

Gene-by-gene or localized dosage compensation on the neo-X chromosome in *Drosophila miranda*

Masafumi Nozawa^{1,2*}, Kazuho Ikeo^{3,4}, and Takashi Gojobori⁵

¹ Department of Biological Sciences, Tokyo Metropolitan University, 1-1 Minamiosawa, Hachioji, Tokyo 192-0397, Japan

² Research Center for Genomics and Bioinformatics, Tokyo Metropolitan University, Hachioji, Tokyo 192-0397, Japan

³ Center for Information Biology, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan

⁴ Department of Genetics, SOKENDAI, 1111 Yata, Mishima, Shizuoka 411-8540, Japan

⁵ King Abdullah University of Science and Technology (KAUST), Computational Bioscience Research Center, Biological and Environmental Science and Engineering, 4700 KAUST, Thuwal 23955-6900, Kingdom of Saudi Arabia

* Corresponding author: Masafumi Nozawa, manozawa@tmu.ac.jp

ABSTRACT

Many organisms have a global mechanism for dosage compensation (DC) operating along the entire male X chromosome, which equalizes gene expression on the male X with that on the two Xs in females and/or on autosomes. At the initial stage of sex chromosome evolution, however, gene-by-gene (or localized) DC may also be necessary because the degeneration of Y-linked genes occurs independently at different times. We therefore tested whether the up-regulation of X-linked genes depends on the status of their Y-linked homologs, using the young sex chromosomes, neo-X and neo-Y, in *Drosophila miranda*. In support of the presence of gene-by-gene DC, the extent of up-regulation in males was indeed higher for neo-X-linked genes with pseudogenized neo-Y-linked homologs than for neo-X-linked genes with functional neo-Y-linked homologs. Further molecular evolutionary analysis also supports the idea that many individual neo-X-linked genes first acquired the potential for up-regulation, which then enabled the pseudogenization of neo-Y-linked homologs, without serious deleterious effects on male fitness.

Y chromosomes often exhibit substantial degeneration (Lander et al. 2001; Zhou and Bachtrog 2012; Papadopoulos et al. 2015). This potential disadvantage of having sex chromosomes is believed to be overcome by dosage compensation (DC), which equalizes gene dosage between one X chromosome in males and two Xs in females as well as between the X and autosomes in males (Ohno 1967; Disteche 2012). Indeed, many organisms have DC mechanisms (Deng et al. 2011; Conrad and Akhtar 2012; Julien et al. 2012; Papadopoulos et al. 2015).

In *Drosophila*, a single X chromosome in males is globally up-regulated to equalize the expression level to that of the two Xs in females as well as that of autosomes (Gelbart and Kuroda 2009; Conrad and Akhtar 2012; Disteche 2012). This global up-regulation of the male X is initiated by the DC complex (DCC). The binding of the DCC to a sequence motif, i.e., chromatin entry sites (CESs), which are scattered across the male X, promotes histone H4 acetylation, which in turn stimulates the hyper-transcription of the many genes on the X chromosome by opening the chromatin structure (Alekseyenko et al. 2012).

During the early evolution of sex chromosomes, however, genes on the Y chromosome are expected to become individually pseudogenized at different times, rather than to become simultaneously non-functionalized. In this case, the global up-regulation of genes on the entire male X may cause the over-expression of X-linked genes whose Y-linked homologs are still functional. This over-expression of X-linked genes may also be deleterious for males. Therefore, a gene-by-gene (or localized) DC mechanism may be necessary at the initial stage of sex chromosome evolution. Indeed, Nozawa et al. (2014) found that DC varies considerably among genes depending on their functional constraints by analyzing the right arm of the X chromosome (or XR), which emerged about 15 million years ago (Mya) in the lineage leading to *D.*

pseudoobscura. However, the XR was too old to study the relationship between the pseudogenization of Y-linked genes and the up-regulation of X-linked homologs; the Y chromosome in *D. pseudoobscura* was almost completely degenerated, and the interplay between the X and Y chromosomes could not be investigated.

In *Drosophila*, even younger sex chromosomes than those of *D. pseudoobscura* emerged independently in different species (Chang et al. 2008). One of such species is *D. miranda* with the so-called neo-sex chromosomes. *D. miranda* is a close relative of *D. pseudoobscura*. After splitting from *D. pseudoobscura* about 2 Mya, an autosome of *D. miranda* became the neo-Y via a fusion with the original Y chromosome about 1 Mya, and consequently, the remaining homologous autosome became the neo-X [see Nozawa et al. (2016) for more detailed processes of karyotype evolution of *D. miranda*]. The neo-Y in *D. miranda* is so young that degeneration is ongoing, providing an excellent target to study the interplay between the sex chromosomes. Indeed, many studies have already been conducted by using this species (e.g., Marin et al. 1996; Zhou and Bachtrog 2012; Ellison and Bachtrog 2013; Zhou et al. 2013; Nozawa et al. 2016). In particular, Zhou et al. (2013) extensively studied the relationship between the functionality of neo-Y-linked genes and the chromatin states of the neo-X-linked homologs. In this study, we therefore focused more on the interplay between the functionality of neo-Y-linked genes and the gene expression level of the neo-X-linked homologs.

Gene-by-gene dosage compensation operates on the neo-X chromosome

To measure the DC level for each neo-X-linked gene in males, we used R_{Lin} , developed by Lin et al. (2012). In short, R_{Lin} computes the ratio of the expression level of an X-linked gene in one species to that of its autosomal ortholog in another species by using the median expression

level of genes located on autosomes in both species as a control (see *Materials and Methods* for more details). A value of 1 indicates perfect DC, whereas a value of 0.5 indicates a lack of DC. If gene-by-gene DC operates on the neo-X and the interplay between the neo-X and neo-Y is fine-tuned at the gene expression level, R_{Lin} is expected to be greater for the neo-X-linked genes whose neo-Y-linked homologs are already pseudogenized than for the neo-X-linked genes whose neo-Y-linked homologs are still functional.

We found that R_{Lin} values are indeed significantly greater for the neo-X-linked genes with pseudogenized neo-Y-linked homologs (X_F-Y_P) than for the neo-X-linked genes with functional neo-Y-linked homologs (X_F-Y_F) (Figs. 1 and S1). The level of DC for X_F-Y_P genes was high in the thorax. In the testis, by contrast, DC for X_F-Y_P genes was less obvious, possibly reflecting meiotic sex chromosome inactivation, a mechanism for repressing gene expression on sex chromosomes during male meiosis (Vibrantovski 2014 for review), or less efficient DC in the testis (Mikhaylova and Nurminsky 2011; Landeen et al. 2016). Consistent with this observation, the abdomen without gonad tissues showed a larger R_{Lin} than the abdomen including gonad tissues (median values for X_F-Y_P genes were 0.84 and 0.73, respectively; Fig. 1). For X_F-Y_F genes, when the expression of both functional neo-X-linked and neo-Y-linked homologs was considered to compute R_{Lin} , the expression levels of the genes in *D. miranda* became similar to those in *D. pseudoobscura*, one of the most closely related species to *D. miranda*, used as a proxy of the ancestral state before neo-sex chromosomes were acquired (i.e., R_{Lin} for X_F+Y_F was ~ 1 ; Fig. S1). Essentially the same pattern of R_{Lin} was observed for all other tissues examined. These results are consistent with the patterns expected under a model of gene-by-gene DC. Note that the X_F-Y_F genes also showed R_{Lin} values significantly greater than 0.5 (Figs. 1 and S1), consistent with global up-regulation on the male neo-X, as previously reported (Ellison and

Bachtrog 2013; Zhou et al. 2013). Therefore, both gene-by-gene and global DC seem to have operated on the male neo-X in *D. miranda*, but gene-by-gene DC may have been particularly important to overcome massive pseudogenization on the neo-Y during the initial stage of sex chromosome evolution.

It should be mentioned that R_{Lin} ranges from zero to infinite, because the expression level of a gene in a species can change even without acquiring sex chromosomes (i.e., without the effect of DC). This is why individual genes often show the R_{Lin} values less than 0.5 or greater than 1. Therefore, considering a representative value (e.g., median) of R_{Lin} values for many genes in each category would be more meaningful to grasp a general pattern of up-regulation than focusing on the R_{Lin} values for individual genes.

It should also be mentioned that R_{Lin} can also be computed for females (see *Materials and Methods*). In females, however, R_{Lin} values for X_F - Y_P genes are not necessarily greater than the values for X_F - Y_F genes (although depending on the tissues examined, Fig. S2). This observation suggests that the gene-by-gene up-regulation is primarily male-specific and operates as a mechanism of DC rather than gene-specific up-regulation operating on both sexes.

Individual neo-X-linked genes seem to have first acquired the potential for up-regulation, which have promoted the pseudogenization of neo-Y-linked homologs

If the up-regulation of neo-X-linked genes in males has occurred in response to the pseudogenization of their neo-Y-linked homologs, DC is expected to be incomplete at the initial stage and gradually fine-tuned over time. In this case, R_{Lin} for the neo-X-linked genes would initially be close to 0.5 and become greater with time after the pseudogenization of the neo-Y-linked homologs. To test this possibility, we estimated the relative time elapsed since the

pseudogenization of neo-Y-linked genes (Fig. S3; Oda et al. 2002). We then classified all X_F - Y_P genes into four bins based on this relative time. Another group for which neo-Y-linked genes were already deleted from the chromosome was also evaluated. The results showed that the DC level is essentially the same, irrespective of the time since the pseudogenization of neo-Y-linked genes (Fig. 2). This observation suggests a general trend that individual neo-X-linked genes first acquired the potential for up-regulation to maintain a stable expression level. After the establishment of such a mechanism for the neo-X-linked genes, the neo-Y-linked homologs could have become easily pseudogenized, without serious deleterious effects on fitness. Yet, our estimates of the relative pseudogenization time contain large sampling errors due to small sequence divergence of neo-X and neo-Y and we cannot still completely rule out other possibilities. For example, it is known that effects of chromosomal aneuploidies on gene expression are often counteracted by a gene-expression buffering mechanism (Stenberg et al. 2009). Therefore, dosage reduction due to pseudogenization of neo-Y-linked genes might have also been buffered to some extent by an intrinsic mechanism that up-regulates their neo-X-linked homologs. However, note that in our definition, this mechanism can also be classified as a gene-by-gene or a localized DC but not a global DC, because in this case up-regulation occurs only on the neo-X-linked genes whose neo-Y-linked homologs are pseudogenized but not on an entire chromosome. In any case, if a genome-editing technique in *D. miranda* is developed, we may be able to evaluate the possibility of this buffering mechanism by making a mutant in which a functional neo-Y-linked gene is knocked-out (i.e., artificially pseudogenized) and by testing whether the expression level of the neo-X-linked homolog is up-regulated in a male.

Genes highly expressed in male gonads tend to remain functional on the neo-Y chromosome

We next examined whether pseudogenization of neo-Y-linked genes depends on the spatiotemporal gene expression pattern. For this analysis, all genes on the neo-sex chromosomes were classified into groups depending on the tissue where the genes are expressed most highly in *D. pseudoobscura*, as a proxy for the ancestor without neo-sex chromosomes. In this way, the kinds of genes in the ancestor that tend to remain functional after becoming neo-sex-linked, if any, were examined. Using a similar approach, Kaiser et al. (2011) reported that male-biased genes have remained functional at a higher rate compared with female-biased genes on the neo-Y. We also observed that the neo-Y-linked genes expressed most highly in the testis or accessory gland had a greater tendency to remain functional than the genes expressed most highly in other somatic tissues (Fig. 3A). Consistent with this observation, the category of genes that are functional only on the neo-Y (X_P - Y_F) contained a significantly greater proportion of genes with high expression in the testis or accessory gland than other categories of genes (Fig. 3B). Qualitatively similar results were obtained when more stringent criteria were applied for defining the tissue with the maximal expression for each gene (Fig. S4). However, we also observed an unexpected trend in which the proportion of genes that are functional was significantly greater for the genes expressed most highly in the ovary than for the genes expressed most highly in other somatic tissues (Figs. 3A). To further inspect this observation, we classified these “ovary” genes into groups based on the tissues in which they exhibit the second-highest expression. The proportion of genes that are functional was significantly greater for neo-Y-linked genes with the second-highest expression in male tissues than for the genes with the second-highest expression in female tissues (0.48 vs. 0.35, respectively; $P = 0.02$ by a Monte-Carlo simulation with 1,000

bootstrap replicates). Therefore, some of these “ovary” genes would have been expressed and function in male tissues as well and have remained functional, even after becoming neo-Y-linked.

An additional possibility is that the neo-X and neo-Y are still recombining or experienced recent recombination during male meiosis at a certain frequency, although it is believed that males of many *Drosophila* species do not exhibit recombination during meiosis (John et al. 2016; Charlesworth 2017). If male recombination indeed occurred in the *D. miranda* lineage, the functional genes on the neo-X that are expressed in the ovary may have moved onto the neo-Y during meiosis. Providing indirect support for this hypothesis, the observed number of shared polymorphisms between the neo-sex chromosomes was much greater than expected under a random mutation model on each chromosome (Table S1). In addition, nucleotide diversity on the neo-Y ($\pi = 6.4 \times 10^{-4}$) was more than half compared with that on the neo-X ($\pi = 10.7 \times 10^{-4}$). This Y/X ratio of nucleotide diversity was greater than one-third, an expected ratio at equilibrium derived from the number of the neo-X and neo-Y in a pair of parents. Although Bachtrog (2004) reported that nucleotide diversity on the neo-Y is ~30 times less than that on the neo-X in *D. miranda*, this conclusion was largely derived from an analysis of non-homologous genes from neo-X and neo-Y. By contrast, our estimates are based on the homologous genes, which would be more appropriate for the analysis.

Note that meiotic male recombination also seems to have resulted in comparable levels of nucleotide diversity and a large number of shared polymorphism between the neo-X and neo-Y, which emerged independently in *D. albomicans* (Satomura and Tamura 2016). With our knowledge, however, male recombination in *D. pseudoobscura*, a closely related species of *D. miranda*, has not been reported in the long research history since 1930's initiated by Dobzhansky. In addition, we used only two *D. miranda* strains to estimate nucleotide diversity, which is much

smaller than the number of strains used in Bachtrog (2004). Therefore, the proposal of this hypothesis should be taken with cautions. Crossing experiments in *D. miranda* are apparently necessary to further evaluate this hypothesis.

Global dosage compensation machinery unlikely contributes to gene-by-gene dosage compensation

If the DCC that initiated global DC also contributed to gene-by-gene DC, the X_F - Y_P genes would have most likely been recognized by male-specific lethal 3 (MSL3), a component of the DCC, and subjected to histone acetylation at H4K16 (H4K16ac hereafter) at a higher frequency than the X_F - Y_F genes. However, Zhou et al. (2013) reported that no significant difference was observed between the X_F - Y_F and X_F - Y_P genes in the proportion of neo-X-linked genes bound by MSL3 or H4K16ac. Nevertheless, they also reported that the neo-X-linked genes bound by MSL3 or H4K16ac tended to be up-regulated to a greater extent than the neo-X-linked genes unbound by these proteins (Zhou et al. 2013). Using our RNA-seq data with the ChIP-seq data from Zhou et al. (2013), we also observed the same tendency (Fig. S5). In addition, the variance of R_{Lin} for the neo-X-linked genes bound by MSL3 or H4K16ac was significantly smaller than that for unbound neo-X-linked genes ($P < 10^{-180}$ for each by the F -test). These observations suggest that the molecular mechanism underlying global DC does not contribute substantially to gene-by-gene DC, but broadly increases the expression of neo-X-linked genes and stabilizes the expression levels within a certain range.

In this study, we were unable to clarify the molecular basis of gene-by-gene (or localized) DC. Nevertheless, if an auto-regulation or a gene-expression buffering mechanism operates on individual neo-X-linked genes, gene-by-gene DC may be quickly established in the early stage of

sex chromosome evolution. Rapidly evolving elements such as satellite DNAs may also contribute to gene-by-gene DC. Indeed, siRNA-directed modification of chromatin structure at the 1.688^X satellite loci seems to be an identifier of the male X chromosome, independent from CESs for DC in *D. melanogaster* (Menon and Meller 2015; Deshpande and Meller 2018). However, whether this type of repetitive elements can function for DC in a gene-by-gene manner is another issue to be clarified.

The interplay between young sex chromosomes at the gene expression level has also been reported in a plant species, *Silene latifolia* (Muyle et al. 2012). According to this study, the expression of X-linked genes increases as the expression of Y-linked homologs decreases in males. Therefore, gene-by-gene or localized DC mechanisms may be present in many organisms with young sex chromosome systems. However, once most of the Y-linked genes are degenerated, global DC would be more effective than gene-by-gene DC. In this way, gene-by-gene DC may disappear as a global DC mechanism is developed. In this sense, the neo-sex chromosome system in *D. miranda* might be a transient stage from gene-by-gene DC to global DC.

In conclusion, our findings support the Ohno's (1967) statement that “*During the course of evolution of this insect (Drosophila), the dosage compensation mechanism must have developed one by one for each individual X-linked gene*”, which is crucial to counteract the deleterious effects due to the massive losses of Y-linked genes at the initial stage of sex chromosome evolution.

MATERIALS AND METHODS

Flies and sequencing data

To estimate the nucleotide polymorphism in *Drosophila miranda*, two strains (14011-0101.16 and 14011-0101.17) were obtained from Drosophila Species Stock Center at UC San Diego (<https://stockcenter.ucsd.edu>). To remove intra-strain polymorphism as much as possible, full-sib mating was repeated for eight generations for each strain before starting experiments. Genome and transcriptome sequences as well as gene expression data for *D. miranda* (strain 14011-0101.17), *D. pseudoobscura* [strain k-s12 (or 14011-0121.94)], and *D. obscura* (strain 14011-0151.01) were obtained from Nozawa et al. (2016).

DNA extraction, PCR, and sequencing

To estimate the nucleotide polymorphism in *D. miranda*, we extracted total DNA from ~10 adult females and males of the two strains, separately, using the Boom et al.'s (1990) method with slight modifications. PCR was conducted in a 25 µl reaction mixture in which 50 mM KCl, 10 mM Tris-HCl (pH8.5), 2 mM MgCl₂, 0.001% gelatin, 0.2 mM each dNTP, 1 µM each primer (forward and reverse), ~5 ng total DNA, and 625 mU *Taq* polymerase were included. Primers used are listed in Table S2. PCR amplification was performed as follows: 30 cycles of 94°C for 30 sec, 60°C for 30 sec, and 72°C for 60 sec. The nucleotide sequences of the amplified fragments were determined by using BigDye Terminator v3.1 Cycle Sequencing Kit and 3130xl Genetic Analyzer (Thermo Fisher Scientific, Waltham, USA). Neo-X-derived sequences were determined from female sequences, whereas neo-Y-derived sequences were determined by subtracting neo-X-derived sequences from male sequences by using MEGA7 (Kumar et al. 2016). It should be mentioned that all primers used were designed to amplify both neo-X- and neo-Y-derived sequences. All sequence data generated in this study are deposited to DDBJ under the accession numbers LC363594-LC363749.

Definition of functional genes and pseudogenes

We obtained the information for the classification of functional genes and pseudogenes from Nozawa et al. (2016). In Nozawa et al. (2016) all genes are classified as follows. (1) If a gene was expressed, contained an initiation as well as a stop codon, and its CDS length was equal to or greater than 80% of the average of other members of an orthologous group, the gene was regarded as ‘functional’. (2) If a gene was expressed, contained an initiation and a stop codon, and its CDS length was less than 80% of the average of other orthologous members, the gene was regarded as being ‘disrupted’. (3) If a gene was not expressed and its CDS length was equal to or greater than 80% of the average of other orthologous members, the gene was being ‘silenced’. (4) If a gene was not expressed and its CDS length was less than 80% of the average of other orthologous members, the gene was regarded as being ‘silenced and disrupted’. (5) If a gene was expressed, lacked either one of an initiation or a stop codon, and its flanking 100 nucleotides did not contain any ambiguous site, the gene was being ‘disrupted’. (6) If a gene was expressed in at least one of the tissues, lacked either one of an initiation or a stop codon, and flanking 100 nucleotides contained at least one ambiguous site (i.e., N), the gene was regarded as being ‘unclassified’.

Estimation of dosage compensation level

To evaluate the extent of dosage compensation (DC), we used R_{Lin} developed by Lin et al. (2012). The equation can be written as

$$R_{\text{Lin}} = \frac{\text{cFPKM}_{X,\text{tar}} / \text{cFPKM}_{\text{Au(proto-X),com}}}{\text{median_cFPKM}_{\text{Au,tar}} / \text{median_cFPKM}_{\text{Au,com}}},$$

where cFPKM is the corrected number of frgments per kilobase of exon per million mapped fragments, which linearly adjusted an individual FPKM to make the median expression identical to 1. Subscripts tar, com, X, Au(proto-X), and Au denote target species (*D. miranda* in this study), comparing species (*D. pseudoobscura* in this study), the X chromosome (the neo-X in this study), an autosome orthologous to the X in comparing species (the chromosome 3 in this study), and autosomes, respectively. The essence of this equation is that we can detect changes in male expression between species due to the emergence of sex chromosomes using expression change of autosomal genes as a control. If there is no DC, the R_{Lin} value will be 0.5 because *D. miranda* males have only one copy of the Muller element C (i.e., the neo-X), whereas males in *D. pseudoobscura* carry two copies (i.e., the chromosome 3). In contrast, the value is expected to be 1 if DC is perfect. For this analysis, only the orthologs that are expressed in *D. pseudoobscura*, are present in *D. obscura*, and have not experienced any inter-chromosomal translocation in the lineage leading to *D. miranda* and *D. pseudoobscura* after splitting from *D. obscura* were used. It should be mentioned that R_{Lin} can be calculated for females as well by using female expression data. All orthologous genes with genomic location, functionality, and expression level are listed in Table S3.

ChIP-seq analysis

The lists of genes that are bound by MSL3 or H4K16ac in the male third instar larvae of *D. miranda* (strain MSH22) were kindly provided by Qi Zhou (Zhou et al. 2013).

ACKNOWLEDGMENTS

We thank the editor and the reviewers for their constructive comments on our manuscript. We also thank Mai Fujimi, Chie Iwamoto, Miu Kubota, Kanako Onizuka, and Yasuko Urabe for their help in conducting experiments and Masanori Toda for species identification. We are also grateful to Yosuke Seto, Aya Takahashi, Koichiro Tamura, Kentaro M. Tanaka, and Ryo Yamaguchi for their comments on earlier versions of the manuscript. This work was supported by grants from Tokyo Metropolitan University and Japan Society for the Promotion of Science (Grant numbers 25711023, 15K14585, and 17H05015) to M.N.

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.
- Bachtrog D. 2004. Evidence that positive selection drives Y-chromosome degeneration in *Drosophila miranda*. *Nat Genet.* 36:518-522.
- Boom R, Sol CJ, Salimans MM, Jansen CL, Wertheim-van Dillen PM, van der Noordaa J. 1990. Rapid and simple method for purification of nucleic acids. *J Clin Microbiol.* 28:495-503.
- Chang T-P, Tsai T-H, Chang H-y. 2008. Fusions of Muller's elements during chromosome evolution of *Drosophila albomicans*. *Zool Stud.* 47:574-584.
- Charlesworth D. 2017. Evolution of recombination rates between sex chromosomes. *Philos Trans R Soc Lond B Biol Sci.* 372.

- Conrad T, Akhtar A. 2012. Dosage compensation in *Drosophila melanogaster*: epigenetic fine-tuning of chromosome-wide transcription. *Nat Rev Genet.* 13:123-134.
- 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.
- Deshpande N, Meller N. 2018. Chromatin that guides dosage compensation is modulated by the siRNA pathway in *Drosophila melanogaster*. *Genetics* In Press.
- Disteche CM. 2012. Dosage compensation of the sex chromosomes. *Annu Rev Genet.* 46:537-560.
- Ellison CE, Bachtrog D. 2013. Dosage compensation via transposable element mediated rewiring of a regulatory network. *Science* 342:846-850.
- Gelbart ME, Kuroda MI. 2009. *Drosophila* dosage compensation: a complex voyage to the X chromosome. *Development* 136:1399-1410.
- John A, Vinayan K, Varghese J. 2016. Achiasmy: Male Fruit Flies Are Not Ready to Mix. *Front Cell Dev Biol.* 4:75.
- Julien P, Brawand D, Soumillon M, Necsulea A, Liechti A, Schutz F, Daish T, Grutzner F, Kaessmann H. 2012. Mechanisms and evolutionary patterns of mammalian and avian dosage compensation. *PLoS Biol.* 10:e1001328.
- Kaiser VB, Zhou Q, Bachtrog D. 2011. Nonrandom gene loss from the *Drosophila miranda* neo-Y chromosome. *Genome Biol Evol.* 3:1329-1337.
- Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol.* 33:1870-1874.

- Landeem EL, Muirhead CA, Wright L, Meiklejohn CD, Presgraves DC. 2016. Sex Chromosome-wide Transcriptional Suppression and Compensatory Cis-Regulatory Evolution Mediate Gene Expression in the *Drosophila* Male Germline. *PLoS Biol.* 14:e1002499.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860-921.
- 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 USA* 109:11752-11757.
- Marin I, Franke A, Bashaw GJ, Baker BS. 1996. The dosage compensation system of *Drosophila* is co-opted by newly evolved X chromosomes. *Nature* 383:160-163.
- Menon DU, Meller VH. 2015. Identification of the *Drosophila* X chromosome: The long and short of it. *RNA Biol* 12:1088-1093.
- 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.
- Muyle A, Zemp N, Deschamps C, Mousset S, Widmer A, Marais GA. 2012. Rapid de novo evolution of X chromosome dosage compensation in *Silene latifolia*, a plant with young sex chromosomes. *PLoS Biol.* 10:e1001308.
- Nozawa M, Fukuda N, Ieko K, Gojobori T. 2014. Tissue- and stage-dependent dosage compensation on the neo-X chromosome in *Drosophila pseudoobscura*. *Mol Biol Evol.* 31:614-624.

- 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.
- Oda M, Satta Y, Takenaka O, Takahata N. 2002. Loss of urate oxidase activity in hominoids and its evolutionary implications. *Mol Biol Evol.* 19:640-653.
- Ohno S. 1967. *Sex chromosomes and sex-linked genes*: Springer-Verlag.
- Papadopulos AS, Chester M, Ridout K, Filatov DA. 2015. Rapid Y degeneration and dosage compensation in plant sex chromosomes. *Proc Natl Acad Sci USA* 112:13021-13026.
- Satomura K, Tamura K. 2016. Ancient Male Recombination Shaped Genetic Diversity of Neo-Y Chromosome in *Drosophila albomicans*. *Mol Biol Evol.* 33:367-374.
- Stenberg P, Lundberg LE, Johansson AM, Ryden P, Svensson MJ, Larsson J. 2009. Buffering of segmental and chromosomal aneuploidies in *Drosophila melanogaster*. *PLoS Genet.* 5:e1000465.
- Vibrantovski MD. 2014. Meiotic sex chromosome inactivation in *Drosophila*. *J Genomics* 2:104-117.
- Zhou Q, Bachtrog D. 2012. Sex-specific adaptation drives early sex chromosome evolution in *Drosophila*. *Science* 337:341-345.
- 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.

FIGURE LEGENDS

Fig. 1: Relationship between the functionality of neo-Y-linked genes and the extent of DC on neo-X-linked genes in the thorax, testis, abdomen without gonad, and abdomen (from left to right). X_F - Y_F , a group of genes for which both neo-X-linked and neo-Y-linked homologs are functional; X_F - Y_P , a group of genes for which neo-X-linked homologs are functional, but neo-Y-linked homologs are pseudogenized. The numbers of genes analyzed are 461 and 757 for the X_F - Y_F and X_F - Y_P genes, respectively. The lines in the boxes represent medians, 50% of values are included in the boxes, and 80% of values are included within the bars. Differences in median values between groups were tested based on a Monte-Carlo simulation with 1,000 bootstrap replicates: ***, $P < 0.001$. A solid line indicates the R_{Lin} value of 1, indicating perfect DC, whereas a broken line corresponds to a value of 0.5, indicating no DC. + below each bar means that the median R_{Lin} value is significantly greater than 0.5 at the 5% significance level.

Fig. 2: Relationship between the relative time after pseudogenization of neo-Y-linked genes and the extent of dosage compensation (DC) on the neo-X-linked homologs. The R_{Lin} values in a tissue where each gene is expressed most highly among the tissues examined were used. The lines in the boxes represent medians, 50% of values are included in the boxes, and 80% of values are included within the bars. Solid and broken lines correspond to the R_{Lin} value of 1 and 0.5, respectively. Differences in median values between groups were tested based on a Monte-Carlo simulation with 1,000 bootstrap replicates; $P \geq 0.05$ for all pairwise comparisons. The number of genes analyzed for each category is shown in parentheses.

Fig. 3: Expression patterns and pseudogenization of neo-Y-linked genes. (A) Relationship between genes showing maximum expression in various tissues and the proportion of functional

genes. Gene expression based on cFPKM (see *Materials and Methods*) in *D. pseudoobscura* was used as a proxy for the ancestor without neo-sex chromosomes. (B) Relationship between the functionality of neo-sex-linked genes and the proportion of genes with the highest expression in the testis or accessory gland. X_F-Y_F , for genes that are functional on both neo-X and neo-Y; X_F-Y_P , for genes that are functional on the neo-X, but pseudogenized on the neo-Y; X_P-Y_F , for genes that are pseudogenized on the neo-X, but functional on the neo-Y; and X_P-Y_P , for genes that are pseudogenized on both the neo-X and neo-Y. Error bars showing the 95% confidence interval and statistical significance were calculated based on a Monte-Carlo simulation with 1,000 bootstrap replicates: ***, $P < 0.001$; *, $P < 0.05$; NS, $P \geq 0.05$. The number of genes analyzed for each category is shown in parentheses.

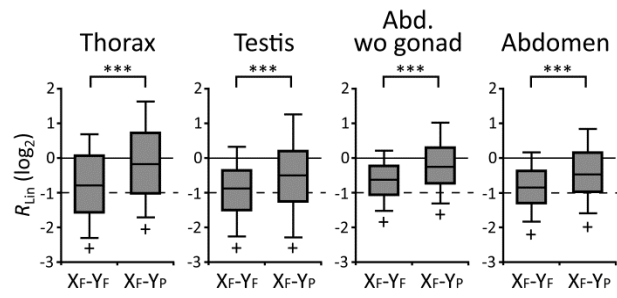


Fig. 1

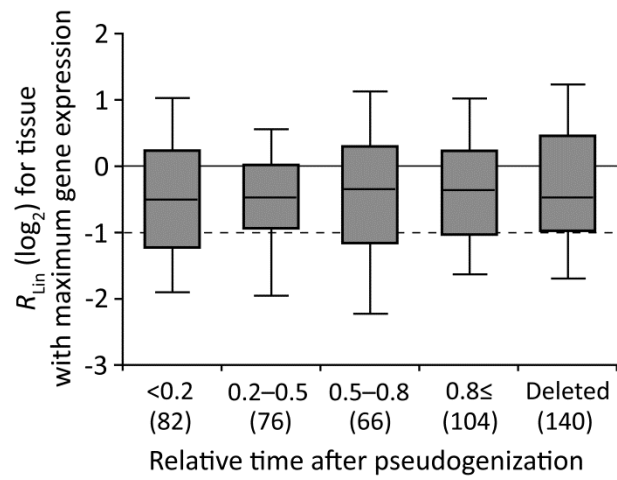


Fig. 2

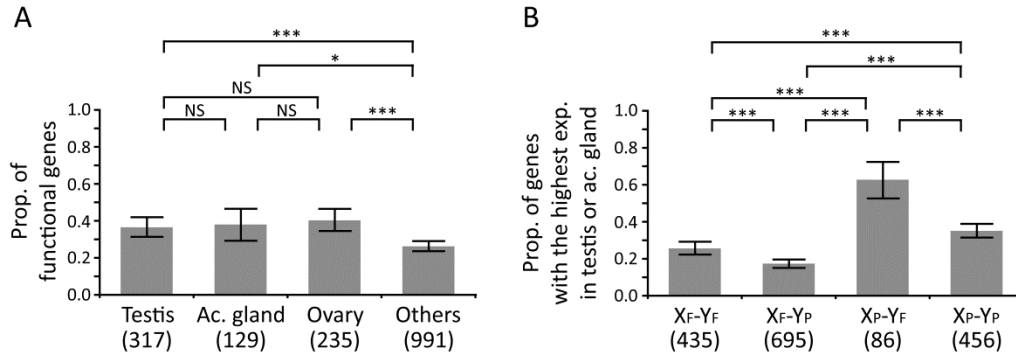


Fig. 3