Supplementary materials

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

Masafumi Nozawa\(^{1,2}\)*, Yohei Minakuchi\(^3\), Kazuhiro Satomura\(^{1,†}\), Shu Kondo\(^{4,‡}\), 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, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan

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

† Present address: Department of Bio-Science, Nagahama Institute of Bioscience and Technology, Nagahama, Shiga 526-0829 Japan

‡ Present address: Department of Biological Science and Technology, Faculty of Advanced Engineering, Tokyo University of Science, Katsushika, Tokyo 125-8585, Japan
SUPPLEMENTARY METHODS

Flies

*D. americana* (strain 15010-0951.03, Millersberg, USA), *D. texana* (strain 15010-1041.00, St. Francisville, USA), and *D. novamexicana* (strain 15010-1031.00, Grand Junction, USA) were obtained from *Drosophila Species Stock Center (https://www.drosophilaspecies.com/). D. albomicans* (strain NG-3, Nago, Japan), *D. nasuta* (strain G-86, Curepipe, Mauritius), and *D. kohkoa* (strain X-145, Kuala Lumpur, Malaysia) were originally collected by Osamu Kitagawa and his colleagues from 1977 to 1979 (Kitagawa et al. 1982; Hatsumi et al. 1988) and have been maintained as living stocks at Tokyo Metropolitan University.

DNA extraction, library construction, and sequencing

To remove the effect of bacterial flora in the gut on the *Drosophila* genome sequences as much as possible, we first treated adult flies under a starvation only with water for three days. All survivors were separated into males and females under a stereomicroscope. High molecular weight genomic DNA was extracted from 1-2 g of the males and females separately by using a standard method with proteinase K and phenol (Sambrook and Russell 2001). The DNA was then further purified with Genomic-tip 100/G (QIAGEN, Venlo, Netherlands).

Whole genome shotgun sequencing was performed using the Illumina HiSeq 2500 (Illumina, San Diego) and PacBio RSII/Sequel (Pacific Biosciences, Menlo Park) platforms. For Illumina sequencing, genomic DNA was sheared by a Covaris S2 ultrasonicator (Covaris, Woburn) and then size-selected on an agarose gel using a MinElute Gel Extraction Kit (QIAGEN). A paired-end library with insert size of ~600 bp from each *Drosophila* species (~400 bp for *D. nasuta*) was constructed using a TruSeq DNA PCR-Free LT Sample Prep Kit (Illumina) according to the manufacturer’s protocol. Each sequencing
library was run on the Illumina HiSeq 2500 platform with rapid mode. Note that for Illumina sequencing we separately sequenced male and female genomes of *D. albomicans* and *D. americana*, whereas only female genomes were sequenced for *D. nasuta*, *D. kohkoa*, *D. texana*, and *D. novamexicana*. For PacBio sequencing, female genomic DNAs of the six *Drosophila* species were sheared with Covaris G-tubes (Covaris). SMRT bell libraries for *D. albomicans*, *D. nasuta*, and *D. americana* were size-selected using a BluePippin (Sage Science, Beverly) with 15-17 kb cut-off. Sequencing run was performed on the PacBio RSII system using P6/C4 chemistry for 6 hours. A 20-kb library for *D. texana* and two 30-kb libraries for *D. kohkoa* and *D. novamexicana* were run on the PacBio Sequel system using Binding Kit1.0/Sequencing Kit1.2.1 for 6 hours and Binding Kit2.0/Sequencing Kit2.0 for 10 hours, respectively. See Table S1 for data summary.

**Genome assembly**

For *D. albomicans*, *D. nasuta*, and *D. americana*, PacBio long reads were assembled *de novo* using the Hierarchical Genome Assembly Process (HGAP) 3 (Chin et al. 2013) in the SMRT Portal (Pacific Biosciences). For *D. kohkoa*, *D. texana*, and *D. novamexicana*, the HGAP4 in SMRT Link (Pacific Biosciences) was used. Illumina sequence reads were then mapped onto the assembled genomes using BWA mem (v0.7.7) with default settings (Li and Durbin 2009). Mismatches and indels between the PacBio assembly and Illumina reads were detected by samtools (0.1.19) and replaced if they were supported by ≥80% of the reads, ≥10 reads, and reads from both directions. All contigs were then screened by the Public database to detect contaminants and mitochondrial sequences. More specifically, using the contigs as queries, we conducted a homology search (BLASTN ver. 2.2.9) against (a) genome sequences of bacteria, archaea, and fungi in the RefSeq database, (b) the mitochondrial genome of the target *Drosophila species*, and (c) the genome sequence of *D. melanogaster* with cut-off a e-value of $10^{-10}$. 

3
The contigs that met the following two criteria were removed as contaminants. (1) Equal to or more than 1% of a contig in length is hit to either (a) or (b). (2) Equal to or more than 1% of a contig in length is not hit to (c). Assembly statistics is given in Table S2.

For each of the genome assemblies of the six species as well as the assembly of D. obscura (Nozawa et al. 2016), we next assigned each contig to a chromosome (a Muller element) as follows. (1) By using all amino-acid sequences (dmel-all-translation-r6.27.fasta from FlyBase) in D. melanogaster as queries, TBLASTN ver. 2.9.0+ (Altschul et al. 1997) was conducted against all contigs in a genome assembly as a database. (2) For each contig region, a best-hit query gene with the lowest e-value was determined. (3) The chromosome number (and the Muller element) on which the gene is located was extracted. 4) If a contig contains 10 or more genes homologous to D. melanogaster and more than two-thirds of the genes are located on a same Muller element in D. melanogaster, we regarded that the contig is a part of the chromosome (Muller element). All other contigs were classified as unassigned (NA).

For D. miranda, we mapped the Illumina sequence reads from the strain 14011-0101.17 (Nozawa et al. 2016) onto the reference genome assembly (D.miranda_PacBio2.0 retrieved from NCBI) by using BWA mem (ver. 0.7.17) (Li and Durbin 2009). Using GATK (ver. 3.7.0) (DePristo et al. 2011), we then detected all variants if the site is homozygous for alternative allele with a minimum coverage of 5× and genotype quality of 20. For autosomes and X, both female and male reads were used to detect variants. For neo-X and neo-Y, only female and male reads were used, respectively. Replacing the nucleotides in the reference genome with the variants detected, we obtained a “pseudo-reference” genome sequence of the strain 14011-0101.17 of D. miranda. For D. pseudoobscura, we used the available genome assembly (dpse-all-chromosome-r3.1.fasta) on FlyBase (https://flybase.org/).

To obtain the neo-Y assemblies of D. americana and D. albomicans, we first filtered out adapter sequences and low quality nucleotides from all reads from Illumina sequencing by using Cutadapt 2.8
and SolexaQA++ v3.1.7.1 (Cox et al. 2010), respectively, and selected read pairs with quality value ≥25 and ≥50-nucleotide long. Those selected read pairs from female and male DNAs were then separately mapped onto the genome assembly obtained above by using BWA mem (ver. 0.7.17) (Li and Durbin 2009). After calling variants by GATK (ver. 3.7.0) (DePristo et al. 2011), if the allele at a nucleotide position on neo-X is heterozygous in males but homozygous for the reference allele in females, the allele that is not detected in females was regarded as the neo-Y allele. When the allele at a position is heterozygous in males but none of them were shared with the allele in females, the allele with the highest frequency was used as the plausible nucleotide on the neo-Y (only 55 and 26 such sites in *D. albomicans* and *D. americana*, respectively). Finally, using an in-house perl script (see SUPPLEMENTARY CODE), neo-Y sequences were obtained by replacing neo-X sequences with those male-specific variants of read depth ≥5 and genotype quality ≥20. See Table S2 for the stats of the genome assemblies in this study.

The genome assemblies of the nine species used in this study were evaluated by BUSCO (benchmarking universal single-copy orthologs) ver. 4.0.2 software (Simao et al. 2015) with the “diptera_odb10” dataset containing 3,285 BUSCOs (Table S3).

**RNA extraction, library construction, and sequencing**

Third instar larvae (before wandering), pupae (48-60 hours after pupation), adults (72-96 hours after eclosion), ovaries (10-14 days old after eclosion), and testes (10-14 days old after eclosion) of *D. albomicans, D. nasuta, D. kohkoa, D. americana, D. texana*, and *D. novamexicana* were used for RNA extraction. Total RNA from each tissue (5 individuals, but ~20 individuals for ovaries and testes) was extracted using PureLink RNA Mini Kit (Thermo Fisher Scientific, Waltham) for males and females separately. mRNA was further purified from the total RNA using the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB, Ipswich). cDNA libraries were then constructed from mRNA using the NEBNext
Ultra Directional RNA Library Prep Kit for Illumina or NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB). Paired-end sequencing of 101 or 151 bp was performed by Macrogen (Seoul, South Korea) with HiSeq 4000 or HiSeq X, respectively. To take fluctuation among samples into account, two independent experiments were conducted for each condition (i.e., two biological replicates were made for each condition). For *D. miranda*, *D. pseudoobscura*, and *D. obscura*, we used the data from Nozawa et al. (2016). See Table S4 for data summary.

**Transcriptome assembly**

To assemble the transcriptome sequence for the nine species, we used female and male RNA-seq data from 3rd instar larvae, pupae, and adults. First, all reads were processed with Cutadapt and SolexaQA++ as mentioned above to remove adapter and low quality sequences. All qualified read pairs from each sample were then mapped onto the genome assembly obtained by using HISAT2-2.1.0 (Kim et al. 2019) and processed with StringTie 1.3.6 (Pertea et al. 2015) to predict the transcriptome for each species.

**Gene annotation**

For expressed genes (i.e., the predicted transcripts), we used TransDecoder-5.5.0 (Haas et al. 2013) to predict coding sequences within each transcript with all amino-acid sequences in *D. melanogaster* (dmel-all-translation-r6.27.fasta retrieved from FlyBase) as guide sequences. We also predicted genes in each genome *ab initio* by using Augustus-3.3.2 (Stanke and Waack 2003) with the option of --species=fly. See Table S5 for the summary of annotated genes. For *D. albomicans* and *D. americana*, if (1) the coding sequences of a homolog on neo-X and neo-Y are completely identical and (2) the male to female coverage ratio of Illumina short reads on the corresponding neo-X positions is within
the bottom 2.5 percentile (i.e., 0.52 for *D. albomicans* and 0.77 for *D. americana*) in the ratio of all nucleotides on autosomes, we regarded that the neo-Y region corresponding to the neo-X homolog is too diverged for neo-Y-derived reads to be mapped, which may have resulted in the identical sequences of neo-X and neo-Y and low male to female coverage ratio. Therefore, we remove these homologs from all analyses.

**Estimation of gene expression level**

All qualified read pairs were mapped onto the genome assembly for each species using STAR 2.7 (Dobin et al. 2013) with the gtf file of the transcriptome sequence obtained above. The mapped data were further processed with RSEM v1.3.1 (Li and Dewey 2011) to obtain FPKM (fragments per kilobase of exon per million mapped reads, Mortazavi et al. 2008) and TPM (transcripts per kilobase million mapped reads, Wagner et al. 2012) for each gene. When we compared the gene expression level among samples from different species, we normalized the FPKM value for each gene by dividing by the median of the FPKM values of a sample (corrected FPKM or cFPKM, Lin et al. 2012) to remove statistical biases as much as possible. When the gene expression level was compared among samples from different species, we also normalized the TPM by dividing by the median of the TPM values (corrected TPM or cTPM). For within-species comparisons, we used TPM as it is. The mapping rate for each sample is shown in Table S21.

**Identification of orthologs and paralogs**

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*), we conducted homology search using BLASTP ver. 2.9.0+ (Altschul et al. 1997). If a gene from each species
was a reciprocal best-hit in all combinations of the species, the genes were regarded as an ortholog. It should be mentioned that when a gene was missing in one species, the remaining genes were still regarded as an ortholog if they were reciprocal best-hits each other.

To identify inparalogs that emerged by gene duplication after splitting from closely related species, we conducted a homology search (BLASTP ver. 2.9.0+) by using all genes in one species (species 1, e.g., D. miranda) as queries, against the database in which all genes in the species (species 1) and its closely related species (species 2, e.g., D. pseudoobscura) were included. For each query, we regarded all species 1 genes that showed lower e-values than the e-value for the species 2 gene (i.e., the ortholog) as inparalogs. Note that for the query genes located on neo-X and neo-Y, we excluded the genes on neo-Y and neo-X, respectively, from the database, because they are not derived from gene duplication but diverged through sex chromosome differentiation. In this way, all genes were classified into 1:1:1 orthologs without paralogs (sometimes missing in one species), other orthologs containing inparalogs, and species-specific genes.

**Gene classification**

To check the integrity of the coding sequence for each gene, we conducted a following procedure.

(1) If either initiation or stop codon was missing in a gene due to the presence of at least one ambiguous site (i.e., N) in the flanking 100 nucleotides or the location of the gene at the end of contigs/scaffolds of the assembly, the gene was regarded to be ‘unclassified’. (2) If either initiation or stop codon was missing in spite of no ambiguous site in the flanking 100 nucleotides nor the location at the end of contigs/scaffolds, the gene was regarded to be ‘disrupted’. For remaining genes in which both initiation and stop codons are present: (3) If a gene had a homolog in D. melanogaster with a cut-off e-value of 10^{-20} by BLASTP search, the nucleotide sequence of the gene was aligned with that of the D. melanogaster
homolog by using MACSE v2.03 (Ranwez et al. 2018) that accounts for frameshifts and internal stop codons. If either frameshifts or internal stop codons were detected in the nucleotide sequence of the gene, the gene was regarded as being ‘disrupted’. However, if frameshifts or internal stop codons were detected in the sequence of the *D. melanogaster* gene, the gene in the target species was regarded as being ‘unclassified’. (4) If a gene did not have a homolog in *D. melanogaster* but a homolog is present in the closely related species (e.g., *D. pseudoobscura* or *D. obscura* for *D. miranda*), we aligned these sequences by MACSE as mentioned above with some manual adjustment. If either frameshifts or internal stop codons were detected in the nucleotide sequence of the gene, the gene was regarded as being ‘disrupted’. (5) If a gene did not have any homolog in closely related species, the coding sequence of the gene was translated into amino-acid sequence. If an internal stop codon was detected, the gene was regarded as being ‘disrupted’. (6) All remaining genes were regarded as being ‘intact’.

To check the expression potential for each gene, we used the following criteria. (7) If a gene (or a transcript of a gene, strictly speaking) was identified by the HISAT-StringTie pipeline, the gene was regarded as being ‘expressed’. (8) For the genes that were not detected by the HISAT-StringTie pipeline but predicted by Augustus, if the gene showed the FPKM value of $\geq 1$ in at least one tissue examined, the gene was regarded to be ‘expressed’. (9) All other genes were regarded as being ‘non-expressed’.

Finally, if a gene is ‘intact’ and ‘expressed’, the gene was regarded as ‘functional’. If a gene is ‘disrupted’ and ‘expressed’, the gene was regarded to be ‘disrupted’. If a gene is ‘intact’ and ‘non-expressed’, the gene was regarded as being ‘silenced’. If a gene is ‘disrupted’ and ‘non-expressed’, the gene was regarded to be ‘silenced and disrupted’. All remaining genes were regarded as ‘unclassified’.

**Estimation of the extent of dosage compensation**
For this analysis, we used only the orthologs that have no inparalogs in any of each trio (i.e., D. miranda-D. pseudoobscura-D. obscura, D. albomicans-D. nasuta-D. kohkoa, or D. americana-D. texana-D. novamexicana) and are located on the same Muller elements in all three species from each trio. In addition, for autosomal and X-linked genes we only used the orthologs that are functional in the three species. For the Muller element that became neo-sex chromosomes, we also analyzed the orthologs that were pseudogenized or deleted on neo-Y.

To examine the level of DC for each gene, we used $R_{\text{Lin}}$ (Lin et al. 2012). The equation of $R_{\text{Lin}}$ can be written as

$$R_{\text{Lin}} = \frac{\text{cFPKM}_{\text{neo-X, tar}}}{\text{median cFPKM}_{A/X, \text{tar}}}/\frac{\text{cFPKM}_{\text{proto-X, com}}}{\text{median cFPKM}_{A/X, \text{com}}}.$$ 

where cFPKM is the corrected number of fragments 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, neo-X, proto-A, and A/X denote target species (e.g., D. miranda), comparing (or different) species (e.g., D. pseudoobscura), the neo-X Chromosome, an autosome orthologous to the neo-X in comparing species (e.g., the Chromosome 3 in D. pseudoobscura), and autosomes/canonical X, respectively. The essence of this equation is to detect changes in male expression between species due to the emergence of sex chromosomes using expression change of autosomal and X-linked genes as a control. If there is no DC, the $R_{\text{Lin}}$ value will be 0.5 because males of target species have only one copy of the neo-X whereas males in comparing species carry two copies. In contrast, the value is expected to be 1 if DC is perfect. We used the cFPKM and the cTPM values as indices of gene expression and check the consistency between the indices.

Note that in the original $R_{\text{Lin}}$, the ratio of expression of a target gene on neo-X in a species (e.g., neo-X in D. miranda) to its ortholog on the orthologous autosome in its close relative (e.g., Chromosome
3 in *D. pseudoobscura* is normalized by the median ratio of expression of all genes on autosomes (see Nozawa et al. 2018 for an equation). However, we used not only autosomal genes but also X-linked genes in both species for normalization, because genes that are located on autosomes or X in both species should theoretically show the same expression level. It should be mentioned that in this study, we used the cFPKM value as an index of gene expression following the original method (Lin et al. 2012) to consider the statistical bias of using the FPKM as much as possible (Wagner et al. 2012). We also used cTPM to check the consistency of our results.

Neo-X-linked and neo-Y-linked genes were classified into four categories depending on the functionality. If a neo-X-linked and a neo-Y-linked homolog are both functional, the homolog was classified into a X<sub>F</sub>–Y<sub>F</sub> gene (i.e., a subscript F indicates functional genes). If a neo-X-linked gene is functional but its neo-Y-linked homolog is a pseudogene, the homolog was grouped into a X<sub>F</sub>–Y<sub>P</sub> gene (i.e., a subscript P indicates pseudogenes). Similarly, homologs can be classified into X<sub>F</sub>–Y<sub>F</sub> and X<sub>F</sub>–Y<sub>P</sub> genes. To test gene-by-gene DC, the *R*<sub>Lin</sub> values for X<sub>F</sub>–Y<sub>F</sub> and X<sub>F</sub>–Y<sub>P</sub> genes were statistically compared by Mann-Whitney *U* test with the correction of multiple testing by the Benjamini and Hochberg method (Benjamini and Hochberg 1995).

**Estimation of pseudogenization events**

We estimated the number of pseudogenization events on each lineage based on the parsimony principle using the gene classification (functional, silenced, disrupted, silenced and disrupted, and unclassified) as mentioned above. For this analysis, we used only the orthologs that are located on the same Muller elements during the evolution of the three species considered, do not have any inparalogs, and functional in the outgroup species (e.g., *D. obscura*, *D. kohkoa*, and *D. novamexicana* for *D. miranda*, *D. albomicans*, and *D. americana*, respectively). For example, if a gene is functional, silenced, disrupted,
and functional on the *D. miranda* neo-X, the *D. miranda* neo-Y, the *D. pseudoobscura* Chromosome 3, and the *D. obscura* Muller element C, respectively, a pseudogenization event for the ortholog is inferred both on the neo-Y and the *D. pseudoobscura* lineages.

The pseudogenization rates between lineages were compared by $\chi^2$ test. The expected numbers of pseudogenization events on the two lineages were computed by multiplying the total number of pseudogenization events on the two lineages by the ratio of the evolutionary time of one lineage to the sum of the evolutionary times of the two lineages. For example, the number of pseudogenization events on the ancestral lineage of the neo-sex chromosomes in *D. miranda* is 31, whereas that on the neo-X lineage is 123. Since the evolutionary time for each lineage is 0.9 and 1.1 Myrs, respectively, the expected number of pseudogenization events on the former and the latter lineages under the equal rate of pseudogenization are $(123+31) \times 0.9/2.0 = 69.3$ and $(123+31) \times 1.1/2.0 = 84.7$, respectively.

**Estimation of evolutionary distance on each branch**

To clarify the functional constraints on neo-X, neo-Y, and its ancestral lineages, we computed the nonsynonymous and synonymous nucleotide divergence per site for each branch of the tree from the three species considered. For this analysis, we used only the orthologs that are located on the same Muller elements during the evolution of the three species considered, do not have any inparalogs, and functional in the outgroup species as mentioned above. Coding sequences for each ortholog were translated into amino-acid sequences and aligned by Muscle v3.8.31 (Edgar 2004). The aligned amino-acid sequences were reversed to nucleotide sequences with keeping alignment gaps by using an in-house perl script (see *SUPPLEMENTARY CODE*). Using another in-house perl script (see *SUPPLEMENTARY CODE*), synonymous and nonsynonymous nucleotide divergence per site between all pairs of sequences were computed based on the modified Nei-Gojobori method (Zhang et al. 1998) with complete deletion option.
and Ti/Tv (transition-transversion ratio) = 2. Finally, synonymous and nonsynonymous nucleotide distances on each branch were computed by the least-squares method (Rzhetsky and Nei 1993).

**Estimation of parallel pseudogenization on neo-sex chromosomes**

In *D. miranda* and *D. albomicans*, the Muller element C independently became neo-sex chromosomes. The number of parallel pseudogenization in these two lineages was therefore examined. First, all orthologs that are located on the same Muller elements during the evolution of the three species considered, do not have any inparalogs, and functional in the outgroup species (*D. obscura* for *D. miranda* and *D. kohkoa* for *D. albomicans*) were collected. To identify the orthologs among the six species (i.e., *D. miranda*, *D. pseudoobscura*, *D. obscura*, *D. albomicans*, *D. nasuta*, and *D. kohkoa*), reciprocal homology search based on BLASTP ver. 2.9.0+ was conducted between the outgroup species of each trio (i.e., *D. obscura* and *D. kohkoa*) with a cut-off e-value of $10^{-10}$. Then, we identified orthologs that show reciprocal best-hit relationship between the species, located on the same Muller elements, and do not have any inparalogs. Finally, the information of orthologous genes for each trio species was added to the orthologs identified in *D. obscura* and *D. kohkoa*. If the pseudogenization for an ortholog was inferred on the same lineages in the two trios (e.g., the neo-Y lineages in *D. miranda* and *D. albomicans*), the ortholog was regarded to have experienced parallel pseudogenization.

To evaluate whether the observed number of parallel pseudogenization events has some biological meaning, we conducted a binomial test. As an example, let us think about Muller element A in *D. miranda* and *D. albomicans* (Fig. 6SA). There were 29 and 38 pseudogenizations in the *D. miranda* and the *D. albomicans* lineages, respectively. Since the number of orthologs analyzed was 678, the expected probability that the same gene is independently pseudogenized in both lineages is $\frac{29}{678} \times \frac{38}{678} = 0.0024$. With this expected probability, the cumulative probability in which the number of
shared pseudogenizations among 678 orthologs exceeds the observed number of shared pseudogenizations (i.e., five in this example) was computed by a binomial test. In this example, the cumulative probability (i.e., $P$-value) was 0.025. However, correction of multiple testing finally gave the $Q$-value of 0.074, i.e., an insignificant value.

**Gene Ontology analysis**

To examine Gene Ontology that are enriched in the functional and nonfunctional genes on neo-sex chromosomes, using the neo-sex-linked genes as queries, we conducted a homology search against the amino-acid sequences in *D. melanogaster* [dmel-all-translation-r6.34.fasta from FlyBase (https://flybase.org/)] based on BLASTP ver. 2.9.0+ with a cut-off e-value of $10^{-10}$. The best-hit gene in *D. melanogaster* was then used for the Gene Ontology analysis by using the GOrilla software (Eden et al. 2009).
Table S1. Summary of the DNA sequencing data of the *Drosophila* species used in this study.

| Species         | Sex   | Sample name | Platform  | Layout/length (bp) | Accession     | Reference          |
|-----------------|-------|-------------|-----------|--------------------|---------------|--------------------|
| *D. miranda*    | Female| mirF        | HiSeq 2000| Paired/101         | DRR055276     | Nozawa et al. (2016) |
|                 | Male  | mirM        | HiSeq 2000| Paired/101         | DRR055277     | Nozawa et al. (2016) |
| *D. albomicans* | Female| albF        | PacBio RS II| Single/-          | DRR075970-99  | This study         |
|                 | Male  | albM        | HiSeq 2500| Paired/500         | DRR076000     | This study         |
| *D. nasuta*     | Female| nasF        | PacBio RS II| Single/-          | DRR125226     | This study         |
|                 |       |             | HiSeq 2500| Paired/400         | DRR061014     | This study         |
| *D. kohkoa*     | Female| kohF        | PacBio Sequel| Single/-       | DRR160727-33  | This study         |
|                 |       |             | HiSeq 2500| Paired/500         | DRR160734     | This study         |
| *D. americana*  | Female| ameF        | PacBio RS II| Single/-          | DRR076002-26  | This study         |
|                 | Male  | ameM        | HiSeq 2500| Paired/500         | DRR076027     | This study         |
| *D. texana*     | Female| texF        | PacBio Sequel| Single/-       | DRR160721-4  | This study         |
|                 |       |             | HiSeq 2500| Paired/500         | DRR160725     | This study         |
| *D. novamexicana*| Female| novF        | PacBio Sequel| Single/-       | DRR160735-9  | This study         |
|                 |       |             | HiSeq 2500| Paired/500         | DRR160740     | This study         |
Table S2. Assembly size, number of contigs, and contig N50 of the *Drosophila* genomes used in this study.

| Species            | Chr. included | Assembly size (bp) | No. of contigs | Contig N50 (bp) | Accession                     | Reference                  |
|--------------------|---------------|--------------------|----------------|-----------------|-------------------------------|---------------------------|
| *D. miranda*       | w/o Y/neo-Y   | 176,610,581        | 40             | 32,547,436<sup>2</sup> | -                             | Mahajan et al. (2018)     |
|                    | with Y/neo-Y  | 287,124,465        | 102            | 35,279,367      | QNQS01000001-010000102        | Mahajan et al. (2018)     |
| *D. pseudoobscura* | w/o Y         | 152,696,384        | 4,790          | 12,541,198      | AADE01000001-01017507         | Richards et al. (2005)    |
| *D. obscura*       | w/o Y         | 181,868,570        | 1,935          | 472,512<sup>2</sup> | BDQP01000001-01001935        | Nozawa et al. (2016)      |
| *D. albomicans*    | w/o Y/neo-Y   | 182,697,126        | 834            | 21,972,321      | BJEI01000001-01000834        | This study                |
|                    | with neo-Y    | 238,238,472        | 836            | 25,367,705      | -                             | This study                |
| *D. nasuta*        | w/o Y         | 173,158,203        | 604            | 17,893,789      | BJEH01000001-010000604       | This study                |
| *D. kohkoa*        | w/o Y         | 170,668,061        | 87             | 3,597,765       | BJEL01000001-01000087        | This study                |
| *D. americana*     | w/o Y/neo-Y   | 177,534,117        | 435            | 1,565,372       | BJEJ01000001-01000435        | This study                |
|                    | with neo-Y    | 207,397,045        | 453            | 1,605,128       | -                             | This study                |
| *D. texana*        | w/o Y         | 165,405,863        | 199            | 2,693,102       | BJEK01000001-01000199        | This study                |
| *D. novamexicana*  | w/o Y         | 173,633,261        | 270            | 2,873,990       | BJEM01000001-01000270        | This study                |

<sup>1</sup> Data are based on the resequence in this study.

<sup>2</sup> Based on scaffolds but not contigs.
Table S3. Assessment of genome assembly of the nine *Drosophila* species used in this study based on the existence of BUSCOs.

| Assembly                                      | Number (proportion) of genes |
|-----------------------------------------------|------------------------------|
|                                               | Complete SC<sup>1</sup> | Complete D<sup>2</sup> | Fragmented | Missing | Total     |
| *D. miranda*                                  | 3,180 (96.8%)             | 78 (2.4%)               | 6 (0.2%)   | 21 (0.6%) | 3,285 (100%) |
| *D. miranda* with Y/neo-Y                     | 2,765 (84.2%)             | 493 (15.0%)             | 6 (0.2%)   | 21 (0.6%) | 3,285 (100%) |
| *D. pseudoobscura*                            | 3,167 (96.4%)             | 54 (1.6%)               | 17 (0.5%)  | 47 (1.4%) | 3,285 (100%) |
| *D. obscura*                                  | 3,100 (94.4%)             | 153 (4.7%)              | 11 (0.3%)  | 21 (0.6%) | 3,285 (100%) |
| *D. albomicans*                               | 3,202 (97.5%)             | 56 (1.7%)               | 5 (0.2%)   | 22 (0.7%) | 3,285 (100%) |
| *D. albomicans* with neo-Y                    | 1,934 (58.9%)             | 1,324 (40.3%)           | 5 (0.2%)   | 22 (0.7%) | 3,285 (100%) |
| *D. nasuta*                                   | 3,230 (98.3%)             | 29 (0.9%)               | 6 (0.2%)   | 20 (0.6%) | 3,285 (100%) |
| *D. kohkoa*                                   | 3,042 (92.6%)             | 210 (6.4%)              | 8 (0.2%)   | 25 (0.8%) | 3,285 (100%) |
| *D. americana*                                | 3,116 (94.9%)             | 103 (3.1%)              | 21 (0.6%)  | 45 (1.4%) | 3,285 (100%) |
| *D. americana* with neo-Y                     | 2,544 (77.4%)             | 678 (20.6%)             | 19 (0.6%)  | 44 (1.3%) | 3,285 (100%) |
| *D. texana*                                   | 3,202 (97.5%)             | 51 (1.6%)               | 7 (0.2%)   | 25 (0.8%) | 3,285 (100%) |
| *D. novamexicana*                             | 3,156 (96.1%)             | 99 (3.0%)               | 7 (0.2%)   | 23 (0.7%) | 3,285 (100%) |

BUSCO ver. 4.0.4 software (Seppey et al. 2019) was used with the “diptera_odb10” dataset.

1 Single copy genes.

2 Duplicated genes. It should be mentioned that when contigs/scaffolds from neo-Y is included in the analysis, the proportion of complete duplicated ORFs considerably increases, because the homologs on neo-X and neo-Y were regarded as duplicated genes.
Table S4. Summary of the RNA sequencing data of the *Drosophila* species used in this study.

| Species       | Sex     | Tissue               | Sample name | Platform | Layout/length (nt) | Accession¹ | Reference         |
|---------------|---------|----------------------|-------------|----------|-------------------|------------|-------------------|
| *D. miranda*  | Female  | Larval whole body    | mirLF       | HiSeq 200| Paired/101        | DRR055176, DRR055177 | Nozawa et al. (2016) |
|               | Male    | mirLM                |             | HiSeq 200| Paired/101        | DRR055178, DRR055179 | Nozawa et al. (2016) |
|               | Female  | Pupal whole body     | mirPF       | HiSeq 200| Paired/101        | DRR055186, DRR055187 | Nozawa et al. (2016) |
|               | Male    | mirPM                |             | HiSeq 200| Paired/101        | DRR055188, DRR055189 | Nozawa et al. (2016) |
|               | Female  | Adult whole body     | mirWF       | HiSeq 200| Paired/101        | DRR055196, DRR055197 | Nozawa et al. (2016) |
|               | Male    | mirWM                |             | HiSeq 200| Paired/101        | DRR055198, DRR055199 | Nozawa et al. (2016) |
|               | Female  | Ovary                | mirO        | HiSeq 200| Paired/101        | DRR055184, DRR055185 | Nozawa et al. (2016) |
|               | Male    | Testis               | mirTe       | HiSeq 200| Paired/101        | DRR055194, DRR055195 | Nozawa et al. (2016) |
| *D. pseudoobscura* | Female | Larval whole body    | pseLF       | HiSeq 200| Paired/100        | DRR055252, DRR055253 | Nozawa et al. (2016) |
|               | Male    | pseLM                |             | HiSeq 200| Paired/100        | DRR055254, DRR055255 | Nozawa et al. (2016) |
|               | Female  | Pupal whole body     | psePF       | HiSeq 200| Paired/100        | DRR055262, DRR055263 | Nozawa et al. (2016) |
|               | Male    | psePM                |             | HiSeq 200| Paired/100        | DRR055264, DRR055265 | Nozawa et al. (2016) |
|               | Female  | Adult whole body     | pseWF       | HiSeq 200| Paired/101        | DRR055272, DRR055273 | Nozawa et al. (2016) |
|               | Male    | pseWM                |             | HiSeq 200| Paired/101        | DRR055274, DRR055275 | Nozawa et al. (2016) |
|               | Female  | Ovary                | pseO        | HiSeq 200| Paired/101        | DRR055260, DRR055261 | Nozawa et al. (2016) |
|               | Male    | Testis               | pseTe       | HiSeq 200| Paired/101        | DRR055270, DRR055271 | Nozawa et al. (2016) |
| *D. obscura*  | Female  | Larval whole body    | obsLF       | HiSeq 200| Paired/101        | DRR055214, DRR055215 | Nozawa et al. (2016) |
|               | Male    | obsLM                |             | HiSeq 200| Paired/101        | DRR055216, DRR055217 | Nozawa et al. (2016) |
|               | Female  | Pupal whole body     | obsPF       | HiSeq 200| Paired/101        | DRR055218, DRR055219 | Nozawa et al. (2016) |
|               | Male    | obsPM                |             | HiSeq 200| Paired/101        | DRR055220, DRR055221 | Nozawa et al. (2016) |
|               | Female  | Adult whole body     | obsWF       | HiSeq 200| Paired/101        | DRR055230, DRR055231 | Nozawa et al. (2016) |
|               | Male    | obsWM                |             | HiSeq 200| Paired/101        | DRR055232, DRR055233 | Nozawa et al. (2016) |
|               | Female  | Ovary                | obsO        | HiSeq 200| Paired/101        | DRR055218, DRR055219 | Nozawa et al. (2016) |
|               | Male    | Testis               | obsTe       | HiSeq 200| Paired/101        | DRR055228, DRR055229 | Nozawa et al. (2016) |
| Species         | Sex                  | Tissue            | Sample name | Platform | Layout/length (nt) | Accession¹ | Reference |
|-----------------|----------------------|-------------------|-------------|----------|-------------------|------------|-----------|
| *D. albomicans* | Female               | Larval whole body | albLF       | HiSeq 4000 | Paired/101        | DRR160743 | This study |
|                 | Male                 |                   | albLM       | HiSeq 4000 | Paired/101        | DRR160744 | This study |
|                 | Female               | Pupal whole body  | albPF       | HiSeq 4000 | Paired/101        | DRR160745 | This study |
|                 | Male                 |                   | albPM       | HiSeq 4000 | Paired/101        | DRR160746 | This study |
|                 | Female               | Adult whole body  | albWF       | HiSeq 4000 | Paired/101        | DRR160741 | This study |
|                 | Male                 |                   | albWM       | HiSeq 4000 | Paired/101        | DRR160742 | This study |
|                 | Female               | Ovary             | albO        | HiSeq 4000 | Paired/101        | DRR160747 | This study |
|                 | Male                 | Testis            | albTe       | HiSeq 4000 | Paired/101        | DRR160748 | This study |
| *D. nasuta*     | Female               | Larval whole body | nasLF       | HiSeq 4000 | Paired/101        | DRR160751 | This study |
|                 | Male                 |                   | nasLM       | HiSeq 4000 | Paired/101        | DRR160752 | This study |
|                 | Female               | Pupal whole body  | nasPF       | HiSeq 4000 | Paired/101        | DRR160753 | This study |
|                 | Male                 |                   | nasPM       | HiSeq 4000 | Paired/101        | DRR160754 | This study |
|                 | Female               | Adult whole body  | nasWF       | HiSeq 4000 | Paired/101        | DRR160749 | This study |
|                 | Male                 |                   | nasWM       | HiSeq 4000 | Paired/101        | DRR160750 | This study |
|                 | Female               | Ovary             | nasO        | HiSeq 4000 | Paired/101        | DRR160755 | This study |
|                 | Male                 | Testis            | nasTe       | HiSeq 4000 | Paired/101        | DRR160756 | This study |
| *D. kohkoa*     | Female               | Larval whole body | kohLF       | HiSeq 4000 | Paired/101        | DRR160759 | This study |
|                 | Male                 |                   | kohLM       | HiSeq 4000 | Paired/101        | DRR160760 | This study |
|                 | Female               | Pupal whole body  | kohPF       | HiSeq 4000 | Paired/101        | DRR160761 | This study |
|                 | Male                 |                   | kohPM       | HiSeq 4000 | Paired/101        | DRR160762 | This study |
|                 | Female               | Adult whole body  | kohWF       | HiSeq 4000 | Paired/101        | DRR160757 | This study |
|                 | Male                 |                   | kohWM       | HiSeq 4000 | Paired/101        | DRR160758 | This study |
|                 | Female               | Ovary             | kohO        | HiSeq X    | Paired/151        | DRR168781 | This study |
|                 | Male                 | Testis            | kohTe       | HiSeq X    | Paired/151        | DRR168782 | This study |
Table S4. Cont’d.

| Species          | Sex     | Tissue            | Sample name | Platform      | Layout/length (nt) | Accession\(^1\) | Reference |
|------------------|---------|-------------------|-------------|---------------|-------------------|----------------|-----------|
| *D. americana*   | Female  | Larval whole body | ameLF       | HiSeq 4000   | Paired/101        | DRR160779      | This study |
|                  | Male    |                   | ameLM       | HiSeq 4000   | Paired/101        | DRR160780      | This study |
|                  | Female  | Pupal whole body  | amePF       | HiSeq 4000   | Paired/101        | DRR160781      | This study |
|                  | Male    |                   | amePM       | HiSeq 4000   | Paired/101        | DRR160782      | This study |
|                  | Female  | Adult whole body  | ameWF       | HiSeq 4000   | Paired/101        | DRR160777      | This study |
|                  | Male    |                   | ameWM       | HiSeq 4000   | Paired/101        | DRR160778      | This study |
|                  | Female  | Ovary             | ameO        | HiSeq 4000   | Paired/101        | DRR160783      | This study |
|                  | Male    | Testis            | ameTe       | HiSeq 4000   | Paired/101        | DRR160784      | This study |
| *D. texana*      | Female  | Larval whole body | texLF       | HiSeq 4000   | Paired/101        | DRR160771      | This study |
|                  | Male    |                   | texLM       | HiSeq 4000   | Paired/101        | DRR160772      | This study |
|                  | Female  | Pupal whole body  | texPF       | HiSeq 4000   | Paired/101        | DRR160773      | This study |
|                  | Male    |                   | texPM       | HiSeq 4000   | Paired/101        | DRR160774      | This study |
|                  | Female  | Adult whole body  | texWF       | HiSeq 4000   | Paired/101        | DRR160769      | This study |
|                  | Male    |                   | texWM       | HiSeq 4000   | Paired/101        | DRR160770      | This study |
|                  | Female  | Ovary             | texO        | HiSeq 4000   | Paired/101        | DRR160775      | This study |
|                  | Male    | Testis            | texTe       | HiSeq 4000   | Paired/101        | DRR160776      | This study |
| *D. novamexicana*| Female  | Larval whole body | novLF       | HiSeq 4000   | Paired/101        | DRR160765      | This study |
|                  | Male    |                   | novLM       | HiSeq 4000   | Paired/101        | DRR160766      | This study |
|                  | Female  | Pupal whole body  | novPF       | HiSeq 4000   | Paired/101        | DRR160767      | This study |
|                  | Male    |                   | novPM       | HiSeq 4000   | Paired/101        | DRR160768      | This study |
|                  | Female  | Adult whole body  | novWF       | HiSeq 4000   | Paired/101        | DRR160763      | This study |
|                  | Male    |                   | novWM       | HiSeq 4000   | Paired/101        | DRR160764      | This study |
|                  | Female  | Ovary             | novO        | HiSeq X      | Paired/151        | DRR168783      | This study |
|                  | Male    | Testis            | novTe       | HiSeq X      | Paired/151        | DRR168784      | This study |
Each accession number newly deposited in this study contains run data for two biological replicates.
Table S5. Numbers of protein-coding genes and their transcripts of the *Drosophila* species used in this study.

| Species          | No. of functional protein-coding genes | No. of transcripts from functional protein-coding genes | No. of non-functional protein-coding genes  
|------------------|----------------------------------------|--------------------------------------------------------|---------------------------------------------|
| *D. miranda*     | 11,299 (13,161)¹                      | 24,541 (28,773)                                        | 4,698 (12,167)                             |
| *D. pseudoobscura* | 10,800                                 | 19,929                                                 | 3,733                                      |
| *D. obscura*     | 11,330                                 | 21,332                                                 | 4,809                                      |
| *D. Albomicans*  | 10,101 (13,834)                        | 19,880 (26,172)                                        | 3,659 (4,685)                             |
| *D. nasuta*      | 9,868                                  | 19,196                                                 | 3,280                                      |
| *D. kohkoa*      | 10,183                                 | 19,865                                                 | 3,176                                      |
| *D. americana*   | 9,615 (11,403)                         | 20,064 (22,916)                                        | 3,364 (3,860)                             |
| *D. texana*      | 9,910                                  | 19,770                                                 | 2,692                                      |
| *D. novamexicana*| 9,722                                  | 20,400                                                 | 2,940                                      |

¹ Numbers in parentheses include the counts of Y/neo-Y-linked genes/transcripts.

² Unclassified genes are included in the counts.
Table S8. Enriched GO terms in functional genes on the neo-X in *Drosophila miranda*.

| GO term          | Description                        | $P$-value$^1$ | FDR $Q$-value | Enrichment | Category   |
|------------------|------------------------------------|---------------|---------------|------------|------------|
| GO:0044237       | cellular metabolic process          | 4.12E-05      | 1.86E-01      | 1.05       | Process    |
| GO:0009987       | cellular process                    | 3.74E-04      | 8.42E-01      | 1.03       | Process    |
| GO:0008152       | metabolic process                   | 1.00E-03      | 1.00E+00      | 1.03       | Process    |
| GO:0003824       | catalytic activity                  | 2.12E-04      | 2.73E-01      | 1.05       | Function   |
| GO:0044424       | intracellular part                  | 1.95E-06      | 1.61E-03      | 1.04       | Component  |
| GO:0044422       | organelle part                      | 1.26E-05      | 5.20E-03      | 1.06       | Component  |
| GO:0044446       | intracellular organelle part        | 2.62E-05      | 7.24E-03      | 1.06       | Component  |
| GO:0044444       | cytoplasmic part                    | 2.85E-05      | 5.90E-03      | 1.06       | Component  |
| GO:0032991       | protein-containing complex          | 1.06E-04      | 1.76E-02      | 1.06       | Component  |
| GO:0005739       | mitochondrion                       | 5.93E-04      | 8.18E-02      | 1.11       | Component  |

$^1$ Only GO terms with $P$-value $\leq$ 0.001 are listed.
Table S9. Enriched GO terms in functional genes on the neo-X in *Drosophila albomicans*.

| GO term               | Description                                               | P-value | FDR Q-value | Enrichment | Category       |
|-----------------------|------------------------------------------------------------|---------|-------------|-------------|----------------|
| GO:0008152            | metabolic process                                          | 2.21E-10| 1.23E-06    | 1.05        | Process        |
| GO:0071704            | organic substance metabolic process                        | 1.22E-08| 3.39E-05    | 1.05        | Process        |
| GO:0044237            | cellular metabolic process                                 | 1.85E-08| 3.42E-05    | 1.05        | Process        |
| GO:0044238            | primary metabolic process                                  | 9.46E-08| 1.32E-04    | 1.05        | Process        |
| GO:0006807            | nitrogen compound metabolic process                        | 2.05E-06| 2.28E-03    | 1.05        | Process        |
| GO:0043170            | macromolecule metabolic process                            | 1.41E-05| 1.31E-02    | 1.05        | Process        |
| GO:1901564            | organonitrogen compound metabolic process                  | 3.51E-05| 2.79E-02    | 1.05        | Process        |
| GO:0009987            | cellular process                                           | 6.97E-05| 4.84E-02    | 1.02        | Process        |
| GO:0044281            | small molecule metabolic process                           | 8.70E-05| 5.37E-02    | 1.09        | Process        |
| GO:0009058            | biosynthetic process                                       | 1.14E-04| 6.36E-02    | 1.08        | Process        |
| GO:0006629            | lipid metabolic process                                    | 1.43E-04| 7.20E-02    | 1.11        | Process        |
| GO:0043603            | cellular amide metabolic process                           | 1.56E-04| 7.21E-02    | 1.13        | Process        |
| GO:1901576            | organic substance biosynthetic process                    | 1.89E-04| 8.07E-02    | 1.07        | Process        |
| GO:0032787            | monocarboxylic acid metabolic process                     | 2.49E-04| 9.89E-02    | 1.16        | Process        |
| GO:0034660            | ncRNA metabolic process                                    | 2.70E-04| 9.99E-02    | 1.12        | Process        |
| GO:0043436            | oxoacid metabolic process                                 | 2.79E-04| 9.70E-02    | 1.11        | Process        |
| GO:0006082            | organic acid metabolic process                             | 2.79E-04| 9.13E-02    | 1.11        | Process        |
| GO:0044255            | cellular lipid metabolic process                           | 2.98E-04| 9.19E-02    | 1.11        | Process        |
| GO:0044249            | cellular biosynthetic process                             | 3.28E-04| 9.60E-02    | 1.07        | Process        |
| GO:0043604            | amide biosynthetic process                                | 3.65E-04| 1.01E-01    | 1.14        | Process        |
Table S9. Cont’d.

| GO term             | Description                                           | \(P\)-value\(^1\) | FDR \(Q\)-value | Enrichment | Category       |
|---------------------|-------------------------------------------------------|--------------------|-----------------|------------|----------------|
| GO:0019752          | carboxylic acid metabolic process                     | 5.40E-04           | 1.43E-01        | 1.11       | Process        |
| GO:0034470          | ncRNA processing                                      | 7.14E-04           | 1.80E-01        | 1.13       | Process        |
| GO:0009056          | catabolic process                                     | 7.22E-04           | 1.74E-01        | 1.07       | Process        |
| GO:0006399          | tRNA metabolic process                                | 7.22E-04           | 1.67E-01        | 1.16       | Process        |
| GO:0034641          | cellular nitrogen compound metabolic process          | 7.42E-04           | 1.65E-01        | 1.05       | Process        |
| GO:1901575          | organic substance catabolic process                   | 8.85E-04           | 1.89E-01        | 1.07       | Process        |
| GO:0034645          | cellular macromolecule biosynthetic process           | 9.07E-04           | 1.87E-01        | 1.10       | Process        |
| GO:0003824          | catalytic activity                                    | 3.60E-09           | 6.49E-06        | 1.06       | Function       |
| GO:0017111          | nucleoside-triphosphatase activity                    | 4.69E-06           | 4.23E-03        | 1.13       | Function       |
| GO:0016817          | hydrolase activity, acting on acid anhydrides         | 1.09E-05           | 6.55E-03        | 1.13       | Function       |
| GO:0016818          | hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides | 1.09E-05 | 4.91E-03 | 1.13 | Function |
| GO:0016462          | pyrophosphatase activity                              | 1.24E-05           | 4.46E-03        | 1.13       | Function       |
| GO:0016787          | hydrolase activity                                    | 1.97E-05           | 5.93E-03        | 1.07       | Function       |
| GO:0043168          | anion binding                                         | 1.30E-04           | 3.36E-02        | 1.07       | Function       |
| GO:0036094          | small molecule binding                                | 4.67E-04           | 1.05E-01        | 1.07       | Function       |
| GO:0043167          | ion binding                                           | 5.59E-04           | 1.12E-01        | 1.05       | Function       |
| GO:0003735          | structural constituent of ribosome                    | 7.22E-04           | 1.30E-01        | 1.16       | Function       |
| GO:1901265          | nucleoside phosphate binding                          | 8.54E-04           | 1.40E-01        | 1.07       | Function       |
| GO:0000166          | nucleotide binding                                   | 8.54E-04           | 1.28E-01        | 1.07       | Function       |
| GO:0019001          | guanyl nucleotide binding                             | 9.78E-04           | 1.36E-01        | 1.16       | Function       |
| GO term         | Description                      | $P$-value$^1$ | FDR $Q$-value | Enrichment | Category  |
|-----------------|----------------------------------|--------------|---------------|------------|-----------|
| GO:0044424      | intracellular part               | 3.53E-10     | 3.60E-07      | 1.04       | Component |
| GO:0044444      | cytoplasmic part                 | 4.24E-07     | 2.16E-04      | 1.05       | Component |
| GO:0044464      | cell part                        | 4.94E-06     | 1.68E-03      | 1.02       | Component |
| GO:0032991      | protein-containing complex       | 8.28E-05     | 2.11E-02      | 1.05       | Component |
| GO:0005737      | cytoplasm                        | 1.95E-04     | 3.97E-02      | 1.05       | Component |
| GO:1902494      | catalytic complex                | 3.48E-04     | 5.91E-02      | 1.07       | Component |
| GO:0044391      | ribosomal subunit                | 5.33E-04     | 7.76E-02      | 1.16       | Component |
| GO:0098588      | bounding membrane of organelle   | 5.54E-04     | 7.05E-02      | 1.14       | Component |
| GO:0098805      | whole membrane                   | 7.22E-04     | 8.18E-02      | 1.16       | Component |
| GO:0005739      | mitochondrion                    | 9.18E-04     | 9.35E-02      | 1.09       | Component |

$^1$ Only GO terms with $P$-value ≤ 0.001 are listed.
Table S10. Enriched GO terms in functional genes on the neo-X in *Drosophila americana*.

| GO term          | Description                                         | $P$-value\(^1\) | FDR $Q$-value | Enrichment | Category        |
|------------------|-----------------------------------------------------|-----------------|---------------|------------|----------------|
| GO:0044237       | cellular metabolic process                          | 8.28E-06        | 3.72E-02      | 1.06       | Process        |
| GO:0006807       | nitrogen compound metabolic process                 | 1.44E-05        | 3.24E-02      | 1.06       | Process        |
| GO:0008152       | metabolic process                                   | 3.43E-05        | 5.14E-02      | 1.05       | Process        |
| GO:0044238       | primary metabolic process                            | 1.68E-04        | 1.89E-01      | 1.05       | Process        |
| GO:0043170       | macromolecule metabolic process                     | 2.21E-04        | 1.98E-01      | 1.06       | Process        |
| GO:0071704       | organic substance metabolic process                 | 2.72E-04        | 2.04E-01      | 1.05       | Process        |
| GO:0044260       | cellular macromolecule metabolic process            | 3.87E-04        | 2.49E-01      | 1.07       | Process        |
| GO:1901360       | organic cyclic compound metabolic process           | 4.92E-04        | 2.77E-01      | 1.08       | Process        |
| GO:0006725       | cellular aromatic compound metabolic process        | 5.91E-04        | 2.95E-01      | 1.08       | Process        |
| GO:0034641       | cellular nitrogen compound metabolic process        | 6.06E-04        | 2.73E-01      | 1.07       | Process        |
| GO:1901564       | organonitrogen compound metabolic process           | 9.14E-04        | 3.73E-01      | 1.06       | Process        |
| GO:0003824       | catalytic activity                                  | 1.43E-04        | 1.83E-01      | 1.05       | Function       |

\(^1\) Only GO terms with $P$-value $\leq 0.001$ are listed.
Table S11. Enriched GO terms in pseudogenes on the neo-X in *Drosophila miranda*.

| GO term            | Description                                           | $P$-value$^1$ | FDR $Q$-value | Enrichment | Category       |
|--------------------|-------------------------------------------------------|---------------|---------------|------------|----------------|
| GO:0009593         | detection of chemical stimulus                        | 7.68E-18      | 3.46E-14      | 6.29       | Process        |
| GO:0051606         | detection of stimulus                                 | 5.08E-12      | 1.14E-08      | 4.40       | Process        |
| GO:0050907         | detection of chemical stimulus involved in sensory perception | 3.73E-10      | 5.60E-07      | 6.77       | Process        |
| GO:0050911         | detection of chemical stimulus involved in sensory perception of smell | 2.63E-09      | 2.96E-06      | 6.72       | Process        |
| GO:0050906         | detection of stimulus involved in sensory perception  | 1.92E-08      | 1.73E-05      | 5.30       | Process        |
| GO:0007606         | sensory perception of chemical stimulus                | 8.17E-06      | 6.13E-03      | 3.04       | Process        |
| GO:0007600         | sensory perception                                    | 2.89E-05      | 1.86E-02      | 2.63       | Process        |
| GO:0050877         | nervous system process                                | 7.04E-05      | 3.96E-02      | 2.30       | Process        |
| GO:0050909         | sensory perception of taste                           | 3.56E-04      | 1.78E-01      | 4.89       | Process        |
| GO:0038023         | signaling receptor activity                           | 2.93E-13      | 3.77E-10      | 4.00       | Function       |
| GO:0004888         | transmembrane signaling receptor activity             | 1.86E-12      | 1.20E-09      | 4.12       | Function       |
| GO:0060089         | molecular transducer activity                         | 1.89E-12      | 8.10E-10      | 3.79       | Function       |
| GO:0004984         | olfactory receptor activity                           | 2.63E-09      | 8.47E-07      | 6.72       | Function       |
| GO:0005549         | odorant binding                                       | 2.08E-06      | 5.37E-04      | 3.80       | Function       |
| GO:0008527         | taste receptor activity                               | 6.10E-06      | 1.31E-03      | 7.33       | Function       |
| GO:0022834         | ligand-gated channel activity                         | 7.07E-06      | 1.30E-03      | 4.58       | Function       |
| GO:0015276         | ligand-gated ion channel activity                     | 7.07E-06      | 1.14E-03      | 4.58       | Function       |
| GO:0022839         | ion gated channel activity                            | 2.32E-05      | 3.32E-03      | 3.40       | Function       |
| GO:0022836         | gated channel activity                                | 2.32E-05      | 2.99E-03      | 3.40       | Function       |
| GO:0004930         | G protein-coupled receptor activity                   | 3.25E-05      | 3.80E-03      | 4.89       | Function       |
Table S11. Cont’d.

| GO term              | Description                                               | $P$-value$^1$ | FDR $Q$-value | Enrichment | Category     |
|---------------------|-----------------------------------------------------------|---------------|---------------|------------|--------------|
| GO:0005216          | ion channel activity                                     | 3.94E-05      | 4.22E-03      | 3.11       | Function     |
| GO:0022838          | substrate-specific channel activity                      | 8.63E-05      | 8.55E-03      | 2.93       | Function     |
| GO:0022803          | passive transmembrane transporter activity               | 3.36E-04      | 3.09E-02      | 2.63       | Function     |
| GO:0015267          | channel activity                                         | 3.36E-04      | 2.88E-02      | 2.63       | Function     |
| GO:0032590          | dendrite membrane                                        | 2.63E-09      | 2.18E-06      | 6.72       | Component    |
| GO:0032589          | neuron projection membrane                               | 2.04E-07      | 8.44E-05      | 5.38       | Component    |
| GO:0031256          | leading edge membrane                                   | 2.04E-07      | 5.63E-05      | 5.38       | Component    |
| GO:0031253          | cell projection membrane                                 | 1.43E-06      | 2.96E-04      | 4.75       | Component    |
| GO:0016021          | integral component of membrane                           | 2.22E-04      | 3.68E-02      | 1.61       | Component    |
| GO:0031224          | intrinsic component of membrane                          | 2.50E-04      | 3.45E-02      | 1.60       | Component    |
| GO:0120038          | plasma membrane bounded cell projection part             | 4.13E-04      | 4.89E-02      | 2.07       | Component    |
| GO:0044463          | cell projection part                                     | 4.13E-04      | 4.28E-02      | 2.07       | Component    |

$^1$ Only GO terms with $P$-value $\leq 0.001$ are listed.
Table S12. Enriched GO terms in pseudogenes on the neo-X in *Drosophila albomicans*.

| GO term                  | Description                                                                 | $P$-value | FDR Q-value | Enrichment | Category |
|--------------------------|------------------------------------------------------------------------------|-----------|-------------|------------|----------|
| GO:0009593               | detection of chemical stimulus                                               | 1.33E-17  | 7.38E-14    | 5.17       | Process  |
| GO:0051606               | detection of stimulus                                                       | 1.31E-12  | 3.65E-09    | 3.58       | Process  |
| GO:0050906               | detection of stimulus involved in sensory perception                        | 9.53E-12  | 1.77E-08    | 4.45       | Process  |
| GO:0050907               | detection of chemical stimulus involved in sensory perception                | 1.35E-11  | 1.87E-08    | 5.13       | Process  |
| GO:0050911               | detection of chemical stimulus involved in sensory perception of smell      | 7.61E-11  | 8.46E-08    | 5.85       | Process  |
| GO:0007606               | sensory perception of chemical stimulus                                      | 6.63E-06  | 6.14E-03    | 2.60       | Process  |
| GO:0007600               | sensory perception                                                           | 7.60E-05  | 6.03E-02    | 2.12       | Process  |
| GO:0050877               | nervous system process                                                       | 8.96E-05  | 6.23E-02    | 1.85       | Process  |
| GO:0006355               | regulation of transcription, DNA-templated                                  | 9.31E-04  | 5.75E-01    | 1.38       | Process  |
| GO:2001141               | regulation of RNA biosynthetic process                                       | 9.31E-04  | 5.17E-01    | 1.38       | Process  |
| GO:1903506               | regulation of nucleic acid-templated transcription                           | 9.31E-04  | 4.70E-01    | 1.38       | Process  |
| GO:0004984               | olfactory receptor activity                                                  | 5.20E-10  | 9.37E-07    | 5.49       | Function |
| GO:0038023               | signaling receptor activity                                                  | 1.96E-08  | 1.76E-05    | 2.39       | Function |
| GO:0060089               | molecular transducer activity                                                | 1.58E-07  | 9.49E-05    | 2.24       | Function |
| GO:0005216               | ion channel activity                                                         | 1.86E-07  | 8.40E-05    | 2.85       | Function |
| GO:0022839               | ion gated channel activity                                                   | 2.45E-07  | 8.85E-05    | 3.14       | Function |
| GO:0022836               | gated channel activity                                                       | 2.45E-07  | 7.37E-05    | 3.14       | Function |
| GO:0005549               | odorant binding                                                             | 3.43E-07  | 8.83E-05    | 3.55       | Function |
| GO:0004888               | transmembrane signaling receptor activity                                    | 4.05E-07  | 9.12E-05    | 2.38       | Function |
| GO:0022838               | substrate-specific channel activity                                          | 4.36E-07  | 8.74E-05    | 2.75       | Function |
| GO term            | Description                                                   | $P$-value | FDR $Q$-value | Enrichment | Category   |
|--------------------|---------------------------------------------------------------|-----------|---------------|------------|------------|
| GO:0022834         | ligand-gated channel activity                                 | 1.27E-06  | 2.28E-04      | 3.62       | Function   |
| GO:0015276         | ligand-gated ion channel activity                             | 1.27E-06  | 2.07E-04      | 3.62       | Function   |
| GO:0022803         | passive transmembrane transporter activity                    | 2.92E-06  | 4.39E-04      | 2.53       | Function   |
| GO:0015267         | channel activity                                              | 2.92E-06  | 4.06E-04      | 2.53       | Function   |
| GO:0003700         | DNA-binding transcription factor activity                     | 3.35E-05  | 4.31E-03      | 1.81       | Function   |
| GO:0005261         | cation channel activity                                       | 8.69E-05  | 1.04E-02      | 2.77       | Function   |
| GO:000981          | DNA-binding transcription factor activity, RNA polymerase II-specific | 1.16E-04  | 1.30E-02      | 1.81       | Function   |
| GO:0140110         | transcription regulator activity                              | 1.50E-04  | 1.59E-02      | 1.62       | Function   |
| GO:000977          | RNA polymerase II regulatory region sequence-specific DNA binding | 1.55E-04  | 1.56E-02      | 1.81       | Function   |
| GO:0001012         | RNA polymerase II regulatory region DNA binding                | 1.55E-04  | 1.48E-02      | 1.81       | Function   |
| GO:0003690         | double-stranded DNA binding                                   | 1.78E-04  | 1.60E-02      | 1.70       | Function   |
| GO:000976          | transcription regulatory region sequence-specific DNA binding | 1.82E-04  | 1.56E-02      | 1.75       | Function   |
| GO:0044212         | transcription regulatory region DNA binding                    | 1.82E-04  | 1.49E-02      | 1.75       | Function   |
| GO:1990837         | sequence-specific double-stranded DNA binding                 | 1.96E-04  | 1.53E-02      | 1.71       | Function   |
| GO:0001067         | regulatory region nucleic acid binding                        | 2.14E-04  | 1.61E-02      | 1.73       | Function   |
| GO:0043565         | sequence-specific DNA binding                                 | 3.32E-04  | 2.39E-02      | 1.63       | Function   |
| GO:0003677         | DNA binding                                                  | 3.82E-04  | 2.65E-02      | 1.48       | Function   |
| GO:0008527         | taste receptor activity                                       | 8.72E-04  | 5.82E-02      | 4.18       | Function   |
| GO:0032590         | dendrite membrane                                            | 1.62E-08  | 1.65E-05      | 5.38       | Component  |
Table S12. Cont’d.

| GO term         | Description                                | \(P\)-value\(^1\) | FDR \(Q\)-value | Enrichment | Category |
|-----------------|---------------------------------------------|-------------------|-----------------|------------|----------|
| GO:0032589      | neuron projection membrane                  | 8.96E-08          | 4.57E-05        | 4.39       | Component |
| GO:0031256      | leading edge membrane                       | 8.96E-08          | 3.04E-05        | 4.39       | Component |
| GO:0031253      | cell projection membrane                    | 2.93E-07          | 7.47E-05        | 3.92       | Component |
| GO:0071683      | sensory dendrite                            | 1.60E-05          | 3.27E-03        | 6.27       | Component |
| GO:0097458      | neuron part                                 | 1.37E-04          | 2.33E-02        | 1.69       | Component |
| GO:0120038      | plasma membrane bounded cell projection part| 2.21E-04          | 3.21E-02        | 1.78       | Component |
| GO:0044463      | cell projection part                        | 2.21E-04          | 2.81E-02        | 1.78       | Component |
| GO:0098590      | plasma membrane region                      | 3.42E-04          | 3.87E-02        | 1.92       | Component |
| GO:0030425      | dendrite                                    | 9.99E-04          | 1.02E-01        | 2.37       | Component |

\(^1\) Only GO terms with \(P\)-value \(\leq 0.001\) are listed.
Table S13. Enriched GO terms in pseudogenes on the neo-X in *Drosophila americana*.

| GO term       | Description                                                                 | \( P \)-value\(^1\) | FDR \( Q \)-value | Enrichment | Category  |
|---------------|------------------------------------------------------------------------------|----------------------|-------------------|------------|-----------|
| GO:0050907    | detection of chemical stimulus involved in sensory perception               | 4.75E-05             | 2.13E-01          | 5.12       | Process   |
| GO:0009593    | detection of chemical stimulus                                              | 4.75E-05             | 1.07E-01          | 5.12       | Process   |
| GO:0050906    | detection of stimulus involved in sensory perception                        | 1.65E-04             | 2.47E-01          | 4.05       | Process   |
| GO:0050911    | detection of chemical stimulus involved in sensory perception of smell      | 2.50E-04             | 2.80E-01          | 4.93       | Process   |
| GO:0004984    | olfactory receptor activity                                                  | 2.50E-04             | 3.18E-01          | 4.93       | Function  |
| GO:0005549    | odorant binding                                                              | 3.30E-04             | 2.11E-01          | 4.19       | Function  |
| GO:0038023    | signaling receptor activity                                                  | 8.90E-04             | 3.79E-01          | 2.24       | Function  |
| GO:0043565    | sequence-specific DNA binding                                                | 9.76E-04             | 3.11E-01          | 1.82       | Function  |
| GO:0032590    | dendrite membrane                                                            | 6.54E-04             | 5.31E-01          | 4.38       | Component |

\(^1\) Only GO terms with \( P \)-value \( \leq 0.001 \) are listed.
Table S14. Enriched GO terms in functional genes on the neo-Y in *Drosophila miranda*.

| GO term       | Description                                               | $P$-value | FDR Q-value | Enrichment | Category     |
|---------------|-----------------------------------------------------------|-----------|-------------|------------|--------------|
| GO:0048523    | negative regulation of cellular process                  | 3.93E-07  | 1.75E-03    | 1.25       | Process      |
| GO:0045184    | establishment of protein localization                     | 1.21E-06  | 2.69E-03    | 1.43       | Process      |
| GO:0009987    | cellular process                                         | 1.21E-06  | 1.80E-03    | 1.07       | Process      |
| GO:0051171    | regulation of nitrogen compound metabolic process         | 1.31E-06  | 1.46E-03    | 1.18       | Process      |
| GO:0080090    | regulation of primary metabolic process                   | 1.63E-06  | 1.45E-03    | 1.18       | Process      |
| GO:0031323    | regulation of cellular metabolic process                  | 1.66E-06  | 1.23E-03    | 1.17       | Process      |
| GO:0046907    | intracellular transport                                   | 1.82E-06  | 1.16E-03    | 1.36       | Process      |
| GO:0009653    | anatomical structure morphogenesis                        | 3.10E-06  | 1.72E-03    | 1.27       | Process      |
| GO:0015031    | protein transport                                         | 3.37E-06  | 1.66E-03    | 1.43       | Process      |
| GO:0050794    | regulation of cellular process                            | 3.80E-06  | 1.69E-03    | 1.12       | Process      |
| GO:0032502    | developmental process                                     | 6.60E-06  | 2.67E-03    | 1.16       | Process      |
| GO:0016043    | cellular component organization                           | 6.60E-06  | 2.45E-03    | 1.16       | Process      |
| GO:0048519    | negative regulation of biological process                | 6.87E-06  | 2.35E-03    | 1.20       | Process      |
| GO:0042886    | amide transport                                           | 7.74E-06  | 2.46E-03    | 1.41       | Process      |
| GO:0071840    | cellular component organization or biogenesis            | 7.94E-06  | 2.36E-03    | 1.16       | Process      |
| GO:0033036    | macromolecule localization                               | 1.09E-05  | 3.02E-03    | 1.29       | Process      |
| GO:0015833    | peptide transport                                         | 1.11E-05  | 2.92E-03    | 1.41       | Process      |
| GO:0008104    | protein localization                                      | 1.24E-05  | 3.05E-03    | 1.31       | Process      |
| GO:0060255    | regulation of macromolecule metabolic process            | 1.75E-05  | 4.10E-03    | 1.15       | Process      |
| GO:0019222    | regulation of metabolic process                           | 1.85E-05  | 4.12E-03    | 1.15       | Process      |
| GO term                  | Description                                             | P-value$^1$ | FDR Q-value | Enrichment | Category              |
|-------------------------|---------------------------------------------------------|-------------|-------------|-------------|------------------------|
| GO:0048518              | positive regulation of biological process               | 2.28E-05    | 4.84E-03    | 1.17        | Process                |
| GO:0048522              | positive regulation of cellular process                 | 2.64E-05    | 5.34E-03    | 1.18        | Process                |
| GO:0044237              | cellular metabolic process                              | 3.49E-05    | 6.74E-03    | 1.10        | Process                |
| GO:0006886              | intracellular protein transport                         | 4.23E-05    | 7.84E-03    | 1.43        | Process                |
| GO:0051649              | establishment of localization in cell                   | 6.40E-05    | 1.14E-02    | 1.27        | Process                |
| GO:0010646              | regulation of cell communication                        | 8.03E-05    | 1.37E-02    | 1.24        | Process                |
| GO:0023051              | regulation of signaling                                 | 8.03E-05    | 1.32E-02    | 1.24        | Process                |
| GO:0051641              | cellular localization                                   | 8.99E-05    | 1.43E-02    | 1.24        | Process                |
| GO:0050789              | regulation of biological process                        | 9.41E-05    | 1.44E-02    | 1.09        | Process                |
| GO:0003006              | developmental process involved in reproduction          | 1.30E-04    | 1.93E-02    | 1.25        | Process                |
| GO:0048585              | negative regulation of response to stimulus            | 1.37E-04    | 1.97E-02    | 1.31        | Process                |
| GO:0010604              | positive regulation of macromolecule metabolic process  | 1.44E-04    | 2.00E-02    | 1.22        | Process                |
| GO:0002181              | cytoplasmic translation                                 | 1.95E-04    | 2.63E-02    | 1.60        | Process                |
| GO:0006412              | translation                                             | 2.21E-04    | 2.88E-02    | 1.46        | Process                |
| GO:0043043              | peptide biosynthetic process                            | 2.21E-04    | 2.80E-02    | 1.46        | Process                |
| GO:0010468              | regulation of gene expression                           | 2.23E-04    | 2.75E-02    | 1.16        | Process                |
| GO:0022414              | reproductive process                                   | 2.98E-04    | 3.58E-02    | 1.18        | Process                |
| GO:0051252              | regulation of RNA metabolic process                     | 3.02E-04    | 3.53E-02    | 1.18        | Process                |
| GO:0044260              | cellular macromolecule metabolic process                | 3.02E-04    | 3.45E-02    | 1.13        | Process                |
| GO:0065007              | biological regulation                                  | 3.13E-04    | 3.48E-02    | 1.08        | Process                |
| GO term                          | Description                                                      | $P$-value$^1$ | FDR Q-value | Enrichment | Category       |
|---------------------------------|------------------------------------------------------------------|--------------|-------------|------------|----------------|
| GO:0031324                      | negative regulation of cellular metabolic process                | 3.57E-04     | 3.87E-02    | 1.25       | Process        |
| GO:0051246                      | regulation of protein metabolic process                          | 3.75E-04     | 3.97E-02    | 1.24       | Process        |
| GO:0048583                      | regulation of response to stimulus                               | 3.90E-04     | 4.03E-02    | 1.19       | Process        |
| GO:0048869                      | cellular developmental process                                   | 3.98E-04     | 4.02E-02    | 1.19       | Process        |
| GO:0023057                      | negative regulation of signaling                                 | 4.26E-04     | 4.21E-02    | 1.31       | Process        |
| GO:2000112                      | regulation of cellular macromolecule biosynthetic process        | 4.62E-04     | 4.47E-02    | 1.16       | Process        |
| GO:0051172                      | negative regulation of nitrogen compound metabolic process       | 4.98E-04     | 4.72E-02    | 1.26       | Process        |
| GO:0022412                      | cellular process involved in reproduction in multicellular organism | 5.33E-04     | 4.93E-02    | 1.23       | Process        |
| GO:0010648                      | negative regulation of cell communication                        | 5.70E-04     | 5.17E-02    | 1.31       | Process        |
| GO:1901566                      | organonitrogen compound biosynthetic process                     | 5.75E-04     | 5.11E-02    | 1.29       | Process        |
| GO:0009966                      | regulation of signal transduction                                | 6.14E-04     | 5.35E-02    | 1.21       | Process        |
| GO:0010556                      | regulation of macromolecule biosynthetic process                 | 6.68E-04     | 5.71E-02    | 1.16       | Process        |
| GO:0071705                      | nitrogen compound transport                                      | 6.92E-04     | 5.81E-02    | 1.26       | Process        |
| GO:0032268                      | regulation of cellular protein metabolic process                 | 7.17E-04     | 5.91E-02    | 1.23       | Process        |
| GO:0043604                      | amide biosynthetic process                                       | 7.48E-04     | 6.05E-02    | 1.37       | Process        |
| GO:0009889                      | regulation of biosynthetic process                               | 7.92E-04     | 6.29E-02    | 1.15       | Process        |
| GO:0019219                      | regulation of nucleobase-containing compound metabolic process   | 7.98E-04     | 6.23E-02    | 1.16       | Process        |
| GO:0010941                      | regulation of cell death                                         | 9.00E-04     | 6.90E-02    | 1.34       | Process        |
| GO:0031326                      | regulation of cellular biosynthetic process                      | 9.42E-04     | 7.10E-02    | 1.15       | Process        |
| GO:0005515                      | protein binding                                                  | 2.31E-09     | 2.90E-06    | 1.20       | Function       |
Table S14. Cont’d.

| GO term             | Description                               | P-value\(^1\) | FDR Q-value | Enrichment | Category  |
|---------------------|-------------------------------------------|----------------|--------------|------------|-----------|
| GO:0003723          | RNA binding                               | 9.19E-05       | 5.77E-02     | 1.26       | Function  |
| GO:0005198          | structural molecule activity              | 4.34E-04       | 1.82E-01     | 1.30       | Function  |
| GO:0005488          | binding                                   | 6.48E-04       | 2.03E-01     | 1.06       | Function  |
| GO:0044424          | intracellular part                        | 1.81E-19       | 1.48E-16     | 1.14       | Component |
| GO:0032991          | protein-containing complex                | 4.49E-16       | 1.84E-13     | 1.25       | Component |
| GO:0044464          | cell part                                 | 4.17E-09       | 1.14E-06     | 1.08       | Component |
| GO:0044446          | intracellular organelle part              | 7.70E-09       | 1.58E-06     | 1.18       | Component |
| GO:0043229          | intracellular organelle                   | 8.34E-09       | 1.37E-06     | 1.14       | Component |
| GO:0043226          | organelle                                 | 1.87E-08       | 2.56E-06     | 1.14       | Component |
| GO:0044444          | cytoplasmic part                          | 2.74E-08       | 3.20E-06     | 1.16       | Component |
| GO:0043227          | membrane-bounded organelle                | 4.22E-08       | 4.32E-06     | 1.15       | Component |
| GO:0044422          | organelle part                            | 9.98E-08       | 9.09E-06     | 1.16       | Component |
| GO:0043231          | intracellular membrane-bounded organelle  | 1.85E-06       | 1.52E-04     | 1.14       | Component |
| GO:0044428          | nuclear part                              | 1.16E-05       | 8.63E-04     | 1.21       | Component |
| GO:0044445          | cytosolic part                            | 1.79E-05       | 1.22E-03     | 1.60       | Component |
| GO:1902494          | catalytic complex                         | 2.41E-05       | 1.52E-03     | 1.24       | Component |
| GO:0005634          | nucleus                                   | 1.71E-04       | 1.00E-02     | 1.15       | Component |
| GO:0098796          | membrane protein complex                  | 3.25E-04       | 1.78E-02     | 1.29       | Component |
| GO:0044451          | nucleoplasm part                          | 4.38E-04       | 2.24E-02     | 1.34       | Component |
| GO:0098798          | mitochondrial protein complex             | 7.48E-04       | 3.61E-02     | 1.37       | Component |
Table S14. Cont’d.

| GO term       | Description                                      | $P$-value$^1$ | FDR $Q$-value | Enrichment | Category          |
|---------------|--------------------------------------------------|---------------|---------------|------------|-------------------|
| GO:0005575    | cellular_component                               | 7.65E-04      | 3.48E-02      | 1.02       | Component         |
| GO:0098800    | inner mitochondrial membrane protein complex     | 9.96E-04      | 4.30E-02      | 1.52       | Component         |

$^1$ Only GO terms with $P$-value ≤ 0.001 are listed.
Table S15. Enriched GO terms in functional genes on the neo-Y in *Drosophila albomicans*.

| GO term          | Description                                              | P-value $^1$ | FDR Q-value | Enrichment | Category            |
|------------------|----------------------------------------------------------|--------------|-------------|-------------|---------------------|
| GO:0044237       | cellular metabolic process                               | 6.06E-11     | 3.41E-07    | 1.06        | Process             |
| GO:0008152       | metabolic process                                        | 9.79E-10     | 2.75E-06    | 1.05        | Process             |
| GO:0071704       | organic substance metabolic process                      | 1.65E-08     | 3.09E-05    | 1.05        | Process             |
| GO:0044238       | primary metabolic process                                | 2.14E-08     | 3.02E-05    | 1.06        | Process             |
| GO:0006807       | nitrogen compound metabolic process                      | 5.74E-08     | 6.46E-05    | 1.06        | Process             |
| GO:0044260       | cellular macromolecule metabolic process                 | 5.96E-07     | 5.59E-04    | 1.08        | Process             |
| GO:0043170       | macromolecule metabolic process                          | 4.48E-06     | 3.60E-03    | 1.05        | Process             |
| GO:0043603       | cellular amide metabolic process                         | 9.48E-06     | 6.67E-03    | 1.16        | Process             |
| GO:0034641       | cellular nitrogen compound metabolic process             | 1.01E-05     | 6.31E-03    | 1.07        | Process             |
| GO:1901564       | organonitrogen compound metabolic process                | 1.02E-05     | 5.76E-03    | 1.06        | Process             |
| GO:0043604       | amide biosynthetic process                               | 5.47E-05     | 2.80E-02    | 1.18        | Process             |
| GO:0044281       | small molecule metabolic process                         | 6.96E-05     | 3.26E-02    | 1.10        | Process             |
| GO:0006518       | peptide metabolic process                                | 7.28E-05     | 3.15E-02    | 1.16        | Process             |
| GO:0009058       | biosynthetic process                                     | 7.72E-05     | 3.10E-02    | 1.08        | Process             |
| GO:0009987       | cellular process                                         | 1.04E-04     | 3.92E-02    | 1.02        | Process             |
| GO:0071702       | organic substance transport                              | 1.20E-04     | 4.22E-02    | 1.10        | Process             |
| GO:1901576       | organic substance biosynthetic process                   | 1.34E-04     | 4.44E-02    | 1.08        | Process             |
| GO:1901575       | organic substance catabolic process                      | 1.61E-04     | 5.03E-02    | 1.09        | Process             |
| GO:0008104       | protein localization                                     | 2.07E-04     | 6.12E-02    | 1.10        | Process             |
| GO:0032787       | monocarboxylic acid metabolic process                    | 2.07E-04     | 5.82E-02    | 1.17        | Process             |
Table S15. Cont’d.

| GO term               | Description                                 | P-value\(^1\) | FDR Q-value | Enrichment | Category  |
|-----------------------|---------------------------------------------|----------------|-------------|------------|-----------|
| GO:0044267            | cellular protein metabolic process          | 2.12E-04       | 5.69E-02    | 1.07       | Process   |
| GO:0071705            | nitrogen compound transport                 | 2.30E-04       | 5.88E-02    | 1.11       | Process   |
| GO:0042886            | amide transport                             | 2.35E-04       | 5.75E-02    | 1.13       | Process   |
| GO:0009056            | catabolic process                           | 2.56E-04       | 5.99E-02    | 1.09       | Process   |
| GO:0033036            | macromolecule localization                  | 2.74E-04       | 6.17E-02    | 1.10       | Process   |
| GO:0044249            | cellular biosynthetic process               | 3.37E-04       | 7.30E-02    | 1.08       | Process   |
| GO:0006412            | translation                                 | 3.39E-04       | 7.06E-02    | 1.17       | Process   |
| GO:0043043            | peptide biosynthetic process                | 3.39E-04       | 6.81E-02    | 1.17       | Process   |
| GO:0015833            | peptide transport                           | 3.43E-04       | 6.66E-02    | 1.12       | Process   |
| GO:0045184            | establishment of protein localization       | 3.55E-04       | 6.66E-02    | 1.12       | Process   |
| GO:0043436            | oxoacid metabolic process                   | 3.58E-04       | 6.50E-02    | 1.12       | Process   |
| GO:0006082            | organic acid metabolic process              | 3.58E-04       | 6.30E-02    | 1.12       | Process   |
| GO:0015031            | protein transport                           | 3.89E-04       | 6.64E-02    | 1.12       | Process   |
| GO:0006631            | fatty acid metabolic process                | 3.97E-04       | 6.57E-02    | 1.19       | Process   |
| GO:0019538            | protein metabolic process                   | 4.46E-04       | 7.17E-02    | 1.05       | Process   |
| GO:0006396            | RNA processing                              | 5.21E-04       | 8.15E-02    | 1.11       | Process   |
| GO:1901566            | organonitrogen compound biosynthetic process| 6.48E-04       | 9.86E-02    | 1.10       | Process   |
| GO:0044248            | cellular catabolic process                  | 6.62E-04       | 9.80E-02    | 1.08       | Process   |
| GO:004282             | small molecule catabolic process            | 7.67E-04       | 1.11E-01    | 1.17       | Process   |
| GO:0019752            | carboxylic acid metabolic process           | 8.14E-04       | 1.15E-01    | 1.11       | Process   |
| GO term          | Description                                                                 | $P$-value $^1$ | FDR $Q$-value | Enrichment | Category                  |
|------------------|------------------------------------------------------------------------------|----------------|---------------|------------|---------------------------|
| GO:0016054       | organic acid catabolic process                                               | 9.72E-04       | 1.33E-01      | 1.19       | Process                   |
| GO:0046395       | carboxylic acid catabolic process                                           | 9.72E-04       | 1.30E-01      | 1.19       | Process                   |
| GO:0003824       | catalytic activity                                                          | 4.84E-07       | 8.84E-04      | 1.05       | Function                  |
| GO:0016887       | ATPase activity                                                              | 2.24E-05       | 2.04E-02      | 1.19       | Function                  |
| GO:0016817       | hydrolase activity, acting on acid anhydrides                                | 4.32E-05       | 2.63E-02      | 1.13       | Function                  |
| GO:0016818       | hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides | 4.32E-05       | 1.97E-02      | 1.13       | Function                  |
| GO:0016462       | pyrophosphatase activity                                                      | 4.93E-05       | 1.80E-02      | 1.13       | Function                  |
| GO:0017111       | nucleoside-triphosphatase activity                                           | 1.24E-04       | 3.76E-02      | 1.13       | Function                  |
| GO:0016740       | transferase activity                                                         | 4.16E-04       | 1.09E-01      | 1.07       | Function                  |
| GO:0042623       | ATPase activity, coupled                                                      | 6.79E-04       | 1.55E-01      | 1.19       | Function                  |
| GO:0043168       | anion binding                                                                | 7.39E-04       | 1.50E-01      | 1.07       | Function                  |
| GO:0003723       | RNA binding                                                                  | 9.08E-04       | 1.66E-01      | 1.09       | Function                  |
| GO:0036094       | small molecule binding                                                       | 9.48E-04       | 1.57E-01      | 1.07       | Function                  |
| GO:0044424       | intracellular part                                                           | 7.52E-14       | 7.68E-11      | 1.05       | Component                 |
| GO:0044444       | cytoplasmic part                                                             | 1.89E-10       | 9.68E-08      | 1.07       | Component                 |
| GO:0032991       | protein-containing complex                                                    | 9.28E-08       | 3.16E-05      | 1.07       | Component                 |
| GO:0044422       | organelle part                                                               | 2.50E-07       | 6.38E-05      | 1.06       | Component                 |
| GO:0044446       | intracellular organelle part                                                 | 2.74E-07       | 5.60E-05      | 1.06       | Component                 |
| GO:0044464       | cell part                                                                    | 1.00E-06       | 1.71E-04      | 1.03       | Component                 |
| GO:0005737       | cytoplasm                                                                    | 9.58E-06       | 1.40E-03      | 1.07       | Component                 |
Table S15. Cont’d.

| GO term     | Description               | $P$-value | FDR $Q$-value | Enrichment | Category |
|-------------|---------------------------|-----------|---------------|------------|----------|
| GO:1902494  | catalytic complex         | 4.50E-05  | 5.75E-03      | 1.09       | Component |
| GO:0098588  | bounding membrane of organelle | 4.69E-04  | 5.33E-02      | 1.16       | Component |
| GO:0044429  | mitochondrial part        | 8.05E-04  | 8.23E-02      | 1.10       | Component |

1 Only GO terms with $P$-value ≤ 0.001 are listed.
Table S16. Enriched GO terms in functional genes on the neo-Y in *Drosophila americana*.

| GO term           | Description                                           | $P$-value | FDR $Q$-value | Enrichment | Category        |
|-------------------|-------------------------------------------------------|-----------|---------------|------------|-----------------|
| GO:0044237        | cellular metabolic process                            | 2.30E-08  | 1.05E-04      | 1.08       | Process         |
| GO:0006807        | nitrogen compound metabolic process                   | 8.66E-07  | 1.97E-03      | 1.08       | Process         |
| GO:0008152        | metabolic process                                     | 1.42E-05  | 2.15E-02      | 1.06       | Process         |
| GO:0043170        | macromolecule metabolic process                       | 2.67E-05  | 3.03E-02      | 1.07       | Process         |
| GO:0044249        | cellular biosynthetic process                         | 2.83E-05  | 2.57E-02      | 1.13       | Process         |
| GO:0034641        | cellular nitrogen compound metabolic process          | 3.99E-05  | 3.02E-02      | 1.10       | Process         |
| GO:0044260        | cellular macromolecule metabolic process              | 5.68E-05  | 3.69E-02      | 1.09       | Process         |
| GO:0009058        | biosynthetic process                                  | 1.23E-04  | 6.99E-02      | 1.11       | Process         |
| GO:1901360        | organic cyclic compound metabolic process             | 1.34E-04  | 6.77E-02      | 1.10       | Process         |
| GO:1901576        | organic substance biosynthetic process                | 1.58E-04  | 7.19E-02      | 1.11       | Process         |
| GO:0006725        | cellular aromatic compound metabolic process          | 1.89E-04  | 7.80E-02      | 1.10       | Process         |
| GO:0044238        | primary metabolic process                             | 2.00E-04  | 7.59E-02      | 1.06       | Process         |
| GO:0009059        | macromolecule biosynthetic process                    | 2.09E-04  | 7.31E-02      | 1.16       | Process         |
| GO:1901564        | organonitrogen compound metabolic process             | 2.11E-04  | 6.85E-02      | 1.08       | Process         |
| GO:0006139        | nucleobase-containing compound metabolic process      | 2.77E-04  | 8.40E-02      | 1.10       | Process         |
| GO:0046483        | heterocycle metabolic process                         | 2.87E-04  | 8.16E-02      | 1.10       | Process         |
| GO:0044271        | cellular nitrogen compound biosynthetic process       | 2.96E-04  | 7.91E-02      | 1.15       | Process         |
| GO:1901566        | organonitrogen compound biosynthetic process          | 3.56E-04  | 8.99E-02      | 1.16       | Process         |
| GO:0016070        | RNA metabolic process                                 | 4.05E-04  | 9.68E-02      | 1.13       | Process         |
| GO:0071704        | organic substance metabolic process                   | 4.36E-04  | 9.91E-02      | 1.05       | Process         |
Table S16. Cont’d.

| GO term       | Description                                | $P$-value$^1$ | FDR $Q$-value | Enrichment | Category    |
|---------------|--------------------------------------------|---------------|---------------|------------|-------------|
| GO:0034645    | cellular macromolecule biosynthetic process | 9.28E-04      | 2.01E-01      | 1.16       | Process     |
| GO:0003723    | RNA binding                                | 1.79E-04      | 2.33E-01      | 1.16       | Function    |
| GO:0044424    | intracellular part                         | 9.72E-10      | 8.04E-07      | 1.06       | Component   |
| GO:0044444    | cytoplasmal part                           | 1.34E-08      | 5.55E-06      | 1.10       | Component   |
| GO:0044422    | organelle part                             | 4.25E-06      | 1.17E-03      | 1.08       | Component   |
| GO:0032991    | protein-containing complex                  | 6.87E-06      | 1.42E-03      | 1.08       | Component   |
| GO:0044446    | intracellular organelle part               | 1.02E-05      | 1.69E-03      | 1.08       | Component   |
| GO:0005737    | cytoplasm                                  | 2.66E-05      | 3.67E-03      | 1.10       | Component   |
| GO:1990904    | ribonucleoprotein complex                   | 2.96E-04      | 3.50E-02      | 1.15       | Component   |
| GO:0044429    | mitochondrial part                          | 5.77E-04      | 5.96E-02      | 1.16       | Component   |

$^1$ Only GO terms with $P$-value ≤ 0.001 are listed.
Table S17. Enriched GO terms in pseudogenes on the neo-Y in *Drosophila miranda*.

| GO term          | Description                                      | P-value\(^1\) | FDR Q-value | Enrichment | Category    |
|------------------|--------------------------------------------------|---------------|-------------|------------|-------------|
| GO:0051606       | detection of stimulus                             | 5.21E-06      | 2.32E-02    | 1.86       | Process     |
| GO:0009593       | detection of chemical stimulus                    | 1.86E-05      | 4.14E-02    | 1.99       | Process     |
| GO:0006811       | ion transport                                     | 4.05E-04      | 6.00E-01    | 1.41       | Process     |
| GO:0030001       | metal ion transport                               | 5.86E-04      | 6.51E-01    | 1.65       | Process     |
| GO:0005216       | ion channel activity                              | 9.91E-08      | 1.24E-04    | 1.91       | Function    |
| GO:0038023       | signaling receptor activity                       | 2.21E-07      | 1.39E-04    | 1.77       | Function    |
| GO:0004888       | transmembrane signaling receptor activity         | 5.62E-07      | 2.35E-04    | 1.80       | Function    |
| GO:0060089       | molecular transducer activity                     | 8.48E-07      | 2.66E-04    | 1.71       | Function    |
| GO:0022838       | substrate-specific channel activity               | 1.20E-06      | 3.00E-04    | 1.82       | Function    |
| GO:0022803       | passive transmembrane transporter activity        | 5.21E-06      | 1.09E-03    | 1.74       | Function    |
| GO:0015267       | channel activity                                  | 5.21E-06      | 9.34E-04    | 1.74       | Function    |
| GO:0022839       | ion gated channel activity                         | 9.89E-06      | 1.55E-03    | 1.85       | Function    |
| GO:0022836       | gated channel activity                            | 9.89E-06      | 1.38E-03    | 1.85       | Function    |
| GO:0015075       | ion transmembrane transporter activity             | 4.17E-05      | 5.24E-03    | 1.48       | Function    |
| GO:0004930       | G protein-coupled receptor activity               | 7.29E-05      | 8.31E-03    | 2.20       | Function    |
| GO:0004252       | serine-type endopeptidase activity                | 7.38E-05      | 7.72E-03    | 1.85       | Function    |
| GO:0008236       | serine-type peptidase activity                    | 7.38E-05      | 7.13E-03    | 1.85       | Function    |
| GO:0017171       | serine hydrolase activity                         | 7.38E-05      | 6.62E-03    | 1.85       | Function    |
| GO:0022834       | ligand-gated channel activity                     | 7.60E-05      | 6.36E-03    | 1.97       | Function    |
| GO:0015276       | ligand-gated ion channel activity                 | 7.60E-05      | 5.96E-03    | 1.97       | Function    |
Table S17. Cont’d.

| GO term               | Description                                           | $P$-value$^1$ | FDR $Q$-value | Enrichment | Category       |
|-----------------------|-------------------------------------------------------|---------------|---------------|-------------|----------------|
| GO:0022857            | transmembrane transporter activity                    | 8.33E-05      | 6.15E-03      | 1.39        | Function       |
| GO:0005215            | transporter activity                                  | 9.73E-05      | 6.78E-03      | 1.37        | Function       |
| GO:0015318            | inorganic molecular entity transmembrane transporter activity | 3.78E-04      | 2.50E-02      | 1.41        | Function       |
| GO:0020037            | heme binding                                         | 4.01E-04      | 2.52E-02      | 1.76        | Function       |
| GO:0046906            | tetrapyrrole binding                                 | 4.01E-04      | 2.40E-02      | 1.76        | Function       |
| GO:0005261            | cation channel activity                              | 4.84E-04      | 2.76E-02      | 1.80        | Function       |
| GO:0016021            | integral component of membrane                       | 7.54E-06      | 6.18E-03      | 1.32        | Component      |
| GO:0005887            | integral component of plasma membrane                | 9.36E-06      | 3.84E-03      | 1.60        | Component      |
| GO:0031226            | intrinsic component of plasma membrane               | 9.36E-06      | 2.56E-03      | 1.60        | Component      |
| GO:0031224            | intrinsic component of membrane                      | 1.07E-05      | 2.19E-03      | 1.32        | Component      |

$^1$ Only GO terms with $P$-value $\leq 0.001$ are listed.
| GO term          | Description                                           | \(P\)-value\(^1\) | FDR \(Q\)-value | Enrichment | Category        |
|------------------|-------------------------------------------------------|---------------------|-----------------|-------------|-----------------|
| GO:0009593       | detection of chemical stimulus                        | 5.77E-17            | 3.25E-13        | 5.06        | Process         |
| GO:0051606       | detection of stimulus                                 | 2.22E-14            | 6.24E-11        | 3.80        | Process         |
| GO:0050907       | detection of chemical stimulus involved in sensory perception | 9.88E-11            | 1.85E-07        | 5.13        | Process         |
| GO:0050906       | detection of stimulus involved in sensory perception  | 1.28E-10            | 1.80E-07        | 4.33        | Process         |
| GO:0050911       | detection of chemical stimulus involved in sensory perception of smell | 7.11E-08            | 8.00E-05        | 5.01        | Process         |
| GO:0007186       | G protein-coupled receptor signaling pathway          | 3.78E-06            | 3.54E-03        | 2.32        | Process         |
| GO:0007606       | sensory perception of chemical stimulus               | 6.17E-05            | 4.96E-02        | 2.38        | Process         |
| GO:0006355       | regulation of transcription, DNA-templated           | 2.11E-04            | 1.49E-01        | 1.39        | Process         |
| GO:2001141       | regulation of RNA biosynthetic process                | 2.11E-04            | 1.32E-01        | 1.39        | Process         |
| GO:1903506       | regulation of nucleic acid-templated transcription     | 2.11E-04            | 1.19E-01        | 1.39        | Process         |
| GO:0007600       | sensory perception                                    | 4.61E-04            | 2.36E-01        | 1.93        | Process         |
| GO:0030154       | cell differentiation                                  | 5.94E-04            | 2.79E-01        | 1.80        | Process         |
| GO:0051252       | regulation of RNA metabolic process                   | 8.68E-04            | 3.76E-01        | 1.33        | Process         |
| GO:0004888       | transmembrane signaling receptor activity             | 4.68E-12            | 8.54E-09        | 2.84        | Function        |
| GO:0038023       | signaling receptor activity                           | 1.45E-11            | 1.33E-08        | 2.60        | Function        |
| GO:0060089       | molecular transducer activity                         | 2.83E-10            | 1.72E-07        | 2.43        | Function        |
| GO:0004930       | G protein-coupled receptor activity                   | 5.64E-09            | 2.57E-06        | 3.44        | Function        |
| GO:0004984       | olfactory receptor activity                           | 7.11E-08            | 2.60E-05        | 5.01        | Function        |
| GO:0022839       | ion gated channel activity                            | 1.59E-07            | 4.84E-05        | 3.10        | Function        |
| GO:0022836       | gated channel activity                                | 1.59E-07            | 4.15E-05        | 3.10        | Function        |

Table S18. Enriched GO terms in pseudogenes on the neo-Y in *Drosophila albomicans*. 

\(^1\) \(P\)-values were corrected for multiple testing using the False Discovery Rate (FDR).
Table S18. Cont’d.

| GO term              | Description                                                                 | $P$-value$^1$ | FDR $Q$-value | Enrichment | Category    |
|---------------------|-----------------------------------------------------------------------------|--------------|---------------|------------|-------------|
| GO:0003700          | DNA-binding transcription factor activity                                   | 7.22E-07     | 1.65E-04      | 1.91       | Function    |
| GO:0022834          | ligand-gated channel activity                                               | 8.38E-07     | 1.70E-04      | 3.57       | Function    |
| GO:0015276          | ligand-gated ion channel activity                                           | 8.38E-07     | 1.53E-04      | 3.57       | Function    |
| GO:0005216          | ion channel activity                                                       | 1.30E-06     | 2.16E-04      | 2.63       | Function    |
| GO:000981           | DNA-binding transcription factor activity, RNA polymerase II-specific       | 2.68E-06     | 4.07E-04      | 1.94       | Function    |
| GO:0022838          | substrate-specific channel activity                                         | 3.15E-06     | 4.42E-04      | 2.52       | Function    |
| GO:0001653          | peptide receptor activity                                                   | 3.20E-06     | 4.17E-04      | 3.86       | Function    |
| GO:000977           | RNA polymerase II regulatory region sequence-specific DNA binding           | 3.33E-06     | 4.06E-04      | 1.94       | Function    |
| GO:0001012          | RNA polymerase II regulatory region DNA binding                             | 3.33E-06     | 3.81E-04      | 1.94       | Function    |
| GO:0004252          | serine-type endopeptidase activity                                          | 4.15E-06     | 4.46E-04      | 2.44       | Function    |
| GO:0022803          | passive transmembrane transporter activity                                  | 4.15E-06     | 4.21E-04      | 2.44       | Function    |
| GO:0015267          | channel activity                                                            | 4.15E-06     | 3.99E-04      | 2.44       | Function    |
| GO:0140110          | transcription regulator activity                                            | 4.26E-06     | 3.89E-04      | 1.71       | Function    |
| GO:0008528          | G protein-coupled peptide receptor activity                                 | 4.71E-06     | 4.10E-04      | 4.01       | Function    |
| GO:000976           | transcription regulatory region sequence-specific DNA binding               | 7.09E-06     | 5.89E-04      | 1.85       | Function    |
| GO:0044212          | transcription regulatory region DNA binding                                 | 7.09E-06     | 5.63E-04      | 1.85       | Function    |
| GO:0001067          | regulatory region nucleic acid binding                                      | 8.86E-06     | 6.74E-04      | 1.84       | Function    |
| GO:1990837          | sequence-specific double-stranded DNA binding                              | 1.08E-05     | 7.90E-04      | 1.80       | Function    |
| GO:0008188          | neuropeptide receptor activity                                              | 1.93E-05     | 1.35E-03      | 3.91       | Function    |
| GO:0008236          | serine-type peptidase activity                                              | 3.61E-05     | 2.44E-03      | 2.21       | Function    |
Table S18. Cont’d.

| GO term                  | Description                                              | $P$-value $^1$ | FDR $Q$-value | Enrichment | Category  |
|--------------------------|----------------------------------------------------------|----------------|---------------|------------|-----------|
| GO:0017171               | serine hydrolase activity                                | 3.61E-05       | 2.35E-03      | 2.21       | Function  |
| GO:0008527               | taste receptor activity                                  | 3.66E-05       | 2.31E-03      | 5.47       | Function  |
| GO:0003690               | double-stranded DNA binding                              | 4.36E-05       | 2.65E-03      | 1.72       | Function  |
| GO:0043565               | sequence-specific DNA binding                            | 4.51E-05       | 2.66E-03      | 1.67       | Function  |
| GO:0005261               | cation channel activity                                  | 1.19E-04       | 6.81E-03      | 2.65       | Function  |
| GO:0003677               | DNA binding                                              | 1.48E-04       | 8.21E-03      | 1.48       | Function  |
| GO:0005549               | odorant binding                                          | 2.26E-04       | 1.21E-02      | 2.73       | Function  |
| GO:0000987               | proximal promoter sequence-specific DNA binding          | 2.45E-04       | 1.28E-02      | 1.80       | Function  |
| GO:0000978               | RNA polymerase II proximal promoter sequence-specific DNA binding | 4.63E-04       | 2.35E-02      | 1.77       | Function  |
| GO:0032590               | dendrite membrane                                        | 3.62E-07       | 3.70E-04      | 4.97       | Component |
| GO:0032589               | neuron projection membrane                                | 3.20E-06       | 1.63E-03      | 3.86       | Component |
| GO:0031256               | leading edge membrane                                    | 3.20E-06       | 1.09E-03      | 3.86       | Component |
| GO:0031253               | cell projection membrane                                  | 1.09E-05       | 2.79E-03      | 3.39       | Component |
| GO:0044459               | plasma membrane part                                     | 4.56E-04       | 9.32E-02      | 1.46       | Component |

$^1$ Only GO terms with $P$-value $\leq$ 0.001 are listed.
Table S19. Enriched GO terms in pseudogenes on the neo-Y in *Drosophila americana*.

| GO term          | Description                                                                 | $P$-value $^1$ | FDR $Q$-value | Enrichment | Category       |
|------------------|----------------------------------------------------------------------------|----------------|---------------|------------|----------------|
| GO:0045935       | positive regulation of nucleobase-containing compound metabolic process    | 2.23E-04       | 1.00E+00      | 1.92       | Process        |
| GO:0050907       | detection of chemical stimulus involved in sensory perception               | 2.34E-04       | 5.32E-01      | 4.04       | Process        |
| GO:0009593       | detection of chemical stimulus                                             | 2.34E-04       | 3.55E-01      | 4.04       | Process        |
| GO:0007606       | sensory perception of chemical stimulus                                     | 3.66E-04       | 4.16E-01      | 2.85       | Process        |
| GO:0009891       | positive regulation of biosynthetic process                                 | 3.92E-04       | 3.56E-01      | 1.84       | Process        |
| GO:0031328       | positive regulation of cellular biosynthetic process                        | 3.92E-04       | 2.97E-01      | 1.84       | Process        |
| GO:0045944       | positive regulation of transcription by RNA polymerase II                   | 3.94E-04       | 2.56E-01      | 2.14       | Process        |
| GO:0007617       | mating behavior                                                             | 6.52E-04       | 3.71E-01      | 3.63       | Process        |
| GO:0051254       | positive regulation of RNA metabolic process                                | 7.12E-04       | 3.59E-01      | 1.85       | Process        |
| GO:0010557       | positive regulation of macromolecule biosynthetic process                   | 9.39E-04       | 4.27E-01      | 1.80       | Process        |
| GO:0050911       | detection of chemical stimulus involved in sensory perception of smell      | 9.67E-04       | 4.00E-01      | 3.89       | Process        |
| GO:0016298       | lipase activity                                                             | 1.32E-05       | 1.72E-02      | 3.46       | Function       |
| GO:0043565       | sequence-specific DNA binding                                               | 7.17E-05       | 4.69E-02      | 1.86       | Function       |
| GO:0000976       | transcription regulatory region sequence-specific DNA binding               | 1.60E-04       | 6.97E-02      | 1.88       | Function       |
| GO:0001067       | regulatory region nucleic acid binding                                      | 2.05E-04       | 6.71E-02      | 1.85       | Function       |
| GO:0044212       | transcription regulatory region DNA binding                                 | 2.05E-04       | 5.37E-02      | 1.85       | Function       |
| GO:0015280       | ligand-gated sodium channel activity                                        | 2.59E-04       | 5.64E-02      | 5.19       | Function       |
| GO:0022834       | ligand-gated channel activity                                               | 2.65E-04       | 4.95E-02      | 3.34       | Function       |
| GO:0015276       | ligand-gated ion channel activity                                           | 2.65E-04       | 4.33E-02      | 3.34       | Function       |
Table S19. Cont’d.

| GO term                        | Description                                                                 | \( P \)-value | FDR \( Q \)-value | Enrichment | Category       |
|--------------------------------|------------------------------------------------------------------------------|---------------|-------------------|------------|----------------|
| GO:1990837 sequence-specific double-stranded DNA binding |                                                                  | 3.32E-04      | 4.83E-02         | 1.81       | Function       |
| GO:0003690 double-stranded DNA binding                           |                                                                              | 4.28E-04      | 5.60E-02         | 1.77       | Function       |
| GO:0060089 molecular transducer activity                        |                                                                              | 4.61E-04      | 5.47E-02         | 2.08       | Function       |
| GO:0009777 RNA polymerase II regulatory region sequence-specific DNA binding |                                                                 | 6.08E-04      | 6.62E-02         | 1.82       | Function       |
| GO:0001012 RNA polymerase II regulatory region DNA binding       |                                                                              | 6.08E-04      | 6.11E-02         | 1.82       | Function       |
| GO:0003700 DNA-binding transcription factor activity              |                                                                              | 6.28E-04      | 5.86E-02         | 1.79       | Function       |
| GO:0052689 carboxylic ester hydrolase activity                    |                                                                              | 6.39E-04      | 5.56E-02         | 2.72       | Function       |
| GO:0038023 signaling receptor activity                           |                                                                              | 6.47E-04      | 5.28E-02         | 2.12       | Function       |
| GO:0004984 olfactory receptor activity                           |                                                                              | 9.67E-04      | 7.44E-02         | 3.89       | Function       |
| GO:0030312 external encapsulating structure                      |                                                                              | 5.51E-05      | 4.56E-02         | 4.15       | Component      |
| GO:0005887 integral component of plasma membrane                 |                                                                              | 7.18E-05      | 2.97E-02         | 2.03       | Component      |
| GO:0031226 intrinsic component of plasma membrane                 |                                                                              | 1.29E-04      | 3.55E-02         | 1.97       | Component      |
| GO:0016021 integral component of membrane                         |                                                                              | 3.18E-04      | 6.58E-02         | 1.51       | Component      |
| GO:0031224 intrinsic component of membrane                        |                                                                              | 4.97E-04      | 8.21E-02         | 1.48       | Component      |
| GO:0044421 extracellular region part                              |                                                                              | 5.41E-04      | 7.45E-02         | 1.71       | Component      |

\(^1\) Only GO terms with \( P \)-value \( \leq 0.001 \) are listed.
Table S20. Possible functions of the 35 genes that have parallelly been pseudogenized on the neo-Ys in *Drosophila miranda* and *D. albomicans* based on the *D. melanogaster* orthologs.

| No.  | Flybase ID       | Symbol | Gene snapshot                                                                 |
|------|------------------|--------|--------------------------------------------------------------------------------|
| 1    | FBgn0001187      | *Hex-C*| A hexokinase involved in glucose homeostasis                                  |
| 2    | FBgn0001319      | *kn*   | A transcription factor required for somatic and alary muscle specification and embryonic head segmentation |
| 3    | FBgn0011659      | *Mlh1* |                                                                                |
| 4    | FBgn0013763      | *Idgf6*| A secreted protein that is highly expressed in larval fat body and adult head, and also involved in molting and egg chamber tube morphogenesis |
| 5    | FBgn0020269      | *mspo* |                                                                                |
| 6    | FBgn0025827      | *CG6421*|                                                                                |
| 7    | FBgn0028424      | *Jhl-26*| A sperm protein, overexpression of which causes upregulation of the male accessory gland protein gene, higher embryonic lethality, and reduction of receptivity to remating in mated females |
| 8    | FBgn0028743      | *Dhit* | A protein that negatively regulates Galpha signaling and is involved in several GPCR-mediated signaling and developmental programs, such as asymmetric cell division in the sensory organ lineage |
| 9    | FBgn0033128      | *Tsp42Eg*|                                                                                |
| 10   | FBgn0033540      | *Elp2* |                                                                                |
| 11   | FBgn0033661      | *CG13185*|                                                                                |
| 12   | FBgn0033708      | *CG8850*|                                                                                |
| 13   | FBgn0033744      | *Dh44-R2*| A protein that exhibits diuretic hormone receptor activity and is involved in G protein-coupled receptor signaling pathways, hormone-mediated signaling pathways, and response to salt stress |
| 14   | FBgn0033948      | *CG12863*|                                                                                |
| 15   | FBgn0034136      | *DAT* | A dopamine transporter that mediates uptake of dopamine from the synaptic cleft, loss of which increases extracellular dopamine and is associated with behavioral phenotypes including increased activity and decreased sleep |
| 16   | FBgn0034166      | *CG6472*|                                                                                |
| 17   | FBgn0034177      | *AsnRS-m*|                                                                                |
| No.  | Flybase ID    | Symbol   | Gene snapshot                                                                 |
|------|---------------|----------|-------------------------------------------------------------------------------|
| 18   | FBgn0034232   | CG4866   |                                                                               |
| 19   | FBgn0034314   | nopo     | A RING domain-containing E3 ubiquitin ligase that is essential for early embryogenesis and positively regulates caspase-dependent cell death |
| 20   | FBgn0034452   | Oseg6    |                                                                               |
| 21   | FBgn0034567   | CG15651  |                                                                               |
| 22   | FBgn0034623   | CG9822   |                                                                               |
| 23   | FBgn0034624   | CG17974  |                                                                               |
| 24   | FBgn0034634   | CG10494  |                                                                               |
| 25   | FBgn0035025   | uri      |                                                                               |
| 26   | FBgn0042083   | Mccc2    |                                                                               |
| 27   | FBgn0046879   | Obp56c   |                                                                               |
| 28   | FBgn0050411   | CG30411  |                                                                               |
| 29   | FBgn0053544   | Vkor     | An enzyme that converts Vitamin-K 2,3-epoxide back to the active form of Vitamin-K by de-epoxidation |
| 30   | FBgn0082585   | sprt     |                                                                               |
| 31   | FBgn0262816   | CG43187  |                                                                               |
| 32   | FBgn0263020   | CG43315  |                                                                               |

1 For other three genes, orthologs were not detected in *D. melanogaster* based on the BLASTP search.

2 Information are retrieved from FlyBase (https://flybase.org/).
Fig. S1: Classification of neo-X and neo-Y gene groups in (A) *Drosophila miranda*, (B) *D. albomicans*, and (C) *D. americana*. For each category, all inparalogs were regarded as a single gene group (see SUPPLEMENTARY METHODS for the procedure to detect inparalogs). Numbers in parentheses indicate the numbers of gene groups 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.
Fig. S2: Relationship between the functionality of neo-Y-linked genes and the DC of neo-X-linked homologs in larva, pupa, adult, and testis of (A) *Drosophila miranda*, (B) *D. albomicans*, and (C) *D. americana*. To compute $R_{\text{Lin}}$, the cTPM value (corrected TPM, TPM 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*, 2,449 and 218 for *D. albomicans*, and 1,204 and 93 for *D. americana*. A box plot is also shown on each violin plot. Differences of $R_{\text{Lin}}$ values between groups were tested based on a permutation test with 10,000 replicates with the correction of multiple testing by the Benjamini and Hochberg method (Benjamini and Hochberg 1995) under the null hypothesis of $R_{\text{Lin}}(X_{F}-Y_{F}) = R_{\text{Lin}}(X_{F}-Y_{P})$: $<<<$, $Q<0.001$; $<<$, $Q<0.01$; $<$, $Q<0.05$; n.s., $Q\geq0.05$. A solid line indicates the $R_{\text{Lin}}$ value of 1 (0 in log$_{2}$) indicating perfect DC, whereas a broken line corresponds to a value of 0.5 (−1 in log$_{2}$) indicating no DC. +, −, and ± along each plot means that the median $R_{\text{Lin}}$ value is $>0.5$, $<0.5$, and 0.5 at the 5% significance level, respectively, based on a bootstrap test with 10,000 replications with the correction of multiple testing.
Fig. S3: 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 (A) Drosophila miranda, (B) D. albomicans, and (C) D. americana. For computing \( R_{\text{Lin}} \), the corrected fragments per kilobase of exon per million mapped reads (cFPKM, FPKM normalized by median) was used. Only uniquely mapped reads were used for computing cFPKM. \( X_F-Y_F \), 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 \( D. miranda \), 2,449 and 218 for \( D. albomicans \), and 1,204 and 93 for \( D. americana \). A box plot is also shown on each violin plot.

Differences of \( R_{\text{Lin}} \) values between groups were tested based on a permutation test with 10,000 replicates with the correction of multiple testing by the Benjamini and Hochberg method (Benjamini and Hochberg 1995) under the null hypothesis of \( R_{\text{Lin}}(X_F-Y_F) = R_{\text{Lin}}(X_F-Y_P) \): <<<, \( Q<0.001 \); <<, \( Q<0.01 \); <, \( Q<0.05 \); n.s., \( Q>0.05 \). A solid line indicates the \( R_{\text{Lin}} \) value of 1 (0 in \( \log_2 \)) indicating perfect DC, whereas a broken line corresponds to a value of 0.5 (−1 in \( \log_2 \)) indicating no DC. +, −, and ± along each plot means that the median \( R_{\text{Lin}} \) value is >0.5, <0.5, and 0.5 at the 5% significance level, respectively, based on a bootstrap test with 10,000 replications with the correction of multiple testing.
Fig. S4: Pseudogenization events in (A) *Drosophila miranda*, (B) *D. albomicans*, and (C) *D. americana* lineages compared with those in their closely related species, *D. pseudoobscura*, *D. nasuta*, and *D. texana*, respectively. Numbers in parentheses indicate the numbers of genes that were pseudogenized in each lineage. The orthologs that were regarded to be functional in the outgroup species (*D. obscura*, *D. kohkoa*, and *D. novamexicana*, respectively) and located on the same Muller element among the three species compared without any inparalogs were used for this analysis. The number of such orthologs is shown above each tree. Statistical significance in the difference of the pseudogenization rate between lineages leading to the species with neo-sex chromosomes and its closely related species was computed by $\chi^2$ test with the correction of multiple testing under the null hypothesis of equal pseudogenization rates between lineages.
Fig. S5: Relationship between spatiotemporal gene expression pattern and pseudogenization. All genes were classified into three groups based on the tissue with the highest expression in (A) *Drosophila pseudoobscura*, (B) *D. nasuta*, and (C) *D. texana*, the closely related species of *D. miranda*, *D. albomicans*, and *D. americana*, respectively, based on the TPM value. However, to remove the genes with similar TPMs in multiple tissues, only genes with at least two-fold TPM in a tissue compared with that in the other tissues examined were used for this analysis. The 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: <<< or >>>, $Q<0.001$; << or >>; $Q<0.01$; < or >, $Q<0.05$; n.s., $Q \geq 0.05$. 

![Graph showing the relationship between spatiotemporal gene expression pattern and pseudogenization.](image-url)
Fig. S6: Lineage-specific and parallel (shared) pseudogenization events on Muller elements (A) A, (B) B, (C-E) D, and (F) E in the lineages of *Drosophila miranda* (*mir*) and *D. albomicans* (*alb*). Muller element D is the Chromosome XR in *D. miranda* whereas it is also the neo-sex chromosomes in *D. albomicans*. Therefore, for Muller element D, branches (C) m and a, (D) m and b, and (E) m and c in Fig. S4 were separately analyzed. Each Venn diagram shows the number of pseudogenization events in the two lineages. The number of orthologs regarded to be functional in the outgroup species (*D. obscura* and *D. kohkooa*, respectively) and located on the same Muller element without any inparalogs is shown below each diagram. Statistical difference between the observed and expected numbers of parallel pseudogenizations was tested by a binomial test with correction for multiple testing: n.s., $Q \geq 0.05$. 

![Venn diagrams showing pseudogenization events](image-url)
Fig. S7: Lineage-specific and parallel (shared) pseudogenization events on Muller elements (A) A, (B-D) B, (E-G) C, (H) D, and (I) E in the lineages of *Drosophila miranda* (*mir*) and *D. americana* (*ame*). Muller element B is the Chromosome 4 in *D. miranda* whereas it is the neo-sex chromosomes in *D. americana*. Therefore, for this Muller element, branches (B) n and g, (C) n and h, and (D) n and i in Fig. S4 were separately analyzed. Similarly, Muller element C is the neo-sex chromosomes in *D. miranda* whereas it is the Chromosome 5 in *D. americana*. Therefore, for this Muller element, branches (E) d and r, (F) e and r, and (G) f and r in Fig. S4 were separately analyzed. Each Venn diagram shows the number of pseudogenization events in the two lineages. The number of orthologs regarded to be functional in the outgroup species (*D. obscura* and *D. novamexicana*, respectively) and located on the same Muller element without any inparalogs is shown below each diagram. Statistical difference between the observed and expected numbers of parallel pseudogenizations was tested by a binomial test with correction for multiple testing: *, \( Q < 0.05 \); n.s., \( Q \geq 0.05 \).
Fig. S8: Lineage-specific and parallel (shared) pseudogenization events on Muller elements (A) A, (B-D) B, (E-G) C, (H-J) D, (K) E, and (L) F in the lineages of *Drosophila albomicans* (*alb*) and *D. americana* (*ame*). Muller element B is the Chromosome 2L in *D. albomicans* whereas it is the neo-sex chromosomes in *D. americana*. Therefore, for this Muller element, branches (B) l and g, (C) l and h, and (D) l and i in Fig. S4 were separately analyzed. Similarly, Muller elements C and D are the neo-sex chromosomes in *D. albomicans* whereas they are the Chromosomes 5 and 3, respectively, in *D. americana*. Therefore, for these Muller elements, branches (E) a and r, (F) b and r, (G) c and r, (H) a and s, (I) b and s, and (J) c and s in Fig. S4 were separately analyzed. Each Venn diagram shows the number of pseudogenization events in the two lineages. The number of orthologs regarded to be functional in the outgroup species (*D. kohkoa* and *D. novamexicana*, respectively) and located on the same Muller element without any inparalogs is shown below each diagram. Statistical difference between the observed and expected numbers of parallel pseudogenizations was tested by a binomial test with correction for multiple testing: *, Q < 0.05; n.s., Q ≥ 0.05.
Fig. S9: Relationship between functionality of neo-Y-linked genes and ratio of nonsynonymous to synonymous nucleotide divergence per site ($d_N/d_S$) in (A) *Drosophila miranda* and (B) *D. albomicans* lineages. Branches c (left), a (center), and b (right) corresponds to those in Fig. 4. The orthologs regarded to be functional in the outgroup species (*D. obscura* and *D. kohkoa*, respectively) and located on Muller element C without any inparalogs were used. Fun, neo-Y-linked genes that are functional at least in (A) *D. miranda* or (B) *D. albomicans*; P$_{mir}$, neo-Y-linked genes that are nonfunctional in *D. miranda* but functional in *D. albomicans*; P$_{alb}$, neo-Y-linked genes that are nonfunctional in *D. albomicans* but functional in *D. miranda*; P$_{share}$, neo-Y-linked genes that are nonfunctional in both *D. miranda* and *D. albomicans*. Differences of $d_N/d_S$ ratios between groups were evaluated by Mann-Whitney U test with correction for multiple testing under the null hypothesis of the equal $d_N/d_S$ ratio. Same roman numerals indicate $Q \geq 0.05$, whereas different numerals mean $Q < 0.05$ between categories.
REFERENCES

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.

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.

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.

Cox MP, Peterson DA, Biggs PJ. 2010. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* 11: 485.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakos AA, del Angel G, Rivas MA, Hanna M et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* 43: 491-498.

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.

Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. 2009. GOriilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10: 48.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32: 1792-1797.

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.
Hatsumi M, Morishige Y, Wakahama K-I. 1988. Metaphase chromosomes of four species of the

*Drosophila nasuta* subgroup. *Jpn J Genet* **63**: 435-444.

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.

Kitagawa O, Wakahama K-I, Fuyama Y, Shimada Y, Takanashi E, Hatsumi M, Uwabo M, Mita Y. 1982.
Genetic studies of the *Drosophila nasuta* subgroup, with notes on distribution and morphology.

*Jpn J Genet* **57**: 113-141.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a
reference genome. *BMC Bioinformatics* **12**: 323.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform.

*Bioinformatics* **25**: 1754-1760.

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.

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.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads.

*EMBnetjournal* **17**: 10-12.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian
transcriptomes by RNA-Seq. *Nat Methods* **5**: 621-628.

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.
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.

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.

Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F. 2018. MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. *Mol Biol Evol* **35**: 2582-2584.

Richards S, Liu Y, Bettencourt BR, Hradecky P, Letovsky S, Nielsen R, Thornton K, Hubisz MJ, Chen R, Meisel RP et al. 2005. Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution. *Genome Res* **15**: 1-18.

Rzhetsky A, Nei M. 1993. Theoretical foundation of the minimum-evolution method of phylogenetic inference. *Mol Biol Evol* **10**: 1073-1095.

Sambrook J, Russell DW. 2001. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, New York.

Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness. *Methods Mol Biol* **1962**: 227-245.

Simao 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.

Stanke M, Waack S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19 Suppl 2**: ii215-225.

Wagner GP, Kin K, Lynch VJ. 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci* **131**: 281-285.
Zhang J, Rosenberg HF, Nei M. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci USA* **95**: 3708-3713.