Evidence of Adaptive Evolution and Relaxed Constraints in Sex-Biased Genes of South American and West Indies Fruit Flies (Diptera: Tephritidae)

Carlos Congrains[1,*], Emeline B. Campanini[1], Felipe R. Torres[1], Víctor B. Rezende[1], Aline M. Nakamura[1], Janaína L. de Oliveira[1], André L.A. Lima[1], Samira Chahad-Ehlers[1], Iderval S. Sobrinho Jr[2], and Reinaldo A. de Brito[1]

[1]Departamento de Genética e Evolução, Universidade Federal de São Carlos, SP, Brazil

[2]Unidade Jatobá, Universidade Federal de Goiás, Jataí, GO, Brazil

*Corresponding author: E-mail: carloscongrains@gmail.com.

Accepted: January 11, 2018

Data deposition: Sequenced reads derived from reproductive tissues of *Anastrepha fraterculus* and *Anastrepha obliqua* have been deposited in the NCBI Sequence Read Archive under the accession SRP128443. The unigenes of *A. fraterculus* and *A. obliqua* assemblies are available upon request.

Abstract

Several studies have demonstrated that genes differentially expressed between sexes (sex-biased genes) tend to evolve faster than unbiased genes, particularly in males. The reason for this accelerated evolution is not clear, but several explanations have involved adaptive and nonadaptive mechanisms. Furthermore, the differences of sex-biased expression patterns of closely related species are also little explored out of *Drosophila*. To address the evolutionary processes involved with sex-biased expression in species with incipient differentiation, we analyzed male and female transcriptomes of *Anastrepha fraterculus* and *Anastrepha obliqua*, a pair of species that have diverged recently, likely in the presence of gene flow. Using these data, we inferred differentiation indexes and evolutionary rates and tested for signals of selection in thousands of genes expressed in head and reproductive transcriptomes from both species. Our results indicate that sex-biased and reproductive-biased genes evolve faster than unbiased genes in both species, which is due to both adaptive pressure and relaxed constraints. Furthermore, among male-biased genes evolving under positive selection, we identified some related to sexual functions such as courtship behavior and fertility. These findings suggest that sex-biased genes may have played important roles in the establishment of reproductive isolation between these species, due to a combination of selection and drift, and unveil a plethora of genetic markers useful for more studies in these species and their differentiation.

Key words: sex-biased gene expression, evolutionary rates, positive selection, relaxed constraints, *fraterculus* group, RNA-seq.

Introduction

Understanding the evolutionary mechanisms underlying sexual dimorphism has been a very challenging task. In this regard, an important question is how two individuals of different sexes in a species may have conspicuous sexual variation, even when both sexes share practically the same genome. Transcriptome studies indicate that most morphological sex differences are caused by divergent patterns of gene expression between sexes (Ellegren and Parsch 2007). These are referred to as sex-biased genes, which have consistently shown rapid sequence evolutionary rates across taxa (Mank et al. 2007; Meisel 2011; Huylmans et al. 2016; Yang et al. 2016; Papa et al. 2017). In *Drosophila*, male-biased expressed genes evolve particularly fast (Ellegren and Parsch 2007), which is mainly caused by adaptive evolution (Pröschel et al. 2006).

Potential explanations for such phenomenon involve sperm competition, sexual selection, and/or sexual conflict (Swanson and Vacquier 2002). If this hypothesis is at least partially true, some of the products of these genes might elicit pre- or postmating barriers which may, ultimately, play important roles reinforcing species boundaries (Snook et al. 2009;

Gavrilets 2014). Indeed, accessory gland proteins secreted in males' seminal fluid in *Drosophila* influence the females' physiology and behavior (Eberhard and Cordero 1995; Ram and Wolfner 2007) and tend to evolve under positive selection (Swanson et al. 2001), which may reflect a role on reproductive isolation in the first stages of speciation. Furthermore, several female reproductive proteins from *Drosophila* have also been reported to evolve under positive selection (Swanson et al. 2004; Panhuis and Swanson 2006). Female proteins that interact with rapidly evolving male proteins may evolve faster because of coevolution (Haerty et al. 2007). In addition, proteins in the external layer of the eggshell (chorion) have been reported to evolve adaptively due to possible role in the sperm–egg and/or egg–environment interactions (Jagadeeshan and Singh 2007).

Despite the evidence of the contribution of positive selection on sex-biased genes in *Drosophila*, there are alternative evolutionary explanations for rapid evolution on such genes. Studies have demonstrated that selection is weakened when trait (or gene) expression is limited to a fraction of individuals such as sex-biased genes, resulting in an increased segregation of slightly deleterious variation, which can reach fixation by genetic drift (Van Dyken and Wade 2010; Purandare et al. 2014). As a consequence, not only polymorphism levels are increased on such genes, but they also evolve faster. In fact, it has been demonstrated that relaxed constraints, genetic drift, or/and an increased segregation of slightly deleterious variation have an important impact on the evolution of male-specific genes (Gershoni and Pietrokovski 2014; Harrison et al. 2015). Furthermore, sex-biased genes tend to have narrower expression pattern than unbiased genes, that may imply less pleiotropy and functional constraints, showing faster evolution because of relaxed purifying selection (Mank et al. 2008).

Such as in reproductive tissues, other tissues may also express reproductive-related proteins. Sex pheromones are important in controlling reproductive behavior (Howard and Blomquist 2005). Pheromones and other environmental olfactory cues are perceived as taste and olfactory stimuli and then processed by the chemosensory system in organs located mainly in the head (such as antennae) (Kohl et al. 2015). Among the genes involved in this process, there are sets of gene families that encode for proteins involved in ligand-binding (odorant binding proteins and chemosensory proteins) and receptor functions (odorant receptors, gustatory receptors, ionotropic receptors, and sensory neuron membrane proteins) (Jin et al. 2008; Sánchez-Gracia et al. 2011). The molecular evolution of these protein families has been widely studied in insects and has revealed that several of these genes evolve under positive selection under a birth-and-death process that leads to a rapid gene turnover (Sanchez-Gracia et al. 2009; Brand et al. 2015; Campanini and de Brito 2016).

Here we investigate genes expressed in reproductive and head tissues of two closely related species, South American fruit flies (*Anastrepha fraterculus*) and West Indies fruit flies (*Anastrepha obliqua*), which belong to the *fraterculus* group (Norrbom et al. 1999). Taxonomic identification of some species within this group based only on morphology is difficult due to overlapping variation even in the aculeus, which is one of the key traits in the systematics of this group (Zucchi 2000; Perre et al. 2016). Molecular phylogeny of the *fraterculus* group based on the mitochondrial gene COI showed polyphyly for these two species (Smith-Caldas et al. 2001). However, phylogenetic analyses using nuclear *loci* revealed that *A. obliqua* is a monophyletic lineage (Scally et al. 2016) and not as closely related to *A. fraterculus* as other species in the group, though there is evidence of historical introgression between these lineages (Scally et al. 2016; Díaz F, Lima ALA, Nakamura AM, Fernandes F, Sobrinho I, de Brito RA, unpublished data). Furthermore, these species may produce viable hybrids in laboratory with descendants of some combinations obeying Haldane's rule (dos Santos et al. 2001), and as they are found in sympatry in several regions, it is possible that current introgression may still occur in nature. Therefore, it is possible that *A. fraterculus* and *A. obliqua* have diverged recently while retaining some gene flow, emphasizing the importance of identifying genomic regions that responded to selection and may have had a leading role on their differentiation as has been proposed for other organisms with similar speciation patterns (Feder et al. 2012).

We generated transcriptomes of reproductive tissues from *A. fraterculus* and *A. obliqua* and compared with RNA-seq data produced from head tissues of the same populations and species (Rezende et al. 2016). In this study, we estimated differentiation indexes and evolutionary rates from pairwise comparisons between both species and among seven species of Tephritidae. In addition, we tested for signals of natural selection and relaxed constraints. These results enabled us to identify which tissues, reproductive or cephalic, and sex, would show genes with higher evolutionary rates and whether this is due to positive selection or nonadaptive evolution. Answers to these questions not only contribute to the understanding of the evolutionary mechanisms affecting sex-biased genes, but also may offer clues to the differentiation process influencing these fruit flies even in the presence of gene flow.

Materials and Methods

Sampling and Laboratory Procedures

Individuals of *A. fraterculus* were collected from the field from guava (Myrtaceae) fruits (22°01′03″S, 47°53′27″W) and *A. obliqua* from jocote (Anacardiaceae) fruits (16°41′58″S, 49°16′35″W). These populations were maintained in laboratory under the following controlled conditions: 26 ± 1°C of temperature, 60–90% of humidity, and natural photoperiod. Reproductive tissues of virgin adult (8–12 days) male (testis, accessory glands, and phallus) and female (ovaries,

accessory glands, spermatheca, uterus, and ovipositor) flies of *A. fraterculus* and *A. obliqua* were collected. Total RNA was extracted from a pool of reproductive tissues of five individuals following the protocol proposed by Chomczynski and Mackey (1995). After extraction, each sample was formed by an equimolar mix of two pools, totaling samples from ten individuals in every mix. Biological replicates were prepared for each profile (species, sex, and tissue), making for a total of eight samples. RNA-seq libraries were constructed using the TruSeq RNA Sample Preparation Kit (Illumina) protocol according to the manufacturer's instructions. Libraries of 2 × 100 bp paired-end reads were sequenced on Illumina HiSeq2000 and HiScan at the Laboratory of Functional Genomics Applied to Agriculture and Agri-energy, ESALQ-USP, Brazil.

Cleaning and Assembly

Reads obtained from sequencing of reproductive tissues as well as published transcriptomes from head tissues of the pair of species studied here (Rezende et al. 2016) were trimmed using the program Trimmomatic v.0.33 (Bolger et al. 2014), setting the parameters LEADING: 5 TRAILING: 5 SLIDINGWINDOW: 5: 20 MINLEN: 50. This program also searches for and removes any remaining TrueSeq Illumina adapters in the reads. Unpaired reads were also discarded. After this censoring, reads from the same species were joined to produce two assemblies. Each group of reads was normalized by coverage and assembled using default parameters of Trinity v.2.4.0 (Grabherr et al. 2011).

Unigene Prediction and Assessment of the Quality of Assemblies

We searched for potential coding sequences (CDSs) in all six open reading frames of each transcript using the software TransDecoder v.3.0.1 (http://transdecoder.github.io) following three steps. First, TransDecoder.LongOrfs was used to retain all potential CDSs coding peptides longer than 100 amino acids. In the second step, these peptides were submitted to the hmmscan tool included in the HMMER v.3.1b2 package (Eddy 2011) to search for protein signatures in the Pfam-A database and BLASTP v.2.6 (Camacho et al. 2009) to search for similar sequences in the nonredundant database of the GenBank (nonredundant [nr]) including only proteins of arthropods. In the third step, the program TransDecoder.Predict uses the information produced by the other steps to predict the CDSs. Redundancy of the obtained CDSs was reduced using Cd-hit-est (Fu et al. 2012) with a similarity threshold of 0.99. To obtain the final set of putative unigenes, transcripts with these CDSs were filtered using the Trinity assembly information and only the isoform with the highest expression per trinity component was retained. For that, the reads from each species were mapped to the respective assembly using Bowtie2 (Langmead and Salzberg 2012) and the abundance of each transcript was estimated by eXpress v.1.5.1 (Roberts and Pachter 2013). These steps were performed by the script align_and_estimate_abundance.pl included in the Trinity package, adding no bias correction option for the eXpress program. The completeness and redundancy level of the raw and filtered assemblies of each species was evaluated by BUSCO (Benchmarking Universal Single-Copy Orthologs) (Simão et al. 2015) using the Arthropoda database as reference.

Functional Annotation

Predicted unigene CDSs were compared against the GenBank nr protein database including only arthropod proteins, the *Drosophila melanogaster* protein database (r6.14), and the Eukaryotic Orthologous Groups of proteins database (KOG) (Koonin et al. 2004) using NCBI BLASTP v.2.6 (Camacho et al. 2009). To all these analyses, we set an *e*-value threshold of 10^{-6}. We also searched for conserved protein domains using InterProScan 5.24-63.0 (Jones et al. 2014). Annotations against nr and conserved protein domains databases were submitted to Blast2GO (Conesa et al. 2005) to obtain a list of gene ontology (GO) terms associated with the annotated genes. Frequencies of GO terms at the level 2 were obtained using the program WEGO (Ye et al. 2006) and their distributions were plotted using GO terms with frequencies greater than 1%.

Identifying Sex- and Tissue-Biased Unigenes

The sex-biased genes were identified by comparing the expressions of male and female profiles in each tissue (reproductive and head) and species (*A. fraterculus* and *A. obliqua*). In addition, tissue-biased genes were identified by comparing the expressions of head and reproductive tissues in each sex and species. Expression analysis was performed by using the scripts align_and_estimate_abundance.pl, abundance_estimates_to_matrix.pl, PtR, and analyze_diff_expr.pl provided by the Trinity package (Grabherr et al. 2011). In the align_and_estimate_abundance.pl, we used Bowtie2 (Langmead and Salzberg 2012) and eXpress v1.5.1 (Roberts and Pachter 2013) to map the reads back to each species' assemblies (set of unigenes) and to estimate abundances of each unigene, respectively. This script was run adding no bias correction option for the eXpress program and very-sensitive option to Bowtie2. The abundance_estimates_to_matrix.pl script put the abundances values estimated to each RNA-seq library in a matrix. The PtR program was used to verify the quality of the biological replicates using Pearson correlation and principal component analysis of the unigenes expression across samples measured as $\log_2$ transformed of counts per million. Differential gene expression analysis among sexes and tissues was performed in edgeR (Robinson et al. 2010) using the TMM (trimmed mean of M-values) normalized abundances. Expression values are shown in transcripts per million (TPM).

Unigenes with fold-changes greater than 4 and a significance of FDR corrected *P*-values smaller than 0.05 were considered as differentially expressed.

Single Nucleotide Polymorphism Calling, Differentiation Indexes, and the McDonald–Kreitman Test

Anastrepha fraterculus unigenes were used as reference for single nucleotide polymorphism (SNP) calling. Filtered reads from each library were mapped to each assembly according to tissue using Bowtie2 (Langmead and Salzberg 2012). Mapped reads were converted to mpileup format and filtered based on minimum mapping quality of 20 and minimum PHRED quality of 30 using mpileup tool provided by Samtools v.1.3.1 package (Li et al. 2009), minimum coverage of 20, minimum reads of 1 to call the variant and strand filter (removed variants with more than 90% supported by only reads of one strand) using the tool mpileup2snp included at VarScan v2.4.2 (Koboldt et al. 2012). We considered only SNPs found in at least two libraries, regardless of sex, for further analysis.

Allelic frequency for each SNP was calculated as the average of the frequencies estimated by VarScan in each library. Hence, the frequency of each SNP was estimated based on 20–40 individuals depending on the number of samples that detected a particular SNP. We determined whether SNPs promoted synonymous or nonsynonymous amino acid changes using the prediction of complete CDSs and a custom python script. Allele frequencies were used to calculate the index of interspecific differentiation (D) defined as the absolute value of the difference in allele frequencies of an SNP in *A. fraterculus* and *A. obliqua* (Renaut et al. 2010; Andrés et al. 2013). The statistical comparison of D distributions inferred for each type of SNP (synonymous, nonsynonymous, and noncoding) was performed by applying Kolmogorov–Smirnov tests. We also estimated the average D using all SNPs in each CDS ($\bar{D}_{CDS}$), using only synonymous SNPs ($\bar{D}_S$), and using only nonsynonymous SNPs ($\bar{D}_{NS}$).

McDonald–Kreitman test (McDonald and Kreitman 1991) (MKT) was performed for each CDS by comparing the number of synonymous and nonsynonymous substitutions of polymorphic and almost fixed SNPs ($D > 0.95$) using a custom python script. We removed variants with a frequency smaller than 0.05 in both species to avoid biases produced by segregation of slightly deleterious mutations (Parsch et al. 2009). Only genes that had a value of at least one in all four classes of SNPs were included in the analysis, and significant departures from neutrality were estimated by Fisher's exact test (two-tail P-value < 0.05). CDSs with significant Fisher's exact test and neutrality index (NI) lower than 1 were considered to evolve under positive selection. The script also calculates the direction of selection (DoS), where a signature of positive selection is observed when $DoS > 0$ (Stoletzki and Eyre-Walker 2011).

Calculating Evolutionary Rates

Complete CDSs were submitted to the reciprocal best hit strategy in BLASTn with an *e*-value threshold of 10^{-6} to obtain the potential pairs of ortholog CDSs between *A. fraterculus* and *A. obliqua*. This strategy seeks to obtain the pairs of genes that produce best hit scores in a bidirectional BLAST comparison (interchanging the CDSs of the species as query and subject). Pairs of sequences that showed a length difference greater than 5% were removed because there was a higher chance of being different isoforms or different genes with only similar domains. We aligned the DNA sequences of putatively orthologs from the two species by their amino acid translations using the MAFFT algorithm (Katoh and Standley 2013) and back converted to DNA implemented in the program TranslatorX (Abascal et al. 2010). Resulted alignments were submitted to KaKs_Calculator (Zhang et al. 2006) to calculate the pairwise nonsynonymous (Ka) to synonymous substitution rate (Ks) ratio of the *fraterculus* group lineage using the Model Selection framework (Posada 2003). To decrease the chance of poorly alignments or saturation, we removed pairs with outlier Ks values, defined as values greater than the average plus three times the standard deviation, which was 0.62. Moreover, all the alignments with $Ka/Ks > 1$ were visually checked.

We also calculated the evolutionary rates of ortholog genes in Tephritidae and tested for selection using a phylogenetic approach. For that, the CDSs of *Ceratitis capitata* (GCF_000347755.2) (Papanicolaou et al. 2016), *Rhagoletis zephyria* (GCF_001687245.1), *Zeugodacus cucurbitae* (GCF_000806345.1) (Sim and Geib 2017), *Bactrocera dorsalis* (GCF_000789215.1), and *Bactrocera oleae* (GCF_001188975.1) were downloaded from GenBank. To avoid using miss-annotated and miss-assembled sequences, we removed CDSs with more than one stop codon and reduced the redundancy using Cd-hit-est (Fu et al. 2012) with a similarity threshold of 0.99. The putative cluster of orthologs was predicted using reciprocal best hit strategy in BLASTn with an *e*-value threshold of 10^{-6} and the CDSs of *A. fraterculus* as reference. The complete clusters (seven sequences) were submitted to the filtering and alignments steps of POTION program (Hongo et al. 2015). This pipeline excludes the sequences based on relative sequence length and identity, then aligns the clusters, trims the alignments using trimAl v.1.2 (Capella-Gutiérrez et al. 2009), and detects recombination using three methodologies (Phi, NSS, and MaxChi2) implemented in PhiPack (Bruen et al. 2006). All parameters used in POTION are available in supplementary file 1, Supplementary Material online. The maximum-likelihood phylogenies were inferred for each remained complete cluster of orthologs using GTRCAT model and 200 bootstrap replicates in the program RAxML v.8.2.9 (Stamatakis 2014).

We used trimmed alignments and the phylogenies to estimate the global nonsynonymous/synonymous rate (d*N*/d*S*)

ratio (ω) and performed the strict branch-site test implemented by CODEML included in the PAML v. 4.9 package (Yang 2007). The ω for the Tephritidae lineage was estimated using the M0 model (model = 0). We removed clusters of orthologs with d*S* higher than the average d*S* plus three times the standard deviation (d$S > 7$). The ancestral branch of *A. fraterculus* and *A. obliqua* in each phylogeny was set as foreground for the branch-site test. In order to statistically test whether a gene is evolving under positive selection, we compared the likelihoods of MA (model = 2, NSsites = 2) and MA1 (model = 2, NSsites= 2, fix_omega = 1) models using likelihood-ratio tests (LRTs) (Zhang et al. 2005). After the LRTs, we used the $\chi 2$ distribution to obtain *P*-values. We also detected variation in selection strength across the cluster of orthologs using RELAX (Wertheim et al. 2015). This program compares LRT of alternative and null ($k = 1$) models, where k is selection intensity defined as $\omega_{\text{foreground}} = \omega_{\text{background}}{}^{k}$. Significant comparisons with $k > 1$ and $k < 1$ indicate selection intensification and relaxation, respectively. To perform the phylogenies and selection tests in parallel, we used a custom python script which uses the script raxml_bs_wrapper.py (Yang and Smith 2014) and functions of the ete3 module (Huerta-Cepas et al. 2016).

Comparing Sexes and Tissues

As *A. fraterculus* and *A. obliqua* are phylogenetically closely related and displayed similar patterns of gene expression, we compared the patterns of sequence evolution of sex-biased genes found in one and both species using ortholog information. This approach allows the evaluation of genes with ancestral expression control and generalizes the results for both species and perhaps to other related species in the *fraterculus* group as well. Besides, the genes with species-specific expression enabled the analysis of recent evolutionary patterns after the change in expression pattern. Statistical comparisons of differentiation indexes and evolutionary rates among biased (sex and tissue) and unbiased categories were performed using Wilcoxon rank-sum test corrected by Holm approach (Holm 1979). Furthermore, differences between proportions of genes evolving under different selective regimes in biased and unbiased categories were estimated using Fisher's exact test.

Results

Sequencing, Cleaning, and Assessment of the Quality of the Assemblies

We produced 28,808,966 × 2 reads for *A. fraterculus* and 28,020,776 × 2 reads for *A. obliqua*, from males and females, two replicates each totalizing eight RNA libraries of reproductive tissue (approximately 7M × 2 reads per library). These libraries along with the previously sequenced samples of head tissue totalized 58,551,775 × 2 reads and 55,626,431 × 2 for *A. fraterculus* and *A. obliqua*, respectively. The cleaning step removed an average of 13.94% of reads for *A. fraterculus* and 14.71% for *A. obliqua*. Summary statistics for the two assemblies produced similar N50, mean, median, and length distributions (supplementary table S1, Supplementary Material online). Furthermore, from 1,066 conserved Arthropoda ortholog groups, BUSCO identified 95% as complete orthologs and around 3% as fragmented orthologs in the transcriptome assemblies of *A. fraterculus* and *A. obliqua* (supplementary table S2, Supplementary Material online). Moreover, roughly 50% of the complete orthologs were duplicated in the raw assembly; however, the redundancy in filtered unigenes was almost zero (approximately 0.5%).

Functional Annotation

Around 70% of the CDSs were annotated using *D. melanogaster* protein database and over 90% were matched with a protein from the GenBank nr protein including only Arthropoda entries (supplementary table S3, Supplementary Material online). The comparison against the nr database showed that most frequent top hits were to Tephritidae species (supplementary fig. S1, Supplementary Material online). Blast2GO successfully mapped approximately 67% of the CDSs. The distributions of the level 2 GO terms of both species and tissues were similar (supplementary fig. S2, Supplementary Material online). KOG functional classification also showed a similar representation of the categories in reproductive tissues of the two species (supplementary fig. S3, Supplementary Material online).

SNP Calling, Differentiation Indexes, and MKT

A total of 226,827 and 140,504 intra- and interspecific SNPs were identified in reproductive and head transcriptomes, respectively. We found 109,828 SNPs (79,947 of them associated with synonymous and 29,881 with nonsynonymous changes) in 3,662 coding regions expressed in reproductive tissues and 63,489 SNPs in 2,602 CDSs expressed in head tissues, of which 48,288 were synonymous and 15,201 nonsynonymous. SNP frequency distribution showed that more than 50% of the SNPs have rare alleles (supplementary fig. S4, Supplementary Material online). We rejected the hypothesis that frequency distributions of *D* for synonymous, nonsynonymous, and noncoding SNPs were drawn from the same distribution (Kolmogorov–Smirnov test, $P < 0.01$ for each of the three pairwise comparisons). These distributions also revealed that over 6% of the total of SNPs were fixed or almost fixed between species ($D > 0.95$) and the proportion of this type of SNPs is greater in nonsynonymous variants in both tissues (fig. 1*A* and *B*). The histograms of $\bar{D}_{\text{CDS}}$, $\bar{D}_{\text{S}}$, and $\bar{D}_{\text{NS}}$ showed that there is a greater proportion of highly differentiated unigenes using nonsynonymous than synonymous SNPs (fig. 1*C* and *D*). After excluding rare alleles, we retained 906 CDSs which met the minimum requisites to perform the MKT,

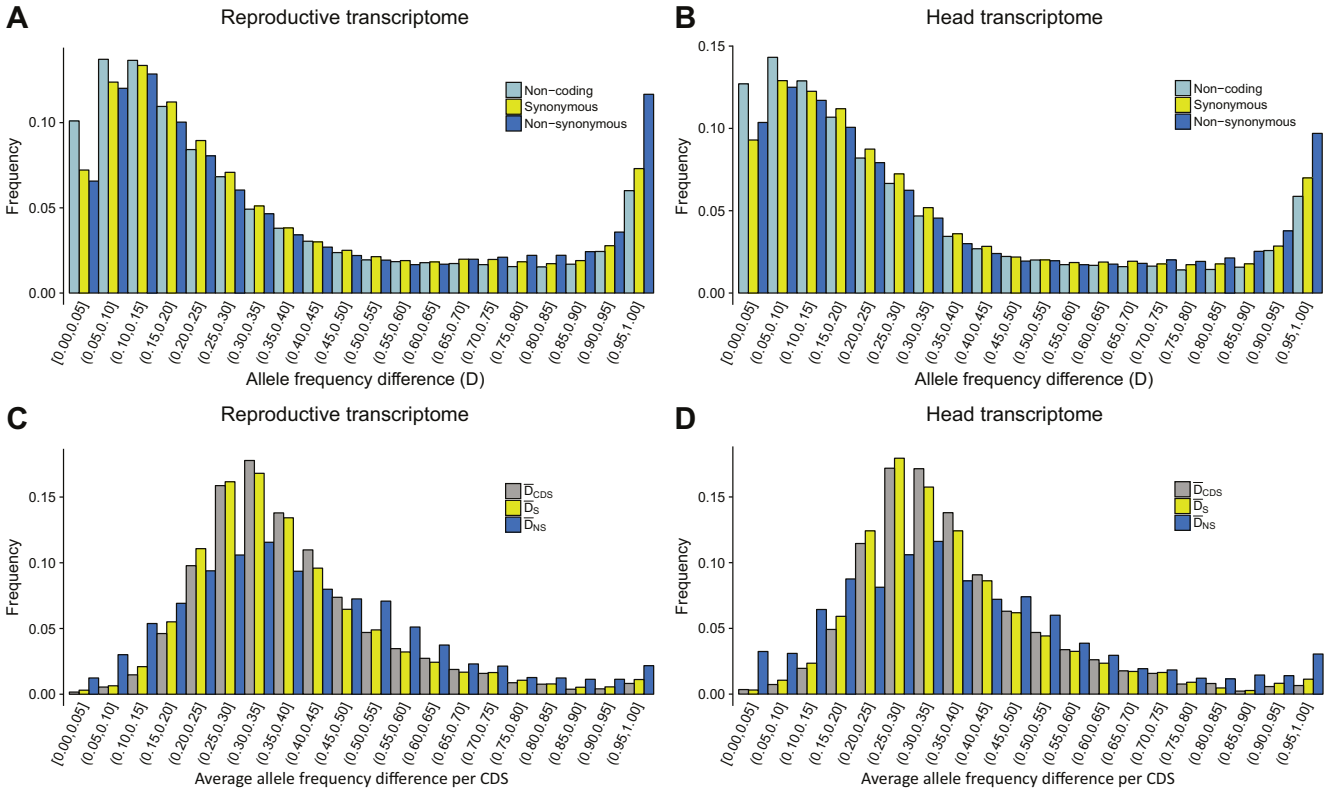

Fig. 1.—Frequency distributions of differentiation index. (*A*) and (*B*) Distributions of *D* (absolute allele frequency differences between *A. fraterculus* and *A. obliqua*) of SNPs found in reproductive and head transcriptomes, respectively. Light blue, yellow, and blue bars represent the distribution of noncoding, synonymous, and nonsynonymous SNPs, respectively. (*C*) and (*D*) Distributions of average *D* per CDS using all SNPs ($\bar{D}_{CDS}$), using only synonymous ($\bar{D}_S$) and only nonsynonymous ($\bar{D}_{NS}$) found in reproductive and head transcriptomes, respectively. Gray, yellow, and blue bars represent $\bar{D}_{CDS}$, $\bar{D}_S$, and $\bar{D}_{NS}$, respectively.

that is, at least one synonymous and one nonsynonymous fixed SNP and one synonymous and one nonsynonymous polymorphic SNP. Fifty-one CDSs showed significant statistical departure from neutrality and NI < 1, thus were considered evolving under positive selection (supplementary file 2, Supplementary Material online).

Patterns of Gene Expression across Sexes and Tissues

There were 12,887 and 13,605 unigenes expressed (TPM > 1) in the *A. fraterculus* transcriptome in head and reproductive tissues, respectively, with similar values also found in *A. obliqua*: 12,073 (head tissue) and 13,455 (reproductive tissue). Biological replicates are strongly correlated, with coefficients ranging from 0.96 to 0.98 and 0.95 to 0.98 for *A. fraterculus* and *A. obliqua*, respectively (fig. 2*A* and *B*). Moreover, female and male samples of both species showed Pearson correlation coefficients higher than 0.96, establishing well-defined clusters in the principal component analysis (fig. 2*C* and *D*). Patterns of differential gene expression across sexes and tissues were similar in both studied species (fig. 3). A total of 21.3% and 28.7% of the unigenes showed biased expression between sexes in *A. fraterculus* and *A. obliqua*, respectively (supplementary table S4, Supplementary Material online). This difference is mainly affected by 6% more genes which are upregulated in *A. obliqua* males. Interestingly, less than 1% of the genes in head transcriptomes are sex-biased in both species (supplementary table S4 and fig. S5, Supplementary Material online). In the comparison across tissues, approximately 27% of the genes were tissue-biased in male transcriptomes of both species (supplementary table S5, Supplementary Material online). Female transcriptomes showed 22% of tissue-biased genes in *A. fraterculus* and 33% in *A. obliqua*. Besides, we also found variation in the magnitude of differential expression of biased expressed genes. Male-biased genes displayed greater fold-change average (measure by $\log_2$) than female-biased genes in both species (Wilcoxon rank-sum test *P*-value < 0.01; supplementary fig. S6, Supplementary Material online). Tissue-biased genes expressed in males and females showed opposite patterns of magnitude of gene expression, whereas in males, the genes with greater differences in gene expression are reproductive-biased, in females, they are head-biased.

Evolutionary Patterns of Sex- and Tissue-Biased Genes

Most comparisons between population differentiation index averages ($\bar{D}_{CDS}$, $\bar{D}_{NS}$, and $\bar{D}_S$) across sex and tissues failed to

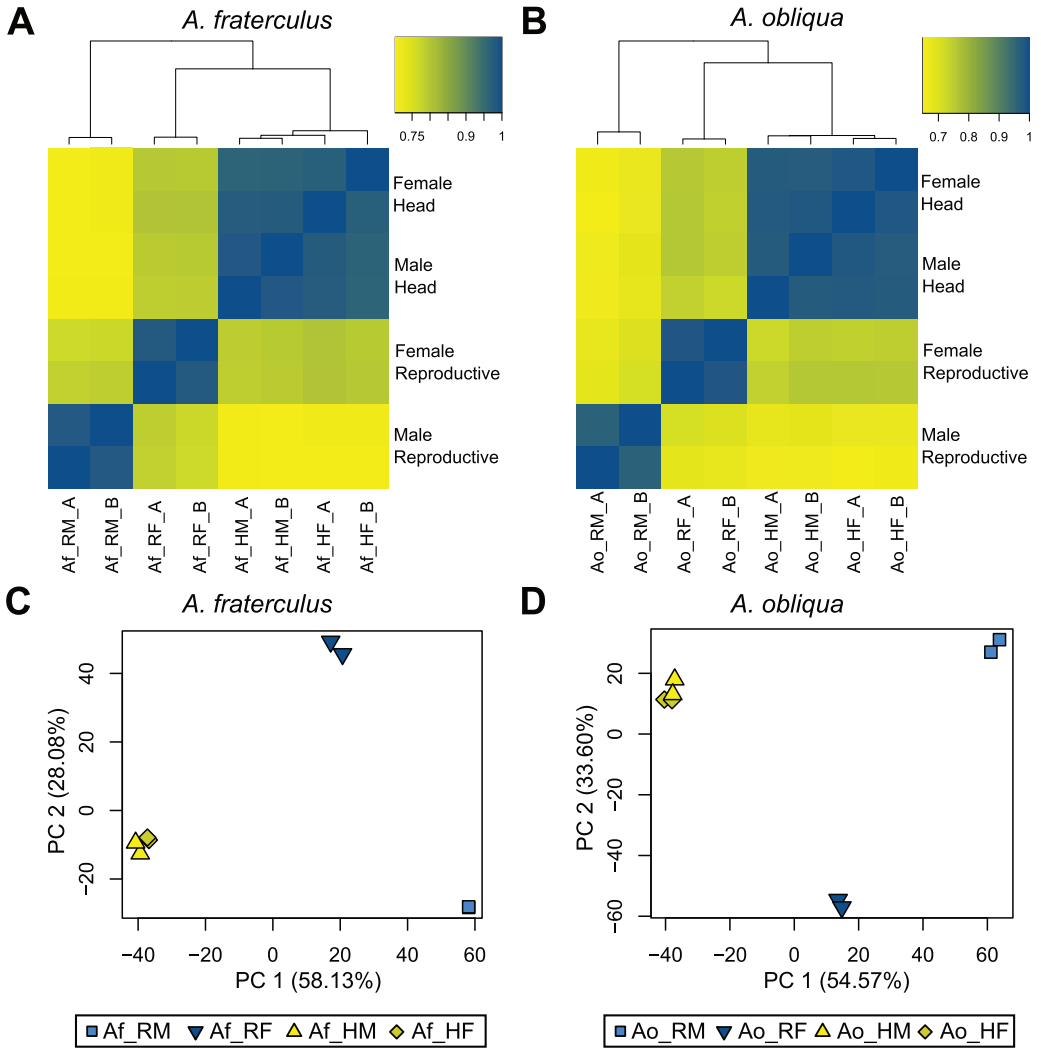

Fig. 2.—Analysis of expression of biological replicates. Heatmap of Pearson correlations and hierarchical cluster of samples from *A. fraterculus* (*A*) and *A. obliqua* (*B*). Principal component analysis of all samples from *A. fraterculus* (*C*) and *A. obliqua* (*D*). RM and RF: Samples from male and female reproductive transcriptomes. HM and HF: Samples from male and female head transcriptomes.

show significant differences between biased and unbiased genes, but the few that did, involved contrasts to nonsynonymous mutations (fig. 4). Male-biased unigenes in both species displayed greater levels of differentiation than unbiased using the parameter $\bar{D}_{NS}$ (Wilcoxon rank-sum test *P*-value < 0.05), whereas male reproductive-biased genes had the highest average $\bar{D}_{NS}$, which was also significantly different from male head-biased and unbiased genes (Wilcoxon rank-sum test *P*-value < 0.01 in both comparisons). In contrast, female transcriptomes failed to show significant differences in any comparison (supplementary fig. S7, Supplementary Material online).

Analysis of approximately 4,000 orthologs between *A. fraterculus* and *A. obliqua* and approximately 3,000 among seven Tephritidae species revealed that male-biased genes in both nonspecies-specific and species-specific groups have significantly higher evolutionary rates than unbiased genes (fig. 5 and supplementary table S6, Supplementary Material online). Likewise, female-biased unigenes also showed significantly higher evolutionary rates than unbiased, though the comparison involving the species-specific expression genes failed to reject the null hypothesis (fig. 5). Moreover, comparisons between tissues revealed that reproductive-biased genes, be it male or female, displayed higher rates of evolution than unbiased in both species (fig. 5).

Male-biased and male reproductive-biased genes displayed significantly greater proportion of genes evolving under positive selection than unbiased as evaluated by MKT and pairwise Ka/Ks (table 1). Moreover, these contrasts showed a positive mean DoS, suggesting adaptive evolution, even though these set of genes exhibited higher values than unbiased, this difference was not statistically significant (supplementary fig. S8, Supplementary Material online). These results

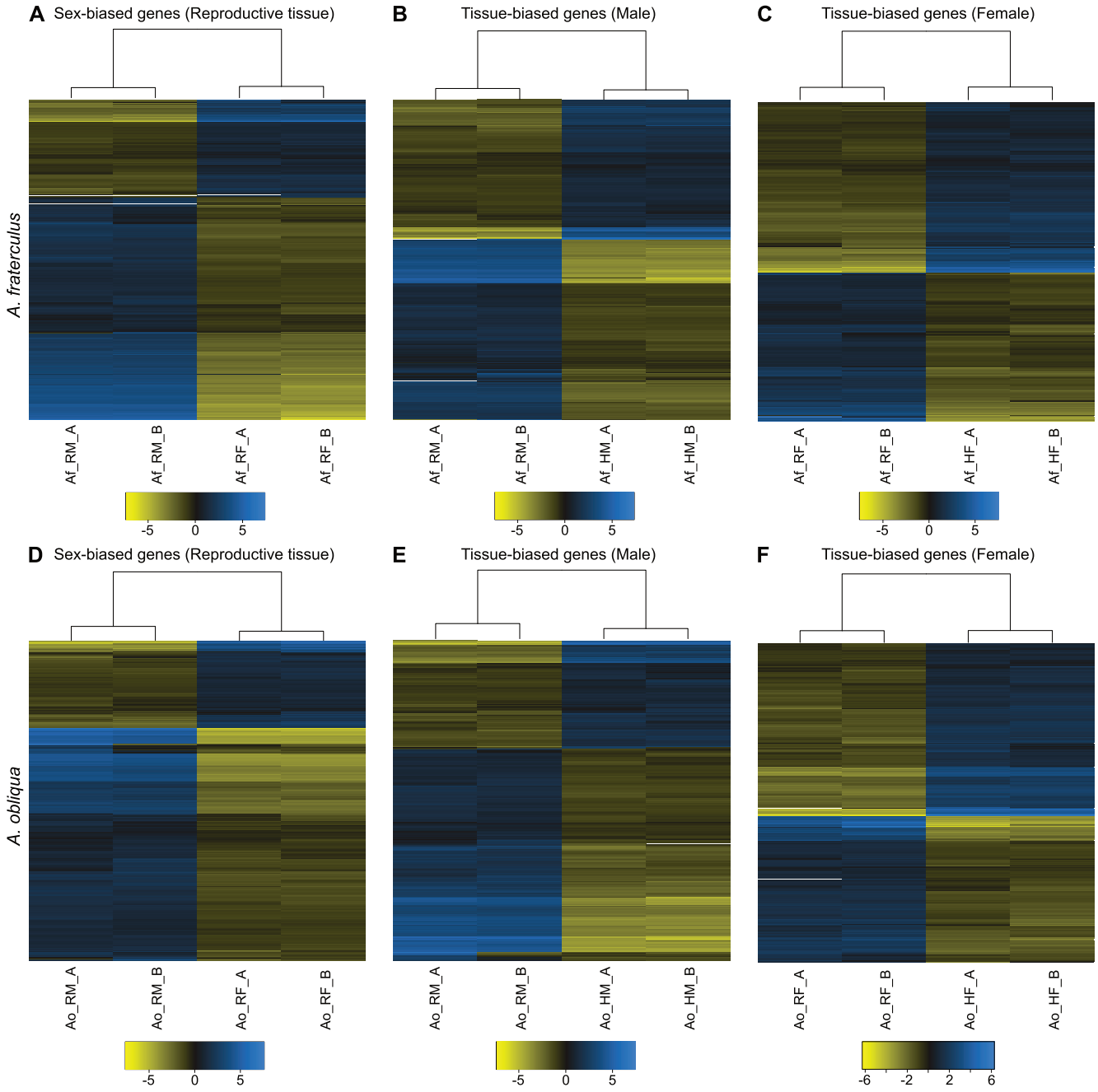

Fig. 3.—Heatmap and hierarchical clustering of differentially expressed genes on head and reproductive transcriptomes from *A. fraterculus* and *A. obliqua*. Differentially expressed genes between male and female reproductive transcriptomes from *A. fraterculus* (*A*) and *A. obliqua* (*D*). Differentially expressed genes between male reproductive and head transcriptomes from *A. fraterculus* (*B*) and *A. obliqua* (*E*). Differentially expressed genes between female reproductive and head transcriptomes from *A. fraterculus* (*C*) and *A. obliqua* (*F*). RM and RF: Samples from male and female reproductive transcriptomes. HM and HF: Samples from male and female head transcriptomes.

diverge from what was found by the branch-site tests, which indicated similar proportion of genes evolving under positive selection among biased and unbiased genes. Interestingly, we noticed that 11% of the male-biased genes with signals of positive selection play important roles in *Drosophila*'s reproduction (table 2), and it is possible that they may retain similar roles in *Anastrepha*. However, we also found a greater proportion of genes with signals of relaxed selection among male-biased and reproductive-biased (both sexes) genes in comparison to unbiased genes, suggesting differences in selective constraints in these contrasts (table 1).

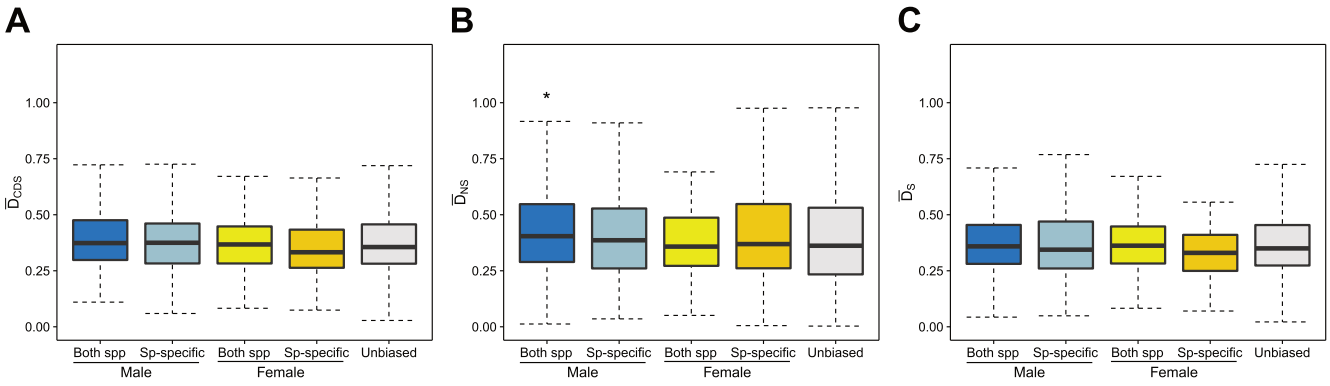

Fig. 4.—Boxplots of differentiation indexes measured among sex-biased and unbiased genes. Differentiation was estimated as average allele frequency differences between *A. fraterculus* and *A. obliqua* using all ($\bar{D}_{CDS}$), nonsynonymous ($\bar{D}_{NS}$) and synonymous ($\bar{D}_{S}$) SNPs. Sex-biased genes are grouped in genes with the same expression pattern in both species (both spp) and biased expression detected in a particular species (sp-specific). Comparison of $\bar{D}_{CDS}$ (*A*), $\bar{D}_{NS}$ (*B*), and $\bar{D}_{S}$ (*C*) among male-, female-biased and unbiased genes expressed in reproductive tissues. *Holm-corrected *P*-value of Wilcoxon rank sum test < 0.05. * just above the box indicates significant level in the comparison to unbiased genes.

Discussion

The RNA-seq data generated high-quality de novo assemblies of *A. fraterculus* and *A. obliqua* as evaluated by length distribution and gene content metrics. The N50 values of around 1,800 bp of these assemblies were in line with equivalent transcriptomes of other available tephritids (Hsu et al. 2012; Morrow et al. 2014; Salvemini et al. 2014). Additionally, we found almost all conserved Arthropoda ortholog clusters in these transcriptome assemblies, suggesting a significant representation and completeness for the panel of genes expressed in head and reproductive tissues of both species. A further indication of their completeness is that most of the CDSs were successfully annotated against proteins of the *D. melanogaster* database (~70%) and the nr GenBank database (~90%). Furthermore, functional annotation using the distribution of GO and KOG categories for reproductive tissues of *A. fraterculus* and *A. obliqua* showed similar distributions, akin to what has been described for head tissues from these species (Rezende et al. 2016).

Transcriptome data studied here enabled identification of thousands of SNPs across *A. fraterculus* and *A. obliqua*. Even though we identified hundreds of SNPs fixed, or nearly fixed, in one or the other species, the most common pattern observed for *A. fraterculus* and *A. obliqua* transcriptomes indicates that the species have diverged recently, because a great number of SNPs show little allele frequency difference across species. However, these results should be interpreted with caution, because we estimated SNP allele frequency distributions from pools of individuals of a single population per species, and they may not represent the whole diversity across the species' geographic distributions. Microsatellite analyses across *A. fraterculus* Brazilian populations showed some evidence of differentiation, but over 90% of variation is intrapopulational (Manni et al. 2015). Furthermore, there is evidence that these species differentiated with gene flow (Scally et al. 2016; Díaz F, Lima ALA, Nakamura AM, Fernandes F, Sobrinho I, de Brito RA, unpublished data), which would make variation in general common to several localities, rather than isolated, even across species boundaries. If this pattern is common across the species' distribution, it might indicate that the diversity distribution here inferred for *A. fraterculus* and *A. obliqua* may hold for the majority of SNPs identified.

The allele frequency distributions are consistent with a scenario where the majority of the genome would be somewhat homogenous, interspersed by highly differentiated regions (Martin et al. 2013), such as what was found in two recently diverged species of *Gryllus* in the presence of gene flow (Andrés et al. 2013). This pattern would hold even if there were no selective forces involved and only drift would be driving the species apart (Cruickshank and Hahn 2014). Here, despite the reduced number of contigs with large allele frequency differences across species, we still detected at least 5% of the SNPs nearly fixed for different alleles in different species. Interestingly, the distributions of *D* (allele frequency difference between species) inferred for synonymous, nonsynonymous, and noncoding SNPs are significantly different from one another. These differences seem to be at least in part due to adaptive evolution as there is a significantly greater proportion of fixed differences across species which are associated with nonsynonymous substitutions, even when contrasted with noncoding substitutions. This increased proportion of nonsynonymous substitutions in fixed differences between species also holds when we consider all substitutions present in a CDS. The unigenes with high $\bar{D}_{NS}$ values may potentially be "islands of divergence," which are genomic regions that remain differentiated between species even in the presence of gene flow due to directional selection

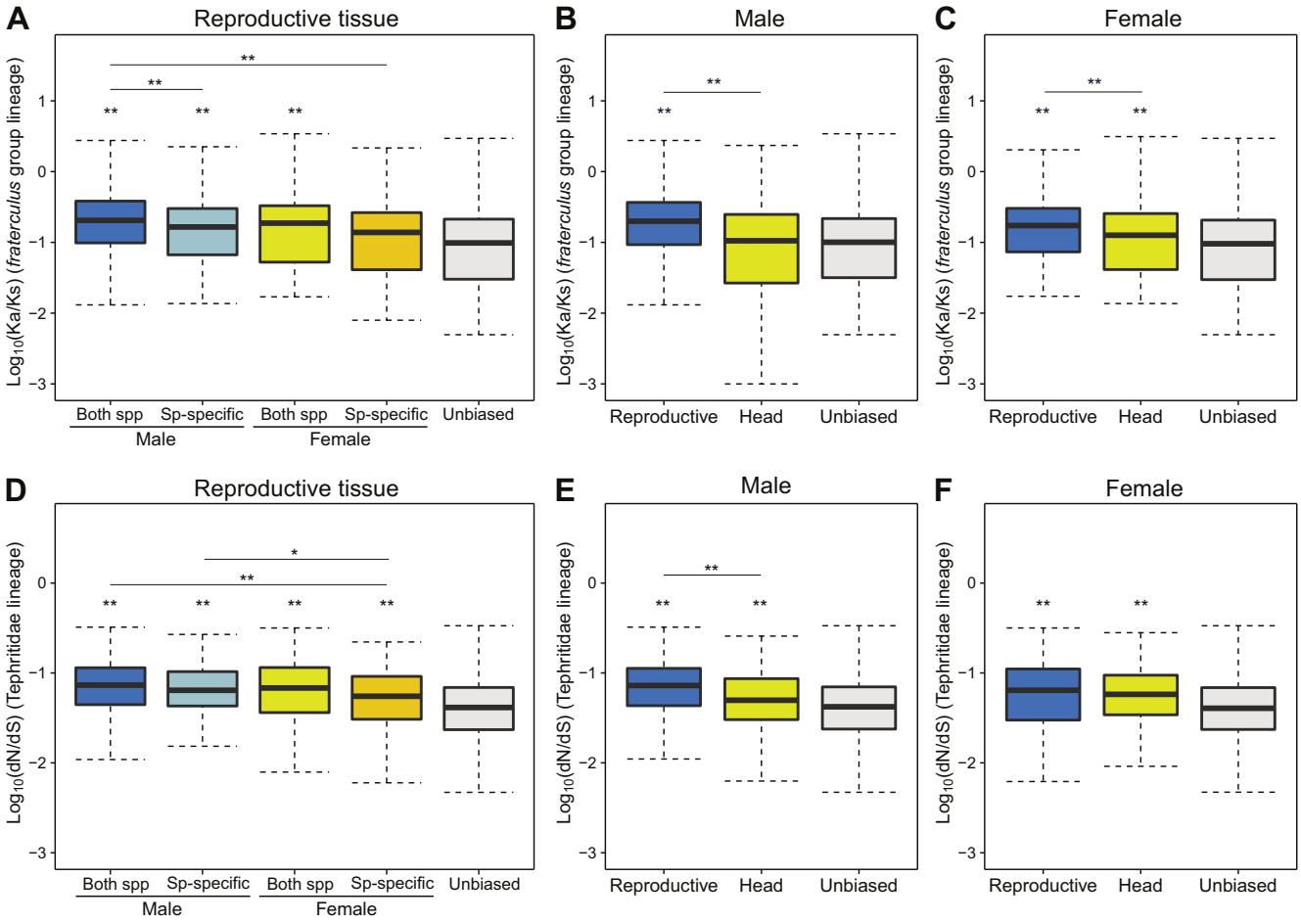

Fig. 5.—Evolutionary rates for sex- and tissue-biased genes estimated based on pairwise comparison (*A. fraterculus* and *A. obliqua*) and seven Tephritidae species (*A. fraterculus*, *A. obliqua*, *Ceratitis capitata*, *Rhagoletis zephyria*, *Zeugodacus cucurbitae*, *Bactrocera dorsalis*, and *Bactrocera oleae*). Sex-biased genes are grouped in genes with the same expression pattern in both species (both spp) and biased expression detected in a particular species (sp-specific). Boxplots of Log_{10}(Ka/Ks) from *A. fraterculus* and *A. obliqua* orthologs for sex-biased genes (*A*) and tissue-biased genes (*B*) and (*C*). Boxplots of Log_{10} (*dN*/*dS*) from seven Tephritidae species orthologs for sex-biased genes (*D*) and tissue-biased genes (*E*) and (*F*). *Holm-corrected *P*-value of Wilcoxon rank sum test < 0.05. **Holm-corrected *P*-value of Wilcoxon rank sum test < 0.01. * or ** just above the box indicates significant level in comparisons with unbiased genes.

(Nosil and Feder 2012). Even though there are other reasons why genomic islands of divergence may occur (Noor and Bennett 2009; Cruickshank and Hahn 2014), many which cannot be tested for the data presented here because of the lack of a reference genome, these results point that at least a portion of the divergence between *A. fraterculus* and *A. obliqua* is due to regions affected by selection. We found some male-biased expressed genes with signals of positive selection involved with male courtship and fertility (table 2), thus these genes may be related to the establishment of prezygotic barriers. This observation agrees with studies on morphotypes of *A. fraterculus* complex species which have suggested that their reproductive isolation is mainly due to prezygotic barriers (Rull et al. 2013; Juárez et al. 2015).

The SNP allele frequency distributions allowed us not only to identify genes potentially involved with species differences but also to investigate general patterns of evolution for genes expressed in reproductive tissues across the two closely related species. In general, SNPs in both species showed a large proportion of rare alleles (supplementary fig. S3, Supplementary Material online), which might be due to demographic expansion and/or selective sweeps or weak purifying selection in particular genes (Fu 1997; Fay et al. 2001). We consider the former to be more likely considering that this pattern seems to be widespread across several genes, and that population expansion due to the increased distribution of host fruits with agriculture has been suggested to have happened to both studied species based on coalescent simulations (Díaz F, Lima ALA, Nakamura AM, Fernandes F, Sobrinho I, de Brito RA, unpublished data).

In general, the expression profiles were similar between *A. fraterculus* and *A. obliqua* showing 20–30% sex-biased

Table 1
Patterns of Evolution for Sex- and Tissue-Biased and Unbiased Genes

	McDonald–Kreitman Test		Pairwise Ka/Ks		Branch-Site Test		RELAX	
	N	$P<0.05$[a]	*N*	Ka/Ks > 1[b]	*N*	$P<0.05$[c]	*N*	$P<0.05$[d]
Reproductive								
Male-biased (both)	155	**16****	488	**23***	272	17	272	**70****
Male-biased (specific)	15	**4****	282	10	184	12	184	22
Female-biased (both)	22	0	115	4	71	3	71	15
Female-biased (specific)	36	1	195	**11***	136	6	136	**33****
Unbiased	407	15	3274	94	2481	120	2481	431
Male								
Reproductive-biased	174	**19***	527	**28****	303	14	303	**84****
Head-biased	25	0	257	9	161	6	161	26
Unbiased	402	17	3321	92	2506	127	2506	432
Female								
Reproductive-biased	40	3	354	13	140	6	140	**36***
Head-biased	37	1	234	10	217	14	217	29
Unbiased	414	17	3129	82	2406	111	2406	416

Note.—*N*, number of unigenes; both, same expression pattern in *A. fraterculus* and *A. obliqua*; specific, sex-biased expressed gene in either *A. fraterculus* or *A. obliqua*.
[a]Significant departure from nonsynonymous and synonymous proportion of polymorphic and fixed SNPs using Fisher's exact test and NI < 1.
[b]Number of orthologs of *A. fraterculus* and *A. obliqua*.
[c]Number of orthologs with significant LRTs between MA and MA1 using the *A. fraterculus* and *A. obliqua* ancestral branch as foreground.
[d]Number of orthologs with $k < 1$ (relaxed selection) and significant LRTs between null ($k = 1$) and alternative using the *A. fraterculus* and *A. obliqua* ancestral branch as foreground.
Fisher's exact test comparing biased with unbiased genes showing $*P<0.05$ and $**P<0.01$.

Table 2
Signals of Positive Selection in Sex-Biased Expressed Genes Potentially Associated with Reproduction

Annotation with *Drosophila melanogaster* Database	Expression Pattern	Signal of Selection	Role	Reference
Neural Lazarillo	Male-biased[a]	Ka/Ks > 1	Fertility and courtship behavior	Ruiz et al. (2011)
Takeout	Male-biased[c]	Ka/Ks > 1	Courtship	Dauwalder et al. (2002)
CG15406	Male-biased[b]	Bst	Influence female's remating	Sitnik et al. (2016)
Kelch-like family member 10	Male-biased[c]	MKT	Spermatogenesis	Arama et al. (2007)
Dynein intermediate chain at 61B	Male-biased[c]	Bst	Spermatogenesis	Fatima (2011)
Hedgehog	Male-biased[b]	Bst	Male's germ line maintenance	Zhang et al. (2013)
Male fertility factor kl5	Male-biased[c]	MKT	Sperm motility	Carvalho et al. (2000)
Lost boys	Male-biased[c]	MKT	Sperm motility	Yang et al. (2011)
Tektin A	Male-biased[c]	MKT	Sperm motility	Dorus et al. (2006)
Egg-derived tyrosine phosphatase	Female-biased[c]	Bst	Oogenesis and embryogenesis	Yamaguchi et al. (2005)
CG14645	Female-biased[b]	Ka/Ks > 1	Courtship	Immonen and Ritchie (2012)
CG14187	Female-biased[a]	Ka/Ks > 1	Chorion protein	Tootle et al. (2011)

Note.—Bst, branch-site test.
[a]Sex-biased expressed gene only in *A. fraterculus*.
[b]Sex-biased expressed gene only in *A. obliqua*.
[c]Sex-biased expressed gene in both species.

unigenes, a pattern similar to what was found in *Drosophila* species (Zhang et al. 2007). This may be due to stability of the gene expression control machinery because of evolutionary constraints (Zhang et al. 2007; He et al. 2011). Data from both species consistently indicate that the majority of sex-biased genes comes from reproductive tissues, which agrees with what was reported for a comparison of somatic tissues and gonad transcriptomes in *D. melanogaster* (Parisi et al. 2004). Furthermore, the approximately 0.5% of sex-biased genes displayed in head tissues of both species of *Anastrepha* contrast with the approximately 16% differentially expressed genes in *D. melanogaster* head (Chang et al. 2011). However, in the latter there is a large difference in the genes with sexually diverged expression between central system and peripheral tissues (Goldman and Arbeitman 2007), so this variation in expression pattern across head organs and structures could obscure the expression of sex-biased genes in the whole head. Besides, our results show a higher number of

upregulated genes and larger magnitudes of their fold changes in males than in females, which could be due to the existence of more male-biased genes or the differences in expression of female-biased are too small, which would require larger statistical power to detect these differences (Assis et al. 2012). The comparison between tissues reveals that most tissue-biased expressed genes are in male reproductive tissues, possibly because most tissue-specific genes are expressed in testis, as indicated for *D. melanogaster* (Meiklejohn and Presgraves 2012).

Our results suggest that male-biased genes have higher ω such as it was found for several lineages (Torgerson et al. 2002; Vacquier and Swanson 2011; Harrison et al. 2015), such as *Drosophila* species (Zhang et al. 2004). In addition, female-biased genes also evolve significantly faster than unbiased genes in *Anastrepha* species, which has also been described for some animal taxa such as birds (Mank et al. 2007), mosquitoes (Papa et al. 2017; Whittle and Extavour 2017), and fishes (Yang et al. 2016). We also found that sex-biased genes also tend to evolve more rapidly, particularly their nonsynonymous rates, but the rate of fixation of nonsynonymous mutations ($\bar{D}_{NS}$) was only significantly greater in males. These findings suggest that male-biased genes may have been evolving under adaptive constraints in the *Anastrepha* species studied here. However, the faster evolution of sex-biased genes may be explained by other factors than sexual selection such as tissue-specific expression, genetic drift, turnovers in expression patterns, and relaxed selective constraints (Mank et al. 2008; Meisel 2011; Gershoni and Pietrokovski 2017; Mank 2017). We found evidence of both selection and relaxed constraints in these genes. Positive selection was detected by greater proportion of genes with significant MKT and high rates of evolution (Ka/Ks $>$ 1) in male-biased genes when compared with unbiased genes. However, branch-site tests displayed a similar proportion of genes evolving under positive selection in male-biased and unbiased genes. It is possible that this is a consequence of the reduced number of substitutions in the short branch between the two recently diverged species that failed to reach the significance level in the branch-site test, but this may also be caused by higher rates of evolution in male-biased genes which would complicate orthology assignment even for species in the same genus (Ellegren and Parsch 2007). In fact, approximately 50% of male-biased genes only showed orthologs in *Anastrepha* but not to other more distantly related species, preventing their analysis in the branch-site test and thus potentially producing a bias in the proportion of positively selected genes.

Our data also suggest signals of relaxed selective constraints, as male-biased and species-specific female-biased genes displayed a greater proportion of genes with signals of relaxed selection than unbiased genes, which would imply that the values are converging to neutrality (Wertheim et al. 2015). As there are many more sites possibly under purifying selection than under positive selection, for instance, we detected only two male-biased genes that displayed Ka/Ks $>$ 1, whereas 68 had Ka/Ks $<$ 1, this pattern would imply that the average ω would tend to increase, helping to explain their higher rates of evolution. On the other hand, genes with tissue-specific expression are probably evolving under relaxed selection. Comparison between head and reproductive transcriptomes from *A. fraterculus* and *A. obliqua* revealed that reproductive genes evolve faster than head genes in both sexes, showing similar patterns to *Drosophila* (Jagadeeshan and Singh 2005). This outcome may be explained by broader patterns of expression in head-biased genes, but as there are no available data for other tissues in these species, we are not able to estimate their actual specificity. Nevertheless, testis typically shows a greater proportion of tissue-specific genes (Baker et al. 2011; Emig et al. 2011; Meiklejohn and Presgraves 2012; Yang et al. 2016), thus it is likely that several reproductive-biased genes would be tissue-specific, particularly for males. These potentially tissue-specific genes would be more likely to evolve under positive or relaxed selection, whereas generally expressed genes seem to experience stronger evolutionary constraints, possibly due to pleiotropy (Mank et al. 2008; Haygood et al. 2010; Kryuchkova-Mostacci and Robinson-Rechavi 2015), or because genes that are expressed in different tissues across the organism are more likely to be part of the constitutive set of essential genes for cell function, thus harboring lower nonsynonymous rates. Finally, gene duplication can reduce functional constraints, so distinct levels of paralogy may also produce this outcome. Nevertheless, the mammalian gene family content is equivalent between housekeeping and tissue-specific genes (Zhang and Li 2004); hence, if this pattern is also true in insects, gene redundancy may not be a plausible explanation.

Our analysis identified hundreds of SNPs associated with unigenes that showed fixed, or nearly fixed, differences between *A. fraterculus* and *A. obliqua*, which have been significantly more associated with nonsynonymous substitutions than to other substitutions and point to an important role for selection in their differentiation. Even though we still lack a formal connection between sex-biased genes and speciation, the first "speciation gene" identified in *Drosophila* was the male-biased *Odysseus site homeo-box*, associated with postzygotic isolation mechanisms which produce sterility in male hybrids (Ting et al. 1998). Furthermore, if a set of sex-biased genes evolve under sexual selection or sexual conflict which may lead to intraspecific intersexual divergence, these same differences may foster differentiation between species, especially when the genes involved are sex-biased in a species-specific manner.

Conclusions

Our work not only contributes to the current functional genomic knowledge on two of the most important fruit pests

from the Neotropics by generating next-generation transcriptome data for reproductive tissues which have been hitherto unavailable, but also explored differences in expression patterns between sexes and tissues. Although several studies are available on this matter for a wide variety of animals, particularly *Drosophila* (Malone et al. 2006; Yang et al. 2006; Perry et al. 2014; Harrison et al. 2015; Dean et al. 2017), little had been known for Tephritidae species. In this matter, our findings indicate that head tissues of *A. fraterculus* and *A. obliqua* exhibit few genes with sex-biased expression. More importantly, sexual dimorphism in expression profiles of reproductive tissues revealed that sex-biased genes evolve faster than unbiased genes, especially in males, a pattern that was associated with signals of positive selection and relaxed constraints. Our results shed some light on the evolution of sex- and tissue-biased genes expressed in reproductive and head tissue of *A. fraterculus* and *A. obliqua* which should be valuable to other species as well. Furthermore, we found a set of sex-biased genes in reproductive tissues that may be candidates to be involved in the differentiation process of *A. fraterculus* and *A. obliqua*. However, further studies that evaluate the populational variation of these genes are necessary to corroborate their role in the differentiation of these and other species of the *fraterculus* group.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

The authors would like to thank the Associate Editor Prof. Judith Mank and three anonymous reviewers for providing invaluable comments and methodological advice that considerably improved the manuscript. This work was supported by the São Paulo Research Foundation (FAPESP, Brazil) (2010/20455-4 to R.A.B. and 2012/17160-8 to E.B.C.); National Council of Scientific and Technological Development (CNPq, Brazil) (160002/2013-3 to C.C. and 150626/2012 to S.C.-E.); and Coordination for the Improvement of Higher Education Personnel (CAPES, Brazil) (PVE 056/13 to R.A.B.).

Literature Cited

Abascal F, Zardoya R, Telford MJ. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res 38(Web Server issue):W7–W13.

Andrés JA, Larson EL, Bogdanowicz SM, Harrison RG. 2013. Patterns of transcriptome divergence in the male accessory gland of two closely related species of field crickets. Genetics 193(2):501–513.

Arama E, Bader M, Rieckhof GE, Steller H. 2007. A ubiquitin ligase complex regulates caspase activation during sperm differentiation in *Drosophila*. PLoS Biol. 5(10):e251.

Assis R, Zhou Q, Bachtrog D. 2012. Sex-biased transcriptome evolution in *Drosophila*. Genome Biol Evol. 4(11):1189–1200.

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

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics 30(15):2114–2120.

Brand P, et al. 2015. Rapid evolution of chemosensory receptor genes in a pair of sibling species of orchid bees (Apidae: Euglossini). BMC Evol Biol. 15(1):176.

Bruen TC, Philippe H, Bryant D. 2006. A simple and robust statistical test for detecting the presence of recombination. Genetics 172(4):2665–2681.

Camacho C, et al. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

Campanini EB, de Brito RA. 2016. Molecular evolution of Odorant-binding proteins gene family in two closely related *Anastrepha* fruit flies. BMC Evol Biol. 16(1):198.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25(15):1972–1973.

Carvalho AB, Lazzaro BP, Clark AG. 2000. Y chromosomal fertility factors kl-2 and kl-3 of *Drosophila melanogaster* encode dynein heavy chain polypeptides. Proc Natl Acad Sci U S A. 97(24):13239–13244.

Chang PL, Dunham JP, Nuzhdin SV, Arbeitman MN. 2011. Somatic sex-specific transcriptome differences in *Drosophila* revealed by whole transcriptome sequencing. BMC Genomics 12:364.

Chomczynski P, Mackey K. 1995. Short technical reports. Modification of the TRI reagent procedure for isolation of RNA from polysaccharide- and proteoglycan-rich sources. Biotechniques 19:942–945.

Conesa A, et al. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21(18):3674–3676.

Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol Ecol. 23(13):3133–3157.

Dauwalder B, Tsujimoto S, Moss J, Mattox W. 2002. The *Drosophila takeout* gene is regulated by the somatic sex-determination pathway and affects male courtship behavior. Genes Dev. 16(22):2879–2892.

Dean R, et al. 2017. Sperm competition shapes gene expression and sequence evolution in the ocellated wrasse. Mol Ecol. 26(2):505–518.

Dorus S, et al. 2006. Genomic and functional evolution of the *Drosophila melanogaster* sperm proteome. Nat Genet. 38(12):1440–1445.

dos Santos P, Uramoto K, Matioli SR. 2001. Experimental hybridization among *Anastrepha* species (Diptera: Tephritidae): production and morphological characterization of F1 hybrids. Ann Entomol Soc Am. 94:717–725.

Eberhard WG, Cordero C. 1995. Sexual selection by cryptic female choice on male seminal products—a new bridge between sexual selection and reproductive physiology. Trends Ecol Evol. 10(12):493–496.

Eddy SR. 2011. Accelerated profile HMM searches. PLoS Comput Biol. 7(10):e1002195.

Ellegren H, Parsch J. 2007. The evolution of sex-biased genes and sex-biased gene expression. Nat Rev Genet. 8(9):689–698.

Emig D, Kacprowski T, Albrecht M. 2011. Measuring and analyzing tissue specificity of human genes and protein complexes. EURASIP J Bioinform Syst Biol. 2011(1):5.

Fatima R. 2011. *Drosophila* dynein intermediate chain gene, *Dic61B*, is required for spermatogenesis. PLoS One 6(12):e27822.

Fay JC, Wyckoff GJ, Wu C-I. 2001. Positive and negative selection on the human genome. Genetics 158(3):1227–1234.

Feder JL, Egan SP, Nosil P. 2012. The genomics of speciation-with-gene-flow. Trends Genet. 28(7):342–350.

Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28(23):3150–3152.

Fu Y-X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147(2):915–925.
Gavrilets S. 2014. Is sexual conflict an "engine of speciation"? Cold Spring Harb Perspect Biol. 6(12):a017723.
Gershoni M, Pietrokovski S. 2014. Reduced selection and accumulation of deleterious mutations in genes exclusively expressed in men. Nat Commun. 5:4438.
Gershoni M, Pietrokovski S. 2017. The landscape of sex-differential transcriptome and its consequent selection in human adults. BMC Biol. 15(1):7.
Goldman TD, Arbeitman MN. 2007. Genomic and functional studies of *Drosophila* sex hierarchy regulated gene expression in adult head and nervous system tissues. PLOS Genet. 3(11):e216.
Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 29(7):644–652.
Haerty W, et al. 2007. Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*. Genetics 177(3):1321–1335.
Harrison PW, et al. 2015. Sexual selection drives evolution and rapid turnover of male gene expression. Proc Natl Acad Sci U S A. 112(14):4393–4398.
Haygood R, Babbitt CC, Fedrigo O, Wray GA. 2010. Contrasts between adaptive coding and noncoding changes during human evolution. Proc Natl Acad Sci U S A. 107(17):7853–7857.
He Q, et al. 2011. High conservation of transcription factor binding and evidence for combinatorial regulation across six *Drosophila* species. Nat Genet. 43(5):414–420.
Holm S. 1979. A simple sequentially rejective multiple test procedure. Scand J Stat. 6:65–70.
Hongo JA, de Castro GM, Cintra LC, Zerlotini A, Lobo FP. 2015. POTION: an end-to-end pipeline for positive Darwinian selection detection in genome-scale data through phylogenetic comparison of protein-coding genes. BMC Genomics 16:567.
Howard RW, Blomquist GJ. 2005. Ecological, behavioral, and biochemical aspects of insect hydrocarbons. Annu Rev Entomol. 50:371–393.
Hsu J-C, et al. 2012. Discovery of genes related to insecticide resistance in *Bactrocera dorsalis* by functional genomic analysis of a *de novo* assembled transcriptome. PLoS One 7:e40950.
Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. Mol Biol Evol. 33(6):1635–1638.
Huylmans AK, López Ezquerra A, Parsch J, Cordellier M. 2016. *De novo* transcriptome assembly and sex-biased gene expression in the cyclical parthenogenetic *Daphnia galeata*. Genome Biol Evol. 8(10):3120–3139.
Immonen E, Ritchie MG. 2012. The genomic response to courtship song stimulation in female *Drosophila melanogaster*. Proc R Soc Lond B Biol Sci. 279(1732):1359–1365.
Jagadeeshan S, Singh RS. 2005. Rapidly evolving genes of *Drosophila*: differing levels of selective pressure in testis, ovary, and head tissues between sibling species. Mol Biol Evol. 22(9):1793–1801.
Jagadeeshan S, Singh RS. 2007. Rapid evolution of outer egg membrane proteins in the *Drosophila melanogaster* subgroup: a case of ecologically driven evolution of female reproductive traits. Mol Biol Evol. 24(4):929–938.
Jin X, Ha TS, Smith DP. 2008. SNMP is a signaling component required for pheromone sensitivity in *Drosophila*. Proc Natl Acad Sci U S A. 105(31):10996–11001.
Jones P, et al. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30(9):1236–1240.
Juárez ML, et al. 2015. Evaluating mating compatibility within fruit fly cryptic species complexes and the potential role of sex pheromones in pre-mating isolation. ZooKeys 540:125–155.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.
Koboldt DC, et al. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 22(3):568–576.
Kohl J, Huoviala P, Jefferis GS. 2015. Pheromone processing in *Drosophila*. Curr Opin Neurobiol. 34:149–157.
Koonin E, et al. 2004. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. Genome Biol. 5(2):R7.
Kryuchkova-Mostacci N, Robinson-Rechavi M. 2015. Tissue-specific evolution of protein coding genes in human and mouse. PLoS One 10(6):e0131673.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods. 9(4):357–359.
Li H, et al. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25(16):2078–2079.
Malone JH, Hawkins DL, Michalak P. 2006. Sex-biased gene expression in a ZW sex determination system. J Mol Evol. 63(4):427–436.
Mank JE. 2017. The transcriptional architecture of phenotypic dimorphism. Nat Ecol Evol. 1(1):0006.
Mank JE, Hultin-Rosenberg L, Axelsson E, Ellegren H. 2007. Rapid evolution of female-biased, but not male-biased, genes expressed in the avian brain. Mol Biol Evol. 24(12):2698–2706.
Mank JE, Hultin-Rosenberg L, Zwahlen M, Ellegren H. 2008. Pleiotropic constraint hampers the resolution of sexual antagonism in vertebrate gene expression. Am Nat. 171(1):35–43.
Manni M, et al. 2015. Relevant genetic differentiation among Brazilian populations of *Anastrepha fraterculus* (Diptera, Tephritidae). ZooKeys 540:157–173.
Martin SH, et al. 2013. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. Genome Res. 23(11):1817–1828.
McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature 351(6328):652–654.
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(10):1007–1016.
Meisel RP. 2011. Towards a more nuanced understanding of the relationship between sex-biased gene expression and rates of protein-coding sequence evolution. Mol Biol Evol. 28(6):1893–1900.
Morrow J, Riegler M, Gilchrist A, Shearman D, Frommer M. 2014. Comprehensive transcriptome analysis of early male and female *Bactrocera jarvisi* embryos. BMC Genet. 15(Suppl 2):S7.
Noor MAF, Bennett SM. 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. Heredity (Edinb.) 103(6):439–444.
Norrbom AL, Zucchi RA, Hernández-Ortiz V. 1999. Phylogeny of the genera *Anastrepha* and *Toxotrypana* (Trypetinae: Toxotrypanini) based on morphology. In: Aluja M, Norrbom AL, editors. Fruit flies (Tephritidae): phylogeny and evolution of behavior. Boca Raton (FL): CRC Press. p. 299–342.
Nosil P, Feder JL. 2012. Genomic divergence during speciation: causes and consequences. Philos Trans R Soc Lond B Biol Sci. 367(1587):332–342.
Panhuis TM, Swanson WJ. 2006. Molecular evolution and population genetic analysis of candidate female reproductive genes in *Drosophila*. Genetics 173(4):2039–2047.
Papa F, et al. 2017. Rapid evolution of female-biased genes among four species of *Anopheles* malaria mosquitoes. Genome Res. 27(9):1536–1548.
Papanicolaou A, et al. 2016. The whole genome sequence of the Mediterranean fruit fly, *Ceratitis capitata* (Wiedemann), reveals insights into the biology and adaptive evolution of a highly invasive pest species. Genome Biol. 17(1):192.

Parisi M, et al. 2004. A survey of ovary-, testis-, and soma-biased gene expression in *Drosophila melanogaster* adults. Genome Biol. 5(6):R40.
Parsch J, Zhang Z, Baines JF. 2009. The influence of demography and weak selection on the McDonald–Kreitman test: an empirical study in *Drosophila*. Mol Biol Evol. 26(3):691–698.
Perre P, et al. 2016. Toward an automated identification of *Anastrepha* fruit flies in the *fraterculus* group (Diptera, Tephritidae). Neotrop Entomol. 45(5):554–558.
Perry JC, Harrison PW, Mank JE. 2014. The ontogeny and evolution of sex-biased gene expression in *Drosophila melanogaster*. Mol Biol Evol. 31(5):1206–1219.
Posada D. 2003. Using MODELTEST and PAUP* to select a model of nucleotide substitution. In: Baxevanis A, Davison D, Page R, Petsko G, Stein L, editors. Current protocols in bioinformatics. Hoboken (NJ): John Wiley & Sons, Inc.
Pröschel M, Zhang Z, Parsch J. 2006. Widespread adaptive evolution of *Drosophila* genes with sex-biased expression. Genetics 174(2):893–900.
Purandare SR, Bickel RD, Jaquiery J, Rispe C, Brisson JA. 2014. Accelerated evolution of morph-biased genes in pea aphids. Mol Biol Evol. 31(8):2073–2083.
Ram KR, Wolfner MF. 2007. Seminal influences: *Drosophila* Acps and the molecular interplay between males and females during reproduction. Integr Comp Biol. 47:427–445.
Renaut S, Nolte AW, Bernatchez L. 2010. Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus* spp. Salmonidae). Mol Ecol. 19:115–131.
Rezende VB, et al. 2016. Head transcriptomes of two closely related species of fruit flies of the *Anastrepha fraterculus* group reveals divergent genes in species with extensive gene flow. G3 6:3283–3295.
Roberts A, Pachter L. 2013. Streaming fragment assignment for real-time analysis of sequencing experiments. Nat Methods. 10(1):71–73.
Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26(1):139–140.
Ruiz M, Sanchez D, Canal I, Acebes A, Ganfornina MD. 2011. Sex-dependent modulation of longevity by two *Drosophila* homologues of human Apolipoprotein D, GLaz and NLaz. Exp Gerontol. 46(7):579–589.
Rull J, et al. 2013. Evolution of pre-zygotic and post-zygotic barriers to gene flow among three cryptic species within the *Anastrepha fraterculus* complex. Entomol Exp Appl. 148(3):213–222.
Salvemini M, et al. 2014. *De Novo* assembly and transcriptome analysis of the Mediterranean Fruit Fly *Ceratitis capitata* early embryos. PLoS One 9(12):e114191.
Sánchez-Gracia A, Vieira FG, Almeida FC, Rozas J. 2011. Comparative genomics of the major chemosensory gene families in Arthropods. In: Encyclopedia of life sciences. Chichester (United Kingdom): John Wiley & Sons, Ltd.
Sanchez-Gracia A, Vieira FG, Rozas J. 2009. Molecular evolution of the major chemosensory gene families in insects. Heredity (Edinb.) 103(3):208–216.
Scally M, et al. 2016. Resolution of inter and intra-species relationships of the West Indian fruit fly *Anastrepha obliqua*. Mol Phylogenet Evol. 101:286–293.
Sim SB, Geib SM. 2017. A chromosome-scale ssembly of the *Bactrocera cucurbitae* genome provides insight to the genetic basis of *white pupae*. G3 7(6):1927–1940.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19):3210–3212.
Sitnik JL, Gligorov D, Maeda RK, Karch F, Wolfner MF. 2016. The female post-mating response requires genes expressed in the secondary cells of the male accessory gland in *Drosophila melanogaster*. Genetics 202(3):1029–1041.
Smith-Caldas MRB, McPheron BA, Silva JG, Zucchi RA. 2001. Phylogenetic relationships among species of the *fraterculus* group (*Anastrepha*: Diptera: Tephritidae) inferred from DNA sequences of mitochondrial cytochrome oxidase I. Neotrop Entomol. 30(4):565–573.
Snook RR, Chapman T, Moore PJ, Wedell N, Crudgington HS. 2009. Interactions between the sexes: new perspectives on sexual selection and reproductive isolation. Evol Ecol. 23(1):71–91.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9):1312–1313.
Stoletzki N, Eyre-Walker A. 2011. Estimation of the neutrality index. Mol Biol Evol. 28(1):63–70.
Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro CF. 2001. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. Proc Natl Acad Sci U S A. 98(13):7375–7379.
Swanson WJ, Vacquier VD. 2002. The rapid evolution of reproductive proteins. Nat Rev Genet. 3(2):137–144.
Swanson WJ, Wong A, Wolfner MF, Aquadro CF. 2004. Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies genes subjected to positive selection. Genetics 168(3):1457–1465.
Ting C-T, Tsaur S-C, Wu M-L, Wu C-I. 1998. A rapidly evolving homeobox at the site of a hybrid sterility gene. Science 282(5393):1501–1504.
Tootle TL, Williams D, Hubb A, Frederick R, Spradling A. 2011. *Drosophila* eggshell production: identification of new genes and coordination by Pxt. PLoS One 6(5):e19943.
Torgerson DG, Kulathinal RJ, Singh RS. 2002. Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes. Mol Biol Evol. 19(11):1973–1980.
Vacquier VD, Swanson WJ. 2011. Selection in the rapid evolution of gamete recognition proteins in marine invertebrates. Cold Spring Harb Perspect Biol. 3(11):a002931.
Van Dyken JD, Wade MJ. 2010. The genetic signature of conditional expression. Genetics 184(2):557–570.
Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. 2015. RELAX: detecting relaxed selection in a phylogenetic framework. Mol Biol Evol. 32(3):820–832.
Whittle CA, Extavour CG. 2017. Rapid evolution of ovarian-biased genes in the yellow fever mosquito *Aedes aegypti*. Genetics 206(4):2119–2137.
Yamaguchi S, Katagiri S, Sekimizu K, Natori S, Homma KJ. 2005. Involvement of EDTP, an egg-derived tyrosine phosphatase, in the early development of *Drosophila melanogaster*. J Biochem. 138(6):721–728.
Yang L, Zhang Z, He S. 2016. Both male-biased and female-biased genes evolve faster in fish genomes. Genome Biol Evol. 8(11):3433–3445.
Yang X, et al. 2006. Tissue-specific expression and regulation of sexually dimorphic genes in mice. Genome Res. 16(8):995–1004.
Yang Y, et al. 2011. Regulation of flagellar motility by the conserved flagellar protein CG34110/Ccdc135/FAP50. Mol Biol Cell. 22(7):976–987.
Yang Y, Smith SA. 2014. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. Mol Biol Evol. 31(11):3081–3092.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24(8):1586–1591.
Ye J, et al. 2006. WEGO: a web tool for plotting GO annotations. Nucleic Acids Res. 34(Web Server):W293–W297.
Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol. 22(12):2472–2479.

Zhang L, Li W-H. 2004. Mammalian housekeeping genes evolve more slowly than tissue-specific genes. Mol Biol Evol. 21(2):236–239.
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(7167):233–237.
Zhang Z, Hambuch TM, Parsch J. 2004. Molecular evolution of sex-biased genes in *Drosophila*. Mol Biol Evol. 21(11):2130–2139.
Zhang Z, et al. 2006. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. Genomics Proteomics Bioinformatics 4(4):259–263.
Zhang Z, Lv X, Jiang J, Zhang L, Zhao Y. 2013. Dual roles of Hh signaling in the regulation of somatic stem cell self-renewal and germline stem cell maintenance in *Drosophila* testis. Cell Res. 23(4):573–576.
Zucchi RA. 2000. Taxonomia. In: Malavasi A, Zucchi RA, editors. Moscas-das-frutas de importância econômica no Brasil: Conhecimento básico e aplicado. Ribeirão Preto (Brazil): Holos. p. 1–24.

Associate editor: Judith Mank