The comparative study of five sex-determining proteins across insects unveils high rates of evolution at basal components of the sex determination cascade

José M. Eirín-López [1,*] and Lucas Sánchez [2]

[1] Department of Biological Sciences, Florida International University, North Miami FL, USA. Email address: jeirinlo@fiu.edu

[2] Centro de Investigaciones Biológicas (C.S.I.C.), Madrid, Spain. Email address: lsanchez@csic.es

** Corresponding author:* Jose M. Eirin-Lopez, Department of Biological Sciences, Florida International University, Marine Sciences Program, Biscayne Bay Campus, 3000 NE 151 St., suite MSB-360, North Miami, FL 33181, USA. Tel: 305-919-4000, Fax: 305-919-4030, Email: jeirinlo@fiu.edu (chromevol.com).

ABSTRACT

In insects, the sex determination cascade is composed of genes that interact with each other in a strict hierarchical manner, constituting a co-adapted gene complex built in reverse order from bottom to top. Accordingly, ancient elements at the bottom are expected to remain conserved ensuring the correct functionality of the cascade. In the present work we have studied the levels of variation displayed by five key components of the sex determination cascade across 59 insect species, including *Sex-lethal*, *transformer*, *transformer-2*, *fruitless*, *doublesex* and *sister-of-Sex-lethal* (a paralog of *Sxl* encompassing sex-independent functions). Surprisingly, our results reveal that basal components of the cascade (*doublesex*, *fruitless*) seem to evolve more rapidly than previously suspected. Indeed, in the case of *Drosophila*, these proteins evolve more rapidly than the master regulator *Sex-lethal*. These results agree with the notion suggesting that genes involved in early aspects of development will be more constrained due to the large deleterious pleiotropic effects of mutations, resulting in increased levels of purifying selection at top positions of the cascade. The analyses of the selective episodes involved in the recruitment of *Sxl* into sex determining functions in *Drosophila* further support this idea, suggesting the presence of bursts of adaptive selection in the common ancestor of drosophilids, followed by purifying selection preserving the master regulatory role of this protein. Altogether, this work underscores the importance of the position of sex determining genes in the cascade, constituting a major constraint shaping the molecular evolution of the insect sex determination pathway.

Keywords: Evolution, development, sex-specific genes, insects, evolutionary rates.

INTRODUCTION

Drosophila melanogaster has been the paradigm for understanding the genetic and molecular basis underlying sex determination (Bopp et al. 2014; Sánchez 2008). In this insect, the program committing the embryo to either the male or the female pathway is under the control of the gene *Sex lethal* (*Sxl*) (Cline 1978; Penalva and Sánchez 2003). The study of the epistatic relationships between *Sxl* and the other genes involved in sex determination in insects [i.e., *transformer* (*tra*), *transformer-2* (*tra-2*), *fruitless (fru)* and *doublesex* (*dsx*)] has revealed a hierarchical interaction among them during development (Baker and Ridge 1980), with the product of one gene controlling the sex-specific splicing of the primary transcript of the gene immediately downstream (reviewed in Sánchez 2008) (Fig. 1A). The search for genes homologous to the sex determination genes of *D. melanogaster* has been undertaken in other insects (reviewed in Gempe and Beye 2011; Sánchez 2008; Verhulst et al. 2010). It has been found a conserved relationship among *dsx*/*tra*/*tra-2* across dipterans, so that this axis represents the ancestral state of the sex determination cascade, with the recruitment of *Sxl* as master regulator constituting an innovation acquired later on in *Drosophila.*

In insects, the sex determination pathway constitutes a regulatory cascade that evolved in reverse order, from the final step in the hierarchy that creates the required product, to the first step in the pathway that allows synthesis of the initial precursor (Wilkins 1995; Bopp et al. 2014; Gempe and Beye 2011)Fig. 1A). This process involved the sequential acquisition of genetic switches each one reversing the action of the previous one, with the final step in the cascade (bottom) representing the oldest (Pomiankowski et al. 2004; Wilkins 1995). Under this model, trans-regulatory elements more recently recruited into sex determining pathways are

expected to cause divergence towards the top because of recent regulatory change (i.e., the *xol-1* and *her-1* genes in *Caenorhabditis elegans*, the *Sxl* gene in *Drosophila*) while ancient elements at the bottom would remain conserved (i.e., the *tra-1* gene in *Caenorhabditis elegans*, the *dsx* gene in *Drosophila*) ensuring the correct functionality of the cascade (Verhulst et al. 2010) (Fig. 1B). On the other hand, an alternative interpretation of the evolution of the cascade can be drawn from the developmental constraint hypothesis (Artieri et al. 2009), suggesting that genes involved in early aspects of development (which, as in the case of *Sxl*, are likely to regulate a large number of downstream effectors through hierarchical regulatory cascades) would be more constrained due to the large deleterious pleiotropic effects of mutations, resulting in increased levels of purifying selection at top positions of the cascade (Fig. 1C).

Overall, the current body of knowledge hints the presence of diverse specific constraints operating at different levels of the cascade, probably imposed by the epistatic interactions of its constituting components with upstream regulators and downstream target genes (Sánchez 2008), as well as by pleiotropic effects (i.e., additional functions unrelated to sex (Kunte et al. 2014)). However, the constraints shaping the evolution of the insect cascade still remain uncertain, mainly because of the lack of comparative studies across different levels of the cascade in diverse insect species. To fill this gap, the present work investigates the levels of variation displayed by five sex determining proteins across 59 insect species. Our results unveil high rates of evolution at basal components of the cascade and provide clues to understand the mechanisms responsible for the recruitment of *Sxl* into sex determination functions at top of the *Drosophila* cascade.

MATERIALS AND METHODS

Evolutionary rates of sex determination proteins

We have performed extensive data mining experiments to build up the dataset of sex determination proteins used in the present work, consisting of 166 sequences (40 SXL, 27 TRA, 30 TRA-2, 25 FRU, 22 DSX-Male and 22 DSX-Female). In addition, 12 SSX protein sequences from *Drosophila* representatives (sex-independent functions) were also included for further comparisons. Altogether, the taxonomic range covered by these sequences spans 6 insect Orders (Coleoptera, Diptera, Hemiptera, Hymenoptera, Lepidoptera and Phthiraptera) encompassing 59 different insect species: *Acromyrmex echinatior, Acyrthosiphon pisum, Aedes aegypti, Anastrepha amita, A. bistrigata, A. fraterculus, A. grandis, A. ludens, A. obliqua, A. serpentina, A. sororcula, A. striata, A. suspensa, Anopheles darlingi, A. gambiae, Antheraea assama, Apis cerana, A. dorsata, A. florea, A. mellifera, Bactrocera oleae, Bombus impatiens, B. terrestris, Bombyx mori, Bradysia coprophila, Camponotus floridanus, Ceratitis capitata, Chrysomya rufifacies, Culex quinquefasciatus, Danaus plexippus, Drosophila ananassae, D. erecta, D. grimshawi, D. hydei, D. mauritiana, D. melanogaster, D. mojavensis, D. persimilis, D. pseudoobscura, D. sechellia, D. simulans, D. subobscura, D. virilis, D. willistoni, D. yakuba, Glossina morsitans, G. morsitans, Harpegnathos saltator, Lucilia cuprina, Megachile rotundata, Megaselia scalaris, Musca domestica, Nasonia vitripennis, Pediculus humanus corporis, Rhynchosciara americana, Sciara ocellaris, Stomoxys calcitrans, Tribolium castaneum, Trichomegalosphys pubescens* (see Supplementary Table 1 for details). Multiple alignments of protein sequences were implemented using the BIOEDIT program (Hall 1999) and visually inspected for errors. Estimations of protein divergence among insects for each component of the sex determination cascade were carried out using *p*-distances with partial deletion (95%), as this approach is known to give better results for distantly related taxa owing to its smaller variance. Estimations

were performed using MEGA version 6 (Tamura et al. 2013). Estimations of divergence times between all pairs of taxa studied were manually retrieved from the TimeTree database (Hedges et al. 2006). Divergence times between taxa are listed together with the corresponding pairwise protein divergences in Supplementary Table 2. Regression analyses describing the relationships between protein divergence estimates and divergence time estimates were implemented for each sex determining protein as well as for SSX (1,439 comparisons in total) using the program STATGRAPHICS Plus version 5.1 (Warrenton, VA). The rates of evolution for the studied proteins (amino acid substitutions/site per million years) were subsequently inferred based on the calculated regression coefficients.

Molecular evolutionary analyses and episodic diversifying selection in *Sxl*

Most part of the molecular evolutionary analyses were carried out using the program MEGA version 6 (Tamura et al. 2013) except where noted. The molecular clock hypothesis was tested in each sex determining protein by using likelihood ratio tests based on the models of evolution defined (see Table 1 for details). Additional tests for the presence of local molecular clocks where carried out in the case of SXL (insects) by using the program HyPhy (Pond et al. 2005). The SXL phylogeny was reconstructed using the maximum-likelihood approach based on the model of evolution that best fit the sets of sequences analyzed. The tree was rooted using the cladocerans *Daphnia pulex*, diverging from the order Diptera approximately 443.2 MYA (Hedges et al. 2006). The reliability of the reconstructed topology was contrasted by nonparametric bootstrap (1,000 replicates) and further examined by bayesian analysis using the program BEAST version 1.7 (Drummond et al. 2012). Three independent Markov chain Monte Carlo (MCMC) runs of 10,000,000 generations each were performed to generate posterior

probabilities, sampling tree topologies every 1,000 generations to ensure the independence of successive trees and discarding the first 1,000 trees of each run as burn-in.

The evolution of *Sxl* was examined for lineages displaying evidence of diversifying selection episodes ($\omega>1$) by using the branch-site Random Effects Likelihood (REL) model (Pond and Frost 2005). Codon positions were examined using the phylogeny of insects as a reference (Wheeler et al. 2001; Wiegmann et al. 2011), without making any prior assumptions about which lineages have been subject to diversifying selection. The proportion of sites inferred to be evolving under diversifying selection at each branch were estimated using likelihood ratio tests (LRTs), resulting in a *p*-value for episodic selection corrected for multiple testing using the Holm-Bonferroni method. The strength of selection was partitioned into three categories ($\omega>5$, $\omega=1$, $\omega=0$) for descriptive purposes, using three different corrected significance levels ($p<0.001$, $p<0.01$ and $p<0.05$) to assess the obtained results. Selection analyses were further expanded to single codon positions in *Sxl* sequences by using a mixed effects model of evolution (MEME), modeling variable ω across lineages at an individual site (Murrell et al. 2012). The numbers of synonymous and nonsynonymous substitutions at these codon positions were estimated and subsequently located in the corresponding functional regions of the SXL protein (N- and C-terminal domains, RNA binding domain). *Sxl* codons subject to diversifying selection were also analyzed in a phylogenetic context, providing information on internal branches accumulating higher numbers of nonsynonymous mutations. All analyses in this section were carried out using the Datamonkey webserver (Delport et al. 2010; Poon et al. 2009).

RESULTS AND DISCUSSION

Rates of evolution in the sex determining proteins from insects

The study of the rates of molecular evolution in key components of the sex determination pathway from insects yielded three interesting results~~:~~. First, the sex determining proteins studied in the present work evolve at constant rates as indicated by global molecular clock tests (Table 1). Second, sex determination proteins located at bottom positions of the cascade (i.e., *DSX* and *FRU*) display relatively high rates of evolution in insects (Fig. 2A). This is specially evident in the case of *Drosophila* (Fig. 2B), where basal proteins display higher evolutionary rates compared with proteins located at top positions (i.e., *SXL*). The high rates of evolution found in DSX might be due, at least in part, to the presence of sexual selection operating on this gene in order to keep up with modifications in downstream components at the bottom of the cascade (e.g., sexual cytodifferentiation genes). Third, TRA represents the fastest evolving protein in the sex determination cascade, encompassing a rate of evolution of approximately 2.57×10^{-3} and 1.25×10^{-2} substitutions/site per MY in insects and in *Drosophila*, respectively (see Table 2 for detailed evolutionary rates). Although unusually high rates of neutral functional evolution have been previously reported for this gene in *Drosophila* (Kulathinal et al. 2003; McAllister and McVean 2000), the present results constitute the first evidence showing rapid evolution of TRA in other insect species. We believe that this observation bears relevance, as *transformer* plays a master regulatory role on top of the sex determination cascade in some non-drosophilid insects.

The high rate of evolution displayed by TRA proteins can be reconciled with its top position in the cascade based on the molecular mechanism of TRA function (Black 2003). Accordingly, TRA participates in splicing regulation through its interaction

(through their SR domains) with other proteins carrying RNA-binding domains, such as Transformer-2. Although SR dipeptide content can vary among TRA proteins, it appears that protein functionality depends on the presence of a minimum number of SR dipeptides located at very conserved positions (Ruiz et al. 2007). Therefore, while SR regions must remain conserved to assure TRA function, this protein can accept high levels of neutral variation on those regions not involved in protein-protein interactions (Kulathinal et al. 2003; McAllister and McVean 2000). Similarly to TRA, TRA-2 also participates in protein-protein interactions. However, this protein constitutes a general splicing factor that also interacts with RNAs, requiring a higher degree of conservation to preserve its functionality, specially at RNA recognition motifs (Sarno et al. 2010). That is mirrored by the low evolutionary rate displayed by this protein in insects (Fig. 2).

SXL (the top component of the *Drosophila* sex determination cascade) constitutes the slowest evolving sex determining protein in drosophilids (Fig. 2B) as well as a slow evolving protein in other insect species (see Table 2 for details). However, there is still the possibility that such a high degree of conservation is a result of the lack of sex-specific functions in insects other than *Drosophila* (Cline et al. 2010; Sánchez 2008). Two approaches were followed in order to explore this scenario: first, the analysis of SXL in non drosophilid insects revealed an evolutionary rate of 0.95×10^{-3} substitutions/site per MY (Table 2), constituting a much lower rate than the one estimated for *Drosophila* (2.80×10^{-3} substitutions/site per MY). Indeed, it seems that all sex determining proteins from *Drosophila* evolve significantly faster than their orthologs in other insects (Fig. 2B, Table 2). These results agree with the rapid evolution of the sex determination cascade in *Drosophila*, with *Sxl* occupying a top position, after medfly and fruitfly diverged (Civetta and Singh 1998; Cline et al. 2010). Second, the analysis of SSX

(a paralog of SXL which took on roles unrelated to sex through a process of neo-functionalized after duplication (Cline et al. 2010)) revealed that this protein evolves almost twice as fast as SXL in drosophilids ($4.32x10^{-3}$ substitutions/site per MY, Table 2), in agreement with previous reports describing a signature of rampant positive selection and relaxation of purifying selection in this gene (Mullon et al. 2012). This result suggests a reinforcement of the selective constraints operating on SXL, most likely resulting from its recruitment into sex-related roles at the top of the *Drosophila* cascade (Cline et al. 2010; Mullon et al. 2012), as well from its role in controlling dosage compensation (reviewed in (Penalva and Sánchez 2003)).

Selective episodes leading to the recruitment of *Sxl* into sex-specific functions

Modifications in the specific components of any network are expected to impact their hierarchical organization and their interactions, especially in those cases where components at top regulatory positions have been modified very recently (Bopp et al. 2014; Gempe and Beye 2011). Since that is precisely the case of *Drosophila* (*Sxl* has been recruited into sex-specific functions at the top of the cascade) this group provides us with a very powerful model to address two important questions: When (during the evolution of insects) and where (in the SXL protein) did the selective episodes responsible for the recruitment of *Sxl* into sex-specific functions take place?. To answer the first question, we screened the phylogeny of insects for lineages at which *Sxl* experienced episodic adaptive selection ($\omega>1$), finding 12 significant branches ($p\leq0.05$) located exclusively within dipterans (Fig. 3A). Interestingly, 8 of these branches fall within the drosophilid subtree, including a highly significant branch at the root of this lineage ($p\leq0.001$). Combined with local molecular clock analyses (Fig. 3B), these results indicate that episodic adaptive selection was probably responsible for the non clock-like

behavior of *Sxl* during its recruitment into sex specific functions in drosophilids (Mullon et al. 2012).

To answer the second part of the question, we studied the specific protein positions targeted by selection in SXL. Significant evidence of adaptive selection was found at 15 codons ($p \leq 0.05$) predominantly located at N- and C-terminal regions (Fig. 4A). These results are consistent with functional studies showing that the sex-specific properties of extant *Drosophila* SXL depend on its global structure, and that modifications at N- and C-terminal domains of SXL in the drosophilid lineage represented co-evolutionary changes determining the appropriate folding of SXL to carry out its sex-specific function (Ruiz et al. 2013). The analysis of the episodes of adaptive selection in *Sxl* revealed significantly higher proportions of nonsynonymous substitutions ($p<0.05$) (Fig. 4B). More specifically, higher numbers of nonsynonymous substitutions were found at 33.3% of the codons subject to episodic adaptive selection in the common ancestor of *Sxl* in Diptera (5 out of 15 codons); 13.3% in the common ancestor of Drosophilidae, Calliphoridae, Muscidae, Tephritidae and Sciaridae (2 out of 15 codons); 53.3% in the common ancestor of Drosophilidae, Calliphoridae, Muscidae and Tephritidae (8 out of 15 codons, highlighted with red boxes in Fig. 4B); 6.7% in the common ancestor of Drosophilidae, Calliphoridae and Muscidae (1 out of 15 codons); and 60% in the common ancestor of drosophilids (9 out of 15 codons, highlighted with red circles in Fig. 4B). Two major conclusions can be drawn from these results: First, the diversification of *Sxl* in dipterans seems to have been driven by episodes of adaptive selection involving amino acid replacements at specific codons in terminal protein domains. Second, the recruitment of *Sxl* into sex-specific roles required bursts of adaptive selection during the evolution of dipterans and most importantly in the common ancestor of drosophilids, probably taking

advantage of its preexisting role as a general splicing factor (Ruiz et al. 2003; Serna et al. 2004).

Conclusions

The rates of evolution observed in sex determining proteins suggest that the position of the different genes in the sex determination cascade has played a very important role shaping the molecular evolution of this pathway in insects. Accordingly, genes involved in early aspects of development (i.e., *Sxl*) will be more constrained than genes expressed later on (i.e., *dsx*, *fru*) due to the large deleterious pleiotropic effects of mutations at top positions of the cascade. Consequently, increased levels of purifying selection will observed at top positions of the cascade, while higher levels of variation will be observed at basal components interacting with diverse downstream factors (e.g., sexual selection). This is nicely illustrated by the recruitment of *Sxl* on top of the *Drosophila* cascade based on bursts of adaptive selection in the common ancestor of drosophilids, followed by purifying selection preserving the master regulatory role of this protein. In addition to providing us with a privileged insight into the mechanisms guiding the evolution of sex determination, the present work constitutes a powerful model for future studies on other functionally relevant co-adapted gene complexes.

ACKNOWLEDGMENTS

The present work was supported by start-up funds from the College of Arts and Sciences at Florida International University (CAS-FIU) to J.M.E.-L.

REFERENCES

Artieri CG, Haerty W, Singh RS (2009) Ontogeny and phylogeny: molecular signatures of selection, constraint, and temporal pleiotropy in the development of Drosophila BMC Biol 7:42 doi:1741-7007-7-42 [pii]
10.1186/1741-7007-7-42

Baker BS, Ridge KA (1980) Sex and the single cell. I. On the action of major loci affecting sex determination in Drosophila melanogaster Genetics 94:383-423

Black DL (2003) Mechanisms of alternative pre-messenger RNA splicing Annu Rev Biochem 72:291-336 doi:10.1146/annurev.biochem.72.121801.161720
121801.161720 [pii]

Bopp D, Saccone G, Beye M (2014) Sex determination in insects: variations on a common theme Sex Dev 8:20-28 doi:000356458 [pii]
10.1159/000356458

Civetta A, Singh RS (1998) Sex-related genes, directional sexual selection, and speciation Mol Biol Evol 15:901-909

Cline TW (1978) Two closely linked mutations in Drosophila melanogaster that are lethal to opposite sexes and interact with daughterless Genetics 90:683-698

Cline TW, Dorsett M, Sun S, Harrison MM, Dines J, Sefton L, Megna L (2010) Evolution of the Drosophila feminizing switch gene Sex-lethal Genetics 186:1321-1336 doi:genetics.110.121202 [pii]
10.1534/genetics.110.121202

Delport W, Poon AF, Frost SD, Kosakovsky Pond SL (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology Bioinformatics 26:2455-2457 doi:btq429 [pii]
10.1093/bioinformatics/btq429

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7 Mol Biol Evol 29:1969-1973 doi:mss075 [pii]
10.1093/molbev/mss075

Gempe T, Beye M (2011) Function and evolution of sex determination mechanisms, genes and pathways in insects Bioessays 33:52-60 doi:10.1002/bies.201000043

Goldman N, Whelan S (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach Mol Biol Evol 18:681-699

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT Nucl Acids Symp Ser 41:95-98

Hedges SB, Dudley J, Kumar S (2006) TimeTree: a public knowledge-base of divergence times among organisms Bioinformatics 22:2971-2972 doi:btl505 [pii]
10.1093/bioinformatics/btl505

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences Comput Applic Biosci 8:275-282

Kulathinal RJ, Skwarek L, Morton RA, Singh RS (2003) Rapid evolution of the sex-determining gene, transformer: structural diversity and rate heterogeneity among sibling species of Drosophila Mol Biol Evol 20:441-452

Kunte K et al. (2014) doublesex is a mimicry supergene Nature 507:229-232 doi:nature13112 [pii]
10.1038/nature13112

McAllister BF, McVean GA (2000) Neutral evolution of the sex-determining gene transformer in Drosophila Genetics 154:1711-1720

Mullon C, Pomiankowski A, Reuter M (2012) Molecular evolution of Drosophila Sex-lethal and related sex determining genes BMC Evol Biol 12:5 doi:1471-2148-12-5 [pii]
10.1186/1471-2148-12-5
Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL (2012) Detecting individual sites subject to episodic diversifying selection PLoS Genet 8:e1002764 doi:10.1371/journal.pgen.1002764
PGENETICS-D-12-00164 [pii]
Penalva LO, Sánchez L (2003) RNA binding protein sex-lethal (Sxl) and control of Drosophila sex determination and dosage compensation Microbiol Mol Biol Rev 67:343-359, table of contents
Pomiankowski A, Nothiger R, Wilkins A (2004) The evolution of the Drosophila sex-determination pathway Genetics 166:1761-1773 doi:166/4/1761 [pii]
Pond SL, Frost SD (2005) A genetic algorithm approach to detecting lineage-specific variation in selection pressure Mol Biol Evol 22:478-485 doi:msi031 [pii]
10.1093/molbev/msi031
Pond SL, Frost SD, Muse SV (2005) HyPhy: hypothesis testing using phylogenies Bioinformatics 21:676-679 doi:bti079 [pii]
10.1093/bioinformatics/bti079
Poon AF, Frost SD, Pond SL (2009) Detecting signatures of selection from DNA sequences using Datamonkey Methods Mol Biol 537:163-183 doi:10.1007/978-1-59745-251-9_8
Ruiz MF, Goday C, Gonzalez P, Sánchez L (2003) Molecular analysis and developmental expression of the Sex-lethal gene of Sciara ocellaris (Diptera order, Nematocera suborder) Gene Expr Patterns 3:341-346 doi:S1567133X03000310 [pii]
Ruiz MF et al. (2007) The gene transformer of anastrepha fruit flies (Diptera, tephritidae) and its evolution in insects PLoS ONE 2:e1239
Ruiz MF, Sarno F, Zorrilla S, Rivas G, Sánchez L (2013) Biochemical and functional analysis of Drosophila-sciara chimeric sex-lethal proteins PLoS ONE 8:e65171 doi:10.1371/journal.pone.0065171
PONE-D-13-08668 [pii]
Sánchez L (2008) Sex-determining mechanisms in insects Int J Dev Biol 52:837-856 doi:072396ls [pii]
10.1387/ijdb.072396ls
Sarno F, Ruiz MF, Eirin-Lopez JM, Perondini AL, Selivon D, Sánchez L (2010) The gene transformer-2 of *Anastrepha* fruit flies (Diptera, Tephritidae) and its evolution in insects BMC Evol Biol 10:140
Serna E, Gorab E, Ruiz MF, Goday C, Eirín-López JM, Sánchez L (2004) The gene Sex-lethal of the Sciaridae family (order Diptera, suborder Nematocera) and its phylogeny in dipteran insects Genetics 168:907-921
Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0 Mol Biol Evol 30:2725-2729 doi:mst197 [pii]
10.1093/molbev/mst197
Verhulst EC, van de Zande L, Beukeboom LW (2010) Insect sex determination: it all evolves around transformer Curr Opin Genet Dev 20:376-383 doi:S0959-437X(10)00079-1 [pii]
10.1016/j.gde.2010.05.001
Wheeler WC, Whiting M, Wheeler QD, Carpenter JM (2001) The phylogeny of the extant hexapod orders Cladistics 12:113-169

Wiegmann BM et al. (2011) Episodic radiations in the fly tree of life Proc Natl Acad Sci U S A 108:5690-5695 doi:1012675108 [pii]
10.1073/pnas.1012675108
Wilkins AS (1995) Moving up the hierarchy: a hypothesis on the evolution of a genetic sex determination pathway Bioessays 17:71-77 doi:10.1002/bies.950170113

FIGURE LEGENDS

Figure 1.- Schematic representation of the hierarchical epistatic interactions constituting the sex determination cascade in *Drosophila* [adapted from (Sánchez 2008)] evolving from bottom to top (*doublesex*, DSX; *fruitless*, FRU; *transformer-2*, TRA-2; *transformer*, TRA; *Sex-lethal*, SXL). **A**, In the absence of X/A signal in males, truncated SXL and TRA proteins will be produced leading to the synthesis of male-specific FRU and DSX that will eventually result in maleness. The major components of the cascade analyzed in the present work are indicated in grey background. **B**, Under the bottom-up hypothesis, genes more recently recruited into sex determining pathways are expected to cause divergence towards the top of the cascade. **C**, According to the developmental constraint hypothesis, genes involved in early aspects of development would be more constrained due to the large deleterious pleiotropic effects of mutations.

Figure 2.- Rates of evolution in sex determination proteins from insects including *Drosophila* (A) and rates of evolution in sex determination proteins exclusive from *Drosophila* (B). Evolutionary rate estimations for DSX have been divided into the male/female common region (c), the female-specific DSX protein (f) and the male-specific DSX protein (m). Evolutionary rates for fast-evolving protamines and slow-evolving cytochrome and histones H2A/H2B are included as references.

Figure 3.- Molecular evolution and diversifying selection in *Sxl* across insects. The taxonomic classification of the insect species studied (Family/Order) is indicated in the right margin of the trees. (A), Episodes of diversifying selection acting on *Sxl* throughout the phylogeny of insects [according to (Wheeler et al. 2001; Wiegmann et al. 2011)]. The ~~the~~ strength of selection at significant branches is represented in red (ω>5), grey (ω=1), and blue (ω=0), with the proportion of sites within each class represented by the color

width. Thicker branches have been classified as undergoing episodic diversifying selection at corrected $p<0.001$ (thickest branches), $p<0.01$ (medium thickness) and $p<0.05$ (thinest branches). (B) Protein phylogeny showing local SXL lineages deviating from a clock-like mode of evolution in insects. Black boxes at internal nodes indicate subtrees at which the molecular clock hypothesis was rejected *(*$p<0.001$). The numbers for interior branches represent bootstrap probabilities (only shown when ≥50%) followed by the corresponding Bayesian posterior probabilities (only shown when ≥0.5). Topologies were rooted using the cladoceran *Daphina* as outgroup.

Figure 4.- Physical position and phylogenetic location of adaptive selection episodes involved in the recruitment of *Sxl* into sex-specific functions. (A) numbers of synonymous (blue bars) and nonsynonymous (red bars) substitutions at codon positions subject to significant episodes of diversifying selection in dipterans ($p<0.05$). (B) Phylogenetic location of the mutations involved in such episodes. Branches in red account for higher numbers of nonsynonymous mutations, branches in blue indicate higher numbers of synonymous mutations, and branches in green represent cases with same numbers of nonsynonymous and synonymous mutations. Red squares indicate codons displaying prevalence of nonsynonymous substitutions in the common ancestor of Drosophilidae, Calliphoridae, Muscidae and Tephritidae. Red circles indicate the same but only in the common ancestor of Drosophilidae.

TABLES

Table 1. Best-fit models of evolution and global molecular clock tests in insect sex determination proteins.

Protein	Model Evolution	lnL	lnL (clock)	*p*-value
SXL (*Drosophila*)	JTT+G	-1225.3	-1234.3	0.0677
SXL (Insects)	WAG+G	-3764.2	-3874.0	4.06×10^{-14}
TRA	JTT+G+F	-3615.2	-3637.3	0.7721
TRA-2	JTT+G	-2146.0	-2184.6	0.0763
FRU	JTT+G	-999.1	-1008.1	0.5633
DSX(c)	JTT+G+I	-3582.5	-3607.5	0.1829
DSX(f)	JTT+G	-2881.0	-2909.5	0.8055
DSX(m)	JTT+G+I	-3017.0	-3042.2	0.1255

Global molecular clock hypothesis rejected (**$p<0.001$). Whelan and Goldman (WAG) and Jones, Taylor and Thornton (JTT) models of protein evolution (Goldman and Whelan 2001; Jones et al. 1992); including gamma distributed variation across sites (G) and invariant sites (I).

Table 2. Rates of protein evolution (amino acid substitutions/site per million years) in the components of the sex determination cascade from insects (excluding *Drosophila*) and from *Drosophila*.

Protein	Insects	*Drosophila*
SXL	$0.95 \times 10^{-3} \pm 1.23 \times 10^{-5}$	$2.80 \times 10^{-3} \pm 3.89 \times 10^{-4}$
SSX	N/A	$4.32 \times 10^{-3} \pm 1.65 \times 10^{-4}$
TRA	$2.57 \times 10^{-3} \pm 1.41 \times 10^{-4}$	$1.25 \times 10^{-2} \pm 3.12 \times 10^{-4}$
TRA-2	$1.00 \times 10^{-3} \pm 2.42 \times 10^{-5}$	$4.56 \times 10^{-3} \pm 2.16 \times 10^{-4}$
FRU	$2.34 \times 10^{-3} \pm 5.02 \times 10^{-5}$	$2.90 \times 10^{-3} \pm 1.47 \times 10^{-4}$
DSX(c)	$1.61 \times 10^{-3} \pm 4.06 \times 10^{-5}$	$5.59 \times 10^{-3} \pm 1.47 \times 10^{-4}$
DSX(f)	$1.73 \times 10^{-3} \pm 3.94 \times 10^{-5}$	$5.59 \times 10^{-3} \pm 1.47 \times 10^{-4}$
DSX(m)	$1.54 \times 10^{-3} \pm 4.22 \times 10^{-5}$	$5.83 \times 10^{-3} \pm 1.58 \times 10^{-4}$

N/A, Not Applicable.

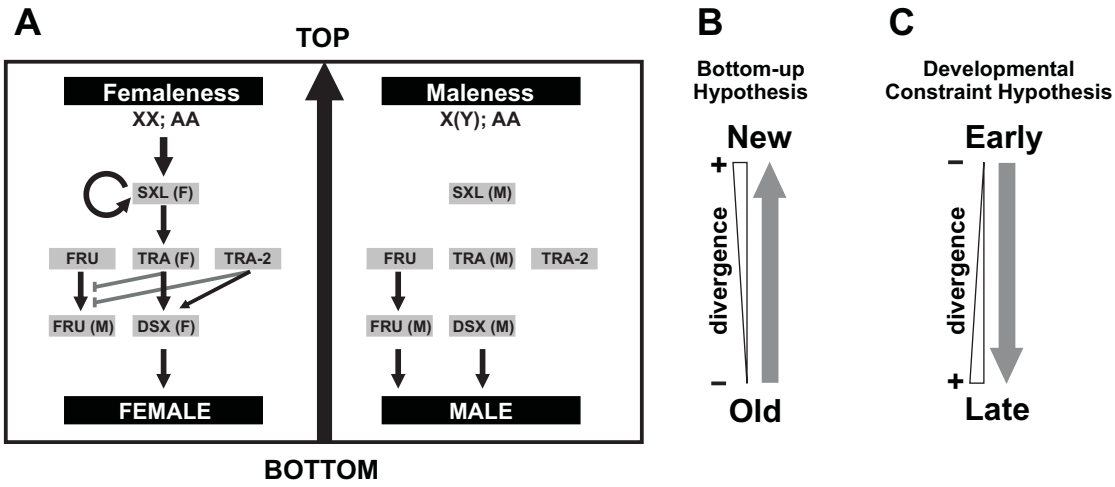

Fig. 1, Eirin-Lopez and Sanchez 2015

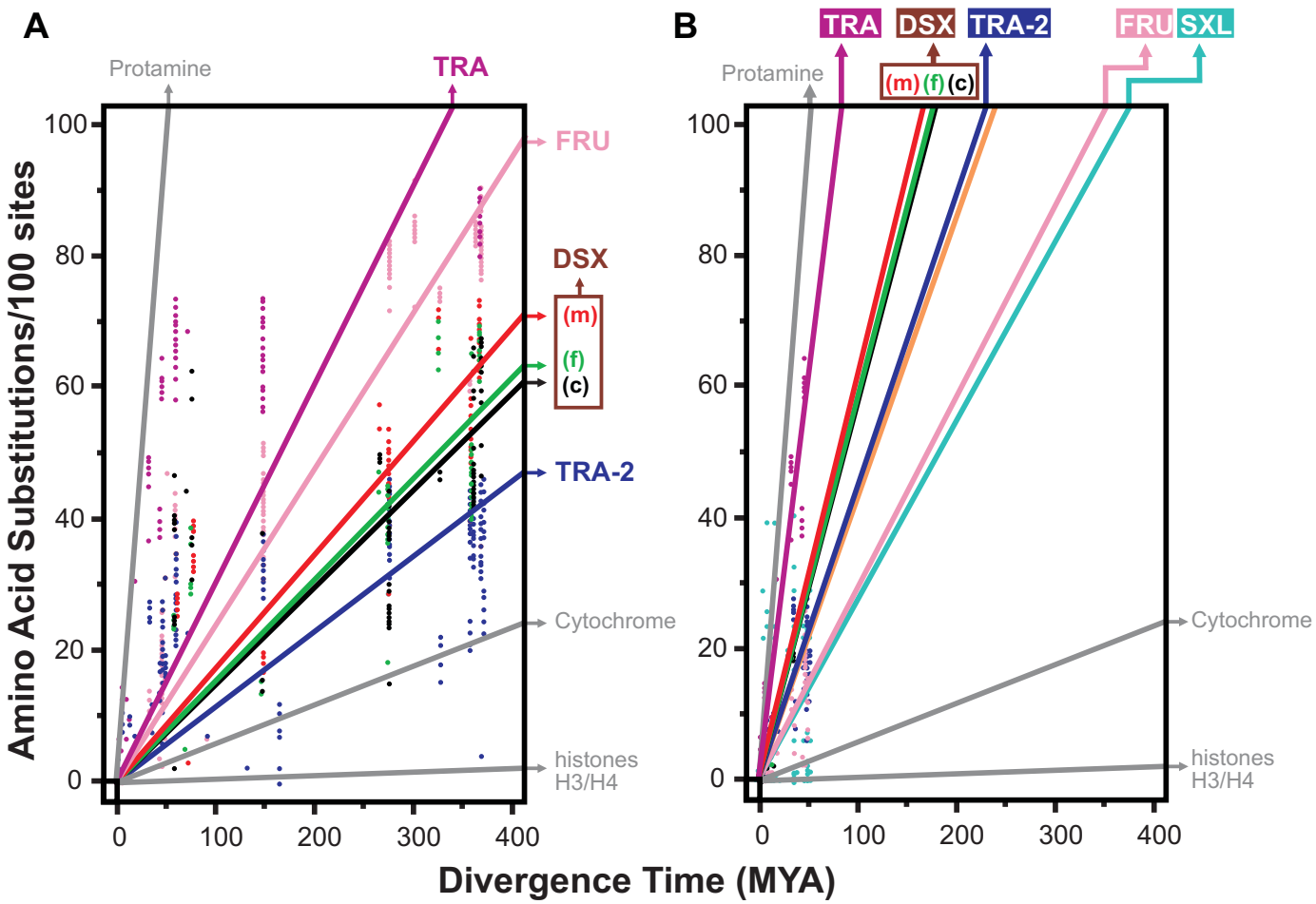

Fig. 2, Eirin-Lopez and Sanchez 2015

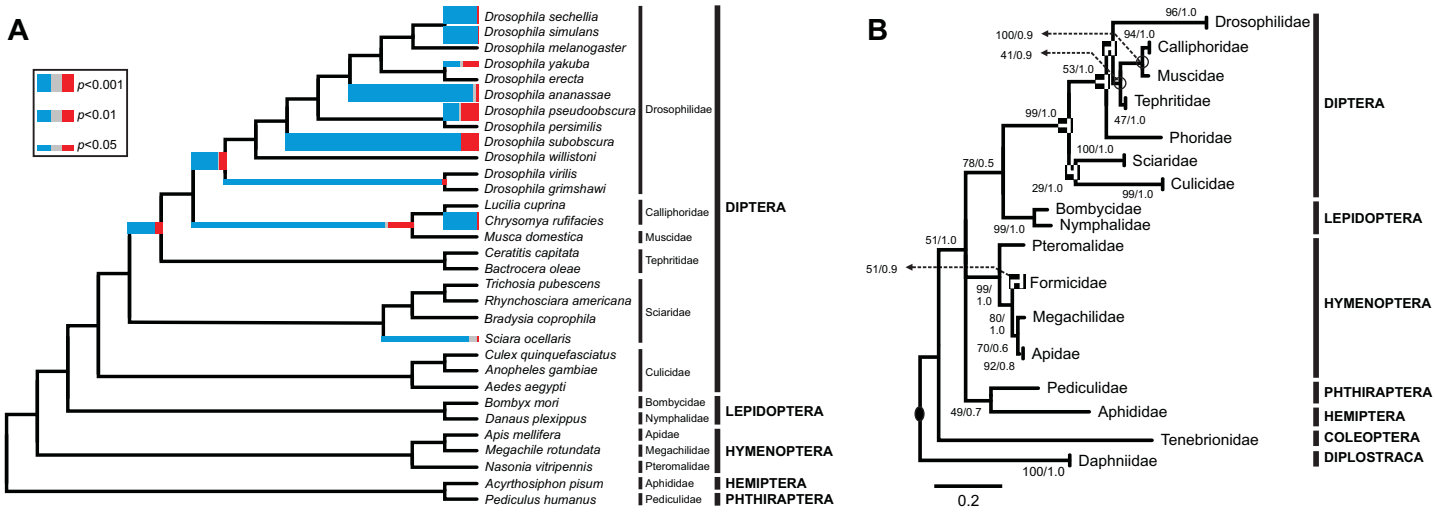

Fig. 3, Eirin-Lopez and Sanchez 2015

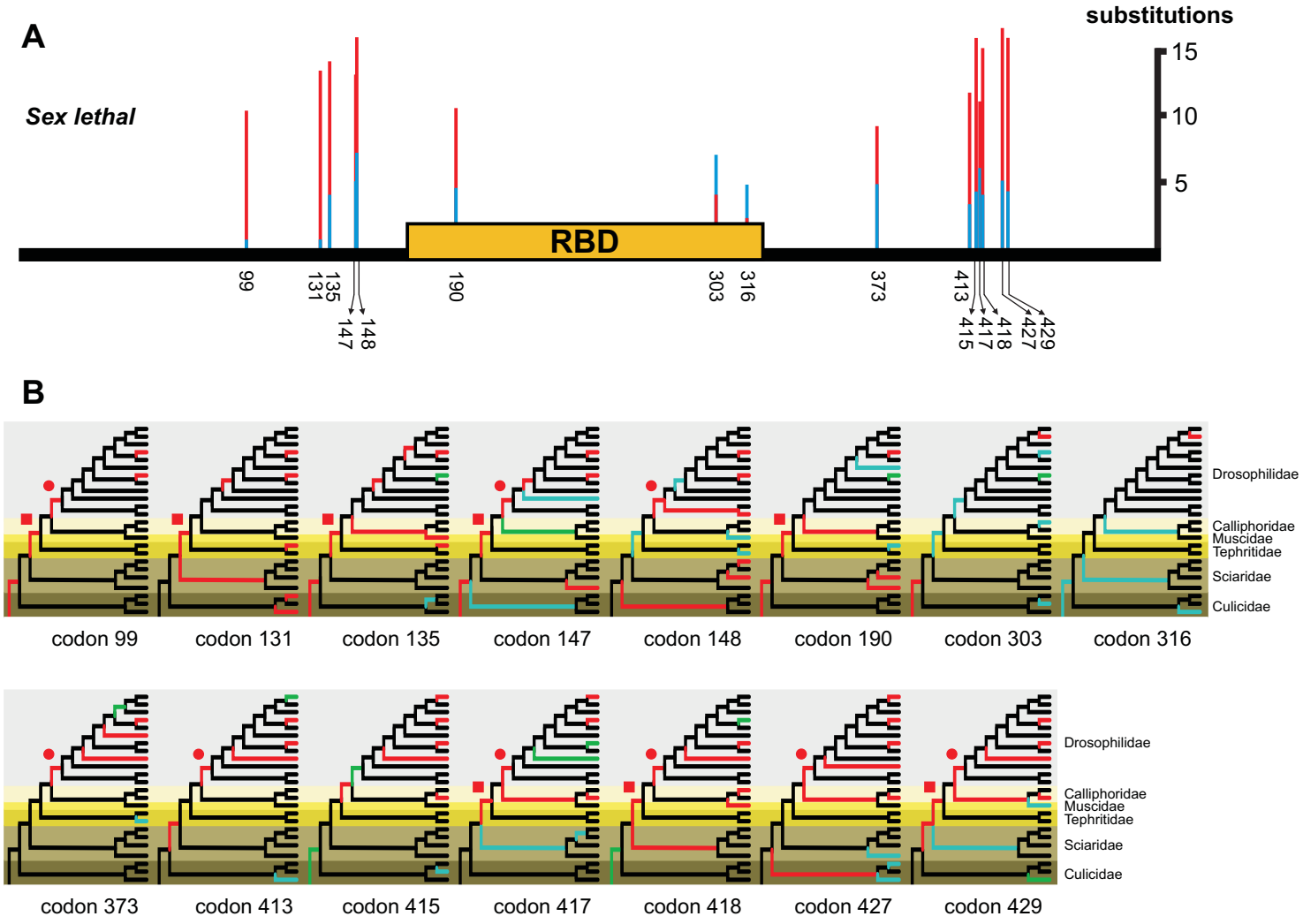

Fig. 4, Eirin-Lopez and Sanchez 2015

Supplementary Table 1.- Insect species and GenBank accession numbers for SXL, SSX, TRA, TRA-2, FRU and DSX sequences analyzed in the present work.

SXL	Taxa	Protein Acc#	Nucleotide Acc#
COLEOPTERA			
Tenebrionidae			
1.	*Tribolium castaneum*	NP_001139415	NM_001145943
DIPTERA			
Calliphoridae			
2.	*Chrysomya rufifacies*	O97018	S79722
3.	*Lucilia cuprina*	AF234184	AF234184
Culicidae			
4.	*Aedes aegypti*	XP_001661445	XM_001661395
5.	*Anopheles gambiae*	XP_560351	XM_560351
6.	*Anopheles darlingi*	EFR22477	
7.	*Culex quinquefasciatus*	XP_001863788	
Drosophilidae			
8.	*Drosophila ananassae*	XP_001963988	XM_001963952
9.	*Drosophila erecta*	XP_001978518	XM_001978482
10.	*Drosophila grimshawi*	XP_001992493	XM_001992457
11.	*Drosophila melanogaster*	NP_727164	NM_167114
12.	*Drosophila mojavensis*	XP_002010681	XM_002010645
13.	*Drosophila persimilis*	XP_002025140	XM_002025104
14.	*Drosophila pseudoobscura*	XP_002133619	XM_002133583
15.	*Drosophila sechellia*	XP_002043208	XM_002043172
16.	*Drosophila simulans*	XP_002106359	XM_002106323
17.	*Drosophila subobscura*	Q24668	X98370
18.	*Drosophila virilis*	XP_002056776	XM_002056740
19.	*Drosophila willistoni*	XP_002067476	XM_002067440
20.	*Drosophila yakuba*	XP_002099784	XM_002099748
Muscidae			
21.	*Musca domestica*	AAB81985	AF025689
Phoridae			
22.	*Megaselia scalaris*	O01671	
Sciaridae			
23.	*Bradysia coprophila*	AAS45603	AY538250
24.	*Rhynchosciara americana*	AAS45604	AY538251
25.	*Sciara ocellaris*	AAO19468	AY178581
26.	*Trichomegalosphys pubescens*	AAS45605	AY538252
Tephritidae			
27.	*Bactrocera oleae*	CAG29242	AJ715415
28.	*Ceratitis capitata*	O61374	AF026145
HEMIPTERA			
Aphididae			
29.	*Acyrthosiphon pisum*	NP_001119609	NM_001126137
HYMENOPTERA			
Apidae			
30.	*Apis mellifera*	XP_003250344	XM_003250296

31.	*Bombus terrestris*	XP_003403084	
32.	*Bombus impatiens*	XP_003489292	
Formicidae			
33.	*Acromyrmex echinatior*	EGI69813	
34.	*Camponotus floridanus*	EFN65860	
35.	*Harpegnathos saltator*	EFN79874	
Megachilidae			
36.	*Megachile rotundata*	XP_003705128	XM_003705080
Pteromalidae			
37.	*Nasonia vitripennis*	XP_003423885	XM_003423837
LEPIDOPTERA			
Bombycidae			
38.	*Bombyx mori*	ABA71352	DQ209269
Nymphalidae			
39.	*Danaus plexippus*	EHJ79210	AGBW01000003
PHTHIRAPTERA			
Pediculidae			
40.	*Pediculus humanus corporis*	XP_002432997	XM_002432952
DIPLOSTRACA			
Daphniidae			
41.	*Daphnia pulex*	EFX75394	GL732575

SSX	**Taxa**	**Flybase Acc#**
DIPTERA		
Drosophilidae		
1.	*Drosophila ananassae*	FBpp0124204
2.	*Drosophila erecta*	FBpp0131393
3.	*Drosophila grimshawi*	FBpp0146022
4.	*Drosophila melanogaster*	FBpp0308821
5.	*Drosophila mojavensis*	FBpp0165544
6.	*Drosophila persimilis*	FBpp0178352
7.	*Drosophila pseudoobscura*	FBpp0272897
8.	*Drosophila sechellia*	FBtr0202116
9.	*Drosophila simulans*	FBpp0214955
10.	*Drosophila virilis*	FBpp0230084
11.	*Drosophila willistoni*	FBpp0245839
12.	*Drosophila yakuba*	FBpp0261693

TRA	**Taxa**	**Protein Acc#**
DIPTERA		
Calliphoridae		
1.	*Lucilia cuprina*	ACS34689
Drosophilidae		

2.	*Drosophila ananassae*	XP_001957652
3.	*Drosophila erecta*	Q23935
4.	*Drosophila hydei*	Q23949
5.	*Drosophila mauritiana_*	AAO38914
6.	*Drosophila melanogaster*	NP_524114
7.	*Drosophila persimilis*	XP_002024880
8.	*Drosophila sechellia_*	AAO38908
9.	*Drosophila simulans_*	AAO38900
10.	*Drosophila virilis*	Q24761
11.	*Drosophila yakuba*	XP_002095112
Glossinidae		
12.	*Glossina morsitans*	ADD19862
Muscidae		
13.	*Musca domestica*	ACY40709
Tephritidae		
14.	*Anastrepha amita*	ABW04175
15.	*Anastrepha bistrigata*	ABW04174
16.	*Anastrepha* fraterculus	ABW04168
17.	*Anastrepha* grandis	ABW04170
18.	*Anastrepha ludens*	ABW04176
19.	*Anastrepha serpentina*	ABW04171
20.	*Anastrepha sororcula*	ABW04172
21.	*Anastrepha striata*	ABW04173
22.	*Anastrepha suspensa*	AET31461
23.	*Bactrocera oleae*	CAG29243
24.	*Ceratitis capitata*	AF434936
HYMENOPTERA		
Apidae		
25.	*Apis cerana*	ABV58876
26.	*Apis dorsata*	ABW36164
27.	*Apis mellifera*	NP_001011569
DIPLOSTRACA		
Daphniidae		
28.	*Daphnia pulex*	AGM48362

TRA-2	**Taxa**	**Protein Acc#**
COLEOPTERA		
Tenebrionidae		
1.	*Tribolium castaneum*	XP_968550
DIPTERA		
Calliphoridae		
2.	*Lucilia cuprina*	ACS34688
Drosophilidae		
3.	*Drosophila ananassae*	XP_001960772
4.	*Drosophila erecta*	XP_001975614
5.	*Drosophila grimshawi*	XP_001985987

6.	*Drosophila melanogaster*	NP_476764
7.	*Drosophila mojavensis*	XP_002006143
8.	*Drosophila persimilis*	XP_002016170
9.	*Drosophila pseudoobscura*	XP_001360605
10.	*Drosophila sechellia*	XP_002033866
11.	*Drosophila simulans*	XP_002081520
12.	*Drosophila virilis*	AAB58112
13.	*Drosophila willistoni*	XP_002063759
14.	*Drosophila yakuba*	XP_002091330
Glossinidae		
15.	*Glossina morsitans*	ADD19377
Muscidae		
16.	*Musca domestica*	AAW34233
17.	*Stomoxys calcitrans*	ADI86271
Sciaridae		
18.	*Bradysia coprophila*	CBX45938
19.	*Sciara ocellaris*	CBX45935
Tephritidae		
20.	*Anastrepha sororcula*	CBJ17287
21.	*Bactrocera oleae*	AAZ14854
22.	*Ceratitis capitata*	ACC68674
HYMENOPTERA		
Apidae		
23.	*Apis florea*	XP_003692251
24.	*Apis mellifera*	NP_001252514
25.	*Bombus terrestris*	XP_003399006
Formicidae		
26.	*Camponotus floridanus*	EFN67401
27.	*Harpegnathos saltator*	EFN80772
Megachilidae		
28.	*Megachile rotundata*	XP_003700631
Pteromalidae		
29.	*Nasonia vitripennis*	XP_001601106
LEPIDOPTERA		
Bombycidae		
30.	*Bombyx mori*	NP_001119707
DIPLOSTRACA		
Daphniidae		
31.	*Daphnia pulex*	EFX90042

FRU	**Taxa**	**Protein Acc#**
COLEOPTERA		
Tenebrionidae		
1.	*Tribolium castaneum*	XP_008200998
DIPTERA		
Culicidae		

2.	*Aedes aegypti*	XP_001657625
3.	*Culex quinquefasciatus*	XP_001860373
Drosophilidae		
4.	*Drosophila ananassae*	XP_001954108
5.	*Drosophila erecta*	XP_001979637
6.	*Drosophila grimshawi*	XP_001990228
7.	*Drosophila melanogaster*	NP_732349
8.	*Drosophila mojavensis*	XP_001998971
9.	*Drosophila persimilis*	XP_002013828
10.	*Drosophila pseudoobscura*	XP_003736513
11.	*Drosophila sechellia*	XP_002038222
12.	*Drosophila simulans*	XP_002102938
13.	*Drosophila virilis*	XP_002056235
14.	*Drosophila willistoni*	XP_002073544
15.	*Drosophila yakuba*	XP_002096204
Muscidae		
16.	*Musca domestica*	XP_005186915
Tephritidae		
17.	*Anastrepha fraterculus*	HQ003715
18.	*Anastrepha obliqua*	HQ003765
19.	*Ceratitis capitata*	XP_004536881
HYMENOPTERA		
Apidae		
20.	*Apis mellifera*	XP_006560820
21.	*Bombus impatiens*	XP_003486291
Megachilidae		
22.	*Megachile rotundata*	XP_003700636
Pteromalidae		
23.	*Nasonia vitripennis*	NP_001157598
LEPIDOPTERA		
Bombycidae		
24.	*Bombyx mori*	XP_004930656
DIPLOSTRACA		
Daphniidae		
25.	*Daphnia pulex*	EFX90042

DSX	**Taxa**	**Protein Acc#**	
		Female-specific	***Male-specific***
	DIPTERA		
Calliphoridae			
1.	*Lucilia cuprina*	ADG37649	ADG37648
Culicidae			
2.	*Aedes aegypti*	ABD96571	ABD96573
3.	*Anopheles gambiae*	AAX48939	AAX48940
Drosophilidae			
4.	*Drosophila erecta*	XP_001979242	XP_001979242

5.	*Drosophila melanogaster*	NP_731198	NP_731197
6.	*Drosophila persimilis*	XP_002013146	XP_002013146
7.	*Drosophila pseudoobscura*	XP_003736648	XP_001359020
8.	*Drosophila sechellia*	XP_002038750	XP_002038750
9.	*Drosophila simulans*	XP_002102542	XP_002102542
10.	*Drosophila yakuba*	XP_002086778	XP_002086778
Muscidae			
11.	*Musca domestica*	AAR23812	AAR23813
Phoridae			
12.	*Megaselia scalaris*	AF283695_1	AF283696_1
Sciaridae			
13.	*Bradysia coprophila*	HG934386	HG934387
14.	*Sciara ocellaris*	HG934388	HG934389
Tephritidae			
15.	*Anastrepha obliqua*	AAY25166	AAY25167
16.	*Bactrocera oleae*	CAD67986	CAD67987
17.	*Ceratitis capitata*	AAN63598	AAN63597
HYMENOPTERA			
Apidae			
18.	*Apis mellifera*	ABV55180	ABV55178
Pteromalidae			
19.	*Nasonia vitripennis*	ACJ65508	ACJ65511
LEPIDOPTERA			
Bombycidae			
20.	*Bombyx mori*	BAB13471	BAB13472
Saturniidae			
21.	*Antheraea assama*	ADL40848	ADL40846
DIPLOSTRACA			
Daphniidae			
22.	*Daphnia magna*	BAJ78307	BAJ78307