

Spermatogenesis Drives Rapid Gene Creation and Masculinization of the X Chromosome in Stalk-Eyed Flies (Diopsidae)

Richard H. Baker^{1,*}, Apurva Narechania¹, Rob DeSalle¹, Philip M. Johns², Josephine A. Reinhardt³, and Gerald S. Wilkinson³

¹Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, NY

²Life Sciences Department, Yale-NUS College, Singapore, Singapore

³Department of Biology, University of Maryland, College Park

*Corresponding author: E-mail: rbaker@amnh.org.

Accepted: February 27, 2016

Data deposition: This project has been deposited at NCBI under the accessions PRJNA240197, PRJNA289400, PRJNA289895 and GSE55601, GSE75872, GSE75873.

Abstract

Throughout their evolutionary history, genomes acquire new genetic material that facilitates phenotypic innovation and diversification. Developmental processes associated with reproduction are particularly likely to involve novel genes. Abundant gene creation impacts the evolution of chromosomal gene content and general regulatory mechanisms such as dosage compensation. Numerous studies in model organisms have found complex and, at times contradictory, relationships among these genomic attributes highlighting the need to examine these patterns in other systems characterized by abundant sexual selection. Therefore, we examined the association among novel gene creation, tissue-specific gene expression, and chromosomal gene content within stalk-eyed flies. Flies in this family are characterized by strong sexual selection and the presence of a newly evolved X chromosome. We generated RNA-seq transcriptome data from the testes for three species within the family and from seven additional tissues in the highly dimorphic species, *Teleopsis dalmanni*. Analysis of dipteran gene orthology reveals dramatic testes-specific gene creation in stalk-eyed flies, involving numerous gene families that are highly conserved in other insect groups. Identification of X-linked genes for the three species indicates that the X chromosome arose prior to the diversification of the family. The most striking feature of this X chromosome is that it is highly masculinized, containing nearly twice as many testes-specific genes as expected based on its size. All the major processes that may drive differential sex chromosome gene content—creation of genes with male-specific expression, development of male-specific expression from pre-existing genes, and movement of genes with male-specific expression—are elevated on the X chromosome of *T. dalmanni*. This masculinization occurs despite evidence that testes expressed genes do not achieve the same levels of gene expression on the X chromosome as they do on the autosomes.

Key words: diopsid, comparative transcriptomes, gene duplication, sex-specific gene expression, meiotic drive, dosage compensation.

Downloaded from https://academic.oup.com/gbe/article/8/3/896/2574146 by guest on 17 March 2021

Introduction

The evolutionary forces associated with sex and reproduction produce dynamic change and can have a profound impact on numerous biological phenomena including differential gene expression (Zhang et al. 2007), rates of protein evolution (Haerty et al. 2007), gene family evolution (Hahn et al. 2007), and chromosome composition (Ranz et al. 2003;

Sturgill et al. 2007). These effects are most pronounced in the male germline, where the production of sperm requires a specialized, and rapidly evolving, developmental program. Spermatozoa are highly differentiated cells that are among the most diversified structures in nature (Pitnick et al. 2009). In generating this diversity, spermatogenesis utilizes a unique

transcriptional profile during the transition to meiosis and sperm differentiation that involves a vast array of genes that are expressed exclusively in the testes (Aoyagi and Wassarman 2000; White-Cooper 2010; White-Cooper and Bausek 2010). Between 10% and 15% of all protein-coding genes in *Drosophila* are expressed specifically in the testes, far more than in any other tissue (Chintapalli et al. 2007; Mikhaylova et al. 2008; Meiklejohn and Presgraves 2012). Similar levels of testes-enriched gene expression have been found in several other organisms (Eddy 2002; Choi et al. 2007; Lo et al. 2008; Baker et al. 2011), but these patterns of testes-specific expression are not conserved across species. Male-biased gene expression, resulting from gonadal differences between the sexes, exhibits significantly more variation across species than does female-biased or unbiased gene expression (Meiklejohn et al. 2003; Ranz et al. 2003; Zhang et al. 2007; Mikhaylova et al. 2008; Llopis 2012). In addition, male-biased genes are more likely to be lost or gained between closely related species than genes with other expression patterns (Proschel et al. 2006; Zhang et al. 2007; Assis et al. 2012).

The genetic and transcriptional novelty associated with spermatogenesis is driven to a large extent by abundant gene duplication (Mikhaylova et al. 2008; White-Cooper and Bausek 2010). Many testes-specific gene copies are derived, via duplication, from ubiquitously expressed paralogs (Hiller et al. 2004; Ting et al. 2004; Belote and Zhong 2009; Dubrulle et al. 2012). In addition to genetic variation created by duplication events, de novo gene creation in *Drosophila* is most prevalent for genes with testes function (Levine et al. 2006; Begun et al. 2007). The pattern of gene creation for testes-enriched genes can influence the chromosomal distribution of sex-biased genes. There has been considerable attention paid to this issue as several studies have found that genes expressed at higher levels in males than females in *Drosophila* are underrepresented on the X chromosome (Parisi et al. 2003; Ranz et al. 2003; Sturgill et al. 2007; Vibranovski, Lopes, et al. 2009). Similar patterns have been found in mosquitos (Magnusson et al. 2012), flour beetles (Prince et al. 2010), and, at least for genes expressed at later stages of spermatogenesis, mice (Khil et al. 2004). In *Drosophila*, this “demasculinization” of the X is driven partially by excessive retrotransposition off of the X chromosome and onto an autosome for genes that subsequently develop testes-specific gene expression (Betran et al. 2002; Meisel et al. 2009; Vibranovski, Zhang, et al. 2009; Han and Hahn 2012). DNA duplications also exhibit increased rates of gene movement off of the X chromosome for relocation events (i.e., when the original paralog is lost) but not when both copies are retained in the genome (Han and Hahn 2012). However, the relationship between movement off the X and testes gene expression is not absolute as retrotranspositions between autosomes also disproportionately involve testes-specific genes (Meisel et al. 2009).

Despite these findings, our understanding of the overall relationship between sex-biased gene expression and chromosomal gene content is still limited. The general nature of X chromosome demasculinization has recently been questioned (Meiklejohn and Presgraves 2012) and the distribution of sex-biased genes on the X chromosome depends on numerous variables. Young genes, arising either de novo or by gene duplication, are more likely to reside on the X chromosome than on an autosome (Zhang, Vibranovski, Krinsky, et al. 2010) suggesting that the X chromosome is not universally hostile to male-biased genes. Furthermore, several recent studies have concluded that the paucity of male-biased genes on the X chromosome is caused by limited dosage compensation in the male germline (Assis et al. 2012; Meiklejohn and Presgraves 2012; Meisel et al. 2012). When testes specificity is measured relative to expression levels in numerous other tissues rather than as a ratio relative to ovary expression, the X chromosome is not depauperate for these genes (Meiklejohn and Presgraves 2012). Overall, the complexity of the relationships requires examining the genomics of sex-biased expression in numerous lineages with different reproductive systems to identify any general patterns and mechanisms. Therefore, we initiated a project to study the interaction among testes gene expression, gene duplication, and sex chromosome content in stalk-eyed flies, a family whose biology is strongly influenced by sexual selection.

Stalk-eyed flies (Diopsidae) are visually remarkable because of the elongation of the head into long stalks with the eyes and antenna laterally displaced at the ends of these stalks. These flies have become a model system for studying sexual selection (Wilkinson and Dodson 1997; Baker et al. 2012). There is substantial variation within the family in eyestalk size and sexual dimorphism in eyespan has evolved independently in several lineages within the family (Baker and Wilkinson 2001). Females in sexually dimorphic species generally exhibit very high mating rates, providing ample opportunity for postcopulatory sperm competition (Baker, Ashwell, et al. 2001; Corley et al. 2006). Similar to eyespan, sperm length and sperm morphology exhibit substantial variation within the family, as well as correlated evolution with female reproductive morphology (Presgraves et al. 1999). Another major component of diopsid reproductive biology is the presence of meiotic drive. In stalk-eyed fly males that carry drive loci, X-bearing sperm incapacitate Y bearing sperm leading to a highly biased sex ratio (Presgraves et al. 1997; Wilkinson and Sanchez 2001). Several species in the genus *Teleopsis* are polymorphic for X chromosome drive (Wilkinson et al. 2003, 2014), and in the dimorphic species, *Teleopsis dalmani*, X drive is associated with reduced eyespan due to linkage (Wilkinson, Presgraves, et al. 1998). Breeding experiments between different populations from peninsular Malaysia have revealed numerous cryptic drive systems (Wilkinson et al. 2014), suggesting that drive has evolved and been suppressed repeatedly within the genus. Genomic

studies have also revealed the presence of a newly derived X chromosome in diopsids that is homologous to chromosome 2L in *Drosophila melanogaster* (Baker and Wilkinson 2010). It is possible that evolutionarily independent X chromosomes may evolve distinct patterns of sex-biased gene expression. Microarray analysis of gene expression in the developing eye-antennal disc of *T. dalmani* showed that female-biased genes were overrepresented on the X chromosome but male-biased genes exhibited no bias (Wilkinson et al. 2013). Here we provide a comprehensive examination of tissue-specific expression in this species and explore, through a comparative approach, the distribution of gene duplication, chromosome location, and gene movement within the family Diopsidae.

Materials and Methods

Sample Preparation

RNA-seq reads were generated from multiple tissues in *T. dalmani* and the testes of two other diopsid species—*Teleopsis quinqueguttata* and *Sphyracephala beccarii*. *Sphyracephala beccarii* represents a basal taxon for the family and *T. quinqueguttata* is the basal representative of the genus *Teleopsis* (Baker, Wilkinson, et al. 2001). The *T. dalmani* and *T. quinqueguttata* flies used for the transcriptome sequencing were chosen from outbred laboratory populations originally collected in 1999 near Ulu Gombak in peninsular Malaysia. The *S. beccarii* flies were collected near Pietermaritzburg, South Africa, in 1994. Tissues sampled in *T. dalmani* included adult head (male and female separate), third instar larvae (sex undetermined), gonadectomized females, gonadectomized males, ovaries and testes from both nondrive and drive X males (Reinhardt et al. 2014). Duplicate samples for each tissue (except adult head which comprised a single male and female sample) were dissected from 5 to 20 flies and RNA was extracted from each using the mirVana RNA Isolation Kit (Invitrogen) according to manufacturer's protocols. The *T. quinqueguttata* and *S. beccarii* samples, along with the *T. dalmani* drive and nondrive testes samples, were sent to Cofactor Genomics (St. Louis, MO) for library preparation and 60-bp paired-end (PE) sequencing on an Illumina Genome Analyzer (GA). We obtained 84-bp PE reads (Illumina GA) for the male and female head samples from the UC Davis Genome Center and 100-bp PE (Illumina Hi-Seq) reads for the remaining tissues (including another nondrive testes sample) from the UMD-IBBR Sequencing Core (supplementary table S1, Supplementary Material online).

Assembly and Annotation

Transcriptome assemblies were generated for all *T. dalmani* tissues combined and the testes of *T. quinqueguttata* and *S. beccarii* with Trinity (Grabherr et al. 2011) using default commands (PE mode, –CPU 24, –kmer_method inchworm –max_memory 190G). The transcriptome for *T. dalmani* was

annotated before initiating transcriptome annotation of the other species to provide a gene reference database for the Diopsidae. All contigs from the *T. dalmani* assembly were blasted (with a BLASTX $> 10^{-5}$ cutoff) against a protein database for *D. melanogaster* (Flybase: dmel-all-translation-r5.51.fa) and a protein sequence file containing five other dipterans (*Drosophila pseudoobscura*, *Drosophila virilis* and three mosquitoes—*Anopheles gambiae*, *Aedes aegypti*, *Culex quinquefasciatus*) downloaded from Flybase (*Drosophila*) and Ensembl (mosquitoes). All hits against *D. melanogaster* were given precedence over other dipteran hits to facilitate homology interpretation. Contigs were also blasted against the nr database. Open reading frame (ORF) sizes for each contig were calculated using the GetORF module of the Mobyle bioinformatics portal (<http://mobyle.pasteur.fr>).

Because transcriptome data lack positional information, definitive identification of all paralogous gene content is not possible. For this study, we applied a conservative approach designed to ensure that all genes scored as paralogs were truly distinct but not to identify every paralogous copy (e.g., very recent duplicates). The general strategy for distinguishing paralogous genes from alternative transcripts and allelic variants was to use the “component” (comp) designation from the Trinity output as the initial demarcation of unique genes and then to identify, through sequence comparison, violations of this clustering rule. Our criteria for separating contigs into different putative paralogs (i.e., before gene tree construction) were 5% divergence at the protein level and at least 60% assembly coverage of each protein relative to their homolog in *D. melanogaster* (supplementary fig. S1, Supplementary Material online). Overall, we found that Trinity tends to be over-inclusive, rarely putting two contigs from the same gene into different components but often combining two or more distinct genes into the same component. We identified 1,132 comps that housed two or more distinct genes (supplementary fig. S2, Supplementary Material online), but only 7 cases where different comps were subsequently clustered together. Further details of gene annotation and orthology assignment are provided in supplementary fig. S1, Supplementary Material online. In order to explore the impact of variation in our paralog criteria on patterns of tissue-specific gene expression and chromosomal location, we also scored duplication events based on putative paralogs requiring 10% protein divergence, and 70% or 80% assembly coverage.

Gene Family Assignment and Assessment of Duplication

We identified duplication events within a phylogenetic context by estimating gene family trees for all unique paralogs identified from the three diopsid species. A single representative sequence, based on the contig with the highest bit score within a comp, was selected for each unique gene

(supplementary fig. S1, Supplementary Material online), including contigs with ORFs greater than 250 amino acids but no significant BLAST hit. Protein translations for all unique genes from the diopsid species were combined with a protein database from eight other insect species (*D. melanogaster*, *D. pseudoobscura*, *D. virilis*, *An. gambiae*, *Ae. aegypti*, *C. quinquefasciatus*, *Bombyx mori* and *Apis mellifera*) and assembled into gene family clusters using OrthologID (Chiu et al. 2006). The proteins from the nondiopsid insects only included the single longest translation for each gene.

OrthologID uses an Markov Cluster algorithm (MCL) in which sequences are initially clustered using BLASTP, and gene clusters are then aligned using MAFFT (Katoh et al. 2002). Phylogenetic trees were generated for all gene family alignments in RaxML (model: PROTGAMMAJTF) (Stamatakis 2006) with 100 bootstrap replicates. Duplication events in stalk-eyed flies were identified by searching for diopsid monophyletic groups that contained more than one terminal sequence from any of the three stalk-eyed fly species. When ancestral gene copies are under strong stabilizing selection but derived duplicate copies have little selective constraint or are under strong positive selection, the ancestral and derived paralogs may not form monophyletic groups despite recent divergence. Therefore, we also searched for monophyletic groups that were supported by bootstrap values greater than 80%, contained only one lineage from each of the nondiopsid species, but had multiple genes for at least one of the three stalk-eyed fly species and scored them as duplicates. Because we lack syntenic information on the duplicates, it is not possible to distinguish the original copy from the duplicates. Therefore, all genes involved in a duplication event were scored as duplicates. Genes that had no significant BLAST hit and were not incorporated into a gene family or belonged to a gene family that contained exclusively diopsid sequences with no significant BLAST hit were designated as orphan genes. As defined in other studies (Domazet-Loso and Tautz 2003; Tautz and Domazet-Lošo 2011; Palmieri et al. 2014), these orphan sequences may represent genes that originated within stalk-eyed flies or are evolving so fast that they no longer exhibit homology to genes in other families. In most analyses presented in this study, duplicate genes and orphan genes are combined into a single category representing “novel” genes. In these cases, however, we also provide supplemental analyses separating duplicate and orphan genes to ensure that any pattern associated with novel genes is not driven entirely by orphan genes. Both duplicate and orphan genes that originated since the split between *Drosophila* and stalk-eyed flies are referred to as novel, lineage-specific genes throughout the article.

The phylogenetic timing of duplication events and the origin of orphan genes within stalk-eyed flies were recorded by examining the composition of orthologous sequences for each species relative to the nodes at which these events were inferred to have occurred. All genes that reside within

monophyletic groups containing multiple sequences from only a single species were scored as species-level duplicates or orphans. Clades that contained two or more lineages that each contained *T. dalmani* and *T. quinqueguttata* sequences, but no *S. beccarii* ortholog, were scored as *Teleopsis*-level duplicates or orphans. Finally, clades that contained two or more lineages in which each clade contained an *S. beccarii* sequence were scored as basal (i.e., occurred before the diversification of the family) duplicates or orphans.

Gene Expression Analysis

Fragments Per Kilobase of transcript per Million mapped reads (FPKM) values were obtained from each tissue-specific library aligned independently to the *T. dalmani* assembly as described in Reinhardt et al. (2014). Briefly, bwa (Li and Durbin 2009) was used to align the first read of a pair, allowing multiple equally good hits, and RNA-seq by Expectation Maximization (RSEM) was used to obtain normalized read counts and FPKM estimates for both “genes” (i.e., individual comps in most cases; supplementary fig. S1, Supplementary Material online) and transcripts. A similar procedure was used to obtain FPKM values for *S. beccarii* and *T. quinqueguttata* by aligning reads to assemblies for each species.

After FPKM values were obtained for all *T. dalmani* genes in the analysis, the FPKM values were extracted for all genes assigned to the orthology analysis and concatenated into a gene expression matrix including all tissue samples. Data were clustered using tools from Bioconductor (Gentleman et al. 2004). The Spearman correlation coefficient was calculated, gene distances were determined using Euclidian methods, and complete hierarchical clustering was used to create a gene expression dendrogram. Tissue-specific gene expression in *T. dalmani* was assessed using Tau (Yanai et al. 2005). Following criteria used in several recent studies (Assis et al. 2012; Meiklejohn and Presgraves 2012; Meisel et al. 2012), a threshold of 0.9 designated tissue-specific expression and only genes with an FPKM value greater than 5 in at least one tissue were scored for specificity. In addition to calculating Tau for *T. dalmani* tissues, we used “gonad specific” to label genes expressed in both testes and ovaries but not somatic tissues. For this analysis, we averaged FPKM values across the testes and ovaries and recalculated Tau for all genes whose expression was not tissue specific based on the initial Tau calculation. In order to compare quantities of genes expressed in the testes among different species, we only included genes that had a testes FPKM greater than 1. Measurements of differential gene expression between testes with and without meiotic drive were taken from Reinhardt et al. (2014). Identification of sex-biased gene expression in somatic tissues was assessed by calculating both fold expression differences and, when duplicate samples were available, identifying significantly differentially expressed genes between males and females using EdgeR (Robinson and Oshlack 2010).

Chromosomal Designation and Gene Movement Analysis

Comparative Genomic Hybridization (CGH) was used to determine whether genes are located on the X chromosome or an autosome in each species. This method hybridizes male and female genomic DNA to a microarray containing thousands of probes and has successfully been used to identify X-linkage in stalk-eyed flies (Baker and Wilkinson 2010). For *T. dalmani*, *T. quinqueguttata*, and *S. beccarii*, 60-mer Agilent oligonucleotide probes were designed based on the annotated contig sequences from the transcriptome assemblies. All protein-coding transcripts were included in the probe design. Four probes were designed for each gene and printed in duplicate on Agilent 4 \times 144 k arrays.

The microarray experiment consisted of four dye-flipped hybridizations for each species. One hybridization each for *T. quinqueguttata* and *S. beccarii* failed and was excluded from further analysis. Each hybridization sample was generated from 5 to 10 adult male or female flies. After removing wings and heads DNA was extracted from macerated fly bodies with Qiagen DNeasy kits using the insect sample protocol. DNA concentration was estimated using an Agilent Bioanalyzer and 3 μ g of DNA was used in each sample. Each DNA sample was fractionated by restriction digestion with Alul and Rsal for 2 h at 37 °C. Each sample was then labeled with either Cy-3 or Cy-5 and processed according to the Agilent array-based CGH protocol. After hybridization for 24 h at 65 °C, arrays were scanned using an Agilent G2505C microarray scanner. Hybridization intensity was measured from array images scanned at each dye wavelength using Agilent's Feature Extraction Software (version 10.7.3.1). Intensity scores were normalized using the linear normalization methodology in the Feature Extraction Software. For each hybridization we calculated the median \log_2 female/male intensity for all probes per gene. Then, the \log_2 ratios for a given hybridization were centralized as in Baker and Wilkinson (2010).

After signal extraction and centralization, we followed the procedure outlined in Baker and Wilkinson (2010) to differentiate autosomal and X-linked genes. In this method, histograms of the gene \log_2 ratios averaged across the hybridizations were generated for each species. Based on this histogram, intervals of size 0.1 were searched every 0.0125 across the \log_2 distribution to determine the interval with the fewest entries. The value in the center of this interval was then used as a cut-off for distinguishing chromosomal categories. We then calculated a 95% confidence interval for each gene based on the variation in \log_2 ratios across all replicate hybridizations. If the confidence interval of a given gene did not contain the cut-off value, then that gene was assigned as autosomal if its average \log_2 ratio value was less than the cut-off and X-linked if its average \log_2 ratio value was greater than the cut-off. If a gene's confidence interval contained the cut-off value then the gene's location was designated as unknown.

Reconstruction of gene movement on or off of the X chromosome was conducted in Mesquite (v 2.75) using parsimony. Chromosomal locations were obtained for *D. melanogaster*, *D. pseudoobscura*, and *An. gambiae* from Flybase and Ensembl. These values were combined with CGH chromosomal designations for the three diopsid species into a single-character matrix for each gene family alignment. The chromosomal character was scored as either autosomal or X-linked for each species. We assumed a syntenic relationship between the X chromosome in stalk-eyed flies and chromosome 3R in *An. gambiae*, chromosome 2L in *D. melanogaster*, and chromosome 4 in *D. pseudoobscura* [70]. Therefore, genes located on these chromosomes were scored as "X" in the character matrix and all other genes were scored as "A." When reconstructions of gene movement were ambiguous, only the minimum required movements were recorded. For instance, if it was equally parsimonious to posit two independent movements onto an autosome versus one movement onto an autosome and one movement onto the X chromosome, then only a single autosomal move was scored. The timing of all diopsid gene movements was scored as either occurring basal to or within the family.

Given a static or equal rate of gene creation on the autosomes or X chromosomes, the expected frequency of movement between the two chromosome categories should be equal in both directions, regardless of size differences between the chromosomes. Essentially, although the X chromosome is smaller than the autosomes, it is less likely to "send out" and "receive" a gene in equal proportion. However, if gene creation of a certain gene type (e.g., gonad specific) is higher on one chromosome, then the expectation of movement off of that chromosome increases. Therefore, we evaluated the pattern of gene movement under two conditions: 1) Assuming equal movement between the chromosomes or 2) modifying the probability of movement off of a chromosome based on the overrepresentation of a given gene type on that chromosome after excluding all movers from the analysis.

Molecular Evolution

The rate of sequence evolution for *T. dalmani* genes was calculated based on their divergence relative to *T. quinqueguttata*. For each *T. dalmani* gene, we identified its homolog in *T. quinqueguttata* using the gene family trees. In cases where a *T. quinqueguttata* gene had undergone a species-specific duplication, both copies were used as homologs and the average divergence value of the *T. dalmani* gene to each *T. quinqueguttata* copy was calculated. Clustal Omega v. 1.2.0 (Sievers et al. 2011) was used to align the translated nucleotide sequences for each *T. dalmani*–*T. quinqueguttata* pair, and the nucleotide sequence was mapped back to the protein alignment. Then, SNAP (Korber et al. 2001) was used to estimate synonymous and nonsynonymous substitution rates. Alignments less than 50aa in length were discarded,

as were any values that indicated saturation of synonymous substitutions.

Overrepresentation of gene ontology terms was evaluated with David (Dennis et al. 2003). Statistical analyses were conducted with JMP v10 (SAS_Institute 2003).

Results

Assembly and Annotation Reveals Extensive Gene Creation in Diopsids

We assembled with Trinity (Grabherr et al. 2011) over 300 million reads, generated from 7 different tissue sources for the dimorphic species, *T. dalmani*, as well as approximately 85 million reads from the testes of each of the monomorphic species, *T. quinqueguttata* and *S. beccarii*. The additional data used in the *T. dalmani* assembly produced substantially longer contigs for this species than were generated for the other two species (supplementary table S2, Supplementary Material online). To assess the quality of the *T. dalmani* assembly, we compared the Trinity contigs with an existing expressed sequence tag (EST) database constructed for the eye-antennal imaginal disc of *T. dalmani* with Sanger sequence data (Baker et al. 2009). A reciprocal best BLAST search produced 9,338 hits (from a total of 11,545 unique ESTs) and revealed a high degree of sequence identity between the Trinity contigs and the EST contigs (supplementary fig. S3, Supplementary Material online). Overall, the Trinity assembly was very accurate in recreating the EST library. Despite the large difference in sequencing depth between *T. dalmani* and the other two species, the number of comps with a significant BLAST hit to another dipteran species was close to 10,000 for all three assemblies (supplementary table S2, Supplementary Material online). In addition to the contigs with BLAST hits, we identified numerous comps with sequences that had no BLAST hit at an e-value cut-off of 10^{-5} but contained an ORF greater than 250 amino acids (supplementary table S2, Supplementary Material online).

Across the three species of stalk-eyed flies, 1,168 genes have undergone duplication since the separation from the common ancestor with *Drosophila* (supplementary table S2 and file S1, Supplementary Material online). On average, each duplication produced approximately 1.5 additional paralogs in each species (supplementary table S1, Supplementary Material online), but 72 genes contained 5 or more paralogous copies in at least one diopsid species (supplementary table S3, Supplementary Material online). The most abundant gene expansion in stalk-eyed flies involves the 14-3-3 protein family. In general, insects possess two, highly conserved, 14-3-3 genes (14-3-3 ϵ and 14-3-3 ζ in *D. melanogaster*) that function in numerous cellular pathways and biological processes (Ferl et al. 2002). Stalk-eyed flies possess up to 21 distinct novel copies that are clearly highly diverged relative to the ancestral genes (supplementary fig. S4, Supplementary Material

online). Across all gene families, *T. dalmani* has approximately twice as many duplicate genes as either of the other two species but this difference is caused to a large extent by the additional transcriptome sampling done for *T. dalmani*. Among genes expressed in the testes (i.e., testes FPKM > 1), there are virtually identical numbers of paralogs in *T. dalmani* and *T. quinqueguttata* (supplementary table S2, Supplementary Material online). As with new duplicates, *T. dalmani* has more orphan genes than the other diopsids but a comparable number of testes-expressed orphan genes (supplementary table S2, Supplementary Material online).

The gene family trees allowed us to estimate the evolutionary timing of novel gene creation in stalk-eyed flies (fig. 1). The large number of novel genes (both duplicates and orphans) on the branch leading to *T. dalmani* is affected by the extra transcriptome sampling for this species. Many of these genes would likely have a more ancestral origin if we had sampled homologous tissues in *T. quinqueguttata* and *S. beccarii*. However, when the comparison is limited to testes genes, there are still differences in gene creation among the diopsid branches. There are over five times as many testes duplicates in the *Teleopsis* lineage compared with the *S. beccarii* branch (fig. 1). Similarly, within *Teleopsis*, the sexually dimorphic species, *T. dalmani*, has 50% more lineage-specific duplicates and a third more orphan genes as the sexually monomorphic species, *T. quinqueguttata*.

Genes that have undergone duplication in stalk-eyed flies were significantly overrepresented for several biological processes and molecular function categories (supplementary table S4, Supplementary Material online). Genes involved in transcription from RNA polymerase II (Rpl1) promoter (GO:0006366; $P < 0.0001$) represent the most noteworthy category of duplicates as this group involves a gene family expansion of nearly every member of the basal transcription machinery complex (table 1). Despite being highly conserved, with 1:1 orthology in most eukaryotes, the basal transcription machinery genes have 20–30 additional paralogs in stalk-eyed flies compared with *D. melanogaster* and *An. gambiae*, including at least one duplicate copy for 10 of the 12 genes that comprise the Rpl1 complex. The largest expansion involves the Rpl1 cofactor *Transcription factor IIB* (*TfIIB*) that has four duplicate, rapidly evolving, paralogs in each of the three diopsid species. The gene tree for this family suggests that duplications have occurred at the base of the family as well as in each of the lineages leading to individual species (fig. 2).

Novel Genes Typically Exhibit Testes-Specific Expression

Gene expression levels in *T. dalmani* were measured for five somatic tissue sources—male and female adult heads, third instar larvae (sexes combined), adult female carcass (gonads removed), and adult male carcass (gonads removed)—and two germline tissues—ovaries and testes (supplementary file S2, Supplementary Material online). Testes exhibited the most

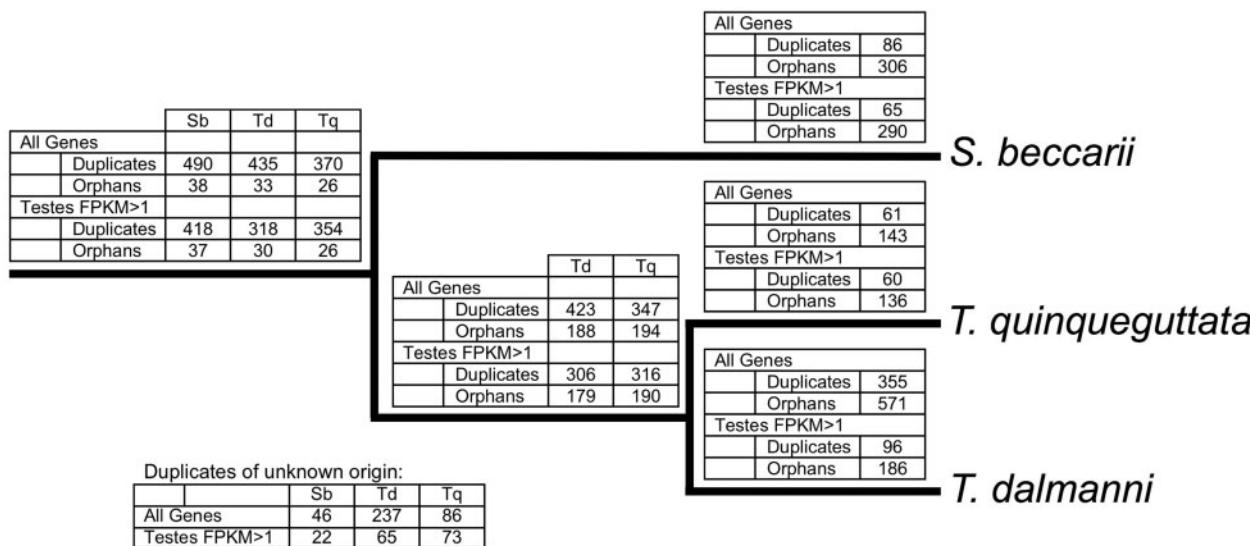


Fig. 1.—Novel gene creation in stalk-eyed flies. Phylogenetic origination of novel genes (duplication events and orphan genes) in stalk-eyed flies for genes with any expression pattern and for genes expressed within the testes (FPKM > 1).

divergent expression patterns from other tissue sources (fig. 3A) with an average between-tissue correlation (i.e., the average correlation of a tissue versus all other tissues) among protein-coding genes of 0.189, while all other tissues had an average between-tissue correlation greater than 0.45 (ovaries: 0.476, larva: 0.523, head: 0.557, female carcass: 0.584, male carcass: 0.612). Much of this divergent expression is driven by transcription that is specific to or highly enriched in the testes. Using a Tau statistic cut-off of 0.9 to designate tissue specificity, we identified 2,286 protein-coding genes with tissue-specific gene expression (fig. 3B). Among all tissue-specific genes, 79.1% are expressed predominantly in the germline and 80.1% of these gonad genes are specific to the testes.

The evolution of tissue-specific gene expression results primarily through the origination of novel genes. Novel genes are five times more likely to be tissue specific than other genes ($\chi^2 = 1,969.9$, $P < 0.0001$). This difference applies to all types of tissue-specific expression but is most pronounced in the testes (fig. 4). Overall, there is a significant relationship between the type of gene origination and tissue-specific expression categories ($\chi^2 = 3,179.9$, $P < 0.0001$; fig. 4 and supplementary table S5, Supplementary Material online). Testes-specific genes comprise 5.1% of single-copy, ancestral genes but 40.3% of novel genes that have arisen by gene duplication and 78.6% of orphan genes (fig. 4). Nearly three-quarters (72%) of all testes-specific genes are novel, while only a quarter of genes for the other types of tissues (somatic: 24.7%, ovaries: 24.8%, testes and ovaries: 21.4%) and 11.2% of ubiquitously expressed genes are novel. As might be expected given the high degree of genetic novelty associated with testes gene expression, there is little evolutionary stability in

testes-specific gene content when compared with *Drosophila*. Of the 1,408 genes that were scored as testes specific in *Drosophila* ($\text{Tau} > 0.9$; Meiklejohn and Presgraves 2012) only 279 (19.8%) have clear homologs in the *T. dalmani* transcriptomes. Of these homologs, 190 are also testes specific in *T. dalmani* and have produced 85 additional testes-specific paralogs through gene duplication. For instance, *Tubulin tyrosine ligase-like 3B*, a glycine ligase that modifies tubulin and impacts sperm individualization in *Drosophila* (Rogowski et al. 2009), has 11 testes-specific duplicate copies in stalk-eyed flies (supplementary table S3, Supplementary Material online).

It has been suggested that abundant gene creation within the testes may serve as a source of novel genetic material for other tissues (Kaessmann 2010). In this scenario, novel genes arise with testes-specific expression, but then develop, over evolutionary time, expression in other tissues. To assess this possibility, we examined the evolutionary transitions in tissue-specific gene expression within the diopsid gene families. We restricted our analysis to gene family clades that contained at least one testes-specific diopsid paralog and scored both the diopsid genes and their homologs in *D. melanogaster* as either testes specific or not testes specific. A total of 142 gene family clades contained a single, novel testes-specific paralog that had acquired this expression pattern within diopsids, while 218 gene family clades contained multiple testes-specific paralogs. Of this latter group, the majority (158 gene family clades) is comprised of testes-specific paralogs exclusively (a total of 427 paralogs), and, therefore, exhibit no evolutionary transition in tissue expression across the members of the gene family. For the remaining gene family clades, only two (supplementary fig. S5, Supplementary Material online) show evidence of a paralog acquiring gene expression in a tissue other

Table 1

Expansion of Basal Transcription Machinery Genes in Stalk-Eyed Flies

Complex	Gene	Sb	Td	Tq	Dm	Ag
RNA Pol II	RpII215	2	2	2	1	1
	RpII140	1	2	2	1	1
	RpII33	3	3	3	1	1
	RPB4	1	0	1	1	1
	RPB5	1	2	1	1	1
	RpII18	3	1	0	1	1
	RPB7	2	2	3	1	1
	RPB8	3	3	3	1	1
	RpII15	2	2	2	1	1
	RPB10	1	1	2	1	1
	RPB11	2	1	1	1	1
	RPB12	1	1	1	1	1
TFIIA	TFIIA-L	1	1	2	1	1
	TFIIA-S	1	5	2	2	1
TFIIB	TFIIB	5	5	5	1	2
TFIID	Tbp	1	2	2	1	1
	Trf	1	4	4	1	1
	Trf2	1	1	0	1	2
	Taf1	1	1	2	1	1
	Taf2	1	1	1	1	1
	Taf4	1	1	1	2	2
	Taf5	4	4	4	2	1
	Taf6	6	1	1	2	1
	Taf7	2	1	2	1	1
	Taf8	1	1	1	2	1
	Taf10	2	2	2	2	1
	Taf11	2	1	1	1	1
	Taf12	1	1	1	2	1
TFIIE	TFIIEalpha	1	1	2	1	1
	TFIIEbeta	2	4	3	1	1
TFIIF	TFIIFalpha	3	1	1	1	1
	TFIIFbeta	3	4	3	1	1
TFIIF	ENL/AF-9	1	1	1	1	1
	Ssl1	1	2	2	1	1
	8 genes	1	1	1	1	1
Total		74	74	73	50	46

NOTE.—The number of paralogous gene copies for each basal transcription machinery gene is presented for five dipteran species. Sb = *Sphyracephala beccarii*; Td = *Teleopsis dalmanni*; Tq = *Teleopsis quinqueguttata*; Dm = *Drosophila melanogaster*; Ag = *An. gambiae*.

than the testes when testes-specific expression was the ancestral condition. Two *T. dalmanni* paralogs belonging to a gene family clade with five duplicates of the *D. melanogaster* genes CG17564 and CG10750 (the pair represent tandem duplicates within *Drosophila*) acquired expression in the male carcass and ovaries, and a *T. dalmanni* duplicate of the *D. melanogaster* genes CG32086 and CG33490 developed expression in the ovaries. In contrast, there are 521 new testes-specific genes in *T. dalmanni* that arise from a testes-specific ancestor, indicating that the transfer of novel expression profiles to other tissues through duplication of testes genes is rare relative to the overall rate of gene creation in this tissue.

The relationship between novel gene creation and sex-biased gene expression is not limited to the gonads but is also evident within somatic tissues. Figure 5 depicts the proportion of genes with sex-biased expression in the adult head and the adult carcass for both ancestral and novel genes. For this analysis, genes that exhibit tissue-specific expression in other tissues were excluded from the analysis. Within the adult head, novel genes are nearly 10 times more likely to exhibit sex-biased expression at a 2-fold level than ancestral genes and over 20 times more likely at a 4-fold difference (fig. 5 and *supplementary tables S6 and S7, Supplementary Material* online). Novel genes are also significantly more likely to exhibit sex-biased expression than ancestral genes in the adult carcass but the magnitude of the difference is less than in the head (fig. 5 and *supplementary tables S6 and S7, Supplementary Material* online).

The Diopsid X Chromosome Is Enriched for Gonad-Specific and Novel Genes

Analysis of the DNA microarray data for over 10,000 genes from 3 species indicates that the diopsid X chromosome arose prior to the diversification of the family. The three species exhibit a similar distribution of X-linked genes (*supplementary fig. S6, Supplementary Material* online) with the size of this chromosome ranging from 18.4% (*T. dalmanni*) to 17.5% (*T. quinqueguttata*) of the genome (*supplementary file S3, Supplementary Material* online). Consistent with a previous study (Baker and Wilkinson 2010), a strong majority of X-linked genes (*S. beccarii*: 83.5%, *T. dalmanni*: 82.8%, *T. quinqueguttata*: 80.3%) are homologous to *D. melanogaster* chromosome 2L genes.

Across a wide range of organisms, sex chromosomes exhibit biased gene content for genes involved in reproduction, with demasculinization or feminization of the X chromosome in XY systems representing the predominant pattern (Parisi et al. 2003; Vicoso and Charlesworth 2006; Ellegren and Parsch 2007; Sturgill et al. 2007; Gao et al. 2014; Vicoso and Bachtrog 2015). We examined the relationship between tissue-specific gene expression and X-linkage in stalk-eyed flies and found a highly biased pattern. The percentage of X chromosome genes varies significantly among tissue expression categories ($\chi^2 = 305.1$; $P < 0.0001$), but contrary to most studies on other taxa, the X chromosome is highly enriched for both male- and female-specific genes. Ubiquitously expressed genes and genes specific to somatic tissues occupy the X chromosome in proportion to its size, while ovary- and testes-specific genes are substantially overrepresented on the X (table 2). Novel gonad-specific genes, in particular, are preferentially located on the X chromosome, occurring at roughly twice the rate expected based on the size of the chromosome (fig. 6 and *supplementary tables S8 and S9, Supplementary Material* online), with novel testes-specific genes exhibiting the highest proportion of X-linkage (35.6%). Conversely,

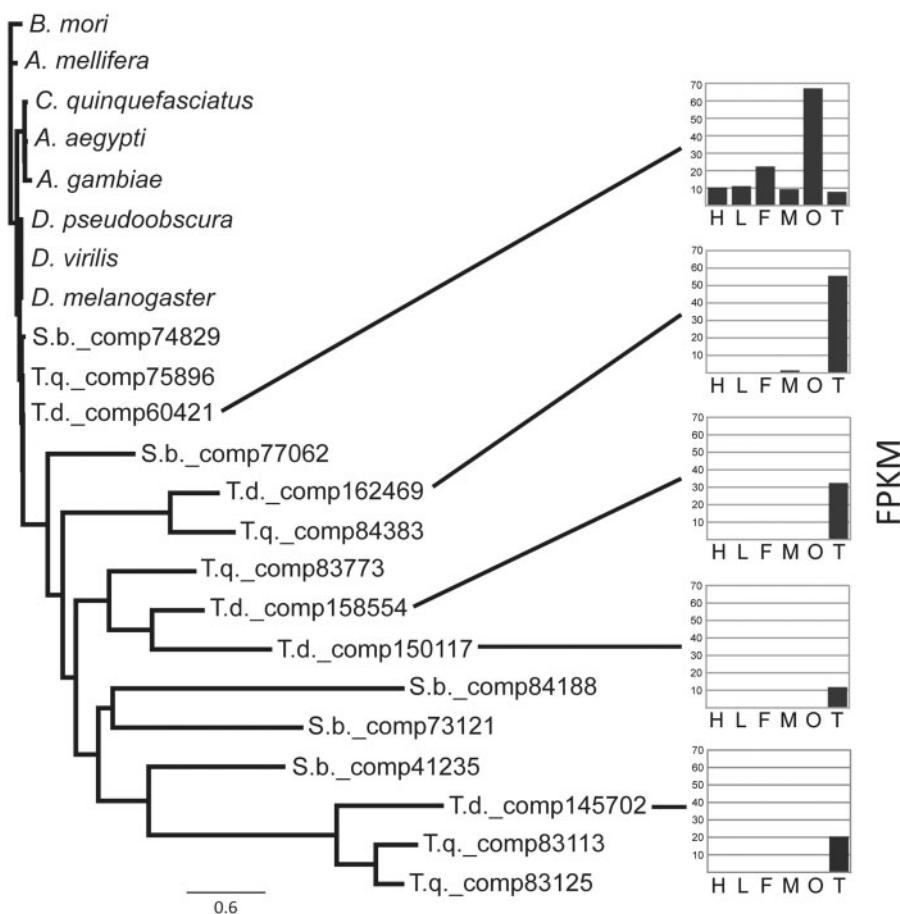


Fig. 2.—TFIIB gene family tree. Expression profiles (FPKM) are provided for each of the paralogous copies in *Teleopsis dalmanni*. H = head; L = larva; F = female carcass; M = male carcass; O = ovaries; T = testes. S.b. = *Sphyracephala beccarii*; T.d. = *Teleopsis dalmanni*; T.q. = *Teleopsis quinqueguttata*.

novel somatic-specific genes (i.e., genes expressed specifically in the head, larva, or carcasses) are significantly more likely to arise on an autosome than the X chromosome (fig. 6 and *supplementary tables S8 and S9, Supplementary Material* online). In *Drosophila*, the age of a gene impacts its chromosomal location with younger male-biased genes more likely to reside on the X and older genes underrepresented on the X (Zhang, Vibranovski, Krinsky, et al. 2010; Gao et al. 2014). In *T. dalmanni*, however, both older (i.e., those that arose prior to the *Sphyracephala*–*Teleopsis* split) novel and younger novel testes-specific genes are enriched on the X (older: 31.8%, $\chi^2 = 29.9$, $P > 0.0001$; younger: 36.9%, $\chi^2 = 34.7$, $P > 0.0001$). There was no significant effect of gene age on the proportion of X-linkage ($\chi^2 = 1.788$, $P = 0.181$).

Although the degree of X-enrichment of novel testes-specific genes is striking, it is also important to note that ancestral testes-specific genes are also overrepresented on the X. This pattern may either reflect an ancestral condition inherited by stalk-eyed flies or be caused by changes in gene expression patterns relative to *Drosophila*. Of the 401 ancestral genes

that are testes specific in *T. dalmanni*, 130 are also testes specific in *Drosophila* (and, therefore, were testes specific prior to the formation of the diopsid X) and these genes are not enriched on the X (18.4%, $\chi^2 = 0.04$, $P > 0.05$), whereas the genes without conserved expression are enriched on the X (29.1%, $\chi^2 = 11.7$, $P < 0.001$). Therefore, the overrepresentation of the ancestral testes-specific genes likely results from X-linked genes in *Teleopsis* acquiring testes-specific expression at a faster rate than genes on an autosome. Despite the strong chromosomal differences for sex-biased gonad genes, somatic sex-biased genes show no chromosomal bias. Male- and female-biased genes in both the adult head ($\chi^2 = 0.3$, $P = 0.566$) and adult carcass ($\chi^2 = 2.8$, $P = 0.092$) exhibit sex linkage proportional to chromosome size.

The ability of the X chromosome to acquire novel genetic material is also evident for the other two stalk-eyed fly species. To facilitate comparison with *T. dalmanni*, we limited this analysis to only genes that had an FPKM > 1 in the testes. In both *S. beccarii* and *T. quinqueguttata*, novel genes are significantly more likely to reside on the X than

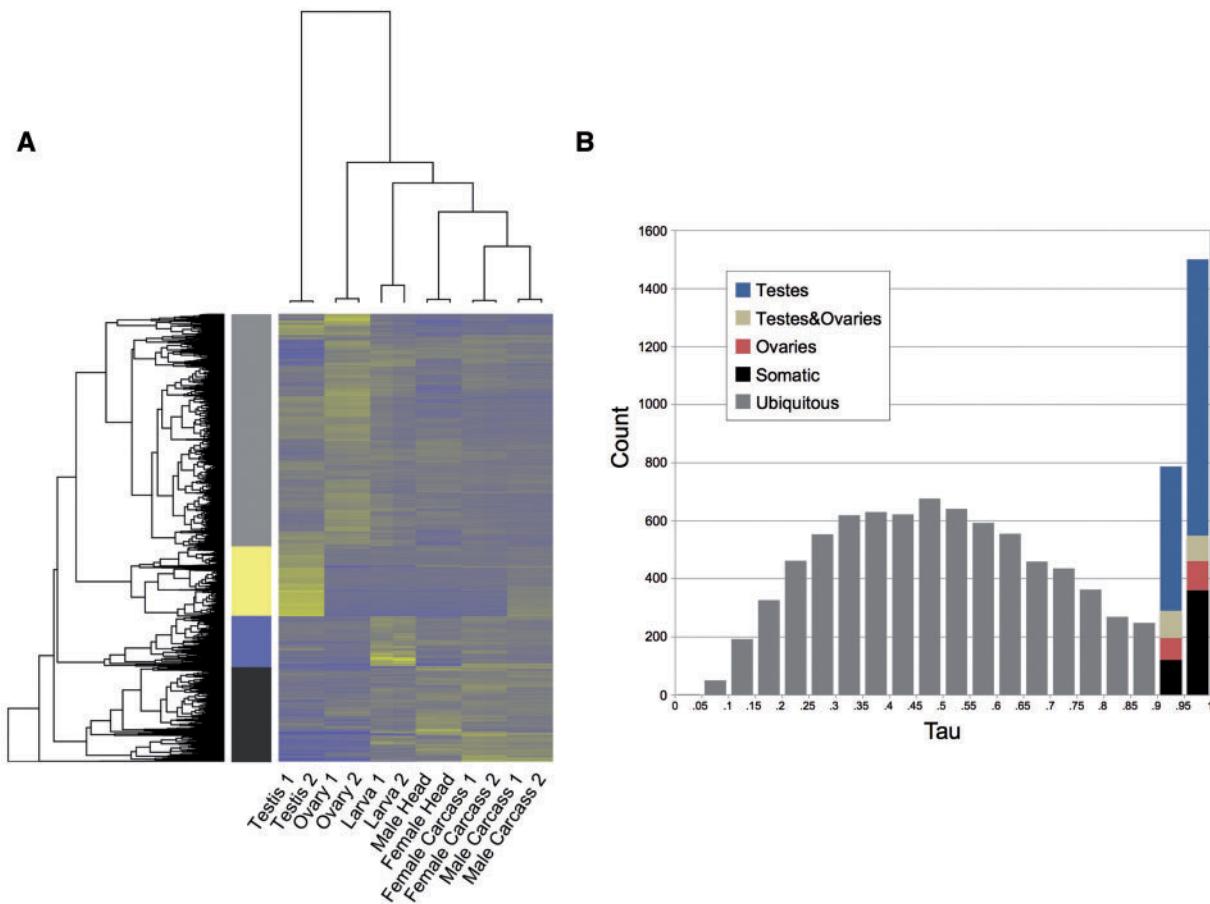


Fig. 3.—*Teleopsis dalmanni* tissue gene expression. (A) Expression cluster heatmap dendrogram illustrating repeatability of expression profiles between samples. (B) Histogram of tissue-specific gene expression values as measured by Tau statistic with expression specificity color coded as indicated in the legend.

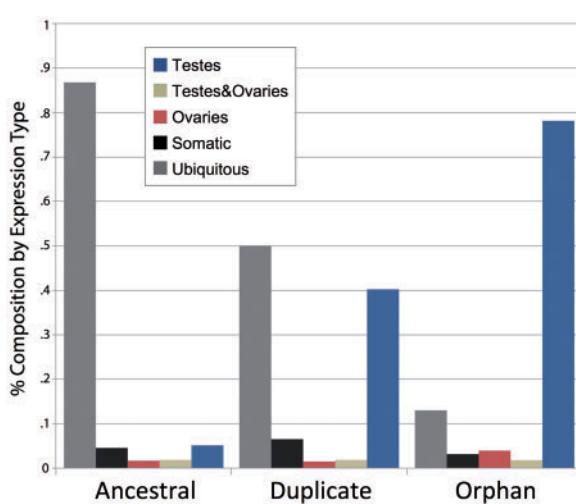


Fig. 4.—Relationship between tissue-specific gene expression and novel gene origination in *Teleopsis dalmanni*. In this analysis, orphan genes that had also undergone duplication were scored as orphan. Somatic-specific genes include genes expressed in the adult head, larva, and adult carcasses and are specific to one of those tissues.

an autosome (*S. beccarii*: $\chi^2 = 98.4$, $P > 0.0001$; *T. quinqueguttata*: $\chi^2 = 101.4$, $P > 0.0001$), and the degree of overrepresentation is similar in all three species (supplementary table S10, Supplementary Material online).

Testes-Specific Genes Move onto the X Chromosome

Biased sex chromosome gene content is often driven by gene movement (Betran et al. 2002; Meisel et al. 2009; Vibranovski, Zhang, et al. 2009) and, despite the strong synteny relationship between diopsids and *Drosophila*, numerous genes have moved on and off the X chromosome in this family. Based on the reconstruction of chromosomal location on the gene family trees, we identified nearly 450 unambiguous gene movements throughout stalk-eyed flies (fig. 7). Genes with testes-specific gene expression were significantly more likely to move on or off of the X chromosome than genes with other expression patterns ($\chi^2 = 1238.6$, $P < 0.0001$; table 2). Other than their association with testes expression, however, gene movers were not overrepresented for any Gene Ontology category. When analyzed across all three species, there is a significant effect of the timing of gene movement (basal to the

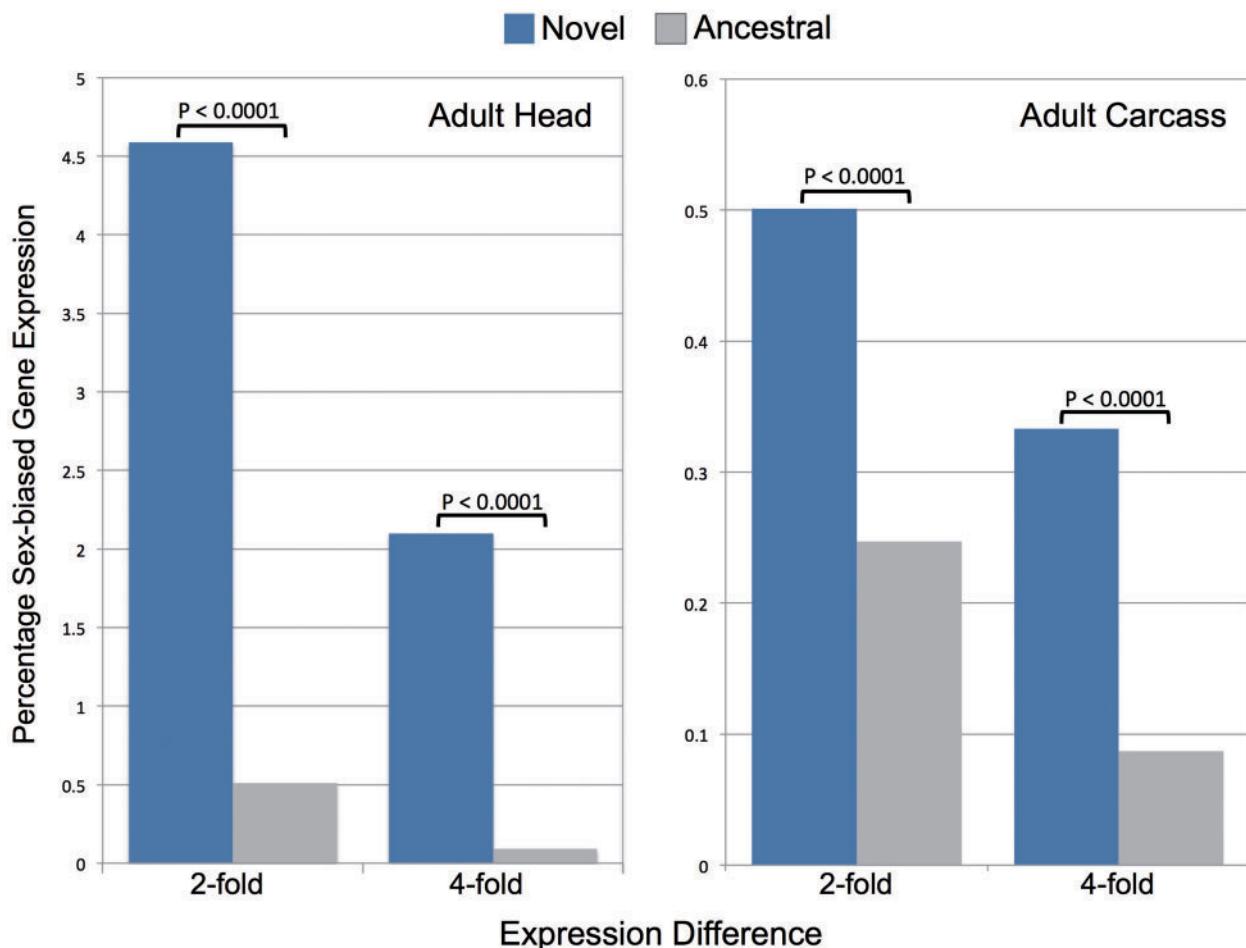


FIG. 5.—Relationship between novel genes and somatic sex-biased gene expression in *Teleopsis dalmanni*. For the adult head, sex-biased gene expression was calculated based on fold differences only; for the adult carcass, sex-biased gene expression was calculated based on fold differences and statistically significant differences between males and females.

Table 2

X-Linkage and Chromosomal Movement of Genes with Different Expression Patterns

Expression Category	No. of Genes	% X-linkage	% Deviation	% Move	% Deviation
Ubiquitous	7,330	15.6	-12.1	1.68	-64.0
Somatic	377	14.3	-19.5	1.07	-77.0
Ovaries	138	29.7	66.9	5.3	10.0
Testes and ovaries	155	19.4	8.7	8.39	81.2
Testes	1,073	32.0	79.6	26.3	438.3
Total	9,073	17.8		4.54	

family vs. within the family) on the rate of movement onto or off of the X ($\chi^2 = 10.3$, $P < 0.005$). Within diopsids, genes are more likely to move onto the X than an autosome ($\chi^2 = 7.4$, $P < 0.01$). Given the masculinization of the X for testes-specific genes, we also assessed the extent gene movement caused this biased distribution. Genes with testes-specific expression are more likely to move onto the X than ubiquitously expressed genes across the entire tree ($\chi^2 = 14.3$, $P < 0.001$; fig. 7). Within diopsids, testes-specific genes are significantly

more likely to move onto the X chromosome ($\chi^2 = 11.9$, $P < 0.001$) if we assume the rate of gene movement off of the X increases with disproportionate creation of testes-specific genes on the X (see Materials and Methods), but not if we expect an equal rate of movement between the chromosomes ($\chi^2 = 3.5$, $P > 0.05$). Overall, however, the impact of movement of testes-specific genes on the masculinization of the X chromosome is minimal. The X chromosome contains 152 more testes-specific genes than expected based

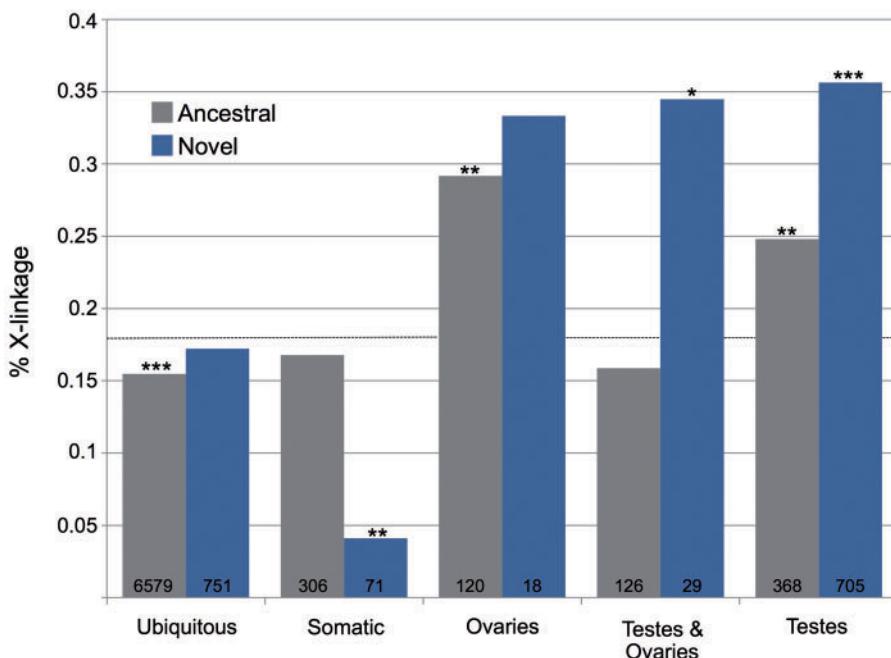


Fig. 6.—Relationship between X-linkage and tissue-specific gene expression in *Teleopsis dalmanni*. Numbers at the bottom of each column indicate the sample size for that category. The dashed line indicates the overall proportion of all X chromosome genes. Somatic-specific genes include genes expressed in the adult head, larva, and adult carcasses and are specific to one of those tissues. Categories with significant under- and overrepresentation on the X chromosome relative to its overall size are indicated. * $P < .05$, ** $P < 0.005$, *** $P < 0.0005$.

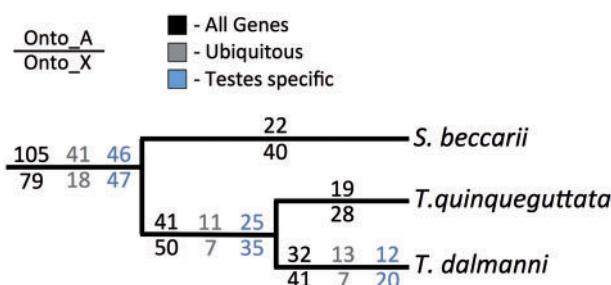


Fig. 7.—Pattern of gene movement in stalk-eyed flies.

on the overall size of the chromosome, but differential gene movement accounts for only 13 of these additional genes.

Novel Genes Exhibit Rapid Evolution

We examined the sequence divergence (dN/dS) of over 7,000 *T. dalmanni* genes that had a clear homolog in *T. quinqueguttata*. An Analysis of Variance (ANOVA) that included chromosome location, gene age, tissue specificity, and their interactions found that both gene age and tissue specificity, but not chromosome location, had significant effects (supplementary table S11, Supplementary Material online). Novel genes evolved significantly faster than ancestral genes, while genes expressed in the gonads evolved faster than somatic-specific genes and ubiquitously expressed genes (supplementary fig. S7, Supplementary Material online).

Novel Genes Are Misexpressed When on a Driving X Chromosome

A previous genomic study quantifying the expression profiles of drive and nondrive testes in *T. dalmanni* revealed that differentially expressed genes were disproportionately testes specific (Reinhardt et al. 2014). Therefore, we investigated whether these genes were also more likely to be diopsid specific. Because the X chromosome contains the drive element and is substantially overrepresented for differentially expressed genes, we analyzed the X chromosome and the autosomes separately. On the X chromosome, novel testes-specific genes were twice as likely to be differentially expressed between drive and nondrive males as genes in other gene categories ($\chi^2 = 36.557$, $P < 0.0001$; supplementary fig. S8, Supplementary Material online). The autosomes exhibited no significant difference in the number of differentially expressed genes between drive and nondrive samples among genes with different lineage- and tissue specificity ($\chi^2 = 5.433$, $P = 0.143$; supplementary fig. S8, Supplementary Material online). To assess whether paralogous gene copies may exist exclusively in one drive/nondrive genotype and contribute to the gene family diversity in *T. dalmanni*, the drive and nondrive testes samples were assembled independently. All *T. dalmanni*-specific duplications, however, were located in both the drive and nondrive assemblies and, therefore, do not result from gene copy or allelic variant differences between the samples.

X-Linkage Constrains Testes Expression

The formation of a novel X chromosome in diopsids has implications regarding the evolution of dosage effects for genes expressed in males. To evaluate the degree of dosage compensation in *T. dalmani*, we compared the average level of gene expression between autosomal and X-linked genes for tissues that we measured in a single sex. In order to minimize the impact of any difference in the distribution of genes with low expression values, we limited our analysis to genes with an FPKM value greater than 5. In addition, the relative distribution of tissue-specific genes on the autosomes and X chromosome can influence the interpretation of A:X expression ratios, particularly if these genes are expressed at substantially different levels than non-tissue-specific genes. Therefore, given the biased distribution of some tissue-specific genes on the X chromosome, we conducted an analysis of variance for each tissue that included chromosome location (X or A), gene age (novel vs. ancestral), tissue-specific expression (or not), and interaction terms (supplementary table S12, Supplementary Material online). In most somatic tissues (except male carcass), novel genes are expressed at a significantly higher level than ancestral genes. For genes expressed in the heads of either sex, those showing head specificity have significantly higher expression values than genes expressed in multiple tissues. In tissues other than the testes, there was little evidence of expression differences between autosomal and X-linked genes (female head showed marginal significance). In the testes, however, there are highly significant effects of chromosome, testes-specific expression, and their interaction, but not gene age (supplementary table S12, Supplementary Material online). Autosomal genes are expressed at a higher level than X chromosome genes and this difference is magnified for testes-specific genes.

The incomplete dosage compensation on the X chromosome may represent an unfavorable environment for genes that require high expression levels during spermatogenesis. Therefore, we examined the relationship between X chromosome gene content and overall expression intensity. Genes highly expressed in the testes are less likely to reside on the X chromosome. This effect is strongest for testes-specific genes but also occurs for non-testes-specific genes (supplementary fig. S9, Supplementary Material online). The X-linkage of highly expressed testes-specific genes (>100 FPKM) is roughly half that of other testes-specific genes (supplementary fig. S9, Supplementary Material online). It is noteworthy, though, that, while the highly expressed testes-specific genes are substantially underrepresented on the X chromosome relative to other testes-specific genes, they are still overrepresented relative to genes expressed in other tissues. Furthermore, the impact of expression on X-linkage is not isolated to novel genes (as might be expected if older genes were more efficiently dosage compensated) as ancestral genes in both the testes-specific and non-testes-specific

categories exhibited significant differences (supplementary fig. S9, Supplementary Material online).

Discussion

Transcriptional Diversity Is Associated with Sperm Production in Stalk-Eyed Flies

In this study, we examined the transcriptional profile associated with spermatogenesis within three stalk-eyed fly species spanning the phylogenetic breadth of the family. As with previous studies on *Drosophila* and other model organisms (Chintapalli et al. 2007; Baker et al. 2011; Meiklejohn and Presgraves 2012; Soumillon et al. 2013), the testes exhibit dynamic transcriptional complexity. For *T. dalmani*, the testes have, far and away, the most divergent expression patterns of all the tissues examined. This tissue has the lowest correlation in expression values with the other tissues and the highest proportion of tissue-specific gene expression, with nearly five times as many genes expressed exclusively in the testes as the next highest sample (larvae). Not only is there a huge catalog of genes that are expressed primarily within the testes, but most of these genes are specific to stalk-eyed flies. Three quarters of all testes-specific genes are novel to diopsids compared with only a quarter for other tissues. This pattern supports analyses in other taxa showing that testes gene function is the primary factor driving gene creation and genetic diversity within the genome (Chintapalli et al. 2007; Mikhaylova et al. 2008; Meiklejohn and Presgraves 2012). Additional sampling of testes gene expression at specific developmental stages will be necessary to determine whether, as in other systems (White-Cooper 2010; Soumillon et al. 2013), novel, testes-specific genes in stalk-eyed flies are disproportionately expressed in later stages of spermatogenesis (i.e., spermatocytes and spermatids).

Overall, the pattern of novel gene creation in stalk-eyed flies is at least similar to and possibly more dynamic than what has been found in *Drosophila*. Zhang, Vibranovski, Krinsky, et al (2010) quantified the number of new *D. melanogaster* genes originating on various branches within the genus and identified 844 duplicates and 103 orphan genes since the split between *D. melanogaster* and *D. virilis*. Although a comprehensive analysis of clade ages has not been conducted for stalk-eyed flies yet, we estimate, based on nucleotide divergence, that the split between *T. dalmani* and *S. beccarii* is similar to that for *D. melanogaster* and *D. virilis*. For genes that have a single representative for all diopsid and *Drosophila* species in a given gene family (and therefore are strict 1-to-1 orthologs), the average level of protein divergence is 0.752 between *T. dalmani* and *S. beccarii* and 0.762 between *D. melanogaster* and *D. virilis*. If we assume the duplicate diopsid genes that have an ambiguous phylogenetic origin are distributed on the branches of the diopsid tree in proportion to the number of genes that are

unambiguously mapped to each branch, then 930 duplicate genes have originated in *T. dalmani* since the split with *S. beccarii*. If we express the duplication quantity relative to the study size (i.e., the number of duplicates divided by the total number of genes identified), the proportion of gene duplication is 6.8% for *D. melanogaster* and 7.8% for *T. dalmani*. Regardless of putative differences in gene creation between the clades, it is clear that the evolutionary turnover in testes-specific gene expression is so great that there is very little similarity in the testes-specific expression profiles of *Teleopsis* and *Drosophila*. Approximately three-quarters of the genes exhibiting testes-specific expression in either *Teleopsis* or *Drosophila* have no direct homolog in the other species. This divergence is not limited to downstream genes but extends to essential components of sperm development. In a recent summary, Catron and Noor (2008) highlighted the interactions among 22 essential regulatory genes involved in the *Drosophila* spermatogenesis pathway. For 17 of these genes, including the 2 primary genes—*bag of marbles* and *benign gonial cell neoplasm*—that initiate spermatocyte formation, we did not identify a direct homolog in stalk-eyed flies, indicating rapid evolutionary change in the core components of the pathway (Flores et al. 2015).

Massive Expansion of Basal Transcription Machinery Genes

Although the high volume of gene creation associated with testes expression appears to be a common feature of organisms, there are numerous patterns identified in this study that are unique to diopsids. The most remarkable class of testes duplicates in stalk-eyed flies involves the basal transcription genes that control the initiation and progression of transcription in nearly all cells. The entire basal transcriptional machinery (BTM) includes the Rpl1 complex and six associated transcription factor complexes—TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIH (Thomas and Chiang 2006). In many species, members of the TFIIA and TFIID complexes contain duplicated genes that function in spermatogenesis (Hiller et al. 2004; Freiman 2009; White-Cooper 2010). For instance, the TFIID complex in *Drosophila* contains several testes-specific paralogs of TAF genes (termed tTAFs) that control the expression of hundreds of genes and are required for progression of the meiotic cell cycle. Expression of the tTAF genes is necessary to activate the large suite of testes-specific genes required for terminal differentiation within spermatocytes and mutations in these genes result in meiotic arrest and sterility (Hiller et al. 2004; Chen et al. 2005). In contrast to TFIIA and TFIID, the genes in the other five complexes are single copy in all eukaryotes, from yeast to humans, and are strongly conserved at the protein level. Stalk-eyed flies, however, possess numerous duplicated copies of genes from these five complexes, nearly all of which are expressed at high levels in the testes. All but 2 of the 12 members of the Rpl1 complex have at least 1 testes-

specific duplicate and, for the 5 genes comprising the TFIIB, TFIIE, and TFIIF complexes, there are at least 13 additional testes-specific duplicates among the 3 diopsid species. Over half of the testes-specific BTM duplicates are expressed at higher levels within the testes than their original copies, suggesting functional significance. It is tempting to speculate that the testes-specific BTM duplicates are integral to proper sperm development and possibly function as an alternative transcription system within the testes, but little is known about the specific roles that these genes play during spermatogenesis. As we discuss in more detail below, it is possible that the BTM diversity has evolved in response to parasitic nucleic acids that utilize their host's transcriptional machinery for their own gene expression (Madhani 2013).

What Drives Novel Gene Creation in Diopsids?

The rapid turnover of genes involved in reproductive biology, particularly those that function in the testes, is driven by a diverse array of evolutionary pressures including sperm competition (Ramm et al. 2014), sexual antagonism (Gallach and Betran 2011), and genomic conflict arising from germline pathogens and selfish genetic elements (Meiklejohn and Tao 2010; Madhani 2013). The results from this study provide insight into the relative importance of these factors in driving genetic and transcriptional diversity. With respect to sperm competition, it is noteworthy that the total number of testes-expressed novel genes for each species is high regardless of the type of mating system they possess. *Teleopsis dalmani* represents a classic sexually selected system characterized by a highly exaggerated, sexually dimorphic male trait (eye-stalks) that functions in a mating system involving both male–male competition and female choice (Wilkinson, Kahler, et al. 1998; Small et al. 2009). This species is highly promiscuous, with both males and females mating several times a day (Kotrba 1996; Baker, Ashwell, et al. 2001; Corley et al. 2006). Alternatively, *T. quinqueguttata* and *S. beccarii* have much smaller eyestalks that are sexually monomorphic or slightly dimorphic and there is little evidence of male competition or female choice in these species. *Teleopsis quinqueguttata* transfers more sperm per mating (Kotrba 1996) and has a substantially lower mating rate (Wilkinson et al. 2003) than *T. dalmani*. No data are available on these traits for *S. beccarii*. Given the more intense sperm competition operating in *T. dalmani*, we might expect significantly more novel testes-specific genes in this species if sperm competition is the primary evolutionary force driving gene creation. Although there are more species-specific testes-specific genes on the branch leading to *T. dalmani* than the branch leading to *T. quinqueguttata*, across the entire tree the species differ by only 2.1% in the total volume of novel gene creation within the testes. Evidence from other dipteran systems also supports the notion that sperm competition alone cannot explain the elevated genetic

diversity associated with testes gene expression. For instance, in *An. gambiae* there is virtually no sperm competition because females mate only once in their lifetime (Triplet et al. 2003), but the testes still exhibit the highest proportion of tissue-specific expression (Baker and Russell 2011).

Whether from sperm competition or other evolutionary forces, the presence of strong selection acting on male reproductive traits can result in substantial conflict between the sexes over how the genome evolves. Males and females share virtually identical genomes and, therefore, the evolutionary trajectory of a gene represents a compromise between the adaptive interests of each sex. In some cases, alleles that are beneficial for females may be harmful to males, and vice versa, leading to antagonism between the sexes over the evolutionary outcome of the gene (Rice 1984). When strong sexual selection is operating on reproductive traits and behavior, sexual conflict will be heightened. The creation of new genes, particularly through gene duplication, provides a fundamental mechanism for resolving sexual conflict by providing separate copies that can each approach the adaptive optima of a given sex (Connallon and Clark 2011; Gallach and Betran 2011), and results from this study provide some support for this hypothesis. In somatic tissue, duplicate genes were significantly more likely to exhibit sex-biased gene expression than single-copy ancestral genes. In the adult heads, novel genes are about ten times more likely to be differentially expressed between the sexes while the difference is about double in the adult carcass. Within the germline, however, although sexual conflict may be responsible for some of the high volume of gene duplication associated with the testes, it is unlikely to represent the primary mechanism driving this diversity because so many of the testes-specific genes are derived from other testes-specific genes. Once a gene achieves male-specific expression, the impact of intralocus sexual antagonism should be minimal because selection cannot operate directly on that gene in females. In *T. dalmanni*, over 70% of all testes-specific duplicates evolved from a testes-specific ancestor suggesting that the vast majority of gene creation in this tissue is driven by selection pressures specific to the male germline.

Genomic conflict is not limited to the interaction between the sexes but may also arise from parasitic and selfish genetic elements, such as meiotic drive, transposable elements, and retroviruses (Meiklejohn and Tao 2010; Madhani 2013). Meiotic drive causes the nonrandom segregation of gametes, promoting themselves at the expense of gametes that do not carry the drive element. In males that carry drive on the X chromosome, most Y-bearing sperm fail to develop properly, resulting in the transfer to females of primarily X-bearing sperm and, consequently, the production of a highly female-biased sex ratio. This shift in the population sex ratio will impact the fitness of autosomal genes (Lande and Wilkinson 1999). Therefore, genes on the autosomes or Y chromosome that suppress the effects of drive are expected to evolve. Recurrent bouts of coevolution between drive elements and

suppressors (Meiklejohn and Tao 2010) may result in the continuous production of new genes to fuel the process. The occurrence of drive systems, both extant and cryptic, is widespread in *Teleopsis* (Wilkinson et al. 2003, 2014) and genes that are differentially expressed between drive and nondrive males are disproportionately testes specific (Reinhardt et al. 2014). Results from this study reveal that these testes-specific genes are also more likely to represent novel diopsid-specific genes providing additional support that the interaction between selfish genetic elements and the rest of the genome may contribute to the rapid turnover of testes-specific genes. Similar processes may operate for other types of parasitic genetic elements. In a recent essay discussing the complexity of eukaryotic gene expression, Madhani (2013) highlighted the potential significance of parasitic nucleic acids in shaping the diversity of the cellular complexes required for proper gene expression. He speculated that, because transposons and retroviruses need to utilize a host's transcriptional machinery for their own gene expression, there will be strong selection to minimize their access to this machinery, particularly in the germline. Therefore, testes-specific copies of essential transcriptional activators, such as *Drosophila* tTAFs and the abundant duplicate BTM genes found in stalk-eyed flies, may have evolved as a defense mechanism and reflect the outcome of an arms race between hosts and parasitic nucleic acids. Given the remarkable diversity of BTM genes found in this study, stalk-eyed flies provide an ideal system in which to investigate this hypothesis.

The extreme volume of testes-specific duplication, combined with the abundant testes expression accompanying novel gene origination (Levine et al. 2006; Begun et al. 2007), has led to the suggestion that the testes may provide a source of genetic diversity for other tissues (Kaessmann 2010). Overall, we found limited evidence for such an "out-of-the testes" pattern. Of 1,064 novel genes expressed in tissues other than the testes, only 3 appear to have evolved from a testes-specific ancestor. However, data on tissue-specific gene expression from several additional species within the family are required to quantify the extent to which testes-specific gene expression shapes genetic diversity throughout the rest of the organism.

X Chromosome Masculinization

Numerous studies on a diverse set of organisms have established that sex chromosomes experience unique evolutionary pressures impacting their gene content and rate of evolutionary change (Ellegren and Parsch 2007; Parsch and Ellegren 2013). In most XY systems studied to date, the X chromosome exhibits some degree of feminization, demasculinization, or both (Parisi et al. 2003; Yang et al. 2006; Sturgill et al. 2007; Zhang, Vibranovski, Krinsky, et al. 2010; Baker et al. 2011; Meisel et al. 2012; Allen et al. 2013; Gao et al. 2014; Vicoso and Bachtrog 2015). Three primary hypotheses have been

proposed to explain these patterns. First, sexual antagonism resulting from an imbalance in the amount of time the X chromosome resides in each sex creates an environment that favors female-biased alleles, or alleles favorable to a female's adaptive interests, if the allelic effects are not recessive in most cases (Rice 1984). Second, expression of the X chromosome may be reduced or shut down during specific stages of spermatogenesis (e.g., meiotic sex chromosome inactivation—MSCI) making it a hostile location for genes required during this developmental process (Betran et al. 2002; Vibranovski, Lopes, et al. 2009). Finally, incomplete dosage compensation, particularly for genes expressed at a high level, may cause autosomes to be a more favorable location for these genes (Vicoso and Charlesworth 2009).

The X chromosome in stalk-eyed flies is derived from an autosome in other flies (e.g., 2L in *D. melanogaster*) and results from this study clearly place the origin of this X chromosome at the base of the diopsid tree prior to the diversification of the family. The most noteworthy feature of the diopsid novel X chromosome is that it is highly masculinized, with nearly twice as many testes-specific genes as expected based on its size. This masculinization appears to be driven by three separate processes—increased gene creation on the X, disproportionate shift in testes-specific expression for ancestral X-linked genes, and biased movement onto the X for testes-specific genes—with novel gene creation representing the most substantial factor. Thus, the gene content of the X chromosome in *T. dalmani* is distinct from the pattern found in numerous other fly species (Parisi et al. 2003; Sturgill et al. 2007; Zhang, Vibranovski, Krinsky, et al. 2010; Baker et al. 2011; Allen et al. 2013; Vicoso and Bachtrog 2015). A detailed examination of the phylogenetic origination and chromosomal location of male-biased genes in *D. melanogaster* (Zhang, Vibranovski, Krinsky, et al. 2010; Gao et al. 2014) revealed a relationship between X-linkage and gene age, with the X chromosome being enriched for young male-biased genes (<10 million years old) and depauperate for older male-biased genes. *Teleopsis dalmani*, however, exhibits strong X chromosome masculinization for novel testes-specific genes of all ages and more movement of these genes onto the X chromosome than off of it for all branches of the diopsid phylogeny. In addition, the X chromosomes of *T. quinqueguttata* and *S. beccarii* are enriched for novel genes expressed in the testes, suggesting that this pattern is consistent throughout the family. It is also noteworthy that novel genes that are expressed specifically in a somatic tissue in *T. dalmani* are significantly underrepresented on the X chromosome. This suggests that the enrichment of the X chromosome with testes-specific genes is not caused by a general relationship between tissue-specific expression and X chromosome linkage, as has been found, albeit as a negative relationship, in *Drosophila* (Mikhaylova and Nurminsky 2011; Meisel et al. 2012).

The disproportionate emergence of novel genes on the diopsid X chromosome may be driven by the fact that the

hemizygous state of this chromosome in males is a more efficient environment for the fixation of new genes (Charlesworth et al. 1987; Meisel et al. 2012). Essentially, new alleles are immediately exposed to selection when on the single male X, whereas any beneficial impact associated with mutations in these genes may be hidden by dominance effects when on an autosome. This process assumes that beneficial fitness effects associated with new genes will generally be recessive, as with faster-X models of chromosome evolution (Meisel et al. 2012), but there is currently little data supporting this assertion in any species, including diopsids. In *Drosophila*, once these genes become fixed, various selective forces—sexual antagonism, MSCI, and dosage effects—appear to reshape the composition of the X chromosome through differential gene survival or gene movement (Meisel et al. 2009, 2012; Vibranovski, Zhang, et al. 2009; Vicoso and Charlesworth 2009; Zhang, Vibranovski, Krinsky, et al. 2010; Assis et al. 2012; Gao et al. 2014). The results from our study suggest that none of these forces are prominent within stalk-eyed flies given the extreme masculinization of the X chromosome throughout the evolutionary history of the family. The reasons for the selective differences between stalk-eyed flies and *Drosophila* are not obvious at this point. As discussed earlier, many of the novel duplicate genes originate with testes-specific expression and, therefore, are unlikely to be affected by sexual antagonism because they have little impact on female fitness. To the extent that sexual antagonism influences the evolution of testes-specific genes, we have no evidence of differences between *Drosophila* and stalk-eyed flies regarding the relative proportion of dominant versus recessive mutational effects for these genes that might cause differences in X chromosome gene content.

The exact nature of X chromosome gene regulation in the *Drosophila* male germline is controversial (Vibranovski, Lopes, et al. 2009; Meiklejohn et al. 2011; Mikhaylova and Nurminsky 2011; Meiklejohn and Presgraves 2012; Vibranovski et al. 2012), with recent studies finding expression differences between the chromosomes consistent with either incomplete dosage compensation (Meiklejohn et al. 2011; Meiklejohn and Presgraves 2012; Mikhaylova and Nurminsky 2012) or MSCI (Hense et al. 2007; Vibranovski, Lopes, et al. 2009; Vibranovski et al. 2012; Kemkemer et al. 2014). Regardless of the mechanism, reduced expression for X-linked genes in the testes may impact the functional viability of genes important in spermatogenesis, making the autosomes a more favorable location for these genes. A comparison of the overall level of gene expression between X-linked and autosomal genes within the testes of *T. dalmani* suggests that germline dosage compensation is incomplete in this species. This difference is most pronounced for testes-specific genes and appears to impact their chromosomal location, as highly expressed testes-specific genes are half as abundant on the X chromosome as testes-specific genes expressed at lower levels. It is also possible that this downregulation is caused by

some form of X chromosome inactivation if the expression of novel testes genes is concentrated during a specific stage of spermatogenesis that coincides with the downregulation of the X chromosome. Overall, given the selective pressure relating to dosage effects influencing X-linked testes genes in *T. dalmani* and the impact these factors have in several model organisms (Vibranski, Lopes, et al. 2009; Vicoso and Charlesworth 2009; Bachtrog et al. 2010; Zhang, Vibranski, Landback, et al. 2010; Baker et al. 2011; Meisel et al. 2012), the extreme masculinization of the diopsid X chromosome is puzzling. In mice, the X chromosome, while inactivated during meiosis, is enriched for testes-specific genes at pre- and postmeiotic stages of spermatogenesis (Khil et al. 2004; Zhang, Vibranski, Landback, et al. 2010; Soumillon et al. 2013). It is possible that a similarly stark divide exists within the testes of stalk-eyed flies in which certain developmental stages are unfavorable to the expression of X-linked genes while others exert no limitations on expression. Overall, research in model organisms such as *Drosophila* and mice have shown that the forces shaping the origin, evolution, and genomic distribution of sex-biased genes are complex and results from this study expand the diversity of patterns associated with this issue. Diopsids exhibit several unique features of sex-biased gene expression and sex chromosome composition that are of evolutionary significance, highlighting the importance of examining nonmodel organism systems to understand the complexity of these interactions.

Supplementary Material

Supplementary tables S1–S12, figures S1–S9, and files S1–S3 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

Acknowledgments

We thank C. Brand and W. Shell for sample preparation, W. Yu and H. Xu for conducting CGH hybridizations, and two anonymous reviewers for providing insightful comments on the manuscript. Funding for this work was provided by the National Science Foundation grants DEB-0951816 to R.H.B. and DEB-0952260 to G.S.W.

Literature Cited

Allen SL, Bonduriansky R, Chenoweth SF. 2013. The genomic distribution of sex-biased genes in *Drosophila serrata*: X chromosome demasculinization, feminization, and hyperexpression in both sexes. *Genome Biol Evol*. 5:1986–1994.

Aoyagi N, Wasserman DA. 2000. Genes encoding *Drosophila melanogaster* RNA polymerase II general transcription factors: diversity in TFIID and TFIID components contributes to gene-specific transcriptional regulation. *J Cell Biol*. 150:F45–F50.

Assis R, Zhou Q, Bachtrog D. 2012. Sex-biased transcriptome evolution in *Drosophila*. *Genome Biol Evol*. 4:1189–1200.

Bachtrog D, Toda NRT, Lockton S. 2010. Dosage compensation and demasculinization of X chromosomes in *Drosophila*. *Curr Biol*. 20:1476–1481.

Baker D, et al. 2011. A comprehensive gene expression atlas of sex- and tissue-specificity in the malaria vector, *Anopheles gambiae*. *BMC Genomics* 12:296.

Baker DA, Russell S. 2011. The role of testis-specific gene expression in sex chromosome evolution of *Anopheles gambiae*. *Genetics* 1117–1120.

Baker RH, Ashwell RIS, et al. 2001. Effects of multiple mating and male eye span on female reproductive output in the stalk-eyed fly, *Cyrtodiopsis dalmani*. *Behav Ecol*. 12:732–739.

Baker RH, Morgan J, Wang X, Boore JL, Wilkinson GS. 2009. Genomic analysis of a sexually-selected character: EST sequencing and microarray analysis of eye-antennal imaginal discs in the stalk-eyed fly *Teleopsis dalmani* (Diopsidae). *BMC Genomics* 10:361.

Baker RH, Narechania A, Johns PM, Wilkinson GS. 2012. Gene duplication, tissue-specific gene expression and sexual conflict in stalk-eyed flies (Diopsidae). *Phil Trans R Soc B*. 367:2357–2375.

Baker RH, Wilkinson GS. 2001. Phylogenetic analysis of sexual dimorphism and eye-span allometry in stalk-eyed flies (Diopsidae). *Evolution* 55:1373–1385.

Baker RH, Wilkinson GS. 2010. Comparative genomic hybridization (CGH) reveals a neo-X chromosome and biased gene movement in stalk-eyed flies (Genus *Teleopsis*). *PLoS Genet*. 6:e1001121.

Baker RH, Wilkinson GS, DeSalle R. 2001. The phylogenetic utility of different types of molecular data used to infer evolutionary relationships among stalk-eyed flies (Diopsidae). *Syst Biol*. 50:87–105.

Begin DJ, Lindfors HA, Kern AD, Jones CD. 2007. Evidence for de novo evolution of testis-expressed genes in the *Drosophila yakuba*–*Drosophila erecta* clade. *Genetics* 176:1131–1137.

Belo JM, Zhong L. 2009. Duplicated proteasome subunit genes in *Drosophila* and their roles in spermatogenesis. *Heredity* 103:23–31.

Betran E, Thornton K, Long M. 2002. Retroposed new genes out of the X in *Drosophila*. *Genome Res*. 12:1854–1859.

Catron DJ, Noor MAF. 2008. Gene expression disruptions of organism versus organ in *Drosophila* species hybrids. *PLoS One* 3:e3009.

Charlesworth B, Coyne JA, Barton NH. 1987. The relative rates of evolution of sex chromosomes and autosomes. *Am Nat*. 130:113–146.

Chen X, Hiller M, Sancak Y, Fuller MT. 2005. Tissue-specific TAFs counteract Polycomb to turn on terminal differentiation. *Science* 310:869–872.

Chintapalli VR, Wang J, Dow JA. 2007. Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nat Genet*. 39:715–720.

Chiu J, et al. 2006. OrthoologID: automation of genome-scale ortholog identification within a parsimony framework. *Bioinformatics* 22:699–707.

Choi E, et al. 2007. Integrative characterization of germ cell-specific genes from mouse spermatocyte UniGene library. *BMC Genomics* 8:256.

Connallon T, Clark AG. 2011. The resolution of sexual antagonism by gene duplication. *Genetics* 187:919–937.

Corley LS, et al. 2006. Highly variable sperm precedence in the stalk-eyed fly, *Teleopsis dalmani*. *BMC Evol Biol*. 6:53.

Dennis G, et al. 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol*. 4:P3.

Domazet-Loso T, Tautz D. 2003. An evolutionary analysis of orphan genes in *Drosophila*. *Genome Res*. 13:2213–2219.

Dubruille R, Marais GA, Loppin B. 2012. Repeated evolution of testis-specific new genes: the case of telomere-capping genes in *Drosophila*. *Int J Evol Biol*. 2012:708980.

Eddy EM. 2002. Male germ cell gene expression. *Endocr Rev*. 57:103–128.

Ellegren H, Parsch J. 2007. The evolution of sex-biased genes and sex-biased gene expression. *Nat Rev Genet*. 8:689–698.

Ferl R, Manak M, Reyes M. 2002. The 14-3-3s. *Genome Biol*. 3:3010–3017.

Flores HA, et al. 2015. Adaptive evolution of genes involved in the regulation of germline stem cells in *Drosophila melanogaster* and *D. simulans*. *G3 (Bethesda)* 5:583–592.

Freiman RN. 2009. Specific variants of general transcription factors regulate germ cell development in diverse organisms. *Biophys Acta*. 1789:161–166.

Gallach M, Betran E. 2011. Intralocus sexual conflict resolved through gene duplication. *Trends Ecol Evol*. 26:222–228.

Gao G, et al. 2014. A long-term demasculinization of X-linked intergenic noncoding RNAs in *Drosophila melanogaster*. *Genome Res*. 24:629–638.

Gentleman RC, et al. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. 5:R80.

Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotech*. 29:644–652.

Haerty W, et al. 2007. Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*. *Genetics* 177:1321–1335.

Hahn MW, Han MV, Han SG. 2007. Gene family evolution across 12 *Drosophila* genomes. *PLoS Genet*. 3:e197.

Han MV, Hahn MW. 2012. Inferring the history of interchromosomal gene transposition in *Drosophila* using n-dimensional parsimony. *Genetics* 190:813–825.

Hense W, Baines JF, Parsch J. 2007. X chromosome inactivation during *Drosophila* spermatogenesis. *PLoS Biol*. 5:2288–2295.

Hiller M, et al. 2004. Testis-specific TAF homologs collaborate to control a tissue-specific transcription program. *Development* 131:5297–5308.

Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. *Genome Res*. 20:1313–1326.

Katoh K, Misawa K, Kuma Ki Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 30:3059–3066.

Kemkemer C, Catalan A, Parsch J. 2014. 'Escaping' the X chromosome leads to increased gene expression in the male germline of *Drosophila melanogaster*. *Heredity* 112:149–155.

Khil PP, Smirnova NA, Romanienko PJ, Camerini-Otero RD. 2004. The mouse X chromosome is enriched for sex-biased genes not subject to selection by meiotic sex chromosome inactivation. *Nat Genet*. 36:642–646.

Korber B, et al. 2001. Evolutionary and immunological implications of contemporary HIV-1 variation. *Br Med Bull*. 58:19–42.

Kotrba M. 1996. Sperm transfer by spermatophore in Diptera: new results from the Diopsidae. *Zool J Linn Soc*. 117:305–323.

Lande R, Wilkinson GS. 1999. Models of sex-ratio meiotic drive and sexual selection in stalk-eyed flies. *Genet Res*. 74:245–253.

Levine MT, Jones CD, Kern AD, Lindfors HA, Begun DJ. 2006. Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. *Proc Natl Acad Sci U S A*. 103:9935–9939.

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

Llopis A. 2012. The rapid evolution of X-linked male-biased gene expression and the large-X effect in *Drosophila yakuba*, *D. santomea*, and their hybrids. *Mol Biol Evol*. 29:3873–3886.

Lo L, Zhang Z, Hong N, Peng J, Hong Y. 2008. 3640 unique EST clusters from the medaka testis and their potential use for identifying conserved testicular gene expression in fish and mammals. *PLoS One* 3:e3915.

Madhani HD. 2013. The frustrated gene: origins of eukaryotic gene expression. *Cell* 155:744–749.

Magnusson K, et al. 2012. Demasculinization of the *Anopheles gambiae* X chromosome. *BMC Evol Biol*. 12:69.

Meiklejohn CD, Landeen EL, Cook JM, Kingan SB, Presgraves DC. 2011. Sex chromosome-specific regulation in the *Drosophila* male germline but little evidence for chromosomal dosage compensation or meiotic inactivation. *PLoS Biol*. 9:e1001126.

Meiklejohn CD, Parsch J, Ranz JM, Hartl DL. 2003. Rapid evolution of male-biased gene expression in *Drosophila*. *Proc Natl Acad Sci U S A*. 100:9894–9899.

Meiklejohn CD, Presgraves DC. 2012. Little evidence for demasculinization of the *Drosophila* X chromosome among genes expressed in the male germline. *Genome Biol Evol*. 4:1007–1016.

Meiklejohn CD, Tao Y. 2010. Genetic conflict and sex chromosome evolution. *Trends Ecol Evol*. 25:215–223.

Meisel RP, Han MV, Hahn MW. 2009. A complex suite of forces drives gene traffic from *Drosophila* X chromosomes. *Genome Biol Evol*. 1:176–188.

Meisel RP, Malone JH, Clark AG. 2012. Disentangling the relationship between sex-biased gene expression and X-linkage. *Genome Res*. 22:1255–1265.

Mikhaylova L, Nurminsky D. 2011. Lack of global meiotic sex chromosome inactivation, and paucity of tissue-specific gene expression on the *Drosophila* X chromosome. *BMC Biol*. 9:29.

Mikhaylova LM, Nguyen K, Nurminsky DI. 2008. Analysis of the *Drosophila melanogaster* testes transcriptome reveals coordinate regulation of paralogous genes. *Genetics* 179:305–315.

Mikhaylova LM, Nurminsky DI. 2012. No severe and global X chromosome inactivation in meiotic male germline of *Drosophila*. *BMC Biol*. 10:50.

Palmieri N, Kosiol C, Schlotterer C. 2014. The life cycle of *Drosophila* orphan genes. *Elife* 3:e01311.

Parisi M, et al. 2003. Paucity of genes on the *Drosophila* X chromosome showing male-biased expression. *Science* 299:697–700.

Parsch J, Ellegren H. 2013. The evolutionary causes and consequences of sex-biased gene expression. *Nat Rev Genet*. 14:83–87.

Pitnick S, Hosken DJ, Birkhead TR. 2009. Sperm morphological diversity. In: Birkhead TR, Hosken DJ, Pitnick S, editors. *Sperm biology: an evolutionary perspective*. San Diego (CA): Academic Press. p. 69–149.

Presgraves DC, Baker RH, Wilkinson GS. 1999. Coevolution of sperm and female reproductive tract morphology in stalk-eyed flies. *Proc R Soc Lond B*. 266:1041–1047.

Presgraves DC, Severence E, Wilkinson GS. 1997. Sex chromosome meiotic drive in stalk-eyed flies. *Genetics* 147:1169–1180.

Prince EG, Kirkland D, Demuth JP. 2010. Hyperexpression of the X chromosome in both sexes results in extensive female bias of X-linked genes in the flour beetle. *Genome Biol Evol*. 2:336–346.

Proschel M, Zhang Z, Parsch J. 2006. Widespread adaptive evolution of *Drosophila* genes with sex-biased expression. *Genetics* 174:893–900.

Ramm SA, Scherer L, Ehmcke J, Wistuba J. 2014. Sperm competition and the evolution of spermatogenesis. *Mol Hum Reprod*. 20:1169–1179.

Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL. 2003. Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science* 300:1742–1745.

Reinhardt JA, et al. 2014. Meiotic drive impacts expression and evolution of x-linked genes in stalk-eyed flies. *PLoS Genet*. 10:e1004362.

Rice WR. 1984. Sex chromosomes and the evolution of sexual dimorphism. *Evolution* 38:735–742.

Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*. 11:R25.

Rogowski K, Juge F, van Dijk J, et al. 2009. Evolutionary divergence of enzymatic mechanisms for posttranslational polyglycation. *Cell* 137:1076–1087.

SAS_Institute. 2003. JMP. Version 5.0.1.2. Cary (NC): SAS Institute, Inc., Cary, NC.

Sievers F, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 7:539.

Small J, Cotton S, Fowler K, Pomiankowski A. 2009. Male eyespan and resource ownership affect contest outcome in the stalk-eyed fly, *Teleopsis dalmanni*. *Anim Behav*. 78:1213–1220.

Soumillon M, et al. 2013. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep.* 3:2179–2190.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.

Sturgill D, Zhang Y, Parisi M, Oliver B. 2007. Demasculinization of X chromosomes in the *Drosophila* genus. *Nature* 450:238–241.

Tautz D, Domazet-Lošo T. 2011. The evolutionary origin of orphan genes. *Nat Rev Genet.* 12:692–702.

Thomas MC, Chiang M-C. 2006. The general transcription machinery and general cofactors. *Crit Rev Biochem Mol.* 41:105–178.

Ting CT, et al. 2004. Gene duplication and speciation in *Drosophila*: evidence from the *Odysseus* locus. *Proc Natl Acad Sci U S A.* 101:12232–12235.

Triplet F, Touré YT, Dolo G, Lanzaro GC. 2003. Frequency of multiple inseminations in field-collected *Anopheles gambiae* females revealed by DNA analysis of transferred sperm. *Am J Trop Med Hyg.* 68:1–5.

Vibranski MD, Lopes HF, Karr TL, Long M. 2009. Stage-specific expression profiling of *Drosophila* spermatogenesis suggests that meiotic sex chromosome inactivation drives genomic relocation of testis-expressed genes. *PLoS Genet.* 5:e1000731.

Vibranski MD, Zhang Y, Long MY. 2009b. General gene movement off the X chromosome in the *Drosophila* genus. *Genome Res.* 19:897–903.

Vibranski MD, et al. 2012. Re-analysis of the larval testis data on meiotic sex chromosome inactivation revealed evidence for tissue-specific gene expression related to the drosophila X chromosome. *BMC Biol.* 10:49.

Vicoso B, Bachtrog D. 2015. Numerous transitions of sex chromosomes in Diptera. *PLoS Biol.* 13:e1002078.

Vicoso B, Charlesworth B. 2006. Evolution on the X chromosome: unusual patterns and processes. *Nat Rev Genet.* 7:645–653.

Vicoso B, Charlesworth B. 2009. The deficit of male-biased genes on the *D. melanogaster* X chromosome is expression-dependent: a consequence of dosage compensation? *J Mol Evol.* 68:576–583.

White-Cooper H. 2010. Molecular mechanisms of gene regulation during *Drosophila* spermatogenesis. *Reproduction* 139:11–21.

White-Cooper H, Bausek N. 2010. Evolution and spermatogenesis. *Phil Trans R Soc B.* 365:1465–1480.

Wilkinson GS, Christianson SJ, Brand CL, Ru G, Shell W. 2014. Haldane's rule is linked to extraordinary sex ratios and sperm length in stalk-eyed flies. *Genetics* 198:1167–1181.

Wilkinson GS, Dodson G. 1997. Function and evolution of antlers and eye stalks in flies. In: Choe J, Crespi B, editors. *The evolution of mating systems in insects and arachnids*. Cambridge: Cambridge University Press. p. 310–328.

Wilkinson GS, Johns PM, Metheny JD, Baker RH. 2013. Sex-biased gene expression during head development in a sexually dimorphic stalk-eyed fly. *PLoS One* 8:e59826.

Wilkinson GS, Kahler H, Baker RH. 1998. Evolution of female mating preferences in stalk-eyed flies. *Behav Ecol.* 9:525–533.

Wilkinson GS, Presgraves DC, Crymes L. 1998. Male eye span in stalk-eyed flies indicates genetic quality by meiotic drive suppression. *Nature* 391:276–278.

Wilkinson GS, Sanchez MI. 2001. Sperm development, age and sex chromosome meiotic drive in the stalk-eyed fly, *Cyrtodiopsis whitei*. *Heredity* 87:17–24.

Wilkinson GS, Swallow JG, Christianson SJ, Madden K. 2003. Phylogeography of sex ratio and multiple mating in stalk-eyed flies from southeast Asia. *Genetica* 117:37–46.

Yanai I, et al. 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* 21:650–659.

Yang X, et al. 2006. Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Res.* 16:995–1004.

Zhang Y, Sturgill D, Parisi M, Kumar S, Oliver B. 2007. Constraint and turnover in sex-biased gene expression in the genus *Drosophila*. *Nature* 450:233–238.

Zhang YE, Vibranski MD, Krinsky BH, Long M. 2010. Age-dependent chromosomal distribution of male-biased genes in *Drosophila*. *Genome Res.* 20:1526–1533.

Zhang YE, Vibranski MD, Landback P, Marais GA, Long M. 2010. Chromosomal redistribution of male-biased genes in mammalian evolution with two bursts of gene gain on the X chromosome. *PLoS Biol.* 8:e1000494.

Associate editor: Esther Betran