Spermatogenesis

Authors: Andrea Lu and Sherry Miller

Introduction

The Sterile Insect Technique (SIT) is one potential method of sustainable pest control. SIT technologies do not use chemical pesticides and are species specific, making them a more environmentally friendly method of insect pest control [1]. In this method, male insects are made sterile by radiation, chemosterilants or gene targeting methods and then released. If the sterile males are able to out-compete wildtype males the insect population declines. Historically SIT technologies have focused on producing sterile males through the use of chemosterilants or radiation. However, these methods, in addition to causing sterility, can compromise the insect's performance, reducing mating competitiveness [2,3]. RNA interference (RNAi) approaches have been suggested as an alternative sterilization approach because if gene targets are chosen carefully reproduction can be impacted without negatively impacting the insect's overall health, competitiveness or longevity [4,5].

Recently, Dong *et al* [6] tested the efficacy of RNAi on eight spermatogenesis related genes in the oriental fruit fly, *Bactrocera dorsalis* to identify potential SIT sterility targets. These genes included *longitudinals lacking* (*lola*), *matotopetli* (*topi*), *Rac*, *rhomboid* (*rho*), *unpaired* (*upd*), *magu*, *always early* (*aly*) and *period* (*per*). Knockdown of each of these genes resulted in a significant decrease in egg hatching rates and knockdown of six of these genes clearly influenced spermatogenesis (*lola*, *topi*, *Rac*, *rho*, *upd*, *magu*). Additionally, greenhouse trials using dsRNA *rho* treated males showed a significant decrease in the number of damaged oranges, illustrating that SIT could be an effective RNAi based method of pest control for agricultural insect pests which have a robust RNAi response. We screened the *Diaphorina citri* v2.0 genome for the presence of the six genes identified as being effective spermatogenesis targets and subsequently identified and manually annotated *lola*, *rac*, *Rho* and *magu* genes in *D. citri*.

Another potential target for SIT has been identified in the brown plant hopper, *Nilaparvata lugens*. Reduced expression of the *N. lugens* ortholog of the mammalian spermatogenesis gene *SPATA5* impaired the male reproductive system without affecting the insect's body weight or longevity [7]. Decreased *SPATA5* expression in males also appeared to effect female fecundity by causing a prolonged pre-oviposition period in females and decreasing the number of eggs laid [7] Its effect on both male and female fecundity makes SPATA5 an excellent target for insect pest control. We identified one *SPATA5 D. citri* ortholog in the genome assembly v2.0.

More recently, Ali *et al* [8], again using *B. dorsalis* as a model pest, tested five more gene targets which had been identified as having a role in spermatogenesis and could act as SIT targets in a variety of other insect pests [4,9]. These gene targets included *boule* (*bol*), *zero population growth* (*zpg*), *doublesex* (*dsx*), *fuzzy onion* (*fzo*) and *growth arrest specific protein 8* (*gas8*). In this study the authors showed all five target genes were affected by a feeding RNAi assay, three genes showed a significant enough reduction in egg hatch rates to be potential SIT targets (*boul*, *zpg*,

dsx) and feeding combinations of dsRNA resulted in a synergistic effect thereby further decreasing the egg hatch rate [8]. We screened the *D. citri* v2.0 genome for the presence of the five spermatogenesis genes tested as SIT targets and subsequently identified and manually annotated the *bol*, *dsx*, *gas8* and *Marf* (a paralog of *fzo*) genes in *D. citri*.

Finally, Cruz *et al* [10] tested the efficacy of *testis-specific serine/threonine protein kinase 1* (*tssk1*), *Thioredoxin T* (*TrxT*) and *topi* as SIT gene targets in the Queensland fruit fly *Bactrocera tryoni*. In *B. tryoni* these three genes are only expressed in adult testes and feeding knockdown assays resulted in decreased fecundity for both *tssk1* and *TrxT* targeted individuals [10]. We identified four tssk genes and four Thioredoxin domain containing genes in the *D. citri* genome assembly v2.0.

Results and Conclusions

longitudinals lacking

Lola is a zinc finger transcription factor that influences reproduction in insects by maintaining germline stem cells in the testis [11]. Lola has also been shown to play an essential role in guiding axons during the innervation of target cells [12]. Both of these processes require the action of multiple Lola isoforms and extensive studies in the dipterans, *Drosophila melanogaster* and *Anopheles gambiae* have established that dozens of *lola* isoforms exist in flies and mosquitoes [11,12]. Lola isoforms all share the same Broad-Complex, Tramtrack, Bric à brac (BTB) domain (also known as the POZ domain) in the N-terminal region of the protein, however, alternative splicing causes their C-terminal regions to vary. For most isoforms, the C-terminal region of each protein contains unique zinc finger domains coded for by alternative exons [11,12].

By BLASTing the MCOT transcriptome we were able to identify 7 MCOT transcripts (MCOT19752.0.CC, MCOT20646.0.CO, MCOT23125.0.CC, MCOT14022.2.CO, MCOT22199.1.CT, MCOT22199.2.CC, MCOT00580.1.CT) with sequence homology to *lola*
isoforms. Four of these transcripts were produced by a *de novo* transcriptome assembly confirming their expression. Unfortunately, many of these transcripts either did not map to the current genome or mapped to very small contigs which likely represent genome misassemblies. The MCOT transcripts identified in *D. citri* contain very similar 5' regions with a BTB domain (Figure 1) while their 3' regions contain unique sequence with at least one zinc finger domain.

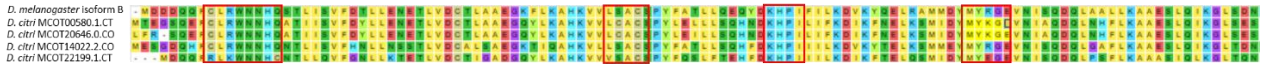

Figure 1: Multiple sequence alignment of the N terminus of de novo MCOT models showing the conserved BTB/POZ domain. Consensus sequences are shown with a red box.

This is similar to the transcript structure seen in *Drosophila* and *Anopheles lola* isoforms, suggesting that these MCOT transcripts in fact represent isoforms as opposed to duplicated genes, despite the fact that they are not mapping to the same loci in the *D. citri* v2.0 genome. However, in dipteran species it is clear that *lola* isoforms have conserved 5' sequence because they share

exons in this region, with alternative splicing only happening for 3' exons. The sequence of our MCOT sequences does not suggested shared 5' exons as the sequences, while conserved, are not identical (Figure 1). Due to the clear assembly and mapping errors in this version of the *D. citri* genome we have only annotated one *lola* isoform corresponding to the transcript represented by MCOT00580.1.CT. PacBio IsoSeq data was used to confirm this manual annotation (Table 1) and domain analysis indicates that the *D. citri* manually annotated Lola contains a BTB domain as expected (Figure 1). While only one isoform has been annotated in this genome, many more are likely to exist as is indicated by transcriptome data.

Gene	**D. citri identifier**	**Gene model**		**Evidence supporting annotation**			
		complete	partial	MCOT	IsoSeq	RNASeq	Ortholog
longitudinals lacking	Dcitr02g16110.1.1	x		x	x		x
magu	Dcitr05g07210.1.1	x		x	x	x	
Rac1	Dcitr07g02580.1.1	x		x	x	x	x
rhomboid2	Dcitr07g08670.1.1	x		x	x	x	
rhomboid4	Dcitr11g01360.1.1	x		x	x	x	
rhomboid5	Dcitr07g08930.1.1 Dcitr07g08940.1.1 Dcitr07g08910.1.1		X(3)	x			x
rhomboid7	Dcitr09g03810.1.1	x		x			x
SPATA5	Dcitr05g04490.1.1	x		x	x	x	x
boule	Dcitr01g09370.1.1 Dcitr01g09370.1.2	x		x	x		
doublesex	Dcitr03g16970.1.1 Dcitr03g16970.1.2	x		x	x		x
Mitochondrial assembly regulatoy factor	Dcitr01g21940.1.1	x		x	x		
growth arrest specific protein 8	Dcitr03g15460.1.1	x		x			
Testis specific serine/threonine protein kinase 1	Dcitr09g02460.1.1	x		x	x		x
Testis specific serine/threonine protein kinase 2	Dcitr05g15000.1.1	x		x	x	x	
Testis specific serine/threonine protein kinase 3	Dcitr10g09840.1.1	x		x			
Testis specific serine/threonine protein kinase 4	Dcitr02g10710.1.1	x		x	x		x
Thioredoxin T	Dcitr05g13300.1.1	x			x		x
Thioredoxin T like	Dcitr01g13370.1.1	x		x	x		
Thioredoxin T mitochondrial 1	Dcitr10g09890.1.1	x					x
Thioredoxin T mitochondrial 2	Dcitr04g03720.1.1	x				x	x

Table 1: Annotated D. citri Genes. Each manually annotated gene has been assigned a gene identifier. For each manually annotated gene it has been denoted as a partial or complete model based on available evidence. Evidence for manual annotation was also recorded. MCOT evidence means a de novo Oases or Trinity model from an independent transcriptome was identified and the sequence from that transcript was used to validate or modify our model. IsoSeq means single reads generated with Pacific Biosciences technology were available and were used to help validate the exon structure of the model. RNASeq means that individually mapped Illumina RNASeq reads and/or StringTie models were used to help validate or modify our model. Ortholog

means ortholog sequences from other insects and information about conserved motifs and domains had to be used to help determine the final annotation.

magu

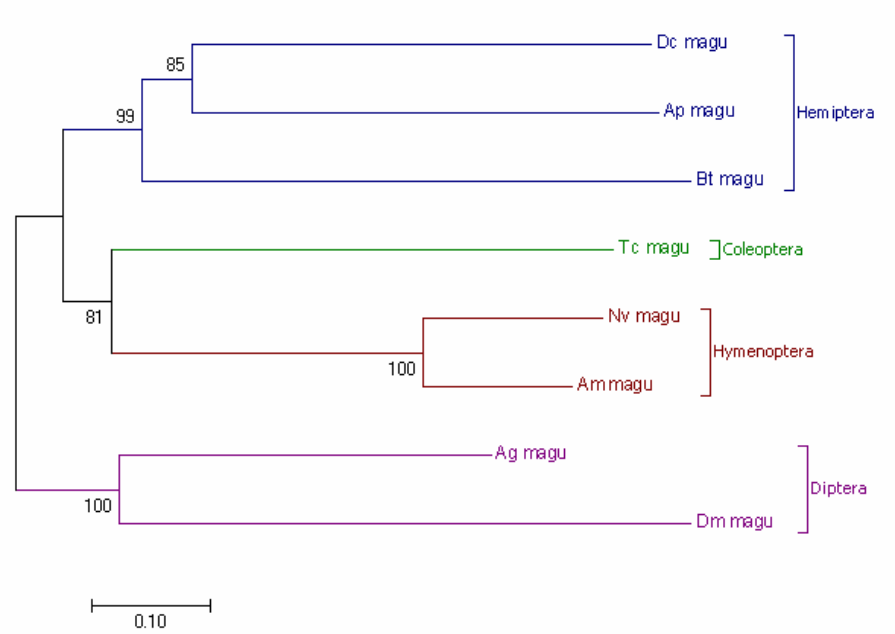

Figure 2: Magu neighbor-joining phylogenetic tree. ClustalW multiple sequence alignments of full length protein sequences were used for phylogenetic analysis. Each insect order is represented with a different color. Species: Dc, Diaphorina citri. Ap, Acyrthosiphon pisum. Bt, Bemisia tabaci. Tc, Tribolium castaneum. Nv, Nasonia vitripennis. Am, Apis mellifera. Ag, Anopheles gambiae. Dm, Drosophila melanogaster.

Magu is a secreted protein involved in the regulation of BMP signaling and as such plays a role in wing development, adult life span and germ-line stem cell maintenance [13]. Studies in *Drosophila* suggest that BMP is the primary pathway involved in germ-line stem cell self-renewal and Magu is essential for the activation of BMP in adjacent germ cells [14]. Overexpression of Magu has been shown to increase fecundity [15] while loss of Magu results in sterility phenotypes [14]. A BLASTp using *Drosophila* Magu protein sequence to the *D. citri* MCOT predicted protein set identified one *D. citri* MCOT predicted protein which reciprocal blasted back to Magu. This MCOT model, MCOT12593.0.CT, was found mapped to the *D. citri* genome and was used as the beginning model for annotation. Multiple IsoSeq and Stringtie datasets confirmed the structure of the gene and alignments of the genome model protein sequence to the MCOT protein sequence suggested changes to the model were not necessary (Table 1). Some IsoSeq models suggested that there may be *magu* isoforms with longer 3'UTRs than the current model but because these slight variations did not affect the protein coding region of the gene multiple isoforms were not annotated. In *Drosophila* four *magu* isoforms have been identified but most of these differences lie in the UTR structure. Only two unique *Drosophila* Magu proteins have been identified. One *Drosophila magu* isoform is lacking an exon present in the other unique *magu* isoform. If *D. citri* were to have two isoforms similar in structure to *Drosophila's* two *magu* isoforms then you would expect to see an absence of some genetic information around exon 7 and 8 in the current *D. citri*

magu model. We found no RNASeq evidence to support a second isoform, therefore, only one isoform has been annotated. For our manually annotated *D. citri magu* model, InterPro analysis confirmed the expected conserved domains (a Kazal domain, a Thyroglobulin type 1 domain and 2 EF-hand calcium-binding domains) and phylogenetic analysis showed expected clustering (by order) of Magu orthologs (Figure 2). While all available data suggests the *D. citri magu* gene model is accurate there does appear to be a genome assembly error within the *magu* gene region. There is a tandem duplication of exons 5 and 6 at positions 735,000 and 737,000. Knowing this error exists it is possible that there may be other errors at this gene locus as well.

Rac1

Rac1 belongs to the Rho family of small GTPases. These proteins are best characterized by their role in regulating actin organization, however, they are also known to play essential roles in many different cellular and developmental processes [13]. Rac1 has specifically been implicated in axon outgrowth, myoblast fusion, dorsal closure, spermatogenesis and phagocytosis of *Staphylococcus aureus* by hemocytes [13]. BLASTs to the *D. citri* MCOT protein database identified one MCOT protein which reciprocal blasted back to insect Rac1 proteins (MCOT21204.0.CO). RNASeq Stringtie and PacBio IsoSeq models confirm the annotation based on MCOT21204.0.CO (Table 1). Domain analysis using EMBL-EBI's Interpro identified the manually annotated *D. citri* Rac1 model as a Rho type small GTPase superfamily protein and detected two conserved domains known to be present in Rac1; a small GTP-binding protein domain and a P-loop containing nucleoside triphosphate hydrolase domain. Phylogenetic analysis showed expected relationships between dipteran, coleopteran, hymenopteran and hemipteran Rac1 proteins (data not shown) and multiple sequence alignments between insect Rac1 orthologs showed extreme conservation of sequence at the amino acid level (Figure 3).

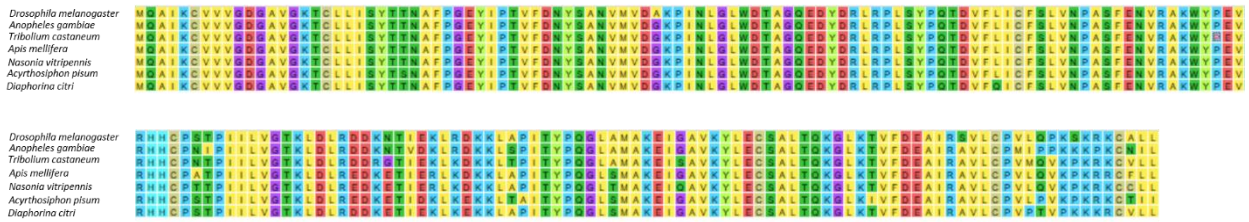

Figure 3: Multiple sequence alignment of Rac1 orthologs. ClustalW alignment shows a high degree of amino acid conservation of Rac1 orthologs in insects suggesting our annotation represents a complete and accurate sequence.

While all current evidence supports an accurate and complete gene model for *D. citri Rac1*, further analysis at this locus suggested some assembly errors in this region. Portions of the *Rac1* genes are duplicated at multiple sites on this contig indicating clear assembly errors. Therefore, caution should be taken when annotating other genes in this genomic region.

Rhomboid

EGF signaling has been heavily implicated in the formation and maintenance of germline stem cell niches [16,17]. In *Drosophila* Rhomboid (Rho1) acts specifically in trafficking EGF ligands and its knockdown in *B. dorsalis* has been shown to reduce male fecundity in the lab and reduce the

efficiency of agricultural pests in the greenhouse [6]. The Rho family is an ancient family that is present in all kingdoms and in eukaryotes there are typically many family members. In *Drosophila*, mice and humans there are at least 7 *rho* genes [18]. These proteins are all transmembrane proteins which, if they contain a catalytic domain, function in proteolysis. While some of their roles within a cell have yet to be elucidated, in general their function is to cleave growth factors from membranes which then initiate various types of signaling pathways. Unfortunately, the expansion of Rho family members in eukaryotes has lessened evolutionary constraints which means that the sequence identity in this family is exceptionally low making homologs difficult to identify [18]. The *Drosophila rho* genes that have been identified include *rho1*, *rho2* (also called *stem cell tumor*), *rho3* (also called *roughoid*) *rho4*, *rho5* (also known as *iRhom*), *rho6* and *rho7*. In the *D. citri* genome v2.0 we were able to identify and annotate 4 *rho* genes. Based on reciprocal blast and phylogenetic analysis these genes appear to be orthologous to *rho2*, *rho4*, *rho5* and *rho7* (Figure 4). Interestingly an ortholog of Rho1, the serine protease first identified in EGF signaling in *Drosophila* was not identified, however an ortholog of Rho2 which has been shown to be essential for germ cell differentiation [16], was identified. This absence of *rho1* appears to be true in pea aphids as well [19]. While it was originally reported that one *rho2*, two *rho4*s and one *rho5* genes were present in the pea aphid genome [19] our analysis of the current pea aphid assembly suggests that, like *D. citri*, pea aphid also contains one *rho2*, one *rho4*, one *rho5* and one *rho7* ortholog (Figure 4).

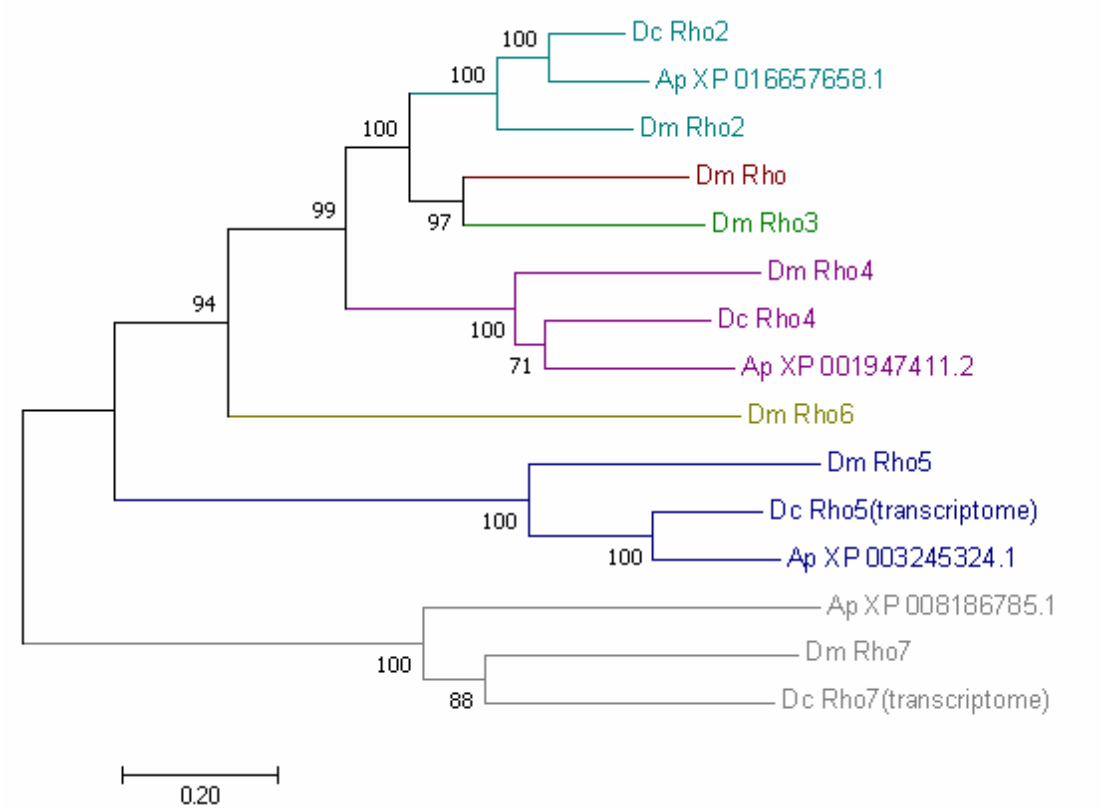

Figure 4: Neighbor-joining tree of rho family members. Full length protein sequences were used for phylogenetic analysis. Due to assembly errors resulting in only partial gene models for Diaphorina citri rho5 and rho7, full length rho5 and rho7 sequences from the MCOT transcriptomes were used. Colors denote clades of different rho proteins. Species: Dc, Diaphorina citri. Ap, Acyrthosiphon pisum. Dm, Drosophila melanogaster.

The *D. citri rho2*, *rho4*, and *rho7* annotations represent full length genes which are supported by *de novo* transcriptome models (Table 1). In contrast, the *rho5* region has many assembly errors. Due to these errors, *rho5* had to be annotated in three pieces. More complete amino acid sequence information can be obtained from MCOT22623.1.CO and MCOT01968.0.CO which likely represent two isoforms of *rho5*.

matotopetli and unpaired

The final two genes identified as potential SIT targets in *B. dorsalis*, *topi* and *upd* [6], appear to be absent in the *D. citri* genome. Upd is the ligand of the JAK-STAT signaling pathway and its absence is unsurprising as it has been previously reported that *upd* is a rapidly evolving gene present only in flies [19]. *topi* was identified in *Drosophila* as a meiotic arrest gene involved in the process of spermatid differentiation [20]. As far as we know its presence has not been investigated outside of flies and therefore its utility in this role across a broader taxa of insects is unknown.

SPATA5

SPATA5 is one of the spermatogenesis-associated proteins first identified in mammals as having a role in testicular biology. While first identified in mammals, SPATA orthologs have been identified in insects [21]. Previous work in *N. lugens* indicates that SPATA5 is upregulated after insecticide treatments and this upregulation is correlated with increased insect fecundity [22]. Additionally, decreased SPATA5 expression in males is associated with decreased fecundity in both males and females [7]. In the *D. citri* genome v2.0 we identified one *SPATA5* ortholog. Our annotation represents a complete gene with strong *de novo* transcriptome (MCOT01561.0.CT), RNASeq, StringTie, PacBio IsoSeq and ortholog support (Table 1).

boule

bol is involved in control of the gap 2 to meiosis transition during the cell cycle and as such has been shown to influence spermatocyte maturation and spermatid differentiation [23]. Because of its essential role in spermatogenesis *bol* has been identified as a potential sterility target and work in several different insect species has already shown that *bol* is an effective SIT target [4,8]. Using the *Drosophila* Boule protein sequence as the query to BLAST the *D. citri* MCOT database identified two *de novo D. citri* MCOT models (MCOT05314.0.CO & MCOT01681.1.CT). These two models both reciprocal blasted back to Bol and likely represent two isoforms which are identical in the N terminus but display differences in the seventh exon near the C terminus of the protein. Errors in genome assembly caused the *bol* locus to be duplicated on two contigs with one MCOT model mapping to each contig, however, complete models were manually annotated on one of the two contigs. The exon structure of the two manually annotated isoforms was confirmed with PacBio IsoSeq models (Table 1).

doublesex

The *dsx* loci produces sex specific transcripts which play an essential role in the sex determination hierarchy in insects [24–30]. Dsx contributes to both morphological and behavioral sexual dimorphism and with respect to spermatogenesis the male version of Dsx has been shown to be required for the proper development of male germline stem cells in insects [31]. BLASTp to the *D. citri* MCOT transcriptome database revealed only one protein model (MCOT04633.0.CT) which reciprocal blasted back to Dsx in *Drosophila*. However, the Maker pipeline produced three different potential isoforms for the *dsx* gene in *D. citri*. One of these Maker produced isoforms (*dsx-RA*) corresponded to our *de novo* MCOT model and PacBio IsoSeq models confirmed the expression of this isoform. Based on length this isoform appears to most closely resemble the female version of *dsx*, although there is little conservation of the female specific sequence meaning functional analysis would be required to determine this transcript's role in sex determination. We tentatively identified one additional isoform (*dsx-RB*) based on RNAseq data. One other potential isoform (*dsx*-RC) (Figure 5) was identified in genome v2.0, but was not supported by evidence mapping to genome v3.0 and was removed.

N. lugens is the only hemipteran insect for which *dsx* homologs have been functionally characterized [30]. Interestingly, in addition to *dsx N. lugens* has three other *dsx-like* genes, *dsx-like 1*, *dsx-like 2* and *dsx-like 3* [30]. Unlike *N. lugens* in the *D. citri* genome we were only able to identify one *dsx* homolog. Interestingly, the *D. citri* and *N. lugens* Dsx proteins do not cluster together in a hemipteran clade (Figure 5). Instead the *D.citri* isoforms form a sister group to the holometabolous Dsx proteins while the *N. lugens* proteins cluster with the *Homo sapiens* Dsx protein outgroup (Figure 5).

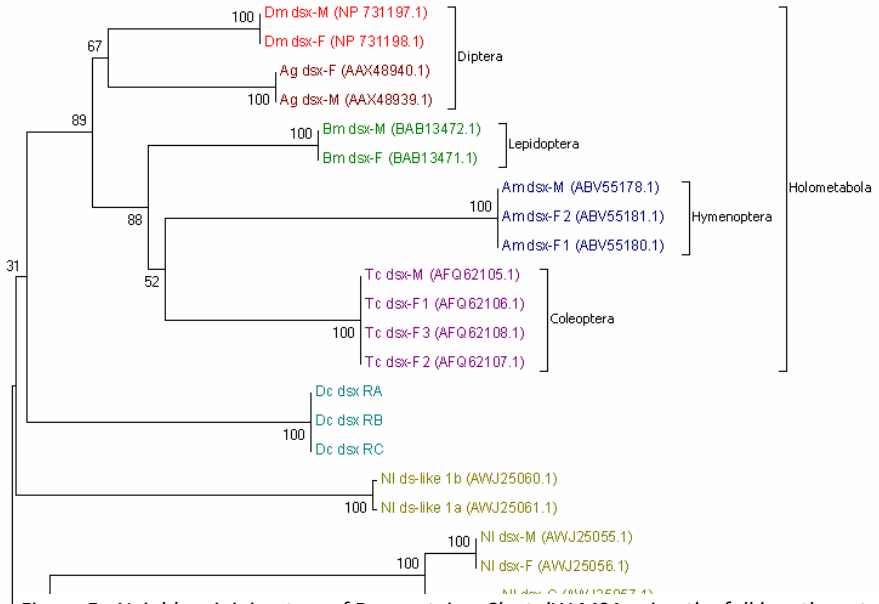

Figure 5: Neighbor-joining tree of Dsx proteins. ClustalW MSA using the full length protein was used for phylogenetic analysis. Male and female specific isoforms are included in the analysis. Species are denoted with different colors. Dm, Drosophila melanogaster. Ag, Anopheles gambiae. Bm, Bombyx mori. Am, Apis mellifera. Tc, Tribolium castaneum. Dc, Diaphorina citri. Nl, Nilaparvata lugens. Hs, Homo sapiens.

growth arrest specific protein 8

gas8 is a gene has been shown to be expressed predominantly in the testes both in mice and insects [8,32]. Work in mice has shown that this expression in the testes increases postmeiotically and the Gas8 protein localizes in the spermatids suggesting a role in spermatid motility [32]. Our analysis of the *D. citri* genome identified one *gas8* ortholog. We used *de novo* transcripts from the MCOT database (MCOT09315.0.TT & MCOT07717.0.OO) to determine the likely exon structure and protein sequence for this manually annotated model. While a complete gene model has been produced the contig *gas8* is located on has assembly errors as *gas8* and its neighboring genes are duplicated within the contig.

fuzzy onion

fzo is a gene which has been found to play a role in spermatogenesis in both mosquitoes and flies [4,8]. *fzo* is the paralog of a gene called *Mitochondrial assembly regulatory factor* (*Marf*) and both *Marf* and *fzo* are homologs of the vertebrate genes *mitofusin 1* (*mfn1*) and *mitofusion 2* (*mfn2*). In flies it has been shown that *fzo* is a testis specific *mfn* homolog [33,34] and as such it has been the focus of sterile insect technologies. BLAST results suggest that a *fzo* ortholog is not present in more basal insects, however, the *Marf* ortholog is present. Expression and functional analysis of *Marf* in insects outside of the dipteran order has not be reported, so whether the *Marf* ortholog could also be involved in spermatogenesis in insects is not known. However, the vertebrate homologs, *mfn1* and *mfn2* are known to be expressed in testes [35] and thus we decided to annotate the *D. citri Marf* gene as it could act a SIT target. In *D. citri* there is one *Marf* ortholog. BLASTs to the MCOT transcriptome identified two *de novo* MCOT models (MCOT14304.3.CO and MCOT14304.1.CT), one being a truncated version of the other. The *Marf* gene was annotated based on the longest MCOT model, MCOT14304.3.CO, and PacBio IsoSeq models support this annotation. While our annotated *Marf* gene appears complete, it is on a small contig which likely represents a duplicated portion of contig 1258. Unfortunately, errors at the *Marf* locus on contig 1258 prevented its annotation on this larger contig. Hopefully, newer versions of the *D. citri* genome will collapse this duplication so that the *Marf* gene can be examined in the context of its genomic position.

zero population growth

zpg is a germline specific gap junction protein responsible for the survival of differentiating early germ cells during gametogenesis in both sexes. *zpg* is a member of the gap junction Innexin gene family and is also known as *Innexin 4* [36]. There are eight identified Innexin proteins in *Drosophila* and phylogenetic analysis of insect Innexin proteins suggests there are six Innexin clades which were likely derived by a duplication event that happened prior to the most recent common ancestor of exopterygotes and endopterygotes [36]. *zpg* is in the same clade as *Inx5* and *Inx6* (the zpg clade) and is the result of two independent duplications in the lineage leading to *Drosophila* [36]. Members of the zpg clade have been shown to be expressed in fly, mosquito and silkworm gonads [36] and thus are the Innexins that could act as SIT targets. BLAST analysis of the *D. citri*

genome v2.0 and the *D. citri* MCOT transcriptome suggest that there is not a *zpg*, or other zpg clade, ortholog in *D. citri*. All significant BLAST hits are ortholgous to other Innexin genes found in other Innexin clades such as *Innexin 2*, *shaking B* and *ogre*. Therefore, a *zpg* gene has not been annotated in *D. ctiri*. Our finding is consistent with the lack of an *zpg*/*Inx5*/*Inx6* ortholog in *A. pisum* [36]. However, the presence of one zpg clade Innexin gene in *P. humanus* suggests a gene loss in the lineage leading to sap/phloem sucking insects.

testis specific serine/threonine protein kinase 1

tssk genes were first identified and cloned in mice [37,38]. Vertebrates appear to have five or six *tssk* genes, however, these genes have not yet been carefully annotated in insect species.

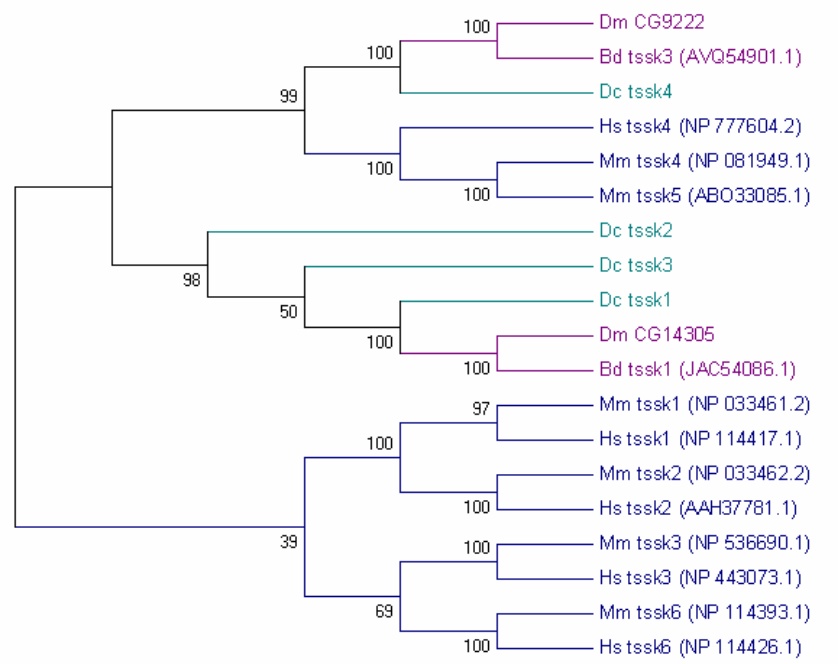

Figure 6: Neighbor-joining tree of tssk family genes. ClustalW MSA alignment was used with full length protein sequences. Blue represents vertebrate genes. Purple represents invertebrates in the dipteran order. Teal represents D. citri manually annotated sequences. Species: Dm, Drosophila melanogaster. Bd, Bactrocera dorsalis. Dc, Diaphorina citri. Hs, Homo sapiens. Mm, Mus musculs.

Drosophila has two genes that share homology with vertebrate *tssk* genes, CG14305 and CG9222 [13]. While little research has been performed in the classic fly model on the *tssk* gene family, in the Oriental and Queensland fruit flies *tssk* orthologs have been annotated and functionally characterized due to their potential role as targets for SIT [10,39]. In the *D. citri* genome we identified four *tssk* genes which have been named *tssk1*, *tssk2*, *tssk3* and *tssk4*. One *tssk* gene, *Dctssk4*, shows clear orthology with the vertebrate *tssk4* gene (Figure 6). The other three tsk genes do not cluster with vertebrate *tssk* genes and instead form a clade with other *tssk* insect homologs. These genes have been named *tssk 1*, *2* and *3*. Note, these names do not suggest one-to-one orthology with the vertebrate *tssk* genes showing the same name. All four *tssk* genes in *D. citri* show evidence of expression as all have *de novo* transcriptome evidence (MCOT22746.2.CO,

MCOT05393.0.CT, MCOT06450.1.CO & MCOT14147.0.CO) and three have PacBio IsoSeq evidence. While errors in genome assembly were identified at each of the four *tssk* loci full length gene models were completed for each gene.

Thioredoxin T

Thioredoxins are ubiquitous proteins involved in cellular redox reactions. As antioxidant enzymes these proteins help the cell deal with oxidative stressors and both cytosolic and mitochondrial Thioredoxins have been identified. *TrxT* is a testis-specific Thioredoxin gene involved in the regulation of intracellular homeostasis in the germline in *Drosophila* [40]. *TrxT* and *deadhead* (*dhd*), another Thioredoxin gene, exist as a gene pair which share a regulatory region in the *Drosophila* genome [40]. While *dhd* and *TrxT* share a regulatory region they are expressed differentially with *TxrT* expressed in the testes and *dhd* being expressed in the ovaries and eggs [40]. Dhd has also been shown to play an essential role in unlocking the sperm chromatin during fertilization by reducing the disulfide crosslinks which keep sperm chromatin in a compacted state [41]. In vertebrates a number of testis-specific and ovary-specific Thioredoxins have been identified as well [42–45], suggesting Thioredoxins play an important role in gametogenesis across animal taxa. While our intention was to annotate *TrxT*, as it is a potential SIT target, the many genome assembly errors which caused full and partial duplications of Thioredoxin family members made the determination of true orthology and the annotation of *TrxT* alone difficult. Given these difficulties and the knowledge that vertebrate Thioredoxin proteins are also involved in gametogenesis as well we embarked on a more complete analysis of Thioredoxin domain containing proteins and annotated four Thioredoxin family members in *D. citri*. *D. citri* has one Thioredoxin protein which clusters with both *Drosophila* TrxT and Dhd. We have named this gene *Dc Thioredoxin T* (*TrxT*), however without expression and functional analysis, what role it plays in spermatogenesis, sperm nuclear decondensation or oogenesis is unclear. The presence of only one gene suggests that *TrxT* in *D. citri* may not have a male-specific role in gametogenesis. In addition to *TrxT* we have also annotated *Dc Thioredxin-like* (*Txl*) which clusters with the *Drosophila* Thioredoxin-like (Txl) protein and the vertebrate Thioredoxin-like protein 1. Additionally, *D. citri* appears to have two mitochondrial Thioredoxin genes, *Thioredoxin mitochondrial 1* (*Trx-mt1*) and *Thioredoxin mitochondrial 2* (*Trx-mt2*). While the phylogenetic tree suggests that the *Drosophila* and *D. citri* mitochondrial Thioredoxin genes are independent duplications, sequence alignments and reciprocal blast analysis suggests that *Dc Trx-mt1* is orthologous to *Drosophila CG8993* and *Dc Trx-mt2* is orthologous to *Drosophila CG8517*. While only one Thioredoxin gene model, *txl*, has *de novo* transcriptome support (MCOT19797.0.CT), all four models have RNASeq expression data to support their annotation (Table 1).

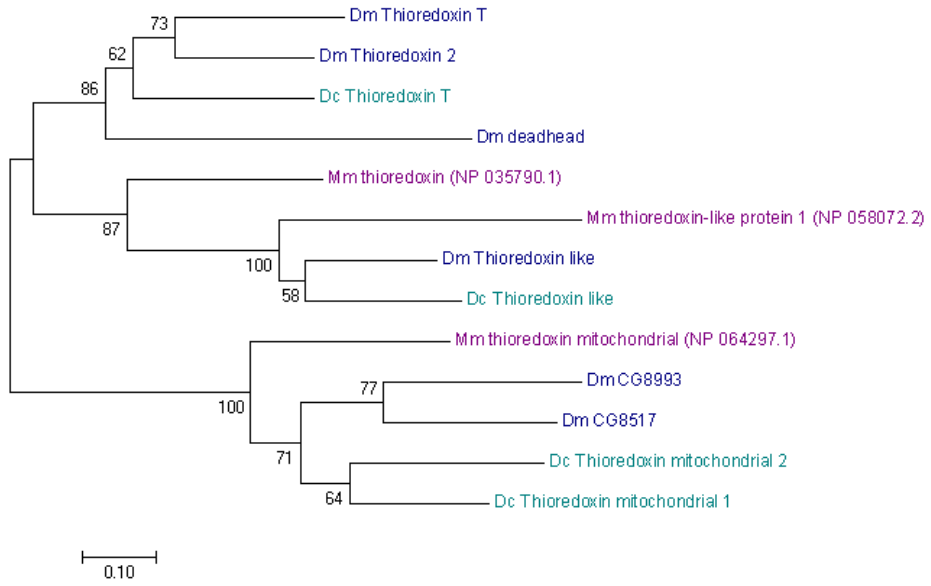

Figure 7: Neighbor-joining tree of thioredoxin family genes. ClustalW MSA alignment was used with full length protein sequences. Species are denoted by color. Blue represents Drosophila melanogaster (Dm) genes. Teal represents Diaphorina citri (Dc) manually annotated sequence. Purple represents Mus musculus (Mm) genes.

References

1. Lees RS, Gilles JR, Hendrichs J, Vreysen MJ, Bourtzis K. Back to the future: the sterile insect technique against mosquito disease vectors. Curr. Opin. Insect Sci. Elsevier Inc.; 2015. p. 156–62.

2. Helinski MEH, Knols BGJ. Mating Competitiveness of Male <I>Anopheles arabiensis</I> Mosquitoes Irradiated with a Partially or Fully Sterilizing Dose in Small and Large Laboratory Cages. J Med Entomol. Oxford University Press (OUP); 2008;45:698–705.

3. Hooper GHS. Sterilization of the Mediterranean Fruit Fly with Gamma Radiation: Effect on Male Competitiveness and Change in Fertility of Females Alternately Mated with Irradiated and Untreated Males1. J Econ Entomol [Internet]. 1972 [cited 2019 Dec 5];65:1–6. Available from: http://academic.oup.com/jee/article/65/1/1/2210435/Sterilization-of-the-Mediterranean-Fruit-Fly-with

4. Whyard S, Erdelyan CNG, Partridge AL, Singh AD, Beebe NW, Capina R. Silencing the buzz: a new approach to population suppression of mosquitoes by feeding larvae double-stranded RNAs. Parasit Vectors [Internet]. 2015 [cited 2019 Dec 5];8:96. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25880645

5. Darrington M, Dalmay T, Morrison NI, Chapman T. Implementing the sterile insect technique with RNA interference - a review. Entomol Exp Appl [Internet]. 2017 [cited 2019 Dec 5];164:155–75. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29200471

6. Dong Y-C, Wang Z-J, Chen Z-Z, Clarke AR, Niu C-Y. Bactrocera dorsalis male sterilization by targeted RNA interference of spermatogenesis: empowering sterile insect technique programs. Sci Rep [Internet]. 2016 [cited 2019 Dec 4];6:35750. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27767174

7. Ge LQ, Xia T, Huang B, Song QS, Zhang HW, Stanley D, et al. Suppressing male spermatogenesis-associated protein 5-like gene expression reduces vitellogenin gene expression and fecundity in Nilaparvata lugens Stål. Sci Rep. Nature Publishing Group; 2016;6.

8. Ali MW, Zheng W, Sohail S, Li Q, Zheng W, Zhang H. A genetically enhanced sterile insect technique against the fruit fly, Bactrocera dorsalis (Hendel) by feeding adult double-stranded RNAs. Sci Rep. Nature Publishing Group; 2017;7.

9. Ant T, Koukidou M, Rempoulakis P, Gong H-F, Economopoulos A, Vontas J, et al. Control of the olive fruit fly using genetics-enhanced sterile insect technique. BMC Biol [Internet]. 2012 [cited 2019 Dec 5];10:51. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22713628

10. Cruz C, Tayler A, Whyard S. RNA Interference-Mediated Knockdown of Male Fertility Genes in the Queensland Fruit Fly Bactrocera tryoni (Diptera: Tephritidae). Insects [Internet]. 2018 [cited 2019 Dec 5];9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30103378

11. Davies EL, Lim JGY, Joo WJ, Tam CH, Fuller MT. The transcriptional regulator lola is required for stem cell maintenance and germ cell differentiation in the Drosophila testis. Dev Biol [Internet]. 2013 [cited 2019 Dec 4];373:310–21. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23159836

12. Goeke S, Greene EA, Grant PK, Gates MA, Crowner D, Aigaki T, et al. Alternative splicing of lola generates 19 transcription factors controlling axon guidance in Drosophila. Nat Neurosci [Internet]. 2003 [cited 2019 Dec 5];6:917–24. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12897787

13. Flybase [Internet]. [cited 2019 Dec 2]. Available from: http://flybase.org

14. Zheng Q, Wang Y, Vargas E, DiNardo S. magu is required for germline stem cell self-renewal through BMP signaling in the Drosophila testis. Dev Biol [Internet]. 2011 [cited 2019 Dec 5];357:202–10. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21723859

15. Li Y, Tower J. Adult-specific over-expression of the Drosophila genes magu and hebe increases life span and modulates late-age female fecundity. Mol Genet Genomics [Internet]. 2009 [cited 2019 Dec 5];281:147–62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19011900

16. Schulz C, Wood CG, Jones DL, Tazuke SI, Fuller MT. Signaling from germ cells mediated by the

rhomboid homolog stet organizes encapsulation by somatic support cells. Development. 2002;129:4523–34.

17. Hudson AG, Parrott BB, Qian Y, Schulz C. A temporal signature of epidermal growth factor signaling regulates the differentiation of germline cells in testes of Drosophila melanogaster. PLoS One [Internet]. 2013 [cited 2019 Dec 5];8:e70678. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23940622

18. Urban S, Dickey SW. The rhomboid protease family: a decade of progress on function and mechanism. Genome Biol [Internet]. 2011 [cited 2019 Dec 5];12:231. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22035660

19. Shigenobu S, Bickel RD, Brisson JA, Butts T, Chang C, Christiaens O, et al. Comprehensive survey of developmental genes in the pea aphid, Acyrthosiphon pisum: frequent lineage-specific duplications and losses of developmental genes. Insect Mol Biol [Internet]. 2010 [cited 2017 Oct 10];19:47–62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20482639

20. Perezgasga L, Jiang J, Bolival B, Hiller M, Benson E, Fuller MT, et al. Regulation of transcription of meiotic cell cycle and terminal differentiation genes by the testis-specific Zn-finger protein matotopetli. Development [Internet]. 2004 [cited 2019 Dec 5];131:1691–702. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15084455

21. Bonilla E, Xu EY. Identification and characterization of novel mammalian spermatogenic genes conserved from fly to human. Mol Hum Reprod. 2008;14:137–42.

22. Ge L-Q, Cheng Y, Wu J-C, Jahn GC. Proteomic analysis of insecticide triazophos-induced mating-responsive proteins of Nilaparvata lugens Stål (Hemiptera: Delphacidae). J Proteome Res [Internet]. 2011 [cited 2019 Dec 5];10:4597–612. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21800909

23. Maines JZ, Wasserman SA. Post-transcriptional regulation of the meiotic Cdc25 protein Twine by the Dazl orthologue Boule. Nat Cell Biol [Internet]. 1999 [cited 2019 Dec 5];1:171–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10559904

24. Baker BS, Wolfner MF. A molecular analysis of doublesex, a bifunctional gene that controls both male and female sexual differentiation in Drosophila melanogaster. Genes Dev [Internet]. 1988 [cited 2019 Dec 5];2:477–89. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2453394

25. Burtis KC, Baker BS. Drosophila doublesex gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. Cell. 1989;56:997–1010.

26. Ohbayashi F, Suzuki MG, Mita K, Okano K, Shimada T. A homologue of the Drosophila doublesex gene is transcribed into sex-specific mRNA isoforms in the silkworm, Bombyx mori. Comp Biochem Physiol - B Biochem Mol Biol. 2001;128:145–58.

27. Scali C, Catteruccia F, Li Q, Crisanti A. Identification of sex-specific transcripts of the Anopheles gambiae doublesex gene. J Exp Biol. 2005;208:3701–9.

28. Cho S, Huang ZY, Zhang J. Sex-specific splicing of the honeybee doublesex gene reveals 300 million years of evolution at the bottom of the insect sex-determination pathway. Genetics [Internet]. 2007 [cited 2019 Dec 5];177:1733–41. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17947419

29. Shukla JN, Palli SR. Doublesex target genes in the red flour beetle, Tribolium castaneum. Sci Rep [Internet]. 2012 [cited 2019 Dec 5];2:948. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23230513

30. Zhuo JC, Hu QL, Zhang HH, Zhang MQ, Jo SB, Zhang CX. Identification and functional analysis of the doublesex gene in the sexual development of a hemimetabolous insect, the brown planthopper. Insect Biochem Mol Biol. Elsevier Ltd; 2018;102:31–42.

31. Le Bras S, Van Doren M. Development of the male germline stem cell niche in Drosophila. Dev Biol. Academic Press Inc.; 2006;294:92–103.

32. Yeh S-D, Chen Y-J, Chang ACY, Ray R, She B-R, Lee W-S, et al. Isolation and properties of Gas8, a growth arrest-specific gene regulated during male gametogenesis to produce a protein associated with the sperm motility apparatus. J Biol Chem [Internet]. 2002 [cited 2019 Dec 5];277:6311–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11751847

33. Hales KG, Fuller MT. Developmentally regulated mitochondrial fusion mediated by a conserved, novel, predicted GTPase. Cell. Cell Press; 1997;90:121–9.

34. Hwa JJ, Hiller MA, Fuller MT, Santel A. Differential expression of the Drosophila mitofusin genes fuzzy onions (fzo) and dmfn. Mech Dev. 2002;116:213–6.

35. Honda S, Hirose S. Stage-specific enhanced expression of mitochondrial fusion and fission factors during spermatogenesis in rat testis. Biochem Biophys Res Commun. Academic Press Inc.; 2003;311:424–32.

36. Hughes AL. Evolutionary diversification of insect innexins. J Insect Sci [Internet]. 2014 [cited 2019 Dec 5];14. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25502029

37. Bielke W, Blaschke RJ, Miescher GC, Zürcher G, Andres AC, Ziemiecki A. Characterization of a novel murine testis-specific serine/threonine kinase. Gene. 1994;139:235–9.

38. Kueng P, Nikolova Z, Djonov V, Hemphill A, Rohrbach V, Boehlen D, et al. A novel family of serine/threonine kinases participating in spermiogenesis. J Cell Biol [Internet]. 1997 [cited 2019 Dec 5];139:1851–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9412477

39. Wei D, Li H-M, Yang W-J, Wei D-D, Dou W, Huang Y, et al. Transcriptome profiling of the testis reveals genes involved in spermatogenesis and marker discovery in the oriental fruit fly,

Bactrocera dorsalis. Insect Mol Biol [Internet]. 2015 [cited 2019 Dec 5];24:41–57. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25255964

40. Svensson MJ, Chen JD, Pirrotta V, Larsson J. The ThioredoxinT and deadhead gene pair encode testis- and ovary-specific thioredoxins in Drosophila melanogaster. Chromosoma [Internet]. 2003 [cited 2019 Dec 5];112:133–43. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14579129

41. Tirmarche S, Kimura S, Dubruille R, Horard B, Loppin B. Unlocking sperm chromatin at fertilization requires a dedicated egg thioredoxin in Drosophila. Nat Commun [Internet]. 2016 [cited 2019 Dec 5];7:13539. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27876811

42. Osborne LJ, Tonissen KF, Tang VH, Clarke FM. Expression and localisation of thioredoxin in mouse reproductive tissues during the oestrous cycle. Mol Reprod Dev [Internet]. 2001 [cited 2019 Dec 5];58:359–67. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11241771

43. Miranda-Vizuete A, Ljung J, Damdimopoulos AE, Gustafsson JÅ, Oko R, Pelto-Huikko M, et al. Characterization of Sptrx, a Novel Member of the Thioredoxin Family Specifically Expressed in Human Spermatozoa. J Biol Chem. 2001;276:31567–74.

44. Miranda-Vizuete A, Tsang K, Yu Y, Jiménez A, Pelto-Huikko M, Flickinger CJ, et al. Cloning and developmental analysis of murid spermatid-specific thioredoxin-2 (SPTRX-2), a novel sperm fibrous sheath protein and autoantigen. J Biol Chem [Internet]. 2003 [cited 2019 Dec 5];278:44874–85. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12909633

45. Jiménez A, Zu W, Rawe VY, Pelto-Huikko M, Flickinger CJ, Sutovsky P, et al. Spermatocyte/spermatid-specific thioredoxin-3, a novel Golgi apparatus-associated thioredoxin, is a specific marker of aberrant spermatogenesis. J Biol Chem [Internet]. 2004 [cited 2019 Dec 5];279:34971–82. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15181017