a young gene, *vismay*, is artificially pseudogenized by the CRISPR-Cas9 system in *D. melanogaster*, its homologous parental gene, *achintya*, is automatically up-regulated, which is also likely an analogy of gene-by-gene DC. Analyses of DC at the

protein level may provide further insights into the mechanism underlying gene-by-gene DC.

Accelerated pseudogenization on the neo-X and neo-Y in *Drosophila*

As discussed above, gene-by-gene DC certainly operates on the neo-X and is likely critical for survival at the early stage of sex chromosome evolution. However, gene-by-gene DC is unlikely to completely mitigate the deleterious effects of neo-Y-linked gene loss because the pseudogenization rate is also accelerated on the neo-X, at least in *D. miranda* and *D. albomicans* (with a similar trend in *D. americana*). If a gene is pseudogenized or lost from the neo-X, gene-by-gene DC cannot operate on its neo-Y-linked homolog. However, neo-X-linked pseudogenes tend to have been under less functional constraints when they were functional on autosomes. Thus, the effects of these pseudogenization events on fitness are apparently weak.

What kinds of evolutionary forces have driven the accelerated rate of pseudogenization on neo-sex chromosomes? For the neo-Y, the suppression of meiotic recombination and small effective population size (one-quarter that of an autosome) would be sufficient explanations. For the neo-X, however, the situation is more complex. The effective population size of the neo-X is three-quarters that of an autosome under equal sex ratios. However, unlike many other organisms, in most *Drosophila* species, there is no recombination at male meiosis (John et al. 2016). Therefore, the proportion of chromosomes that recombine at meiosis is two-thirds for the neo-X, greater than two-quarters for an autosome (Charlesworth 2012). Therefore, the inclusive efficacies of natural selection on the neo-X and an autosome are expected to be similar (i.e., $3/4 \times 2/3 = 1/2$ for the neo-X and $4/4 \times 2/4 = 1/2$ for an autosome). Therefore, other forces, such as sexual conflict, may influence the fates of genes on neo-sex chromosomes. Indeed, genes likely involved in sexual conflict show a biased pattern of pseudogenization on the neo-X in *D. miranda*, and this may have reduced sexual conflict between males and females owing to the acquisition of neo-sex chromosomes (Nozawa et al. 2016; but for neo-sex chromosomes as a new battleground for sexual conflict, see also Bachtrog et al. 2019).

The duration of accelerated pseudogenization on the neo-X after the emergence of neo-sex chromosomes would also be an intriguing point for future research. In the dipteran lineage, dozens of examples of sex chromosome turn-over have been detected (Vicoso and Bachtrog 2015). Because most Y-linked genes in dipteran species, as well as probably many other species, are pseudogenized, the X, but not Y, will be reversed to an autosome

Figure 7. Lineage-specific and parallel (shared) pseudogenization events on the neo-sex chromosomes in *Drosophila miranda* (*mir*) and *D. albomicans* (*alb*) lineages. A total of 814 orthologs regarded to be functional in the outgroup species (*D. obscura* and *D. kohkoa*, respectively) and located on Muller element C without any inparalogs was used for the analysis. Each Venn diagram shows the number of lineage-specific or parallel pseudogenization events on the neo-X (*A*), neo-Y (*B*), and ancestral lineages (*C*; i.e., branches *a*, *b*, and *c*, respectively, in Fig. 4A,C) of *D. miranda* and *D. albomicans*. Statistical difference between the observed and expected numbers of shared pseudogenizations was tested by a binomial test with correction for multiple testing: (**) $Q < 0.01$; (n.s.) $Q \geq 0.05$.

(Meisel 2020). Indeed, a large number of genes remain functional on the X in many species (Koerich et al. 2008; Cortez et al. 2014; Zhou et al. 2014; Dupim et al. 2018; Bracewell and Bachtrog 2020). Because chromosome-level assemblies of X at different ages have recently become available in several *Drosophila* species (Mahajan et al. 2018; Bracewell et al. 2019; Wei and Bachtrog 2019; Bracewell and Bachtrog 2020), it may be possible to estimate transitions in the pseudogenization rate on the X after its emergence in the future.

Pseudogenization pattern on neo-sex chromosomes in *Drosophila*

We found that the genes under less functional constraint before neo-sex chromosome emergence are frequently pseudogenized after becoming neo-X-linked. This finding suggests that the deleterious effect of acquiring sex chromosomes is weak. In contrast, this trend does not necessarily hold for neo-Y-linked genes, particularly for genes pseudogenized only on the neo-Y. Instead, the spatiotemporal gene expression pattern is likely to be a primary factor determining the fate of neo-Y-linked genes. If a gene is highly expressed in the testis of the closely related species without neo-sex chromosomes (possibly reflecting the gene expression pattern in the ancestor), it tends to remain functional after becoming neo-Y-linked. This observation is reasonable because the Y is transmitted only through males; accordingly, the genes whose functions are specific or biased to males, such as genes highly expressed in the testis, are likely to be maintained.

In this context, however, our finding that genes with the highest expression in the ovary of closely related species are less likely to be pseudogenes on the neo-Y than are genes with the highest expression in other somatic tis-

sues is counterintuitive because these genes are quite unlikely to be necessary for males. It is possible that the genes with the highest expression in the ovary are also important in male tissues. Indeed, the genes with the highest and second highest expression in the ovary and male tissues, respectively, were more likely to remain functional compared with those with the highest and second highest expression in the ovary and other female tissues, respectively, on the *D. miranda* neo-Y (Nozawa et al. 2018). Although these patterns were not reproduced in our updated data set for *D. miranda* with more stringent criteria for gene selection (proportions of pseudogenes were 0.43 and 0.37 for the former and the latter, respectively, with $P = 1.00$ by Fisher's exact test), we observed a weak but nonsignificant pattern in *D. albomicans* (0.14 and 0.17, respectively, with $P = 0.52$) and a marginally significant pattern in *D. americana* (0.02 and 0.09, respectively, with $P = 0.06$). Therefore, some genes with the highest expression in the ovary may also be important for males and remain functional on the neo-Y.

If the neo-X and neo-Y recombine at male meiosis, a higher retention rate of functionality for genes expressed the highest in
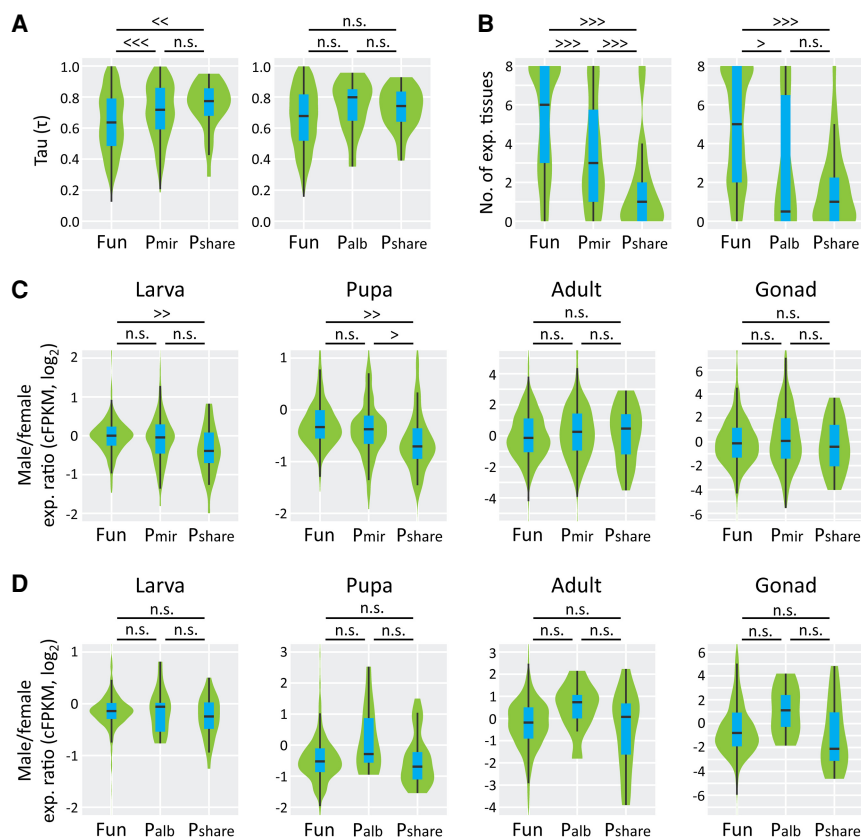
Figure 8. Relationship between parallel pseudogenization on the neo-Y and expression patterns. (*A*) Relationship between pseudogenization pattern and tissue specificity (τ). τ was computed for orthologs in *D. pseudoobscura* (*left* panel) and *D. nasuta* (*right* panel), closely related species of *D. miranda* and *D. albomicans*, respectively, as proxies of ancestral species. (*B*) Relationship between the pseudogenization pattern and the number of tissues in which each gene was expressed. A gene was regarded to be expressed in a tissue if cFPKM was one or more in *D. pseudoobscura* (*left* panel) and *D. nasuta* (*right* panel). (*C,D*) Relationship between the pseudogenization pattern and the male-to-female gene expression ratio based on the cFPKM value. (Fun) Neo-Y-linked genes functional at least in *D. miranda* (*left* panel in *A* and *B* and all panels in *C*) or *D. albomicans* (*right* panel in *A* and *B* and all panels in *D*); (P$_{mir}$) neo-Y-linked genes nonfunctional in *D. miranda* but functional in *D. albomicans*; (P$_{alb}$) neo-Y-linked genes nonfunctional in *D. albomicans* but functional in *D. miranda*; and (P$_{share}$) neo-Y-linked genes nonfunctional in both *D. miranda* and *D. albomicans*. Differences between groups were evaluated by a Mann–Whitney *U* test with correction for multiple testing: (<<< or >>>) $Q < 0.001$; (<< or >>) $Q < 0.01$; (< or >) $Q < 0.05$; (n.s.) $Q \geq 0.05$.

ovary on the neo-Y may also be possible, because pseudogenes on the neo-Y can theoretically be replaced with neo-X-derived functional genes. Although recombination at meiosis does not occur in males in *D. melanogaster* and many other *Drosophila* species (John et al. 2016), *D. nasuta*, closely related to *D. albomicans*, shows recombination during male meiosis (Satomura and Tamura 2016). In addition, the neo-X and neo-Y in *D. albomicans* likely recombined until recently, supported by a large number of shared polymorphisms, although male recombination does not occur at present in *D. albomicans* (Satomura and Tamura 2016; see also Wei and Bachtrog 2019). A large number of shared polymorphisms between neo-sex chromosomes has also been reported in *D. miranda* (Nozawa et al. 2018), suggestive of male recombination in *D. miranda* and/or its recent ancestor. In *D. americana*, sufficient polymorphism data are lacking. However, the geographical distributions of *D. americana* and the closely related *D. texana* partly overlap (Sillero et al. 2014). Therefore, if a hybrid male is backcrossed with a *D. americana* female, a neo-Y pseudogene can be replaced with a functional gene that is highly expressed in the ovary, on the orthologous autosome (i.e., Chromosome 4) from *D. texana*.

Parallel pseudogenization of neo-Y-linked genes in *D. miranda* and *D. albomicans*

As we have discussed, there is a common evolutionary trajectory of neo-sex chromosomes in *Drosophila* species. One of the most striking findings among the share features may be the parallel pseudogenization of neo-Y-linked genes in *D. miranda* and *D. albomicans*. We did not detect any enriched GO terms in genes with parallel pseudogenization, possibly owing to the small number of genes (only 35 of the 814 genes analyzed). However, a detailed inspection revealed that sexual conflict may be involved in at least some of the parallel pseudogenization events (Supplemental Table S20). For example, *Idgf6* is involved in egg chamber tube morphogenesis (Zimmerman et al. 2017). In addition, *JhI-26*, originally identified as a juvenile hormone-inducible protein, is a sperm protein shown to reduce male fertility under overexpression (Liu et al. 2014). Therefore, these genes may be under sexual conflict and thus may have been pseudogenized after neo-Y-linkage.

Conclusions

In conclusion, neo-sex chromosomes in *Drosophila* with independent origins share evolutionary trajectories, including gene-by-gene DC, accelerated pseudogenization on the neo-X, and similar pseudogenization patterns. This evolutionary path may be a key for sex chromosomes to avoid becoming an evolutionary dead end. Further studies of a broad range of organisms with young sex chromosomes are required to evaluate whether the patterns reported in this study can be generalized to other groups of organisms.

Methods

Flies

D. americana (strain 15010-0951.03, Millersberg, United States), *D. texana* (strain 15010-1041.00, St. Francisville, United States), and *D. novamexicana* (strain 15010-1031.00, Grand Junction, United States) were obtained from the *Drosophila* Species Stock Center (https://www.drosophilaspecies.com/). *D. albomicans* (NG-3, Nago, Japan), *D. nasuta* (strain G-86, Curepipe, Mauritius), and

D. kohkoa (strain X-145, Kuala Lumpur, Malaysia) have been maintained as living stocks at Tokyo Metropolitan University since the late 1970s.

Genome and RNA sequencing

For genome sequencing, high-molecular-weight genomic DNA was extracted from males and females separately. Whole-genome shotgun sequencing was performed using the Illumina HiSeq 2500 and PacBio RSII/Sequel platforms (Pacific Biosciences). PacBio long-reads were used to assemble the genome using the Hierarchical Genome Assembly Process (HGAP) 3 or HGAP4 (Chin et al. 2013). Illumina short-reads were then used to correct errors using BWA (Li and Durbin 2009). To obtain the neo-Y assemblies for *D. americana* and *D. albomicans*, the neo-X sequences were replaced with male-specific variants. The genome assemblies of the nine species used in this study were evaluated using BUSCO ver. 4.0.2 (Simão et al. 2015). For details, see "DNA extraction, library construction, and sequencing" and "Genome assembly" in the Supplemental Methods.

For RNA sequencing, RNA was extracted from third instar larvae, pupae, adults, ovaries, and testes, and a cDNA library from each sample was sequenced using HiSeq 4000 or HiSeq X (Illumina). We prepared two biological replicates for each condition to account for variation among samples. The transcriptome assembly for each species was obtained using HISAT2-2.1.0 (Kim et al. 2019) and StringTie 1.3.6 (Pertea et al. 2015). TransDecoder-5.5.0 (Haas et al. 2013) and AUGUSTUS v.3.3.2 (Stanke and Waack 2003) were used to annotate coding sequences of expressed and nonexpressed genes, respectively. The expression level for each gene was estimated using the STAR 2.7 (Dobin et al. 2013)-RSEM v1.3.1 (Li and Dewey 2011) pipeline (for mapping rates, see Supplemental Table S21). For details, see "RNA extraction, library construction, and sequencing," "Transcriptome assembly," "Gene annotation," and "Estimation of gene expression level" in the Supplemental Methods.

Gene classification

To identify orthologs in the three closely related species (i.e., *D. miranda*–*D. pseudoobscura*–*D. obscura*, *D. albomicans*–*D. nasuta*–*D. kohkoa*, or *D. americana*–*D. texana*–*D. novamexicana*) and inparalogs for each species, we conducted a BLASTP (ver. 2.9.0+) search (Altschul et al. 1997). All identified genes were classified as functional, silenced, disrupted, silenced-and-disrupted, and unclassified based on coding sequence integrity and expression level. For details, see "Gene classification" in the Supplemental Methods.

Estimation of the extent of dosage compensation

To examine the level of DC for each gene, R_{Lin} was used (Lin et al. 2012). R_{Lin} was defined as follows:

$$R_{Lin} = \frac{cFPKM_{neo-X,tar}/cFPKM_{proto-X,com}}{median\_cFPKM_{A/X,tar}/median\_cFPKM_{A/X,com}},$$

where cFPKM is the corrected number of fragments per kilobase of exon per million mapped fragments, in which an individual FPKM is linearly adjusted to make the median expression 1.0. Subscripts tar, com, neo-X, proto-A, and A/X denote target species (e.g., *D. miranda*), comparing (or closely related) species (e.g., *D. pseudoobscura*), the neo-X Chromosome, an autosome orthologous to the neo-X in the comparing species (e.g., the Chromosome 3 in *D. pseudoobscura*), and autosomes/canonical X, respectively. This index detects changes in male expression between species owing to the emergence of sex chromosomes using changes in the expression of autosomal and X-linked genes as a control. If there is no

DC, the R_{Lin} value will be 0.5 because males of target species have only one copy of the neo-X, whereas males in related species carry two copies. In contrast, the value is expected to be 1.0 if DC is perfect. For details, see "Estimation of the extent of dosage compensation" in the Supplemental Methods.

Data access

All sequence data generated in this study have been submitted to the DDBJ Sequence Read Archive (DRA; https://ddbj.nig.ac.jp/DRASearch/) under the accession numbers DRA005245–DRA005248, DRA007619–DRA007626, DRA008052, DRA006582, and DRA004735. Genome sequences assembled in this study have been submitted to the NCBI GenBank database (https://www.ncbi.nlm.nih.gov/genbank/) under accession numbers BJEI01000001–BJEI01000834, BJEH01000001–BJEH01000604, BJEL01000001–BJEL01000087, BJEJ01000001–BJEJ01000435, BJEK01000001–BJEK01000199, and BJEM01000001–BJEM01000270. The gene annotation files for the genome sequences and in-house scripts are provided as Supplemental Material and Supplemental Code, respectively.

Competing interest statement

The authors declare no competing interests.

References

Alekseyenko AA, Ho JW, Peng S, Gelbart M, Tolstorukov MY, Plachetka A, Kharchenko PV, Jung YL, Gorchakov AA, Larschan E, et al. 2012. Sequence-specific targeting of dosage compensation in *Drosophila* favors an active chromatin context. *PLoS Genet* **8:** e1002646. doi:10.1371/journal.pgen.1002646

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389–3402. doi:10.1093/nar/25.17.3389

Argyridou E, Parsch J. 2018. Regulation of the X chromosome in the germline and soma of *Drosophila melanogaster* males. *Genes (Basel)* **9:** 242. doi:10.3390/genes9050242

Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman TL, Hahn MW, Kitano J, Mayrose I, Ming R, et al. 2014. Sex determination: why so many ways of doing it? *PLoS Biol* **12:** e1001899. doi:10.1371/journal.pbio.1001899

Bachtrog D, Mahajan S, Bracewell R. 2019. Massive gene amplification on a recently formed *Drosophila* Y chromosome. *Nat Ecol Evol* **3:** 1587–1597. doi:10.1038/s41559-019-1009-9

Benjamini H, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B* **57:** 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

Bracewell R, Bachtrog D. 2020. Complex evolutionary history of the Y chromosome in flies of the *Drosophila obscura* species group. *Genome Biol Evol* **12:** 494–505. doi:10.1093/gbe/evaa051

Bracewell R, Chatla K, Nalley MJ, Bachtrog D. 2019. Dynamic turnover of centromeres drives karyotype evolution in *Drosophila*. *eLife* **8:** e49002. doi:10.7554/eLife.49002

Celniker SE, Wheeler DA, Kronmiller B, Carlson JW, Halpern A, Patel S, Adams M, Champe M, Dugan SP, Frise E, et al. 2002. Finishing a whole-genome shotgun: release 3 of the *Drosophila melanogaster* euchromatic genome sequence. *Genome Biol* **3:** research0079.1. doi:10.1186/gb-2002-3-12-research0079

Charlesworth B. 2012. The role of background selection in shaping patterns of molecular evolution and variation: evidence from variability on the *Drosophila* X chromosome. *Genetics* **191:** 233–246. doi:10.1534/genetics.111.138073

Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. *Heredity (Edinb)* **95:** 118–128. doi:10.1038/sj.hdy.6800697

Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, et al. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* **10:** 563–569. doi:10.1038/nmeth.2474

Cortez D, Marin R, Toledo-Flores D, Froidevaux L, Liechti A, Waters PD, Grützner F, Kaessmann H. 2014. Origins and functional evolution of Y chromosomes across mammals. *Nature* **508:** 488–493. doi:10.1038/nature13151

Deng X, Hiatt JB, Nguyen DK, Ercan S, Sturgill D, Hillier LW, Schlesinger F, Davis CA, Reinke VJ, Gingeras TR, et al. 2011. Evidence for compensatory upregulation of expressed X-linked genes in mammals, *Caenorhabditis elegans* and *Drosophila melanogaster*. *Nat Genet* **43:** 1179–1185. doi:10.1038/ng.948

Disteche CM. 2012. Dosage compensation of the sex chromosomes. *Annu Rev Genet* **46:** 537–560. doi:10.1146/annurev-genet-110711-155454

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29:** 15–21. doi:10.1093/bioinformatics/bts635

Dupim EG, Goldstein G, Vanderlinde T, Vaz SC, Krsticevic F, Bastos A, Pinhão T, Torres M, David JR, Vilela CR, et al. 2018. An investigation of Y chromosome incorporations in 400 species of *Drosophila* and related genera. *PLoS Genet* **14:** e1007770. doi:10.1371/journal.pgen.1007770

Ellegren H, Hultin-Rosenberg L, Brunström B, Dencker L, Kultima K, Scholz B. 2007. Faced with inequality: Chicken do not have a general dosage compensation of sex-linked genes. *BMC Biol* **5:** 40. doi:10.1186/1741-7007-5-40

Ellison CE, Bachtrog D. 2013. Dosage compensation via transposable element mediated rewiring of a regulatory network. *Science* **342:** 846–850. doi:10.1126/science.1239552

Ellison CE, Bachtrog D. 2019. Contingency in the convergent evolution of a regulatory network: dosage compensation in *Drosophila*. *PLoS Biol* **17:** e3000094. doi:10.1371/journal.pbio.3000094

Flores SV, Evans AL, McAllister BF. 2008. Independent origins of new sex-linked chromosomes in the *melanica* and *robusta* species groups of *Drosophila*. *BMC Evol Biol* **8:** 33. doi:10.1186/1471-2148-8-33

Gartler SM. 2014. A brief history of dosage compensation. *J Genet* **93:** 591–595. doi:10.1007/s12041-014-0360-5

Graves JA. 2016. Evolution of vertebrate sex chromosomes and dosage compensation. *Nat Rev Genet* **17:** 33–46. doi:10.1038/nrg.2015.2

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nat Protoc* **8:** 1494–1512. doi:10.1038/nprot.2013.084

Hayakawa T, Suzuki-Hashido N, Matsui A, Go Y. 2014. Frequent expansions of the bitter taste receptor gene repertoire during evolution of mammals in the Euarchontoglires clade. *Mol Biol Evol* **31:** 2018–2031. doi:10.1093/molbev/msu144

John A, Vinayan K, Varghese J. 2016. Achiasmy: male fruit flies are not ready to mix. *Front Cell Dev Biol* **4:** 75. doi:10.3389/fcell.2016.00075

Kaiser VB, Zhou Q, Bachtrog D. 2011. Nonrandom gene loss from the *Drosophila miranda* neo-Y chromosome. *Genome Biol Evol* **3:** 1329–1337. doi:10.1093/gbe/evr103

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37:** 907–915. doi:10.1038/s41587-019-0201-4

Koerich LB, Wang X, Clark AG, Carvalho AB. 2008. Low conservation of gene content in the *Drosophila* Y chromosome. *Nature* **456:** 949–951. doi:10.1038/nature07463

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* **12:** 323. doi:10.1186/1471-2105-12-323

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25:** 1754–1760. doi:10.1093/bioinformatics/btp324

Lin F, Xing K, Zhang J, He X. 2012. Expression reduction in mammalian X chromosome evolution refutes Ohno's hypothesis of dosage compensation. *Proc Natl Acad Sci* **109:** 11752–11757. doi:10.1073/pnas.1201816109

Liu C, Wang JL, Zheng Y, Xiong EJ, Li JJ, Yuan LL, Yu XQ, Wang YF. 2014. Wolbachia-induced paternal defect in *Drosophila* is likely by interaction with the juvenile hormone pathway. *Insect Biochem Mol Biol* **49:** 49–58. doi:10.1016/j.ibmb.2014.03.014

Mahajan S, Wei KH, Nalley MJ, Gibilisco L, Bachtrog D. 2018. De novo assembly of a young *Drosophila* Y chromosome using single-molecule sequencing and chromatin conformation capture. *PLoS Biol* **16:** e2006348. doi:10.1371/journal.pbio.2006348

Mai D, Nalley MJ, Bachtrog D. 2020. Patterns of genomic differentiation in the *Drosophila nasuta* species complex. *Mol Biol Evol* **37:** 208–220. doi:10.1093/molbev/msz215

McElroy KA, Kang H, Kuroda MI. 2014. Are we there yet? Initial targeting of the Male-Specific Lethal and Polycomb group chromatin complexes in *Drosophila*. *Open Biol* **4:** 140006. doi:10.1098/rsob.140006

Meisel RP. 2020. Evolution of sex determination and sex chromosomes: a novel alternative paradigm. *Bioessays* **42:** 1900212. doi:10.1002/bies.201900212

Muller HJ. 1932. Further studies on the nature and causes of gene mutations. In *Proceedings of the Sixth International Congress of Genetics*, Ithaca, NY, pp. 213–255.

Nei M, Niimura Y, Nozawa M. 2008. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nat Rev Genet* **9:** 951–963. doi:10.1038/nrg2480

Niimura Y, Nei M. 2007. Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLoS One* **2:** e708. doi:10.1371/journal.pone.0000708

Nozawa M, Nei M. 2007. Evolutionary dynamics of olfactory receptor genes in *Drosophila* species. *Proc Natl Acad Sci* **104:** 7122–7127. doi:10.1073/pnas.0702133104

Nozawa M, Fukuda N, Ikeo K, Gojobori T. 2014. Tissue- and stage-dependent dosage compensation on the neo-X chromosome in *Drosophila pseudoobscura*. *Mol Biol Evol* **31:** 614–624. doi:10.1093/molbev/mst239

Nozawa M, Onizuka K, Fujimi M, Ikeo K, Gojobori T. 2016. Accelerated pseudogenization on the neo-X chromosome in *Drosophila miranda*. *Nat Commun* **7:** 13659. doi:10.1038/ncomms13659

Nozawa M, Ikeo K, Gojobori T. 2018. Gene-by-gene or localized dosage compensation on the neo-X chromosome in *Drosophila miranda*. *Genome Biol Evol* **10:** 1875–1881. doi:10.1093/gbe/evy148

Ohno S. 1967. *Sex chromosomes and sex-linked genes*. Springer-Verlag, Berlin.

Olvera O, Powell JR, de la Rosa ME, Salceda VM, Gaso MI, Guzman J, Anderson WW, Levine L. 1979. Population genetics of Mexican *Drosophila* VI: cytogenetic aspects of the inversion polymorphism in *Drosophila pseudoobscura*. *Evolution* **33:** 381–395. doi:10.1111/j.1558-5646.1979.tb04691.x

Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33:** 290–295. doi:10.1038/nbt.3122

Satomura K, Tamura K. 2016. Ancient male recombination shaped genetic diversity of neo-Y chromosome in *Drosophila albomicans*. *Mol Biol Evol* **33:** 367–374. doi:10.1093/molbev/msv221

Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness. *Methods Mol Biol* **1962:** 227–245. doi:10.1007/978-1-4939-9173-0_14

Sillero N, Reis M, Vieira CP, Vieira J, Morales-Hojas R. 2014. Niche evolution and thermal adaptation in the temperate species *Drosophila americana*. *J Evol Biol* **27:** 1549–1561. doi:10.1111/jeb.12400

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31:** 3210–3212. doi:10.1093/bioinformatics/btv351

Stanke M, Waack S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19** (Suppl. 2): ii215–ii225. doi:10.1093/bioinformatics/btg1080

Valsecchi CIK, Basilicata MF, Georgiev P, Gaub A, Seyfferth J, Kulkarni T, Panhale A, Semplicio G, Manjunath V, Holz H, et al. 2021. RNA nucleation by MSL2 induces selective X chromosome compartmentalization. *Nature* **589:** 137–142. doi:10.1038/s41586-020-2935-z

Vibranovski MD. 2014. Meiotic sex chromosome inactivation in *Drosophila*. *J Genomics* **2:** 104–117. doi:10.7150/jgen.8178

Vicoso B, Bachtrog D. 2011. Lack of global dosage compensation in *Schistosoma mansoni*, a female-heterogametic parasite. *Genome Biol Evol* **3:** 230–235. doi:10.1093/gbe/evr010

Vicoso B, Bachtrog D. 2015. Numerous transitions of sex chromosomes in Diptera. *PLoS Biol* **13:** e1002078. doi:10.1371/journal.pbio.1002078

Vieira CP, Coelho PA, Vieira J. 2003. Inferences on the evolutionary history of the *Drosophila americana* polymorphic *X/4* fusion from patterns of polymorphism at the *X*-linked *paralytic* and *elav* genes. *Genetics* **164:** 1459–1469. doi:10.1093/genetics/164.4.1459

Wei KH, Bachtrog D. 2019. Ancestral male recombination in *Drosophila albomicans* produced geographically restricted neo-Y chromosome haplotypes varying in age and onset of decay. *PLoS Genet* **15:** e1008502. doi:10.1371/journal.pgen.1008502

Wilson Sayres MA, Makova KD. 2013. Gene survival and death on the human Y chromosome. *Mol Biol Evol* **30:** 781–787. doi:10.1093/molbev/mss267

Xia S, VanKuren NW, Chen C, Zhang L, Kemkemer C, Shao Y, Jia H, Lee U, Advani AS, Gschwend A, et al. 2021. Genomic analyses of new genes and their phenotypic effects reveal rapid evolution of essential functions in *Drosophila* development. *PLoS Genet* **17:** e1009654. doi:10.1371/journal.pgen.1009654

Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, Bar-Even A, Horn-Saban S, Safran M, Domany E, et al. 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21:** 650–659. doi:10.1093/bioinformatics/bti042

Zha X, Xia Q, Duan J, Wang C, He N, Xiang Z. 2009. Dosage analysis of Z chromosome genes using microarray in silkworm, *Bombyx mori*. *Insect Biochem Mol Biol* **39:** 315–321. doi:10.1016/j.ibmb.2008.12.003

Zhou Q, Bachtrog D. 2015. Ancestral chromatin configuration constrains chromatin evolution on differentiating sex chromosomes in *Drosophila*. *PLoS Genet* **11:** e1005331. doi:10.1371/journal.pgen.1005331

Zhou Q, Ellison CE, Kaiser VB, Alekseyenko AA, Gorchakov AA, Bachtrog D. 2013. The epigenome of evolving *Drosophila* neo-sex chromosomes: dosage compensation and heterochromatin formation. *PLoS Biol* **11:** e1001711. doi:10.1371/journal.pbio.1001711

Zhou Q, Zhang J, Bachtrog D, An N, Huang Q, Jarvis ED, Gilbert MT, Zhang G. 2014. Complex evolutionary trajectories of sex chromosomes across bird taxa. *Science* **346:** 1246338. doi:10.1126/science.1246338

Zimmerman SG, Merrihew GE, MacCoss MJ, Berg CA. 2017. Proteomics analysis identifies orthologs of human chitinase-like proteins as inducers of tube morphogenesis defects in *Drosophila melanogaster*. *Genetics* **206:** 973–984. doi:10.1534/genetics.116.199323

Shared evolutionary trajectories of three independent neo-sex chromosomes in *Drosophila*

Masafumi Nozawa, Yohei Minakuchi, Kazuhiro Satomura, et al.

Supplemental Material	http://genome.cshlp.org/content/suppl/2021/10/08/gr.275503.121.DC1
References	This article cites 64 articles, 7 of which can be accessed free at: http://genome.cshlp.org/content/31/11/2069.full.html#ref-list-1
Open Access	Freely available online through the *Genome Research* Open Access option.
Creative Commons License	This article, published in *Genome Research*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**