Edinburgh Research Explorer

Gene expression differentiation in the reproductive tissues of Drosophila willistoni subspecies and their hybrids

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Gene expression differentiation in the reproductive tissues of *Drosophila willistoni* subspecies and their hybrids

José M. Ranz[1] | **Alwyn C. Go[2]** | **Pablo M. González[3]** | **Bryan D. Clifton[1]** | **Suzanne Gomes[2]** | **Amirali Jaberyzadeh[1]** | **Amanda Woodbury[1]** | **Carolus Chan[1]** | **Kania A. Gandasetiawan[1]** | **Suvini Jayasekera[1]** | **Chelsea Gaudreau[2]** | **Hsiu-Ching Ma[1]** | **Victor M. Salceda[1,4]** | **Cei Abreu-Goodger[3]** | **Alberto Civetta[2]**

[1]Department of Ecology and Evolutionary Biology, University of California Irvine, California, USA

[2]Department of Biology, University of Winnipeg, Winnipeg, Manitoba, Canada

[3]Unidad de Genómica Avanzada (Langebio), CINVESTAV, Irapuato GTO, Mexico

[4]Departamento de Biología, Instituto Nacional de Investigaciones Nucleares, Ocoyoacac, Mexico

Correspondence
José M. Ranz, Department of Ecology and Evolutionary Biology, University of California Irvine, CA 92697, USA.
Email: jranz@uci.edu

Alberto Civetta, Department of Biology, University of Winnipeg, Winnipeg, MB R3B 2E9, Canada.
Email: a.civetta@uwinnipeg.ca

Present address
Pablo M. González, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK

Carolus Chan, Department of Integrative Biology, University of California Berkley, Berkley, California 94720, USA

Cei Abreu-Goodger, Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, UK

Funding information
National Science Foundation, Grant/Award Number: MCB1157876 and MCB2129845; Natural Sciences and Engineering Research Council of

Abstract

Early lineage diversification is central to understand what mutational events drive species divergence. Particularly, gene misregulation in interspecific hybrids can inform about what genes and pathways underlie hybrid dysfunction. In *Drosophila* hybrids, how regulatory evolution impacts different reproductive tissues remains understudied. Here, we generate a new genome assembly and annotation in *Drosophila willistoni* and analyse the patterns of transcriptome divergence between two allopatrically evolved *D. willistoni* subspecies, their male sterile and female fertile hybrid progeny across testis, male accessory gland, and ovary. Patterns of transcriptome divergence and modes of regulatory evolution were tissue-specific. Despite no indication for cell-type differences in hybrid testis, this tissue exhibited the largest magnitude of expression differentiation between subspecies and between parentals and hybrids. No evidence for anomalous dosage compensation in hybrid male tissues was detected nor was a differential role for the neo- and the ancestral arms of the *D. willistoni X* chromosome. Compared to the autosomes, the *X* chromosome appeared enriched for transgressively expressed genes in testis despite being the least differentiated in expression between subspecies. Evidence for fine genome clustering of transgressively expressed genes suggests a role of chromatin structure on hybrid gene misregulation. Lastly, transgressively expressed genes in the testis of the sterile male progeny were enriched for GO terms not typically associated with sperm function, instead hinting at anomalous development of the reproductive tissue. Our thorough tissue-level portrait of transcriptome differentiation between recently diverged *D. willistoni* subspecies and their hybrids provides a more nuanced view of early regulatory changes during speciation.

KEYWORDS
dosage compensation, hybrid dysfunction, neo-*X* chromosome, regulatory evolution, transcriptome divergence

1 | INTRODUCTION

A central goal in evolutionary biology is finding the genetic alterations underlying ecological, physiological, and morphological change (Lewontin, 1974). Mutations affecting regulation of gene expression can impact how the species blueprint becomes implemented molecularly during life cycle, across cell types, or under particular external cues, mediating phenotypic change during adaptation and species divergence (Glaser-Schmitt & Parsch, 2018; Hagen et al., 2019; Kratochwil et al., 2018). Early stages during lineage differentiation merit special attention as it is when the differential presence of genetic regulatory variants between two decreasingly connected gene pools can influence the consolidation of phenotypic differentiation and speciation. This differential presence of regulatory variants is ultimately reflected in diverging expression, resulting in some cases in disrupted regulatory networks, hybrid misexpression and dysfunction (Civetta, 2016; Cutter & Bundus, 2020; Mack & Nachman, 2017; Signor & Nuzhdin, 2018).

Different methodologies have been used to uncover the regulatory changes underlying changes in mRNA levels (Gilad et al., 2008; Wittkopp et al., 2004). Particularly, allele-specific expression (ASE) studies have revealed pervasive *trans*-regulatory variation within species, a more preeminent role of *cis*-regulatory changes in the interspecific evolution of mRNA levels, and normally larger size effects on expression differences in the case of *cis*- as opposed to *trans*-acting changes (Coolon et al., 2014; Emerson et al., 2010; Kopania et al., 2022; Zhang et al., 2011). Notably, the interpretation of results from ASE studies involving different species can be distorted by developmental defects that result in cellular tissue composition differences (Brawand et al., 2011; Good et al., 2010; Hunnicutt et al., 2022). Therefore, observations drawn from ASE studies using early divergent lineages in which cellular tissue composition differences are minimal are particularly informative (Mugal et al., 2020).

A key aspect of expression divergence is how lineage differentiation is distributed across the genome, in particular on the sexual chromosomes versus the autosomes. Different studies point towards a larger interspecific differentiation on the *X* (Brawand et al., 2011; Dean et al., 2015; Llopart, 2012; Meisel et al., 2012b), a pattern known as faster-*X* evolution, which can result from both neutral and adaptive evolution (Vicoso & Charlesworth, 2009). In *Drosophila*, adaptive expression evolution underlying the faster-*X* effect has been reported in embryos (Kayserili et al., 2012), for male-biased genes in adults (Llopart, 2012), and for tissue-restricted expressed genes also in adults (Meisel et al., 2012b). Importantly, the *X* chromosome contributes disproportionately to postzygotic reproductive isolation between diverging lineages (Coyne, 1992;

Haldane, 1922). This prominent role in hybrid dysfunction in general, and hybrid male sterility in particular, can have several nonmutually exclusive explanations (Coyne, 2018; Coyne & Orr, 2004), including *X*-linked gene misregulation, the disruption of the dosage compensation, and inactivation of the *X* chromosome during spermatogenesis (Good et al., 2010; Johnson & Lachance, 2012; Presgraves, 2008; Rodriguez et al., 2007). Nevertheless, the *X* chromosome does not always show enrichment for misregulated genes compared to the autosomes (Gomes & Civetta, 2015; Mugal et al., 2020). This picture might be compounded by the acquisition of a new *X* chromosome (aka neo-*X*), which will impact the patterns of molecular evolution of the resident genes. Depending on the time of origination of the neo-*X* relative to the ancestral *X* (anc-*X*) chromosome, the two chromosomes might play a different role in the divergence of expression profiles and regulatory networks as the properties of many of the genes harboured in the two chromosomes can substantially differ (Assis et al., 2012; Meisel et al., 2012b). How an anc- and a neo-*X* chromosome can impact expression differentiation across different reproductive tissues between recently diverged lineages and their hybrids remains largely unexplored.

The fly *Drosophila willistoni* has a metacentric *X* chromosome that resulted from the fusion between the chromosomal elements A and D of the ancestral *Drosophila* karyotype (Muller, 1940). Specifically, Muller's element A corresponds to the ancestral *X* chromosome in the genus *Drosophila* whereas Muller's element D, an autosome in most *Drosophila* species (Table S1), is an acquired neo-sex chromosome (Spassky & Dobzhansky, 1950). *D. willistoni* includes three allopatric subspecies, all of them found in the American continent: *D. willistoni willistoni*; *D. w. winge*; and *D. w. quechua* (Mardiros et al., 2016). The first two subspecies are not only indistinguishable at the morphological level, including fast evolving traits such as the male genitalia (Civetta & Gaudreau, 2015), but also in terms of sequence differentiation as revealed by the gene mtCOI (Mardiros et al., 2016). Although no fixed premating isolation has been found between these subspecies (Davis et al., 2020), they do exhibit unidirectional hybrid male sterility (Mardiros et al., 2016). Notably, and unlike other sterile hybrids between recently diverged *Drosophila* lineages that show sperm defects and reduced sperm motility (Gomes & Civetta, 2014; Kulathinal & Singh, 1998; Moehring et al., 2006), these sterile hybrid males produce typical numbers of motile sperm but cannot transfer them during copulation (Civetta & Gaudreau, 2015; Gomes & Civetta, 2014). This failed sperm transfer results from a bulging of the seminal vesicle, which hinders sperm migration towards the *vas deferens*, and therefore sperm incorporation into the ejaculate, which is otherwise successfully transferred to the

female (Davis et al., 2020). Collectively, the features of the *D. w. willistoni* and *D. w. winge* subspecies are largely suggestive of a very early stage in their process of consolidation as distinct species, offering an excellent opportunity to detect misregulation in their hybrids while discerning the role of different regulatory elements on expression divergence under minimal biases introduced by dissimilar cell type composition. Moreover, this subspecies pair is an ideal system to analyse the role of an anc- and neo-*X* chromosomes during early stages of transcriptome differentiation.

Here, we have constructed a new reference-quality genome assembly using long-sequencing reads and upgraded its annotation. With these new 'omic resources, we have characterized the transcriptome of three tissues relevant for reproductive functions (testis, male accessory gland, and ovary) using RNA-seq in *D. w. willistoni* and *D. w. winge*, aiming at: (i) uncovering transcriptome differentiation at the tissue-level between parental subspecies and their sterile F1 hybrid; (ii) assessing whether the tissue-specific expression patterns of the two arms of the *X* chromosome have converged while testing if this chromosome shows disrupted gene dosage compensation in testis; (iii) scrutinizing the types of transgressive expression across tissues between the sterile hybrid and the parental subspecies while determining how they are distributed in the genome; (iv) deciphering the architecture of regulatory evolution behind the differentiation of the subspecies by conducting ASE analyses; and (v) extracting functional interpretable patterns from transgressively expressed genes in interspecific hybrids to determine how gene functional attributes and protein–protein interaction networks might have been impacted early in speciation. The new omic resources generated, as well as the portrait of expression differentiation between the subspecies and their hybrids, make the *D. willistoni* subspecies system particularly appealing for gaining key insight into the interface between transcriptome evolution and the early stages of speciation.

2 | MATERIALS AND METHODS

2.1 | Fly husbandry

Strains from the National Drosophila Species Stock Centre used: *D. w. willistoni*, 14030-0811.24 (Guadeloupe Island) or standard; *D. w. willistoni*, 14030-0814.10 (Guadeloupe island); and *D. w. winge*, 14030-0811.16 (Rocha, Uruguay). Hybrids were generated by mating *D. w. willistoni* Guadeloupe females (14030-0814.10) with *D. w. winge* Uruguay males; in this cross, hybrid males are sterile (Mardiros et al., 2016). Vials were maintained at room temperature, under 24 h light conditions, and on standard corn meal-agar-dextrose medium. Adult flies were collected as newly emerged within 4 h to ensure virginity. Separate sets of flies were collected: 3–5 day-old virgin flies for setting crosses or for total RNA extraction with annotation purposes; 5 day-old virgin flies for tissue-level total RNA extraction utilized in differential expression assays; and 0–4 day-old individuals for DNA extraction to be utilized for genome sequencing.

2.2 | Genomic DNA extraction and de novo genome assembly construction

Genomic DNA from females of the standard strain was extracted for Illumina paired-end (PE; 150 bp) sequencing (Ranz et al., 2021). Library preparation was performed using the NEXTFLEX DNA kit (insert size = 403 bp), and sequencing was done over a quarter of a lane on an Illumina HiSeq 2500 instrument. Further, genomic DNA for Single Molecule Real-Time (SMRT) sequencing from PacBio technologies was prepared from 160 2 hr-starved males per column of the same strain (Ranz et al., 2021). Sheared genomic DNA was size selected from 20–80 kb long fragments using the BluePippin system. The library was sequenced using PacBio's P6-C4 chemistry across 17 cells. For assembly generation, we implemented a computational pipeline that integrates Illumina PE150 and PacBio sequencing reads (Ranz et al., 2021) (Supporting Information S1).

2.3 | Physical mapping

Fifty-two protein-coding genes were PCR-amplified (Table S2 for primer sequences), cloned, and mapped by in situ hybridization on polytene chromosomes of the standard strain (Ranz et al., 1997). Mapping information from eight additional genes was added (Chan et al., 2015).

2.4 | RNA sequencing

Tissue dissection (testis, male accessory gland, and ovary) was performed daily during 2 h windows in ice-cold 1 × PBS and always trying to randomize (e.g., the order of dissection of different genotypes) or minimize (e.g., the time of the day the dissection was started) any possible experimental error. Total RNA extraction and quality control were performed as reported (Clifton et al., 2017; Ranz et al., 2021). Twenty-seven ribodepleted, strand-specific libraries (3 biological replicates × 3 genotypes × 3 tissues) were prepared (Ranz et al., 2021). The resulting libraries were multiplexed and PE sequenced (100 bp) over one lane per tissue on an Illumina HiSeq 2500 instrument. An additional library was identically prepared using total RNA from whole bodies of naïve males and virgin females, and sequenced over one lane on the same instrument. All RNA and genome sequencing was performed at the UCI Genomics High-Throughput Facility.

2.5 | Repeat and gene annotation

Repeat content was predicted, combined with existing information in Diptera, and used to soft mask the genome assembly generated. Protein-coding gene models in the release dwil_r1.05 (FB2017_04) of FlyBase (dos Santos et al., 2015) were identified using Exonerate 2.47.3 (Slater & Birney, 2005) and upgraded by leveraging the

RNA-seq libraries generated. This was done using StringTie version 1.3.2d (Pertea et al., 2016) and PASA version 2.3.3 (Haas et al., 2003). For miRNAs, tRNAs, rRNAs and other types of short noncoding RNAs, we combined existing information from different databases with dedicated computational tools. Lastly, lncRNA identification required removing protein coding gene models from the StringTie merged predictions and using FEELnc (Wucher et al., 2017), demanding that lncRNA gene models should have associated >200 nt long transcripts, include ≥1 splicing junction, do not overlap with rRNA gene models, and be antisense if they overlapped protein-coding gene models (Text S2).

2.6 | Homology identification

The longest isoform per protein-coding gene was extracted with AGAT (Dainat et al., 2021), and fed into Orthofinder version 2.5.2 (Emms & Kelly, 2015) using the "diamon_ultra_sens" parameter to establish orthologous calls across protein sets from several drosophilids: *D. buscki*, *D. melanogaster*, *D. pseudoobscura*, *D. simulans*, and *D. yakuba* (Gonzalez-De-la-Rosa et al., 2021). Orthogroups with other drosophilids were identified independently for our gene predictions and previous annotation (dwil_r1.05).

2.7 | Differential gene expression analysis and tissue preferential expression

RNA-sequencing reads were examined for quality control and mapped against our de novo genome assembly with STAR (Dobin et al., 2013) (Supporting Information S3). Pairwise differential expression analysis across the parental subspecies and their hybrids was performed per tissue using both edgeR (Robinson et al., 2010) and DESeq2 (Love et al., 2014). A $\log_2$ fold-change threshold of 0.5 was further applied to increase the true positive rate (Schurch et al., 2016), and a 5% false-discovery rate (FDR) implemented (Benjamini & Hochberg, 1995). The intersect between the lists of differentially expressed genes from the two methods was used. All tools used for the differential gene expression analysis were ran on UseGalaxy (http://usegalaxy.org). Further, for deeming a gene as tissue-preferentially expressed, we used several criteria: a minimum threshold expression value; a minimum fold-change in expression between the tissues compared; and a minimum expression-specificity index ≥0.9 (Supporting Information S3).

2.8 | Allele-specific expression analysis to identify *cis*- and *trans*- regulatory incompatibilities

Fixed subspecies-specific single nucleotide polymorphisms (SNPs) were identified (Supporting Information S4) using Naïve variant caller followed by the Variant annotator (Blankenberg et al., 2014), and used to estimate their relative presence in hybrids. Criteria for the identification of fixed SNPs, procedures followed to measure allele specific expression in the hybrids, and the mode of regulatory evolution for each gene were determined according to the outcome across three statistical tests implemented (Fraser, 2019; Go & Civetta, 2020; Gomes & Civetta, 2015; McManus et al., 2010). Briefly, significant expression differences between the parental subspecies (P_{GUA} vs. P_{URU}) and between their alleles in the hybrid (H_{PGUA} vs. H_{PURU}) were tested using a binomial exact test. To keep consistency with the differential expression analysis, a 0.5 $\log_2$ fold-change threshold was applied in addition to the binomial exact test. To detect significant differences between the ratio of parental expression with the ratio of each parental allele in the hybrid (P_{GUA}/P_{URU} vs. H_{PGUA}/H_{PURU}), a Fisher's exact test was used. A 5% FDR was applied to the *p*-values resulting from both the binomial exact test and Fisher's exact test (Benjamini & Hochberg, 1995).

2.9 | Functional analyses

Protein–protein interactions among focal genes were predicted using STRING version 11.0 (Szklarczyk et al., 2019) under the medium and high confidence thresholds, and using all active interaction sources. Significant enrichment for gene ontology (GO) terms was determined with g:Profiler (Raudvere et al., 2019). In both cases, a 5% FDR threshold was applied (Benjamini & Hochberg, 1995).

2.10 | Statistical analyses

We used built-in statistical functions in R (R Development Core Team, 2016). In the case of chi-square tests for contingency tables, the null hypothesis was systematically tested by performing 2000 Monte Carlo simulations. In the case of the differential expression analyses performed, the lists of differentially expressed genes according to the combination of criteria indicated above resulted always to be more conservative than just the application of a joint 5% FDR across tissues (Supporting Information S5).

3 | RESULTS

3.1 | De novo genome assembly and annotation

We generated Illumina PE-150 and PacBio sequencing reads for the standard *D. w. willistoni* strain, resulting in a ~85× and ~65× coverage, respectively. Both sequencing outputs were used to generate a de novo genome assembly (dwil_UCI; Supporting Information S1 and Figure S1), which is 5% larger than the current reference assembly (dwil_caf1), with a substantially smaller gap size, fewer scaffolds, and twice the N50 value (Table 1). This enhanced contiguity is epitomized by the longest scaffold, which corresponds to the entire chromosome *3* (~34 Mb). Further, gene-level completeness was ascertained using Benchmarking universal single-copy orthologues (BUSCO)

TABLE 1 Salient features of an existing assembly for *D. willistoni* and the newly generated dwil_UCI.

Identifier	Assembly	
	dwil_caf1[a]	dwil_UCI
Total assembly size (bp)	235,516,348	248,362,691
Gap size (bp)	11,906,320	277
GC (%)	37.25	37.45
Number of contigs	20,358	758
Number of scaffolds	14,838	755
Scaffold N50 (bp)	4,511,350	9,374,611
Scaffold L50	15	7
Largest scaffold (bp)	16,660,200	33,854,152
BUSCO (n = 3285)[b]	C: 3222 (98.1%)	C: 3223 (98.1%)
	D: 23 (0.7%)	D: 26 (0.8%)
	F: 8 (0.2%)	F: 9 (0.3%)
	M: 32 (1.0%)	M: 27 (0.8%)

[a]GenBank Assembly Accession GCA_000005925.1 (Drosophila 12 Genomes et al., 2007; Zimin et al., 2008).

[b]BUSCO version 5.1.3 proteome results for diptera_odb10 (last accessed in July 2021). C, complete (uni and multicopy); D, duplicated or multicopy; F: fragmented; M: missing. The number of complete unicopy BUSCOs can be calculated as the difference between the total number of complete BUSCOs and the number of multicopy BUSCOs.

version 2.0.1 (Simao et al., 2015), recovering 98.1% (n = 3285) of Diptera genes as single copy and in complete form (Table 1). Internal assembly validation involved examining the relative order of two gene sets, including cytological mapping information, finding no disagreement (Supporting Information S6; Table S2; Figure S2). The new genome assembly was virtually complete, substantially more contiguous than the existing one, and highly reliable in terms of internal sequence order.

We annotated repetitive and low-complexity regions in the dwil_UCI assembly. In total, 102.5 Mb were populated by repeats, with 91.5 Mb (37.01%) corresponding to interspersed repeats. Our assembly includes 8.7% more repeats (41.5% vs. 32.8%, Table S3), particularly from the long terminal repeat (LTR) class (Table S3). This is expected as a more contiguous assembly is less likely to collapse repeat sequences compared to more fragmented assemblies. Further, we upgraded the models of protein-coding genes in the annotation dwil_r1.05 by incorporating RNA-seq data primarily from the reproductive tissues (testis, accessory gland, and ovary), which are the focus of our transcriptome comparative analysis (Supporting Information S7; Tables S4 and S5). The improvements affect many gene features, including an increase in the number of transcripts per gene and number and span of UTRs and exons, resulting in for example the correction of previously truncated gene models (Figure S3) and a more reliable quantification of expression levels in downstream analyses. The new annotation (dwil_UCI_r1.0) also increases substantially the number of lncRNAs, from 120 to 1679 (Table S4; Supporting Information S7). Overall, 28.51% of the genome (70.8 Mb) is transcribed into primary transcripts with

10.3% associated with mature transcripts, and coding sequences representing 8.88% of the genome. Lastly, we used protein-coding gene models from the two annotations to anchor the contigs of the new assembly to the Muller's chromosomal elements in *D. willistoni* (Supporting Information S8; Table S6).

3.2 | Transcriptome sequencing

We sequenced in triplicates the transcriptome of testis, accessory gland, and ovary of the parental subspecies *D. w. willistoni* and *D. w. winge*, and their F_1 hybrids, generating 335 million of RNA-seq reads (Table S7). We found no evidence of bias between the transcripts of the parental subspecies when mapped against the reference genome (Supporting Information S9; Table S8). Principal component analysis largely substantiated the expected grouping of samples (Figure S4), confirming the reliability of the sequencing results. Further, we found significant differences in the fraction of uniquely mapped reads across tissues (Kruskal-Wallis rank of sums, $p < .0001$), with ovary showing the highest numbers (nonparametric pair-wise Steel-Dwass test; ovary vs testis, $P_{adj} = .0012$; ovary vs accessory gland, $P_{adj} = .0012$; testis vs. accessory gland, $P_{adj} = .0012$). This could be due to an incomplete ribosomal depletion in the accessory gland and testis samples, an unanticipated biological difference across tissues, or both (Supporting Information S10; Figure S5).

3.3 | Tissue transcriptome differentiation between parental subspecies

The magnitude of transcriptome differentiation is heavily dependent on the functional constraints associated with different tissues (Brawand et al., 2011; Khaitovich et al., 2005). We tested this possibility between the subspecies *D. w. willistoni* and *D. w. winge*. We found 12,078 gene models expressed across tissues, with very similar numbers per tissue across subspecies and hybrids (Tables S9 and S10). Compared to accessory gland and ovary, testis expressed more genes, including those with tissue-preferential expression, and with a higher proportion of lncRNA genes (Supporting Information S11; Tables S11–S13; Figures S6 and S7). Twenty-four percent of the genes were differentially expressed between the parental subspecies (5% FDR; *Materials and Methods*), with testis harbouring more expression differences than accessory gland and ovary (18.85% vs. 6.28% vs. 9.70%, respectively; three-sample test for equality of proportions, $\chi^2 = 604.24$, d.f. = 2, $P < 2.2 \times 10^{-16}$; Table 2; Figure 1a). This pattern was consistent to both coding and lncRNA (3-sample test for equality of proportions; coding: $\chi^2 = 480.22$, d.f. = 2, $P_{adj} < 2.2 \times 10^{-16}$; lncRNA: $\chi^2 = 75.00$, d.f. = 2, $P_{adj} < 2.2 \times 10^{-16}$), and when the fold-change in expression was increased to 2, although not when increased to 4 (Tables S14–S16). Further, neither coding nor lncRNA genes showed any bias in their patterns of differential expression towards any particular parental subspecies across tissues (3-sample test for equality of proportions;

TABLE 2 Salient patterns from comparative expression analysis between *D. willistoni* subspecies.

Tissue and pattern	Gene category		
	Coding	Noncoding[a]	All
Testis	10,298	972 (932, 28, 4, 8)	11,270
Gua = Uru	8749	622 (589, 24, 1, 8)	9371 (83.15%)
Gua > Uru	775	214 (209, 2, 3, 0)	989 (8.78%)
Gua < Uru	774	136 (134, 2, 0, 0)	910 (8.07%)
Accessory Gland	8945	596 (548, 33, 1, 14)	9541
Gua = Uru	8437	505 (462, 29, 1, 13)	8942 (93.72%)
Gua > Uru	276	66 (63, 3, 0, 0)	342 (3.59%)
Gua < Uru	232	25 (23, 1, 0, 1)	257 (2.69%)
Ovary	7833	332 (324, 2, 0, 6)	8165
Gua = Uru	7139	234 (227, 1, 0, 6)	7373 (90.30%)
Gua > Uru	328	62 (61, 1, 0, 0)	390 (4.78%)
Gua < Uru	366	36 (36, 0, 0, 0)	402 (4.92%)
All 3 samples[b]	10,840	1238 (1184, 33, 5, 16)	12,078
Consistent pattern Gua = Uru	5686	118 (117, 1, 0, 0)	5804 (48.05%)
	5655	113 (113, 0, 0, 0)	5768 (99.38%)
Gua > Uru	12	4 (3, 1, 0, 0)	16 (0.28%)
Gua < Uru	19	1 (1, 0, 0, 0)	20 (0.34%)
Inconsistent pattern[c]	5154	1120 (1067, 32, 5, 16)	6274 (52.95%)

Note: Gua, *D. w. willstoni*; Uru, *D. w. winge*. Differentially expression was set at a 5% FDR and a 0.5-fold change in expression. Direction of the differential expression between the two subspecies: > overexpression, < underexpression.

[a]In parenthesis the number of lncRNAs, rRNAs, tRNAs, and snoRNA, respectively.

[b]Only genes expressed across the three types of biological samples.

[c]Genes that show differences in mRNA levels for at least one tissue in a given direction between the subspecies that are not observed in at least one other tissue.
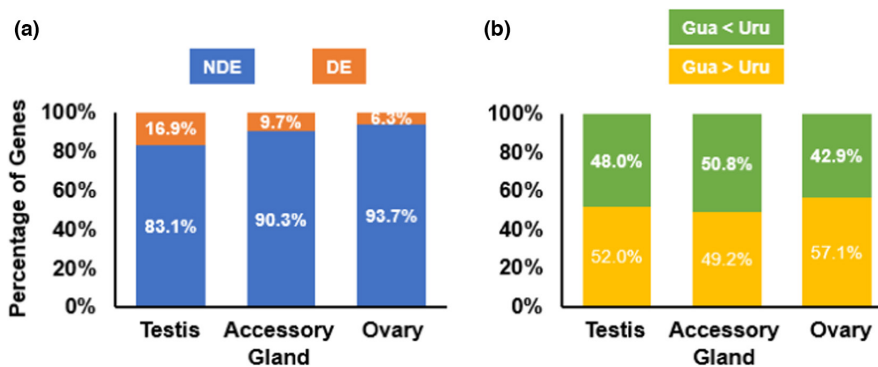

FIGURE 1 Relationship between tissue types and patterns of differential expression between the parental subspecies. (a) Percentage of nondifferentially (NDE) and differentially (DE) expressed genes between the parental subspecies. Testis harbour significantly more differentially expressed genes than accessory gland and ovary. (b) Percentage of differentially expressed genes relative to the subspecies in which they exhibit significantly higher (>) or lower (<) expression. Only coding and lncRNA genes were considered. Gua, *D. w. willstoni*; Uru, *D. w. winge*.

coding: $\chi^2 = 5.88$, d.f. $= 2$, $P_{adj} = 0.106$; lncRNA, $\chi^2 = 4.43$, d.f. $= 2$, $P_{adj} = 0.109$; Figure 1b). In terms of genomic distribution, and unlike other reports (Brawand et al., 2011; Llopart, 2012; Meisel et al., 2012b), we did not find evidence that the X chromosome was enriched for differentially expressed genes relative to the autosomes in any of the tissues assayed (Table S17). Notably, when considering the patterns of expression across all three tissues, we found only 36 (0.30%) differentially expressed genes displaying identical patterns between the parental subspecies. Most of the remaining differentially expressed genes (6274 or 52.95%) exhibited inconsistencies across tissues in the directionality of expression between the subspecies.

3.4 | Tissue-preferential expression and the demasculinization of the *X* chromosome

Previous studies in *Drosophila* documented an underrepresentation of male-biased –demasculinization– and an overrepresentation of female-biased –feminization– genes in expression on the *X* chromosome (Assis et al., 2012; Parisi et al., 2003; Sturgill et al., 2007). The *X* chromosome of *D. willistoni* is the result of a fusion between an anc-*X* (left arm or *XL*) and a neo-*X* (right arm or *XR*). As the pattern of masculinization and feminization of the gene content of the anc- and neo-sex chromosomes can differ in magnitude (Assis et al., 2012; Meisel et al., 2012b), *D. willistoni* offers the opportunity to test for such differences. In fact, previous analyses with this in mind were performed in this species (Meisel et al., 2012a), but involving anatomical sections (abdomen, thorax, and head), which are amalgams of different tissues. By using our tissue-specific data and more complete gene annotation, we revisited the extent to which the anc- and neo-*X* chromosomes of *D. willistoni* differ in their degree of feminization/demasculinization.

Different types of tissue-preferentially expressed genes in the parental subspecies showed very different chromosomal distributions ($\chi^2 = 93.59$, d.f. = 2, $P_{adj} = 3.3 \times 10^{-16}$; Table S18). Ovary-preferentially expressed genes are overrepresented on the anc-*X* but not on the neo-*X* (post hoc tests, $P_{adj} < .001$ and $P_{adj} > .05$, respectively; Table S18). Accessory gland-preferentially expressed genes were significantly underrepresented on the anc-*X* and overrepresented on the autosomes, particularly on the *2*R, as reported in other *Drosophila* species (Meiklejohn & Presgraves, 2012; Mikhaylova & Nurminsky, 2011). Testis-preferentially expressed genes were not significantly depleted from the anc-*X*, being also present according to random expectation on the neo-*X*. Grouping the Muller's elements based on whether they form part of the *X* chromosome or the autosomal complement did not change these conclusions nor did when more restricted fold changes in expression, four-fold instead of two-fold, were applied. Lastly, we re-examined our findings by considering lncRNA and protein-coding genes separately (Table S19). These two classes of genes did not necessarily show consistent patterns of chromosome distribution. For example, testis-preferentially expressed lncRNA genes were enriched in the anc-*X* while they were either underrepresented (*2*R) or found according to random expectation in the neo-*X* and the autosomes (*2*L and *3*). In contrast, testis-preferentially expressed coding genes do not show any chromosomal bias. Overall, these results highlight the very different dynamics between lncRNA and coding genes in *D. willistoni* while uncovering notable differences with previous observations in other *Drosophila* species (Supporting Information S12).

3.5 | Chromosome-level mechanisms of expression in the hybrid and parental subspecies

Whole-chromosome-based regulatory mechanisms such as incomplete dosage compensation and meiotic sex chromosome inactivation, if disrupted, can contribute to sterility in interspecific hybrids (Good et al., 2010; Johnson & Lachance, 2012; Presgraves, 2008; Rodriguez et al., 2007). To evaluate this possibility, we performed two analyses (Supporting Information S13). We first ascertained whether the X:A expression ratios was lower than 1 by calculating absolute median expression levels in testis and accessory gland (X:AA) as well as in ovary (XX:AA). We found significantly lower median expression levels in the two arms of the *X* chromosome relative to the autosomes in testis and accessory gland. In testis, the neo-*X* arm was expressed 28%–32% less, depending on the genotype, than the autosomes, with the accessory gland always showing a more moderate reduction in expression (11%–14%). The anc-*X* shows a similar significantly reduced expression relative to the autosomes in testis (27%–30%) as well as in accessory gland (20%–26%). Conversely, in ovary, we did not detect statistically significant differences in expression among the anc-*X*, the neo-*X*, and the autosomes (Figure 2; Figure S8). Overall, we found virtually the same patterns of interchromosomal expression across three minimum expression thresholds considered (Figures S9 and S10). This is suggestive of no overt difference in the degree of downregulation of the *X*-chromosome in the two tissues examined in sterile hybrid males in relation to the males of the parental subspecies.

Further, the analysis of the degree of expression equalization between the sexes, that is, that based on the comparison of ratios between the two male tissues to ovary for both the autosomes (AA:AA) and the two arms of the *X* (anc-*X*:anc-*X*anc-*X*; neo-*X*:neo-*X*neo-*X*), allowed to further assess the dose effect in the face of incomplete dosage compensation (Figures S11–S13) (Mank, 2013). Additionally, it also revealed that the anc-*X* displays a significantly lower expression ratio in both the comparison testis to ovary and that of accessory gland to ovary, regardless of the minimum expression threshold considered. For the neo-*X*, the pattern is similar in the comparison testis to ovary while it is dependent on the minimum expression threshold considered in the case of the comparison accessory gland to ovary. These results mirror the unequal distribution of tissue-preferentially expressed genes between the two arms (Table S18), further suggesting that the neo-*X* chromosome is in a delayed stage of demasculinization compared to the anc-*X* (Supporting Information S14).

3.6 | Patterns of differential expression in hybrids across tissues

As transgressive expression can be relevant for hybrid sterility (Brill et al., 2016; Civetta, 2016; Go & Civetta, 2020; Mack & Nachman, 2017; Moehring et al., 2007), we identified those genes that were either over or underexpressed in the hybrid in relation to both parental subspecies. In addition, we also identified genes showing additive expression patterns, that is, those expressed higher in the hybrid than in one of the parental subspecies but lower relative to the other (Figure 3a; Table S20).

Among the tissues assayed, testis (2.4%, 263/11,045) and ovary (2.8%, 228/8083) were particularly enriched for differentially
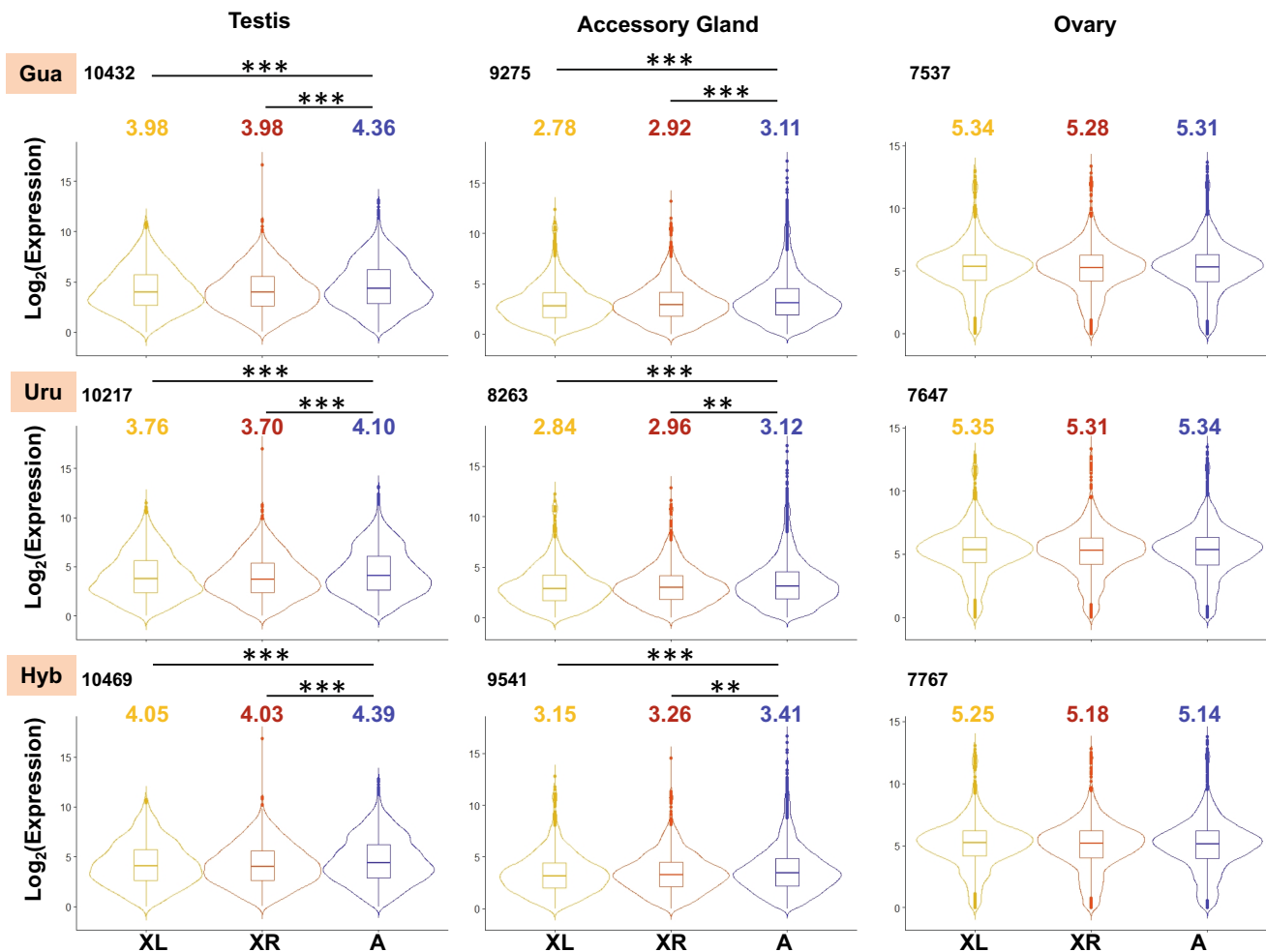
FIGURE 2 Absolute expression levels of the *XL*, *XR*, and the autosomes in three tissues and three genotypes of *D. willistoni*. The violin plots and nested box plots show the distribution of the $\log_2$(expression values) for genes expressed >1 TPM (transcripts-per-million). Autosomes are indicated in blue while the ancestral and neo-*X* chromosomes are in orange (*XL*) and red (*XR*), respectively. Medians are indicated on top of each plot. Statistical significance of differences between median values was determined using Kruskal-Wallis tests, followed by pairwise comparisons using Wilcoxon rank sum tests and correcting for multiple tests, which was followed by an additional correction that considered the number of tests performed (Benjamini & Hochberg, 1995). Solid lines indicate particular chromosomal contrasts entailing a statistically significant difference in expression: $P_{adj} < .05$ (*); $P_{adj} < .01$ (**); and $P_{adj} < .001$ (***). The number of genes considered is shown on the top left of each plot. Gua, *D. w. willistoni*; Uru, *D. w. winge*; Hyb, hybrid resulting from a Gua female and Uru male.

expressed genes relative to the total number of expressed genes in those tissues (accessory gland: 0.3%, 25/9330; 3-sample test for equality of proportions, $\chi^2 = 191.24$, d.f. $= 2$, $P < 2.2 \times 10^{-16}$). The accessory gland also departed from the gonadal tissues in showing the highest proportion of additive expression (64% vs. 33% -testis- and 46% -ovary-) and therefore the lowest proportion of transgressive expression, just the opposite pattern to that shown by testis and ovary (3-sample test for equality of proportions, $\chi^2 = 15.782$, d.f. $= 2$, $P < 3.7 \times 10^{-4}$). Further, we examined the degree of commonality in the identity of genes showing misexpression across tissues. Only 1% (3/305; FBgn0216635, FBgn0226669, FBgn0225891) of the genes showed transgressive expression in more than one tissue (Figure 3b), underscoring that transgressive expression in the hybrids examined is effectively tissue-dependent.

Next, we analysed whether transgressively expressed genes were preferentially expressed in particular tissues as, due to their narrower expression breadth, they are thought to be less pleiotropic and under weaker selective constraints, which might result in a higher rate of evolution (Assis et al., 2012; Khaitovich et al., 2005; Meisel, 2011). Under this assumption, we expect that tissue-preferentially expressed genes in the parental subspecies would be more likely to show transgressive expression in the hybrids. We did not find a prevalence of tissue-preferentially expressed genes among those transgressively expressed, with testis showing even a lower proportion than ovary (Supporting Information S15 and Table S20). Further, to discard any transgressive expression in testis as the result of differential cell-type composition in the hybrid (Good et al., 2010; Mugal et al., 2020), we examined the expression levels of 10 known cell types marker genes (Witt et al., 2019; Witt, Shao, et al., 2021) across the hybrid and parental subspecies. None of the genes shows statistically significant differential expression between the parental subspecies (Figure S14). More importantly, only one
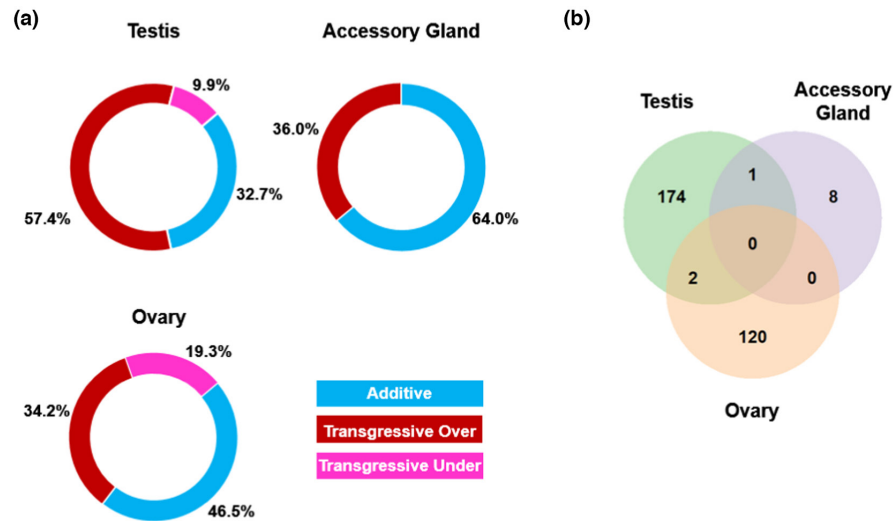
FIGURE 3 Patterns of differential expression in hybrids relative to the parental subspecies of *D. willistoni*. (a) Doughnut charts showing the percentage of genes showing different patterns of differential expression across the three tissues assayed. Additive, when hybrid gene expression falls within the ranges of the expression levels of the parental subspecies; transgressive over and transgressive under, when hybrid gene expression is above or below the expression levels of the parental subspecies, respectively. (b) Venn diagram showing the level of overlap among transgressively expressed genes in the hybrids across the tissues assayed.

TABLE 3 Chromosomal distribution of the differentially expressed genes in the hybrids in relation to the parental subspecies.

Tissue (# genes)	Differential expression pattern	XL	XR	Autosomes
Testis (262)	Additive	$0 \downarrow$ ($<1.0 \times 10^{-5}$)	$2 \downarrow$ (0.002)	$83 \uparrow$ ($<1.0 \times 10^{-5}$)
	Transgressive	$30 \uparrow$ ($<1.0 \times 10^{-5}$)	$31 \uparrow$ (0.002)	$116 \downarrow$ ($<1.0 \times 10^{-5}$)
		$\chi^2 = 32.68, P_{adj} = 1.5 \times 10^{-3}$		
Accessory gland (25)	Additive	0	1	15
	Transgressive	0	3	6
		G-test $= 3.05, P_{adj} = 0.327$ [a]†		
Ovary (220)	Additive	18	16	65
	Transgressive	20	17	84
		$\chi^2 = 0.36, P_{adj} = 0.841$		

Note: P_{adj} values are provided in parenthesis next to the trend symbol; ↑, enrichment; ↓, depletion; ns, no deviation in relation to the random expectation. P_{adj} values correspond to the implementation of the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995).

[a]As the chi-square test cannot be computed with zero marginals, a G-test of independence was performed instead.

marker gene (*Fas3*; preferentially expressed in the hub cells) was differentially expressed in the sterile hybrids relative to the parental subspecies (transgressive-over). Collectively, these results are suggestive of a testis cell-type composition largely similar between the sterile hybrids and the parental subspecies.

3.7 | Genomic distribution of hybrid differential expression

The disproportionate role of the X chromosome to hybrid disfunction through gene misregulation has been documented in some but not all cases (Gomes & Civetta, 2015; Llopart, 2012; Masly & Presgraves, 2007; Mugal et al., 2020). We examined whether

differentially expressed genes in the hybrids relative to the parental subspecies were equally abundant on the X chromosome and autosomes, and whether there was any difference among tissues. For 493 genes with reliably determined chromosomal location, we found an uneven distribution of the additive and transgressive patterns in testis but not in ovary or accessory gland, in this last case possibly because of limited statistical power (Table 3). In testis, the two arms of the X chromosome seemed enriched for transgressively expressed genes, while the autosomes were enriched for additively expressed genes ($\chi^2 = 32.68, P_{adj} = 1.5 \times 10^{-3}$). The two types of transgressively expressed genes were not found to be differentially associated at the chromosomal level ($\chi^2 = 0.778, p = .752$).

We also sought for evidence of clustering of transgressively expressed genes at a finer scale, which could denote misregulation

of a regulatory element of regional effect or differential chromatin-based regulation (Degner et al., 2012; McVicker et al., 2013). We used three arbitrary cutoff values (5, 10, and 15 kb) to determine the share of genes dubbed as clustered, as well as the number of clusters. We found abundant clustering at the three cutoff values (Table S21). Notably, 12 clusters remained detected across cutoff values, seven of them being relevant for ovary, and another five for testis. When examined these clusters for the presence of gene models belonging to the same orthogroup, we found that two of the clusters relevant for ovary included paralogues.

3.8 | Identification of genome-wide regulatory mismatches in interspecific hybrids

Divergent regulatory elements in a hybrid genome can cause differential expression associated with the different effects of the parental alleles (Coolon et al., 2015; Landry et al., 2005; McManus et al., 2010; Wittkopp et al., 2004). To determine the extent of such regulatory mismatches, we used putatively fixed SNPs between the parental subspecies to identify allele-specific gene expression in the hybrids. Upon identifying usable SNPs for the three tissues assayed (Table S22), we inferred modes of regulatory divergence (*Materials and methods*; Table 4). As this requires the presence of both parental alleles in the hybrid, the analysis could only be performed on autosomal genes for testis and accessory gland. For ovary, the analysis was done separately for the autosomal and *X*-linked genes. Altogether, we could investigate the mode of regulatory divergence for 1626 genes in testis, for 1821 in accessory gland, and for 6384 in ovary.

Regardless of the tissue, 81%–88% of the analysable genes showed no evidence of regulatory divergence (Table 4). Of the remaining genes, regulatory divergence in the gonads (i.e., testis and ovary) appear to be more significantly impacted by *cis*- than by *trans*-changes, with no significant differences being observed in accessory gland (Table S23). Notably, 89.8% of the genes analysable in at least two tissues showed consistency in their mode of regulatory evolution, predominantly in association with the conserved category. Further, we detected very little evidence of compensatory evolution as well as very limited presence of artefactual significant negative correlations between *cis* and *trans* effects (Supporting Information S16; Table S24; Fraser, 2019; Zhang & Emerson, 2019). We then analysed the relationship between the mode of regulatory evolution and transgressive expression in the hybrids, which was possible for testis and ovary but not for accessory gland due to the limited number of genes showing transgressive and additive expression patterns (five and four genes, respectively). For both tissues, we found evidence of nonrandom associations between particular modes of regulatory evolution and patterns of differential

TABLE 4 Modes of regulatory divergence between the parental *D. willistoni* subspecies.

Regulatory category[a]	Testis		Accessory gland		Ovary	
	N	Percent	N	Percent	N	Percent
Ambiguous	224	13.8%	171	9.4%	433	10.8%
Compensatory	1	0.1%	2	0.1%	18	0.4%
cis & trans	2	0.1%	3	0.2%	21	0.5%
trans only	6	0.4%	17	0.9%	61	1.5%
cis only	53	3.3%	25	1.4%	218	5.4%
Conserved	1340	82.4%	1603	88.0%	3266	81.3%
Total	1626		1821		218	

[a]Different modes of regulatory divergence as inferred from the allele specific expression analysis, which entails the comparison of expression levels between the parental subspecies (P_{GUA} vs. P_{URU}), the comparison of their alleles in the hybrid progeny (H_{PGUA} vs. H_{PURU}), and the comparison of the expression ratios between the parental alleles in parental individuals and in the hybrids (P_{GUA}/P_{URU} vs. H_{PGUA}/H_{PURU}). Six modes of regulatory divergence are considered. Conserved refers to genes for which there is no evidence of regulatory evolution in either *cis* or *trans*. No significant differential expression is observed between parental subspecies ($P_{GUA}=P_{URU}$), between their alleles in the hybrids ($H_{PGUA}=H_{PURU}$), and between the ratio of parental alleles expression and the ratio between the parental alleles in the hybrids ($P_{GUA}/P_{URU}=H_{PGUA}/H_{PURU}$). *Cis* only refers to genes whose regulatory evolution is driven by genetic differences in their *cis* regulatory elements. Significant differential expression is detected between the parental subspecies and between their alleles in the hybrid but no significant differences are detected between the ratio of parental alleles expression and the ratio of the parental alleles in the hybrid. *Trans* only refers to genes whose expression evolution is driven by changes affecting the trans-acting factors that regulate them. Significant differential expression is detected between the parental subspecies but not between the parental alleles in the hybrid. Significant differences between the ratio of parental allele expression and that of the parental alleles in the hybrid are detected. *Cis & trans* refers to genes that are evolving at the regulatory level through changes at both levels. Significant differences in expression are detected between parental subspecies, in the expression of the parental alleles in the hybrid, and between the ratio of parental allele expression and the ratio of the parental alleles in the hybrid. Compensatory refers to genes that show no significant expression differences between the parental subspecies despite having evidence of both *cis* and *trans* divergence. Regulatory divergence in both *cis*- and *trans*- perfectly compensate for each other, resulting in no significant differential expression between the parental subspecies. Significant differences in expression between alleles in the hybrid and between the ratio of parental expression and the ratio of alleles in the hybrid are nevertheless detected. Ambiguous refers to genes for which the expression patterns observed in the parental subspecies and the hybrid do not fall into any of the above patterns. Since males are hemizygous for the X chromosome, only autosomal genes were available for analysis in testis and accessory gland samples. For ovary, both autosomal and X-linked are analysable (Table S22). Only comparable results (i.e., autosomal genes) across tissues are shown.

expression. Bootstrapping showed that in ovary, but not testis, genes experiencing *cis*-regulatory divergence exhibit the most significant enrichment among those showing additive expression (Figure 4). In contrast, genes displaying transgressive expression, both in ovary and testis, were primarily characterized by conserved regulatory evolution (Figure 4).

Lastly, we tested the expectation of a comparatively larger effect on expression of *cis* relative to *trans* changes due to the presumed more limited detrimental pleiotropic effects of the former (Coolon et al., 2014; Hill et al., 2021; Kopania et al., 2022). For testis, the size effect of the expression difference induced by *cis* changes is significantly larger than that of *trans* changes, while for accessory gland and ovaries we found no significant difference (Wilcoxon rank sum test: testis, $P_{adj}=0.0429$; accessory gland, $P_{adj}=0.536$; ovary, $P_{adj}=0.536$; Table S25). This lack of consistency across tissues might reflect differences in the strength of purifying selection operating on them. Collectively, and despite the prevalence of the conserved mode of regulatory evolution across the tissues of these two *D. willistoni* subspecies, subtle differences in the role and magnitude effects of *cis*- and *trans*- regulatory changes, as well as different associations to additive and transgressive expression across tissues, highlight the importance of the tissue context to properly uncover patterns of regulatory evolution.

3.9 | Functional patterns of enrichments among differentially expressed genes in interspecific hybrids

C*is*- and *trans*-regulatory incompatibilities in a hybrid background could lead to a cascade of differential expression (Go & Civetta, 2020), being preferentially associated with particular functional rubrics or part of a shared network of interactions. We examined these two features (*Materials and methods*) assuming functional

conservation between *D. willistoni* and *D. melanogaster* for the genes examined.

For accessory gland, neither significant protein–protein interaction (PPI) network nor overrepresentation of biological process were detected. In ovary, 55 out of the 228 differentially expressed genes in the hybrids were predicted to have more protein–protein interactions than randomly expected (PPI enrichment $P_{adj}=4.34\times10^{-13}$). In addition, this set of genes were disproportionally enriched for the GO biological process term *chorion-containing eggshell formation* ($P_{adj}=0.0307$). In testis, we also found evidence for a significant PPI network involving 64 out of the 263 differentially expressed genes (PPI enrichment $P_{adj}=8.11\times10^{-9}$), and up to 12 functionally enriched biological process terms, including *cell adhesion* ($P_{adj}=0.0307$), which has been also found enriched in the transcriptome analysis of sterile male hybrids between two other very closely related *Drosophila* subspecies (Go et al., 2019; Gomes & Civetta, 2015) (Table 5; Figures S15–S16; Supporting Information S17). Restricting these analyses to the 122 and 177 transgressively expressed genes in ovary and testis, respectively, does not essentially alter the patterns of enrichment found for biological processes, and many shared interaction networks remained (Figure 5; Tables 5 and S26). A qualitative difference between these most restricted interaction networks in ovary and testis is the prevalence of transgressive overexpression in the latter. Notably, in the case of transgressively expressed genes that are part of PPI networks, we found that seven of them were part of three physical gene clusters.

4 | DISCUSSION

We have comprehensively compared the transcriptome of two *D. willistoni* subspecies, *D. w. willistoni* and *D. w. winge*, and the hybrids resulting from the cross between them that yields sterile males
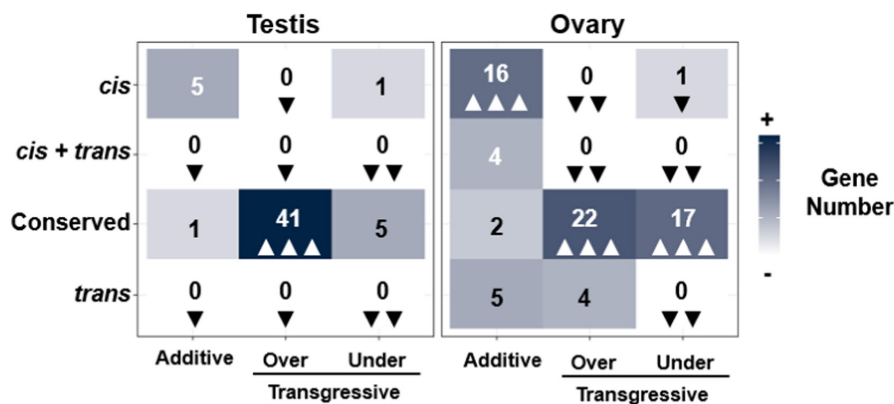

FIGURE 4 Nonrandom association between regulatory divergence between subspecies and differential expression in the hybrid in relation to those subspecies. Heat maps for 53 and 71 genes differentially expressed in testis and ovary between the hybrid and the parental subspecies, and in relation to different types of regulatory evolution, are shown. The category of ambiguous expression divergence is excluded, and those corresponding to "compensatory" and "*cis* and *trans*" are combined and labelled as "*cis* + *trans*". Bootstrapping ($n=1\times10^{5}$) was used to assess departures from the random expectation: $p<.05$, one arrowhead; $p<.01$, two arrowheads; and $p<.001$, three arrowheads. The directionality of the arrow denotes whether the significant association represents enrichment or depletion.

TABLE 5 Patterns of enrichment of gene ontology terms among genes differentially expressed in hybrids in relation to the parental subspecies.

Term id	Gene ontology: Biological processes	All[a]		Only transgressive	
		Gene count	P_{adj}[b]	Gene count	P_{adj}[b]
Testis					
GO:0120192	Tight junction assembly	7 of 31	0.0077	7 of 31	3.70E-04
GO:0010466	Negative regulation of peptidase activity	8 of 73	0.0307		
GO:0007155	Cell adhesion	10 of 139	0.0307	9 of 139	0.0154
GO:0009605	Response to external stimulus	26 of 660	0.0307	21 of 660	0.0154
GO:0009611	Response to wounding	8 of 92	0.0307		
GO:0022610	Biological adhesion	10 of 139	0.0307	9 of 139	0.0154
GO:0035001	Dorsal trunk growth; open tracheal system	3 of 7	0.0307	3 of 7	0.0154
GO:0042060	Wound healing	7 of 69	0.0307	6 of 69	0.0154
GO:0060039	Pericardium development	2 of 2	0.0307	2 of 2	0.0154
GO:0009617	Response to bacterium	9 of 120	0.0307		
GO:0022404	Moulting cycle process	6 of 53	0.0307	5 of 53	0.0227
GO:0007229	Integrin-mediated signalling pathway	3 of 10	0.0405	3 of 10	0.0154
Ovary					
GO:0007304	Chorion-containing eggshell formation	8 of 75	0.0307	6 of 75	0.0189

[b]As determined by g:Profiler (Raudvere et al., 2019); redundant terms were excluded using REVIGO (Supek et al., 2011). P_{adj} values correspond to the implementation of the Benjamini-Hochberg correction to the number of terms and tissues considered (Benjamini & Hochberg, 1995).
[a]It includes both transgressive and additive patterns of differential expression in the hybrid relative to the parental subspecies.
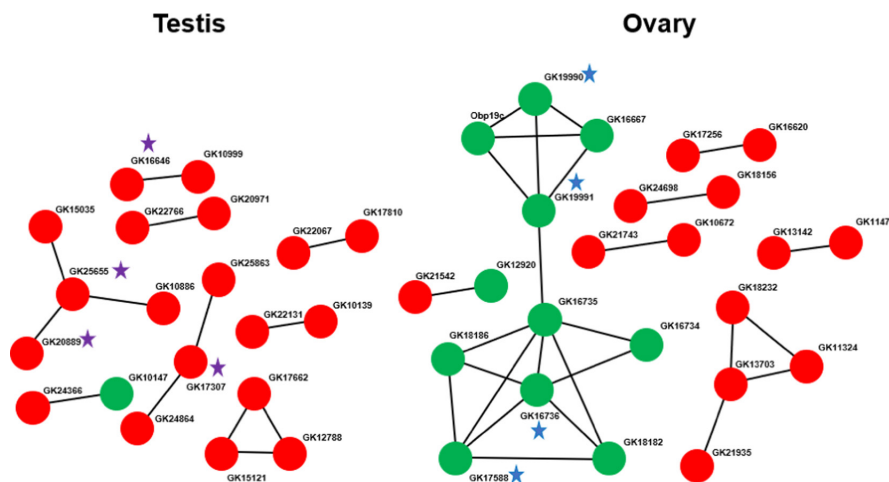

FIGURE 5 STRING protein–protein interactions (PPI) networks for transgressively expressed genes in the testis and ovary of hybrids relative to the parental subspecies. Lines between nodes denote interacting proteins. Only high-confidence interactions are considered. PPI enrichment: testis, $P_{adj} = 4.44 \times 10^{-08}$; ovary, $P_{adj} < 3 \times 10^{-16}$. A colour code identifies different types of differential expression: red, overexpression; green, underexpression. Purple stars, *cell adhesion* related genes. Blue stars, *chorion-containing eggshell formation* related genes.

(Davis et al., 2020; Mardiros et al., 2016). This unidirectional hybrid male sterility plus a negligible cell-type tissue composition differentiation in the sterile hybrid make this subspecies pair well poised to address how regulation in gene expression evolves during the early stages of speciation across tissues, resulting in hybrid dysfunction. Further, we have generated a more contiguous genome assembly, which should facilitate the study of the rampant structural variation in the *D. willistoni* species group (Rohde & Valente, 2012), and an

upgraded annotation with more reliable gene models and a substantially higher number of lncRNA genes, which are relevant both in transcriptional regulation and phenotypic change (Wen et al., 2016; Zhu et al., 2021).

4.1 | Expression profile differences across tissues in *D. willistoni*

Not many studies have focused on the early lineage divergence, at the onset of hybrid breakdown, comparing gene expression across tissues (Mugal et al., 2020). Prior comparative expression analyses among male-, female-, and nonbiased genes have shown that the former category evolves the fastest in *Drosophila* (Llopart, 2012; Ranz et al., 2003). Here, we also found that testis exhibit not only a higher fraction of differentially expressed genes between subspecies than ovary but also in relation to the male accessory gland, which although somatic, it is also part of the reproductive system. Roughly, one fourth of all the genes consistently expressed across tissues are differentially expressed between *D. w. willistoni* and *D. w. winge* in at least one of the tissues assayed. Notably, the degree of consistency in relation to the identity of the genes that are differentially expressed between the subspecies was minimal, stressing the need of tissue-level characterization even during early lineage differentiation.

4.2 | The *X* chromosome arms show very distinctive patterns of gene content differentiation

Early results pointed to the demasculinization of the gene content of the *X* chromosome relative to the autosomes (Assis et al., 2012; Parisi et al., 2003; Sturgill et al., 2007). A key explanation is that the *X* chromosome spends more time in females than in males, which should result in an overrepresentation of genes beneficial for females at the expenses of males (Rice, 1984; Vicoso & Charlesworth, 2006). Our results reveal that the *X* chromosome of *D. willistoni* is not depleted of testis-preferentially expressed genes. In fact, when expression breadth is considered, we find no evidence of depletion of testis-specific genes on the *X* chromosome as previously shown in *D. melanogaster* (Meiklejohn & Presgraves, 2012; Meisel et al., 2012a). Further, when considering lncRNA and protein-coding genes separately, the former are overrepresented on the anc-*X* while the second follow the random expectation. This differs from observations on the *X* of *D. pseudoobscura*, species in which lncRNAs follow the random expectation (Nyberg & Machado, 2016). The reason for this interspecific difference is not apparent currently.

The anc- and neo-*X* chromosomes of *D. willistoni* exhibit additional differences. The anc-*X* shows depletion of accessory gland-preferentially expressed genes (demasculinization) and an overrepresentation of ovary-preferentially expressed genes (feminization). These findings agree with previous observations on

the *X* of *D. melanogaster* (Meiklejohn & Presgraves, 2012; Meisel et al., 2012a). The neo-*X* chromosome does not deviate from the random expectation for these two gene classes. Collectively, the anc- and neo-*X* chromosomes of *D. willistoni* are characterized by marked differences in gene content composition, consistent with a neo-*X* chromosome that has not yet converged functionally with the anc-*X* chromosome, similarly to *D. pseudoobscura* (Assis et al., 2012).

4.3 | No evidence of disrupted dosage compensation in male sterile hybrids

In several *Drosophila* species, lower *X*-to-autosome expression ratio in testis and accessory gland has been reported (Assis et al., 2012; Meiklejohn et al., 2011; Meisel et al., 2012a; Parisi et al., 2003) while a higher *X*-to-autosome expression ratio has been found in *D. melanogaster* ovary (Assis et al., 2012; Parisi et al., 2003) although not in *D. pseudoobscura* (Assis et al., 2012). The patterns detected here for *D. willistoni* mirrored those observed in *D. pseudoobscura* for all the tissues and genotypes assayed.

In the case of testis, a lower *X*-to-autosome expression ratio could result from gene silencing on the *X* chromosome as spermatogenesis progresses (Mahadevaraju et al., 2021), lack of dosage compensation (Meiklejohn et al., 2011), or demasculinization of its gene content due to intralocus sexual antagonism (Charlesworth et al., 1987; Rice, 1984; Vicoso & Charlesworth, 2006). This last mechanism though does not seem to be that relevant in *D. willistoni* as we do not find a depletion of testis-preferentially expressed genes on the *X*. Assuming similar mechanisms between *D. willistoni* and *D. melanogaster*, canonical dosage compensation could be operating in spermatogonia and whole *X*-chromosome silencing, or incomplete noncanonical dosage compensation, could be operating in spermatocytes (Gupta et al., 2006; Mahadevaraju et al., 2021). Notably, the reduced gene expression on the *X* chromosome in *D. willistoni* testis is not as pronounced as in *D. melanogaster* (Meiklejohn & Presgraves, 2012). While the reasons for this interspecific discrepancy remain unclear, both in hybrids and parental subspecies, our results support an overall downregulation of the *X* chromosome during spermatogenesis –although not to the extent of *D. melanogaster*– and a reduced *X* chromosome compensation in male accessory gland similarly to *D. melanogaster* (Meisel et al., 2012a). Therefore, dosage compensation mechanism is not disrupted in reproductive tissues of *D. willistoni* sterile hybrid males.

4.4 | Distinct patterns of hybrid misexpression across tissues and gene ontologies

Transgressive expression varied substantially across tissues in *D. willistoni* hybrids. Testis exhibited the largest proportion of transgressively expressed genes. Interestingly, testis showed the lowest proportion of tissue-preferentially expressed genes among those

with transgressive expression. While spermatogenesis genes are commonly misregulated in sterile hybrids (Larson et al., 2022; Li et al., 2016; Moehring et al., 2007), we did not find them overrepresented among misregulated genes in hybrids. This result agrees well with the high numbers of motile sperm produced by the sterile hybrid males (Gomes & Civetta, 2014).

The pattern of gene ontology enrichment among the transgressively expressed testis genes did not reveal an overt relationship with testis function. Instead, we found several GO terms related to different aspects of cell adhesion, which is important for the proper assembly of the muscle sheaths in the *D. melanogaster* male reproductive system (Kuckwa et al., 2016; Susic-Jung et al., 2012). The low proportion of misregulated accessory gland genes in hybrid males aligns with the absence of differences in seminal fluid composition of the sterile males, as these are able to induce the typical changes in uterus morphology (Davis et al., 2020). Lastly, in ovary, chorion-containing eggshell formation genes were over-represented among those showing transgressive expression, possibly reflecting a fast regulatory divergence between subspecies, similarly to their evolutionary rate at the protein sequence level (Jagadeeshan & Singh, 2007).

4.5 | Evidence of a faster-*X* effect on hybrid differential expression

Prior expression studies in hybrids of the *D. melanogaster* species group, as well in hybrids between *D. pseudoobscura* subspecies, found fewer than expected *X*-linked chromosome genes misexpressed (Gomes & Civetta, 2015; Moehring et al., 2007). In contrast, sterile male hybrids in mice showed predominant overexpression of *X*-linked spermatogenesis genes (Good et al., 2010) and sterile male hybrids between closely related species of flycatchers had limited over and under-expression of *Z*-linked testis expressed genes (Mugal et al., 2020). We found a greater proportion of *X*-linked genes with transgressive expression in hybrids for testis than for other tissues.

It is unclear why *X*-linked testis-expressed genes show enrichment for transgressive expression in hybrids, but previous work has shown faster *X* evolution effects for male-biased and tissue-restricted genes in expression in adults (Llopart, 2012; Meisel et al., 2012b). Differences in the number of regulatory elements affecting gene expression in testis could explain differences in rate of evolution relative to other tissues. In fact, gene expression in ovary is more heavily dependent on *trans*-regulatory factors than in testis, which relies additionally on open promoter chromatin state (Witt, Svetec, et al., 2021). Faster rate of change in *X*-linked regulatory elements that preferentially regulate *X*-linked genes could cause a higher expression divergence of these genes relative to those on the autosomes (Coolon et al., 2015). However, we do not find the *X* chromosome enriched relative to the autosomes for differentially expressed genes between the subspecies, which argues against an elevated presence of *cis*-regulatory changes within the *X* chromosome. Divergence of autosomal *trans*-acting factors preferentially regulating the expression of *X*-linked genes could therefore explain

our observations. Interestingly, the contribution of *trans*-regulatory changes to faster-*X* expression changes is more common within than between species in the *D. melanogaster* species group (Coolon et al., 2015), which suggests that the two *D. willistoni* subspecies analysed here mimic more closely intraspecific patterns of regulatory differentiation. Lineage-specific effects across phylogenies, developmental or age stage-specific sample effects, and whole individuals vs. tissue sampling can explain discrepancies in results among different sets of species.

Notably, we find evidence of clusters of misexpressed genes in hybrids for testis and ovary, which is suggestive of an incipient subspecies differentiation at the chromatin structure level, impacting sets of nearby genes. Although tandem duplicates are more frequently colocalized on the same topologically chromatin domain (Ibn-Salem et al., 2017; Makova & Li, 2003), we found very little evidence for their presence in the clusters identified.

4.6 | *Cis*-regulatory changes of variable size effects across tissues prevail during incipient speciation stages

Previous analysis on the tempo of accumulation of *cis*- and *trans*-acting regulatory changes during lineage divergence supported the prevalence of the former between species (Coolon et al., 2014; Gomes & Civetta, 2015; Metzger et al., 2016; Tirosh et al., 2009). Our results in testis and ovary agree well with this observation; for accessory glands, we did not find significant differences in the occurrence of *cis* and *trans* changes. Previous work has suggested that expression changes involving diffusible *trans*-regulatory factors encoding genes could be buffered by selection due to possible pleiotropic effects (Lemos et al., 2008; Prud'homme et al., 2007). Further, genes expressed in testis and diverging in *cis* showed a larger average size effect on expression differentiation between *D. willistoni* subspecies than genes diverging in *trans*, which is compatible with more intense selective constrains acting on *trans*-regulatory changes. This difference was not observed in ovary and accessory gland. Overall, the preponderance of *cis* relative to *trans* changes for testis and ovary but not for accessory gland, and the different size effect that these types of changes have on expression differences between the subspecies, highlight how regulatory evolution differs across reproductive tissues. The lack of constrain observed in the accumulation of *trans*- relative to *cis*- regulatory changes in the accessory gland agrees well the majority of seminal fluid protein encoding genes evolving interspecifically under relaxed selection (Patlar et al., 2021).

A common outcome in several ASE studies has been the prevalence of compensatory changes, that is, mutations in *cis* and *trans* that differ in the direction of their effect on gene expression, thus resulting in no difference between the populations or species compared (Fear et al., 2016; Goncalves et al., 2012; Mack et al., 2016; Tirosh et al., 2009). This type of changes is expected under stabilizing selection, which preserves expression levels despite regulatory

differentiation (Signor & Nuzhdin, 2018). Nevertheless, the role of compensatory evolution in ASE assays might have been overstated (Fraser, 2019; Zhang & Emerson, 2019) (but see Signor & Nuzhdin, 2019). Between the *D. willistoni* subspecies, we found scarce evidence of compensatory changes, which is compatible with the limited divergence time between the subspecies as well as with the high proportion of conserved genes found at the regulatory level. This differs from other intraspecific analyses in mice (Goncalves et al., 2012), *D. melanogaster* (Fear et al., 2016), and during cotton domestication (Bao et al., 2019). Differences in the rate of deleterious mutations or in the efficacy of selection to eliminate them between *D. willistoni* and those other species could contribute to the observed difference (Andolfatto et al., 2011).

4.7 | Limitations and future directions

Our work highlights the need of investigating tissue-pattern expression when characterizing the transcriptome of interspecific hybrids while controlling for cell-type composition differences. Evidence for such differences should prompt the adoption of methodological approaches such as single-cell RNA-seq (Hunnicutt et al., 2022). Further, our study is not exempt of limitations. Both the expression changes and nucleotide differences used in the ASE analysis are not necessarily fixed, as we included one line per subspecies. Additionally, we were prone to overestimate the true divergence. Conversely, the combination of working with early diverged lineages and only one line per subspecies can also result in genes that harbour no informative difference, limiting the power of some analyses. Lastly, the genome sequencing of these two subspecies will allow to directly analyse nucleotide changes impacting regulatory elements as well as to identify structural changes responsible for expression differences unrelated to regulatory evolution. Such 'omics resources, plus the ones generated here, will enhance the study of speciation in the *D. willistoni* subgroup and across the *Drosophila* genus.

CONFLICT OF INTEREST STATEMENT
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT
All raw sequencing data have been deposited as part of the NCBI BioProject PRJNA670571. The genomic assembly UCI-dwil_1.1 is deposited at DDBJ/ENA/GenBank under the accession number GCA_018902025.1. The annotation file for the sequenced assembly is available at Zenodo (Gonzalez-De-la-Rosa et al., 2021) (https://doi.org/10.5281/zenodo.6557588). Additional description of the scripts used can be found at https://github.com/pgonzale60/Nonmodel-Species/tree/main and Supporting Information S1–S4.

BENEFIT-SHARING STATEMENT
Benefits from this research accrue from the sharing of our data and results on public databases as described above.

ORCID
José M. Ranz https://orcid.org/0000-0003-3585-3129
Alberto Civetta https://orcid.org/0000-0002-9122-5621

REFERENCES
Andolfatto, P., Wong, K. M., & Bachtrog, D. (2011). Effective population size and the efficacy of selection on the X chromosomes of two closely related drosophila species. *Genome Biology and Evolution*, *3*, 114–128. https://doi.org/10.1093/gbe/evq086

Assis, R., Zhou, Q., & Bachtrog, D. (2012). Sex-biased transcriptome evolution in drosophila. *Genome Biology and Evolution*, *4*(11), 1189–1200. https://doi.org/10.1093/gbe/evs093

Bao, Y., Hu, G., Grover, C. E., Conover, J., Yuan, D., & Wendel, J. F. (2019). Unraveling cis and trans regulatory evolution during cotton domestication. *Nature Communications*, *10*(1), 5399. https://doi.org/10.1038/s41467-019-13386-w

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate – a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Blankenberg, D., Von Kuster, G., Bouvier, E., Baker, D., Afgan, E., Stoler, N., Galaxy Team, Taylor, J., & Nekrutenko, A. (2014). Dissemination of scientific software with Galaxy ToolShed. *Genome Biology*, *15*(2), 403. https://doi.org/10.1186/gb4161

Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csardi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A., Kircher, M., Albert, F. W., Zeller, U., Khaitovich, P., Grützner, F., Bergmann, S., Nielsen, R., Pääbo, S., & Kaessmann, H. (2011). The evolution of gene expression levels in mammalian organs. *Nature*, *478*(7369), 343–348. https://doi.org/10.1038/nature10532

Brill, E., Kang, L., Michalak, K., Michalak, P., & Price, D. K. (2016). Hybrid sterility and evolution in Hawaiian drosophila: Differential gene and allele-specific expression analysis of backcross males. *Heredity (Edinb)*, *117*(2), 100–108. https://doi.org/10.1038/hdy.2016.31

Chan, C., Jayasekera, S., Kao, B., Paramo, M., von Grotthuss, M., & Ranz, J. M. (2015). Remodelling of a homeobox gene cluster by multiple independent gene reunions in drosophila. *Nature Communications*, 6, 6509. https://doi.org/10.1038/ncomms7509

Charlesworth, B., Coyne, J. A., & Barton, N. H. (1987). The relatie rates of evolution of sex chromosomes and autosomes. *The American Naturalist*, 130(1), 113–146. https://www.jstor.org/stable/2461884

Civetta, A. (2016). Misregulation of gene expression and sterility in interspecies hybrids: Causal links and alternative hypotheses. *Journal of Molecular Evolution*, 82(4–5), 176–182. https://doi.org/10.1007/s00239-016-9734-z

Civetta, A., & Gaudreau, C. (2015). Hybrid male sterility between *Drosophila willistoni* species is caused by male failure to transfer sperm during copulation. *BMC Evolutionary Biology*, 15, 75. https://doi.org/10.1186/s12862-015-0355-8

Clifton, B. D., Librado, P., Yeh, S. D., Solares, E. S., Real, D. A., Jayasekera, S. U., Zhang, W., Shi, M., Park, R. V., Magie, R. D., Ma, H.-C., Xia, X.-Q., Marco, A., Rozas, J., & Ranz, J. M. (2017). Rapid functional and sequence differentiation of a tandemly repeated species-specific multigene family in drosophila. *Molecular Biology and Evolution*, 34(1), 51–65. https://doi.org/10.1093/molbev/msw212

Coolon, J. D., McManus, C. J., Stevenson, K. R., Graveley, B. R., & Wittkopp, P. J. (2014). Tempo and mode of regulatory evolution in drosophila. *Genome Research*, 24(5), 797–808. https://doi.org/10.1101/gr.163014.113

Coolon, J. D., Stevenson, K. R., McManus, C. J., Yang, B., Graveley, B. R., & Wittkopp, P. J. (2015). Molecular mechanisms and evolutionary processes contributing to accelerated divergence of gene expression on the drosophila X chromosome. *Molecular Biology and Evolution*, 32(10), 2605–2615. https://doi.org/10.1093/molbev/msv135

Coyne, J. A. (1992). Genetics and speciation. *Nature*, 355(6360), 511–515. https://doi.org/10.1038/355511a0

Coyne, J. A. (2018). "two rules of speciation" revisited. *Molecular Ecology*, 27(19), 3749–3752. https://doi.org/10.1111/mec.14790

Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sinauer Associates.

Cutter, A. D., & Bundus, J. D. (2020). Speciation and the developmental alarm clock. *eLife*, 9, e56276. https://doi.org/10.7554/eLife.56276

Dainat, J. D., Hereñú, D., & pascal-git. (2021). *NBISweden/AGAT: AGAT-v0.8.0*. Zenodo. https://doi.org/10.5281/zenodo.5336786

Davis, H., Sosulski, N., & Civetta, A. (2020). Reproductive isolation caused by azoospermia in sterile male hybrids of drosophila. *Ecology and Evolution*, 10(12), 5922–5931. https://doi.org/10.1002/ece3.6329

Dean, R., Harrison, P. W., Wright, A. E., Zimmer, F., & Mank, J. E. (2015). Positive selection underlies faster-Z evolution of gene expression in birds. *Molecular Biology and Evolution*, 32(10), 2646–2656. https://doi.org/10.1093/molbev/msv138

Degner, J. F., Pai, A. A., Pique-Regi, R., Veyrieras, J. B., Gaffney, D. J., Pickrell, J. K., De Leon, S., Michelini, K., Lewellen, N., Crawford, G. E., Stephens, M., Gilad, Y., & Pritchard, J. K. (2012). DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature*, 482(7385), 390–394. https://doi.org/10.1038/nature10808

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, 29(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

dos Santos, G., Schroeder, A. J., Goodman, J. L., Strelets, V. B., Crosby, M. A., Thurmond, J., Emmert, D. B., & Gelbart, W. M. (2015). FlyBase: Introduction of the Drosophila melanogaster release 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic Acids Research*, 43, D690–D697. https://doi.org/10.1093/nar/gku1099

Drosophila 12 Genomes Consortium, Clark, A. G., Eisen, M. B., Smith, D. R., Bergman, C. M., Oliver, B., Markow, T. A., Kaufman, T. C., Kellis, M., Gelbart, W., Iyer, V. N., Pollard, D. A., Sackton, T. B., Larracuente, A. M., Singh, N. D., Abad, J. P., Abt, D. N., Adryan, B., Aguade, M., … MacCallum, I. (2007). Evolution of genes and genomes on the drosophila phylogeny. *Nature*, 450(7167), 203–218. https://doi.org/10.1038/nature06341

Emerson, J. J., Hsieh, L. C., Sung, H. M., Wang, T. Y., Huang, C. J., Lu, H. H., Lu, M. J., Wu, S., & Li, W. H. (2010). Natural selection on cis and trans regulation in yeasts. *Genome Research*, 20(6), 826–836. https://doi.org/10.1101/gr.101576.109

Emms, D. M., & Kelly, S. (2015). OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, 16, 157. https://doi.org/10.1186/s13059-015-0721-2

Fear, J. M., Leon-Novelo, L. G., Morse, A. M., Gerken, A. R., Van Lehmann, K., Tower, J., Nuzhdin, S. V., & McIntyre, L. M. (2016). Buffering of genetic regulatory networks in *Drosophila melanogaster*. *Genetics*, 203(3), 1177–1190. https://doi.org/10.1534/genetics.116.188797

Fraser, H. B. (2019). Improving estimates of compensatory cis-trans regulatory divergence. *Trends in Genetics*, 35(1), 3–5. https://doi.org/10.1016/j.tig.2018.09.003

Gilad, Y., Rifkin, S. A., & Pritchard, J. K. (2008). Revealing the architecture of gene regulation: The promise of eQTL studies. *Trends in Genetics*, 24(8), 408–415. https://doi.org/10.1016/j.tig.2008.06.001

Glaser-Schmitt, A., & Parsch, J. (2018). Functional characterization of adaptive variation within a cis-regulatory element influencing Drosophila melanogaster growth. *PLoS Biology*, 16(1), e2004538. https://doi.org/10.1371/journal.pbio.2004538

Go, A., Alhazmi, D., & Civetta, A. (2019). Altered expression of cell adhesion genes and hybrid male sterility between subspecies of Drosophila pseudoobscura. *Genome*, 62(10), 657–663. https://doi.org/10.1139/gen-2019-0066

Go, A. C., & Civetta, A. (2020). Hybrid incompatibilities and transgressive gene expression between two closely related subspecies of drosophila. *Frontiers in Genetics*, 11, 599292. https://doi.org/10.3389/fgene.2020.599292

Gomes, S., & Civetta, A. (2014). Misregulation of spermatogenesis genes in drosophila hybrids is lineage-specific and driven by the combined effects of sterility and fast male regulatory divergence. *Journal of Evolutionary Biology*, 27(9), 1775–1783. https://doi.org/10.1111/jeb.12428

Gomes, S., & Civetta, A. (2015). Hybrid male sterility and genome-wide misexpression of male reproductive proteases. *Scientific Reports*, 5, 11976. https://doi.org/10.1038/srep11976

Goncalves, A., Leigh-Brown, S., Thybert, D., Stefflova, K., Turro, E., Flicek, P., Brazma, A., Odom, D. T., & Marioni, J. C. (2012). Extensive compensatory cis-trans regulation in the evolution of mouse gene expression. *Genome Research*, 22(12), 2376–2384. https://doi.org/10.1101/gr.142281.112

Gonzalez-De-la-Rosa, P. M., Abreu-Goodger, C., & Ranz, J. (2021). *Drosophila willistoni* genome annotation. https://doi.org/10.5281/zenodo.4498543

Good, J. M., Giger, T., Dean, M. D., & Nachman, M. W. (2010). Widespread over-expression of the X chromosome in sterile F(1)hybrid mice. *PLoS Genetics*, 6(9), e1001148. https://doi.org/10.1371/journal.pgen.1001148

Gupta, V., Parisi, M., Sturgill, D., Nuttall, R., Doctolero, M., Dudko, O. K., Malley, J. D., Eastman, P. S., & Oliver, B. (2006). Global analysis of X-chromosome dosage compensation. *Journal of Biology*, 5(1), 3. https://doi.org/10.1186/jbiol30

Haas, B. J., Delcher, A. L., Mount, S. M., Wortman, J. R., Smith, R. K., Jr., Hannick, L. I., Maiti, R., Ronning, C. M., Rusch, D. B., Town, C. D., Salzberg, S. L., & White, O. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research*, 31(19), 5654–5666. https://doi.org/10.1093/nar/gkg770

Hagen, J. F. D., Mendes, C. C., Blogg, A., Payne, A., Tanaka, K. M., Gaspar, P., Figueras Jimenez, J., Kittelmann, M., McGregor, A. P., & Nunes, M. D. (2019). Tartan underlies the evolution of drosophila male genital morphology. *Proceedings of the National Academy of Sciences of the United States of America*, 116(38), 19025–19030. https://doi.org/10.1073/pnas.1909829116

Haldane, J. B. S. (1922). Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics*, 12(2), 101–109. https://doi.org/10.1007/BF02983075

Hill, M. S., Vande Zande, P., & Wittkopp, P. J. (2021). Molecular and evolutionary processes generating variation in gene expression. *Nature Reviews. Genetics*, 22(4), 203–215. https://doi.org/10.1038/s41576-020-00304-w

Hunnicutt, K. E., Good, J. M., & Larson, E. L. (2022). Unraveling patterns of disrupted gene expression across a complex tissue. *Evolution*, 76(2), 275–291. https://doi.org/10.1111/evo.14420

Ibn-Salem, J., Muro, E. M., & Andrade-Navarro, M. A. (2017). Co-regulation of paralog genes in the three-dimensional chromatin architecture. *Nucleic Acids Research*, 45(1), 81–91. https://doi.org/10.1093/nar/gkw813

Jagadeeshan, S., & Singh, R. S. (2007). Rapid evolution of outer egg membrane proteins in the Drosophila melanogaster subgroup: A case of ecologically driven evolution of female reproductive traits. *Molecular Biology and Evolution*, 24(4), 929–938. https://doi.org/10.1093/molbev/msm009

Johnson, N. A., & Lachance, J. (2012). The genetics of sex chromosomes: Evolution and implications for hybrid incompatibility. *Annals of the new York Academy of Sciences*, 1256, E1–E22. https://doi.org/10.1111/j.1749-6632.2012.06748.x

Kayserili, M. A., Gerrard, D. T., Tomancak, P., & Kalinka, A. T. (2012). An excess of gene expression divergence on the X chromosome in drosophila embryos: Implications for the faster-X hypothesis. *PLoS Genetics*, 8(12), e1003200. https://doi.org/10.1371/journal.pgen.1003200

Khaitovich, P., Hellmann, I., Enard, W., Nowick, K., Leinweber, M., Franz, H., Weiss, G., Lachmann, M., & Paabo, S. (2005). Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science*, 309(5742), 1850–1854. https://doi.org/10.1126/science.1108296

Kopania, E. E. K., Larson, E. L., Callahan, C., Keeble, S., & Good, J. M. (2022). Molecular evolution across mouse spermatogenesis. *Molecular Biology and Evolution*, 39(2), msac023. https://doi.org/10.1093/molbev/msac023

Kratochwil, C. F., Liang, Y., Gerwin, J., Woltering, J. M., Urban, S., Henning, F., Machado-Schiaffino, G., Hulsey, C. D., & Meyer, A. (2018). Agouti-related peptide 2 facilitates convergent evolution of stripe patterns across cichlid fish radiations. *Science*, 362(6413), 457–460. https://doi.org/10.1126/science.aao6809

Kuckwa, J., Fritzen, K., Buttgereit, D., Rothenbusch-Fender, S., & Renkawitz-Pohl, R. (2016). A new level of plasticity: Drosophila smooth-like testes muscles compensate failure of myoblast fusion. *Development (Cambridge, England)*, 143(2), 329–338. https://doi.org/10.1242/dev.126730

Kulathinal, R., & Singh, R. S. (1998). Cytological characterization of premeiotic versus postmeiotic defects producing hybrid male sterility among sibling species of the Drosophila melanogaster complex. *Evolution*, 52(4), 1067–1079. https://doi.org/10.2307/2411237

Landry, C. R., Wittkopp, P. J., Taubes, C. H., Ranz, J. M., Clark, A. G., & Hartl, D. L. (2005). Compensatory cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of drosophila. *Genetics*, 171(4), 1813–1822. https://doi.org/10.1534/genetics.105.047449

Larson, E. L., Kopania, E. E. K., Hunnicutt, K. E., Vanderpool, D., Keeble, S., & Good, J. M. (2022). Stage-specific disruption of X chromosome expression during spermatogenesis in sterile house mouse hybrids.

G3 (Bethesda, Md), 12(2), jkab407. https://doi.org/10.1093/g3journal/jkab407

Lemos, B., Araripe, L. O., Fontanillas, P., & Hartl, D. L. (2008). Dominance and the evolutionary accumulation of cis- and trans-effects on gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 105(38), 14471–14476. https://doi.org/10.1073/pnas.0805160105

Lewontin, R. C. (1974). *The genetic basis of evolutionary change*. Columbia University Press.

Li, R., Ren, X., Bi, Y., Ho, V. W., Hsieh, C. L., Young, A., Zhang, Z., Lin, T., Zhao, Y., Miao, L., Sarkies, P., & Zhao, Z. (2016). Specific down-regulation of spermatogenesis genes targeted by 22G RNAs in hybrid sterile males associated with an X-chromosome introgression. *Genome Research*, 26(9), 1219–1232. https://doi.org/10.1101/gr.204479.116

Llopart, A. (2012). The rapid evolution of X-linked male-biased gene expression and the large-X effect in drosophila yakuba, D. santomea, and their hybrids. *Molecular Biology and Evolution*, 29(12), 3873–3886. https://doi.org/10.1093/molbev/mss190

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Mack, K. L., Campbell, P., & Nachman, M. W. (2016). Gene regulation and speciation in house mice. *Genome Research*, 26(4), 451–461. https://doi.org/10.1101/gr.195743.115

Mack, K. L., & Nachman, M. W. (2017). Gene regulation and speciation. *Trends in Genetics*, 33(1), 68–80. https://doi.org/10.1016/j.tig.2016.11.003

Mahadevaraju, S., Fear, J. M., Akeju, M., Galletta, B. J., Pinheiro, M., Avelino, C. C., Cabral-de-Mello, D. C., Conlon, K., Dell'Orso, S., Demere, Z., Mansuria, K., Mendoca, C. A., Palacios-Gimenez, O. M., Ross, E., Savery, M., Yu, K., Smith, H. E., Sartorelli, V., Yang, H., … Oliver, B. (2021). Dynamic sex chromosome expression in drosophila male germ cells. *Nature Communications*, 12(1), 892. https://doi.org/10.1038/s41467-021-20897-y

Makova, K. D., & Li, W. H. (2003). Divergence in the spatial pattern of gene expression between human duplicate genes. *Genome Research*, 13(7), 1638–1645. https://doi.org/10.1101/gr.1133803

Mank, J. E. (2013). Sex chromosome dosage compensation: Definitely not for everyone. *Trends in Genetics*, 29(12), 677–683. https://doi.org/10.1016/j.tig.2013.07.005

Mardiros, X. B., Park, R., Clifton, B., Grewal, G., Khizar, A. K., Markow, T. A., Ranz, J. M., & Civetta, A. (2016). Postmating reproductive isolation between strains of *Drosophila willistoni*. *Fly (Austin)*, 10(4), 162–171. https://doi.org/10.1080/19336934.2016.1197448

Masly, J. P., & Presgraves, D. C. (2007). High-resolution genome-wide dissection of the two rules of speciation in drosophila. *PLoS Biology*, 5(9), e243. https://doi.org/10.1371/journal.pbio.0050243

McManus, C. J., Coolon, J. D., Duff, M. O., Eipper-Mains, J., Graveley, B. R., & Wittkopp, P. J. (2010). Regulatory divergence in drosophila revealed by mRNA-seq. *Genome Research*, 20(6), 816–825. https://doi.org/10.1101/gr.102491.109

McVicker, G., van de Geijn, B., Degner, J. F., Cain, C. E., Banovich, N. E., Raj, A., Lewellen, N., Myrthil, M., Gilad, Y., & Pritchard, J. K. (2013). Identification of genetic variants that affect histone modifications in human cells. *Science*, 342(6159), 747–749. https://doi.org/10.1126/science.1242429

Meiklejohn, C. D., Landeen, E. L., Cook, J. M., Kingan, S. B., & Presgraves, D. C. (2011). Sex chromosome-specific regulation in the drosophila male germline but little evidence for chromosomal dosage compensation or meiotic inactivation. *PLoS Biology*, 9(8), e1001126. https://doi.org/10.1371/journal.pbio.1001126

Meiklejohn, C. D., & Presgraves, D. C. (2012). Little evidence for demasculinization of the drosophila X chromosome among genes

expressed in the male germline. *Genome Biology and Evolution*, *4*(10), 1007–1016. https://doi.org/10.1093/gbe/evs077

Meisel, R. P. (2011). Towards a more nuanced understanding of the relationship between sex-biased gene expression and rates of protein-coding sequence evolution. *Molecular Biology and Evolution*, *28*(6), 1893–1900. https://doi.org/10.1093/molbev/msr010

Meisel, R. P., Malone, J. H., & Clark, A. G. (2012a). Disentangling the relationship between sex-biased gene expression and X-linkage. *Genome Research*, *22*(7), 1255–1265. https://doi.org/10.1101/gr.132100.111

Meisel, R. P., Malone, J. H., & Clark, A. G. (2012b). Faster-X evolution of gene expression in drosophila. *PLoS Genetics*, *8*(10), e1003013. https://doi.org/10.1371/journal.pgen.1003013

Metzger, B. P. H., Duveau, F., Yuan, D. C., Tryban, S., Yang, B., & Wittkopp, P. J. (2016). Contrasting frequencies and effects of cis- and trans-regulatory mutations affecting gene expression. *Molecular Biology and Evolution*, *33*(5), 1131–1146. https://doi.org/10.1093/molbev/msw011

Mikhaylova, L. M., & Nurminsky, D. I. (2011). Lack of global meiotic sex chromosome inactivation, and paucity of tissue-specific gene expression on the drosophila X chromosome. *BMC Biology*, *9*, 29. https://doi.org/10.1186/1741-7007-9-29

Moehring, A. J., Llopart, A., Elwyn, S., Coyne, J. A., & Mackay, T. F. C. (2006). The genetic basis of prezygotic reproductive isolation between drosophila santomea and D-yakuba due to mating preference. *Genetics*, *173*(1), 215–223. https://doi.org/10.1534/genetics.105.052993

Moehring, A. J., Teeter, K. C., & Noor, M. A. (2007). Genome-wide patterns of expression in drosophila pure species and hybrid males. II. Examination of multiple-species hybridizations, platforms, and life cycle stages. *Molecular Biology and Evolution*, *24*(1), 137–145. https://doi.org/10.1093/molbev/msl142

Mugal, C. F., Wang, M., Backstrom, N., Wheatcroft, D., Alund, M., Semon, M., McFarlane, S. E., Dutoit, L., Qvarnstrom, A., & Ellegren, H. (2020). Tissue-specific patterns of regulatory changes underlying gene expression differences among Ficedula flycatchers and their naturally occurring F1 hybrids. *Genome Research*, *30*(12), 1727–1739. https://doi.org/10.1101/gr.254508.119

Muller, H. J. (1940). Bearings of the drosophila work on systematics. In J. Huxley (Ed.), *The new systematics* (pp. 185–268). Clarendon Press.

Nyberg, K. G., & Machado, C. A. (2016). Comparative expression dynamics of intergenic long noncoding RNAs in the genus drosophila. *Genome Biology and Evolution*, *8*(6), 1839–1858. https://doi.org/10.1093/gbe/evw116

Parisi, M., Nuttall, R., Naiman, D., Bouffard, G., Malley, J., Andrews, J., Eastman, S., & Oliver, B. (2003). Paucity of genes on the drosophila X chromosome showing male-biased expression. *Science*, *299*(5607), 697–700.

Patlar, B., Jayaswal, V., Ranz, J. M., & Civetta, A. (2021). Nonadaptive molecular evolution of seminal fluid proteins in drosophila. *Evolution*, *75*(8), 2102–2113. https://doi.org/10.1111/evo.14297

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., & Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and ballgown. *Nature Protocols*, *11*(9), 1650–1667. https://doi.org/10.1038/nprot.2016.095

Presgraves, D. C. (2008). Sex chromosomes and speciation in drosophila. *Trends in Genetics*, *24*(7), 336–343. https://doi.org/10.1016/j.tig.2008.04.007

Prud'homme, B., Gompel, N., & Carroll, S. B. (2007). Emerging principles of regulatory evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(Suppl 1), 8605–8612. https://doi.org/10.1073/pnas.0700488104

R Development Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from. https://www.R-project.org/

Ranz, J. M., Castillo-Davis, C. I., Meiklejohn, C. D., & Hartl, D. L. (2003). Sex-dependent gene expression and evolution of the drosophila

transcriptome. *Science*, *300*(5626), 1742–1745. https://doing.org/10.1126/science.1085881

Ranz, J. M., Gonzalez, P. M., Clifton, B. D., Nazario-Yepiz, N. O., Hernandez-Cervantes, P. L., Palma-Martinez, M. J., Valdivida, D. I., Jimenez-Kaufman, A., Lu, M. M., Markow, T. A., & Abreu-Goodger, C. (2021). A de novo transcriptional atlas in Danaus plexippus reveals variability in dosage compensation across tissues. *Communications Biology*, *4*(1), 791. https://doi.org/10.1038/s42003-021-02335-3

Ranz, J. M., Segarra, C., & Ruiz, A. (1997). Chromosomal homology and molecular organization of Muller's elements D and E in the Drosophila repleta species group. *Genetics*, *145*(2), 281–295. https://doi.org/10.1093/genetics/145.2.281

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). G:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*, *47*(W1), W191–W198. https://doi.org/10.1093/nar/gkz369

Rice, W. R. (1984). Sex chromosomes and the evolution of sexual dimorphism. *Evolution*, *38*(4), 735–742. https://doi.org/10.1086/595754

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Rodriguez, M. A., Vermaak, D., Bayes, J. J., & Malik, H. S. (2007). Species-specific positive selection of the male-specific lethal complex that participates in dosage compensation in drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(39), 15412–15417. https://doi.org/10.1073/pnas.0707445104

Rohde, C., & Valente, V. L. (2012). Three decades of studies on chromosomal polymorphism of *Drosophila willistoni* and description of fifty different rearrangements. *Genetics and Molecular Biology*, *35*(4 (suppl)), 966–979. https://doi.org/10.1590/s1415-47572012000600012

Schurch, N. J., Schofield, P., Gierlinski, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G. G., Owen-Hughes, T., Blaxter, M., & Barton, G. J. (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA*, *22*(6), 839–851. https://doi.org/10.1261/rna.053959.115

Signor, S. A., & Nuzhdin, S. V. (2018). The evolution of gene expression in cis and trans. *Trends in Genetics*, *34*(7), 532–544. https://doi.org/10.1016/j.tig.2018.03.007

Signor, S. A., & Nuzhdin, S. V. (2019). Compensatory evolution of gene expression. *Trends in Genetics*, *35*(12), 890–891. https://doi.org/10.1016/j.tig.2019.09.008

Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics (Oxford, England)*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Slater, G. S., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, *6*, 31. https://doi.org/10.1186/1471-2105-6-31

Spassky, B., & Dobzhansky, T. (1950). Comparative genetics of *Drosophila willistoni*. *Heredity (Edinb)*, *4*(2), 201–215. https://doi.org/10.1038/hdy.1950.15

Sturgill, D., Zhang, Y., Parisi, M., & Oliver, B. (2007). Demasculinization of X chromosomes in the drosophila genus. *Nature*, *450*(7167), 238–241. https://doi.org/10.1038/nature06330

Supek, F., Bosnjak, M., Skunca, N., & Smuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*, *6*(7), e21800. https://doi.org/10.1371/journal.pone.0021800

Susic-Jung, L., Hornbruch-Freitag, C., Kuckwa, J., Rexer, K. H., Lammel, U., & Renkawitz-Pohl, R. (2012). Multinucleated smooth muscles and mononucleated as well as multinucleated striated muscles develop during establishment of the male reproductive organs of

Drosophila melanogaster. *Developmental Biology*, *370*(1), 86–97. https://doi.org/10.1016/j.ydbio.2012.07.022

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Von Mering, C. (2019). STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, *47*(D1), D607–D613. https://doi.org/10.1093/nar/gky1131

Tirosh, I., Reikhav, S., Levy, A. A., & Barkai, N. (2009). A yeast hybrid provides insight into the evolution of gene expression regulation. *Science*, *324*(5927), 659–662. https://doi.org/10.1126/science.1169766

Vicoso, B., & Charlesworth, B. (2006). Evolution on the X chromosome: Unusual patterns and processes. *Nature Reviews. Genetics*, *7*(8), 645–653. https://doi.org/10.1038/nrg1914

Vicoso, B., & Charlesworth, B. (2009). Effective population size and the faster-X effect: An extended model. *Evolution*, *63*(9), 2413–2426. https://doi.org/10.1111/j.1558-5646.2009.00719.x

Wen, K., Yang, L., Xiong, T., Di, C., Ma, D., Wu, M., Xue, Z., Zhang, X., Long, L., Zhang, W., Zhang, J., Bi, X., Dai, J., Zhang, Q., Lu, Z. J., & Gao, G. (2016). Critical roles of long noncoding RNAs in drosophila spermatogenesis. *Genome Research*, *26*(9), 1233–1244. https://doi.org/10.1101/gr.199547.115

Witt, E., Benjamin, S., Svetec, N., & Zhao, L. (2019). Testis single-cell RNA-seq reveals the dynamics of de novo gene transcription and germline mutational bias in drosophila. *eLife*, *8*, e47138. https://doi.org/10.7554/eLife.47138

Witt, E., Shao, Z., Hu, C., Krause, H. M., & Zhao, L. (2021). Single-cell RNA-sequencing reveals pre-meiotic X-chromosome dosage compensation in drosophila testis. *PLoS Genetics*, *17*(8), e1009728. https://doi.org/10.1371/journal.pgen.1009728

Witt, E., Svetec, N., Benjamin, S., & Zhao, L. (2021). Transcription factors drive opposite relationships between gene age and tissue specificity in male and female drosophila gonads. *Molecular Biology and Evolution*, *38*(5), 2104–2115. https://doi.org/10.1093/molbev/msab011

Wittkopp, P. J., Haerum, B. K., & Clark, A. G. (2004). Evolutionary changes in cis and trans gene regulation. *Nature*, *430*(6995), 85–88. https://doi.org/10.1038/nature02698

Wucher, V., Legeai, F., Hedan, B., Rizk, G., Lagoutte, L., Leeb, T., Jagannathan, V., Cadieu, E., David, A., Lohi, H., Cirera, S., Fredholm, M., Botherel, N., Leegwater, P. A. J., Le Beguec, C., Fieten, H., Johnson, J., Alfoldi, J., Andre, C., ... Derrien, T. (2017). FEELnc: A tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Research*, *45*(8), e57. https://doi.org/10.1093/nar/gkw1306

Zhang, X., Cal, A. J., & Borevitz, J. O. (2011). Genetic architecture of regulatory variation in Arabidopsis thaliana. *Genome Research*, *21*(5), 725–733. https://doi.org/10.1101/gr.115337.110

Zhang, X., & Emerson, J. J. (2019). Inferring compensatory evolution of cis- and trans-regulatory variation. *Trends in Genetics*, *35*(1), 1–3. https://doi.org/10.1016/j.tig.2018.11.003

Zhu, B., Li, L., Wei, R., Liang, P., & Gao, X. (2021). Regulation of GSTu1-mediated insecticide resistance in Plutella xylostella by miRNA and lncRNA. *PLoS Genetics*, *17*(10), e1009888. https://doi.org/10.1371/journal.pgen.1009888

Zimin, A. V., Smith, D. R., Sutton, G., & Yorke, J. A. (2008). Assembly reconciliation. *Bioinformatics (Oxford, England)*, *24*(1), 42–45. https://doi.org/10.1093/bioinformatics/btm542

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.