

**Additional File 1. Supplementary Figures and Legends**

**Figure S1. Adult *Anopheles stephensi* salivary gland lobe and cellular morphometry.** (A) Graph of average length (blue) and width (red) of each lobe from *An. stephensi*. Ten samples were measured per lobe. Results indicate that lobe length varies more than lobe width, but variations are similar across lobes (standard deviations are shown on the graph). (B) Graph of adult *An. stephensi* SG cellular morphometry. Each data point represents 50 cellular measurements among 10 lobes. Results show low variability in both female and male gland cellular dimensions, except for proximal lateral cell length, which varies with asymmetric positioning of the duct within that part of the gland. (C) Cross-sections from Nile Red and DAPI stained SGs illustrating the number of nuclei in circumference in the female distal lateral (DL), proximal lateral (PL), and medial lobe (M), and male gland. Proximal (P) and distal (D) are labeled on each lobe/gland for orientation. Arrow in (PL) marks the PL cross-section. The number of circumferential nuclei varies along the length of each lobe. These cell shape and duct observations agree well with previously published electron microscopy cross-sectional views [11]. Scale bars are 50 microns long.

**Figure S2. Additional details regarding adult *Drosophila* salivary gland cellular architecture.** Adult *D. melanogaster* (Canton S) SGs stained with DAPI (nuclei), phalloidin (actin), and WGA (chitin; A) or Nile Red (lipids; B-D). A) Evidence of cell turnover was observed, including DNA fragmentation (white arrow), basal cell detachment, and actin accumulation (yellow arrow). (B-D) Female (B) and Male (C-D) SG cell shape. Results show Nile Red staining throughout the cell membranes, highlighting the cuboidal shape of the

secretory cells (e.g. Bii, white arrows). Actin is well organized into apico-basal fibers (Biii), and debris can sometimes be visualized in the lumen (Biv, white arrow). The luminal surface appears irregular [compare Cii to D (proximal portion from another gland)]. Images not labeled with “MIP” (maximum intensity projections), are single slices. D is a MIP of two slices. Scale bar lengths are: 50 microns – A, Bi, Ci; 20 microns – Bii, Biii, Biv, Cii, D.

**Figure S3. Larval *Drosophila* salivary gland nuclear dimension correlate with known ploidy.** (A) Male *D. melanogaster* (Oregon R) larval SGs stained with DAPI (nuclei). Three classes of nuclei are visible: 1) imaginal nuclei (blue arrow), 2) duct nuclei (red arrow), and 3) secretory nuclei (green arrow). Imaginal nuclei are known to be diploid; duct nuclei are weakly polyploid and secretory nuclei are very highly polyploid (polytene). (B) Graph of nuclear diameters of these three cell types. Sample size is given at the base of each bar. Note that nuclear size increases with an increase in ploidy. (C) *An. stephensi* nuclear diameter measurements. A graph (Ci) of nuclear diameter measurements from two populations of cells (duct, and secretory cells), reveals that both cell types had consistently sized nuclei; secretory cell nuclei are a little more than twice the diameter of duct nuclei. Secretory cell and duct cell measurements were taken from Fig. 2A and similar images. Small DAPI body (micronucleus) measurements were taken from Cii. In this female DL lobe, micronuclei are abundant (sites of enriched DAPI and Nile Red colocalization). Micronucleus morphology was often elongated, not spherical, so maximal length was measured. Micronucleus length distribution is plotted in Ciii, sorted from smallest to largest. Discontinuous outlier values are boxed. (D) Visual comparison of *D. melanogaster* imaginal cell (Di) and *An. stephensi* micronucleus (Dii) size. Images are the same

scale, and nuclei of interest are marked by white arrows. \*: p-value < 0.001 by pair-wise Mann-Whitney U tests. Scale bar lengths are: 50 microns – A, Dii; 20 microns – Ci, Cii.

**Figure S4. Controls for fluorescence microscopy in *Anopheles stephensi* salivary glands.** A-B) Consistency in immunofluorescence results when performed on frozen versus freshly dissected *An. stephensi* SG tissue. Shown are maximum intensity projections from either freshly dissected (A) or frozen, then thawed (B) adult female and male *An. stephensi* SGs stained with DAPI and Rh-WGA (chitin, O-GlcNAc groups). Similar staining effectiveness was observed on both tissue preparations. C) Secondary antibody background signal controls. Shown are maximum intensity projections from adult *An. stephensi* female or male SGs stained without adding primary antibody. When rabbit secondary antibodies were used in the absence of primary antibody, we detected weak duct staining and punctate cellular staining. This background signal disappeared in the presence of primary antisera against a variety of proteins (See Methods). Mouse secondary antibodies alone showed very little signal accumulation. Scale bars are 50 microns in length.

**Figure S5. Conservation of the mosquito protein AAPP.** A) Multiple sequence alignment of AAPP homologs from multiple mosquito species. Green: identical residue in all proteins. Yellow: conserved residue. Teal: similar residue. B) AAPP protein conservation tree diagram. AAPP is conserved within mosquitoes, but is not present in *Drosophila*. See Methods for additional details.

**Figure S6. Conservation of the microtubule component protein  $\alpha$ -tubulin.** Multiple sequence alignment of  $\alpha$ -tubulin homologs in mosquitoes and *Drosophila*. Green: identical residue in all proteins. Yellow: conserved residue. Teal: similar residue.

**Figure S7. Conservation of the cis-Golgi-associated protein GM130.** Multiple sequence alignment of GM130 homologs in mosquitoes and *Drosophila*. Green: identical residue in all proteins. Yellow: conserved residue. Teal: similar residue.

**Figure S8. Conservation of the ER-associated KDEL receptor.** Multiple sequence alignment of KDEL receptor homologs in mosquitoes and *Drosophila*. Green: identical residue in all proteins. Yellow: conserved residue. Teal: similar residue.

**Figure S9. Conservation of the mitochondrial transcription factor mtTFA.** Multiple sequence alignment of mtTFA homologs in mosquitoes and *Drosophila*. Green: identical residue in all proteins. Yellow: conserved residue. Teal: similar residue.

**Figure S10. Conservation of the mosquito protein SG6.** A) Multiple sequence alignment of SG6 homologs in several mosquito species. Green: identical residue in all proteins. Yellow: conserved residue. Teal: similar residue. B) SG6 protein conservation tree diagram. SG6 is conserved within mosquitoes, but is not present in *Drosophila*. See Methods for additional details.

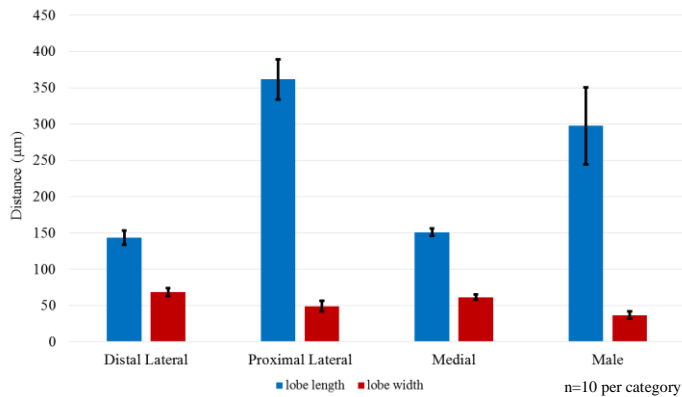
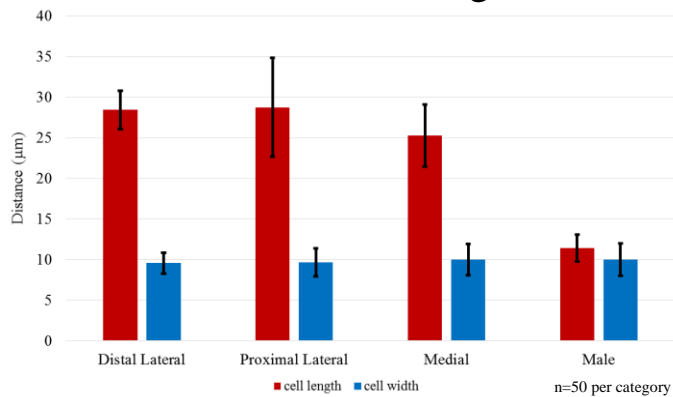
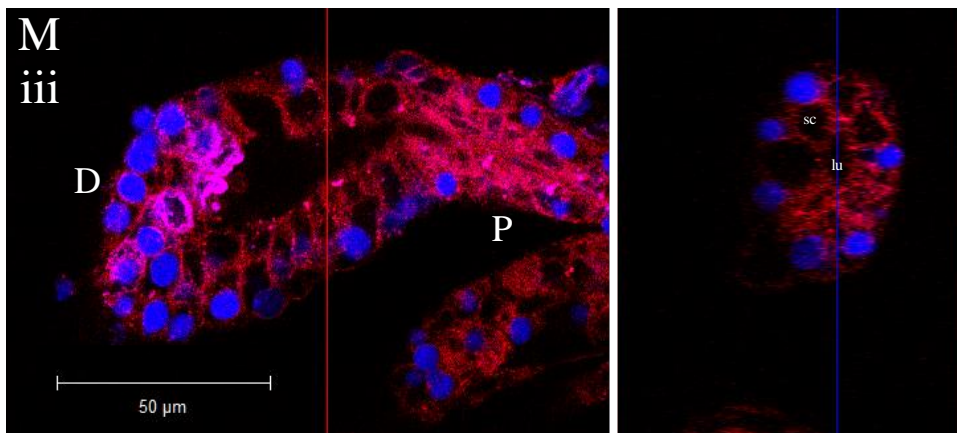
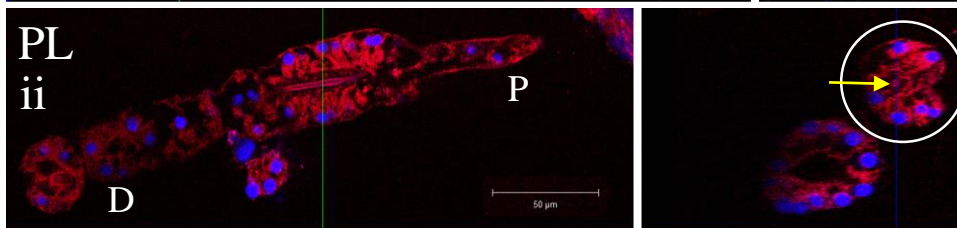
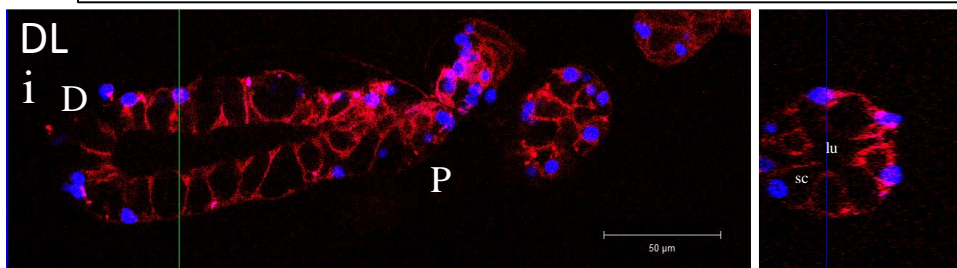
**A.****B.**

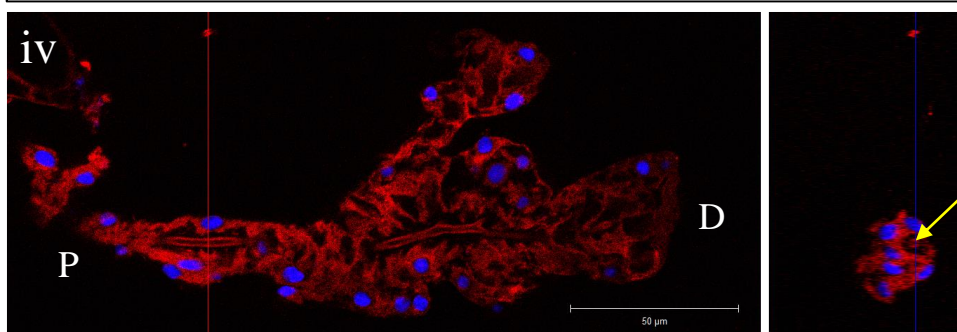
Figure S1

**C.**

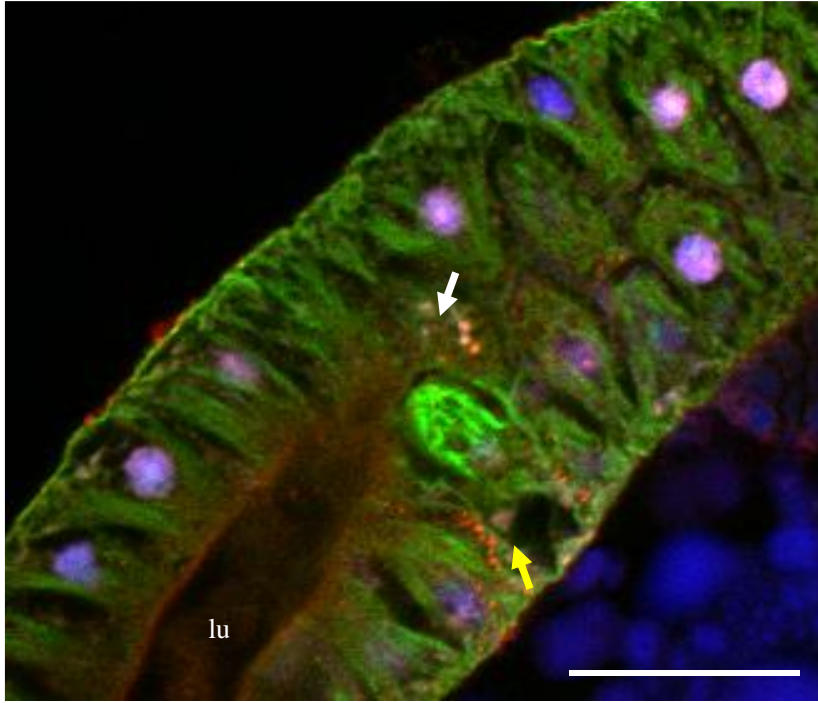
Female DAPI Nile Red



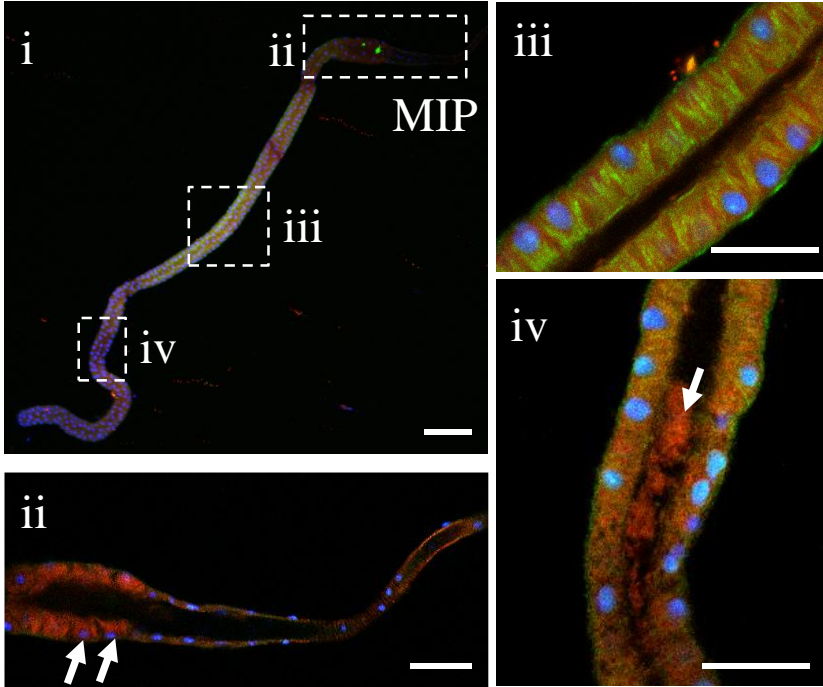
Male DAPI Nile Red



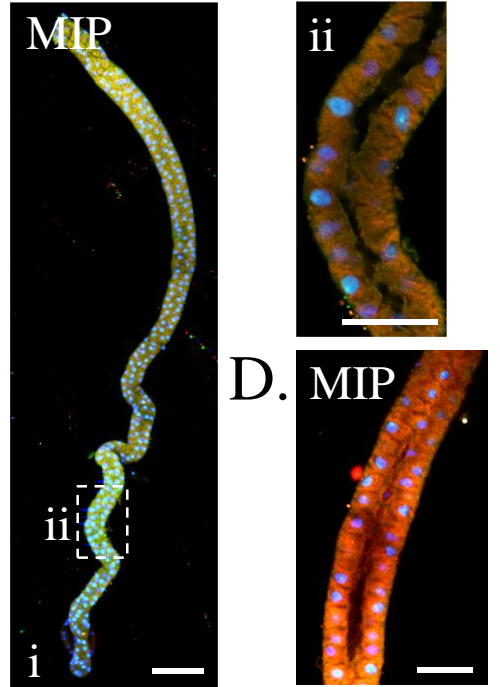
A. Female DAPI Phalloidin WGA



B. Female DAPI Phalloidin Nile Red



C. Male DAPI Phalloidin Nile Red



D. MIP

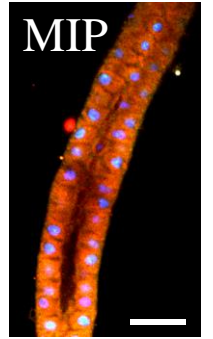
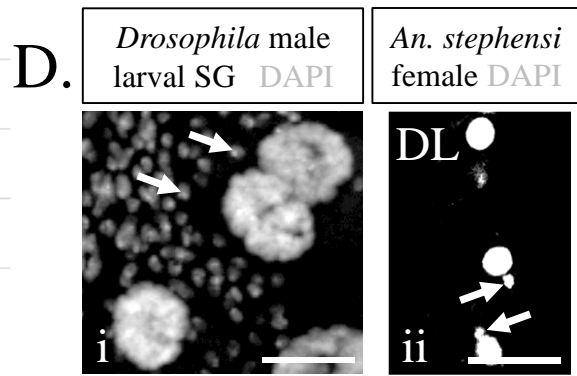
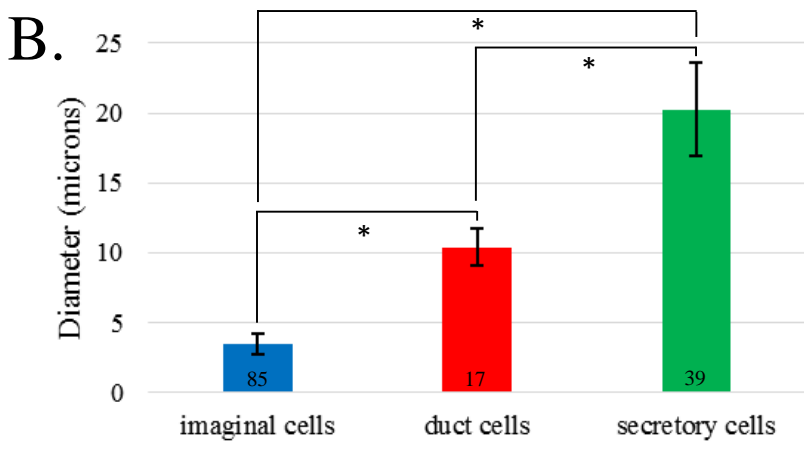
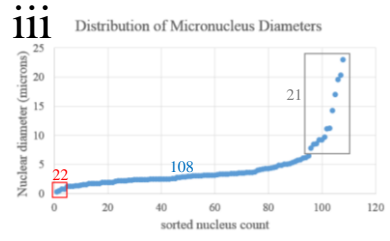
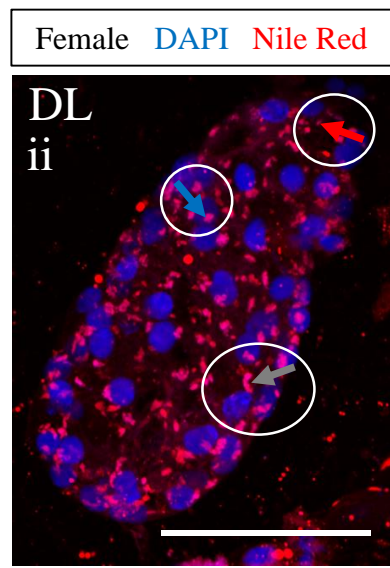
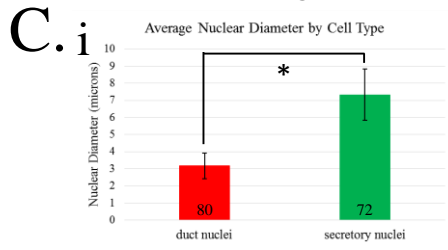
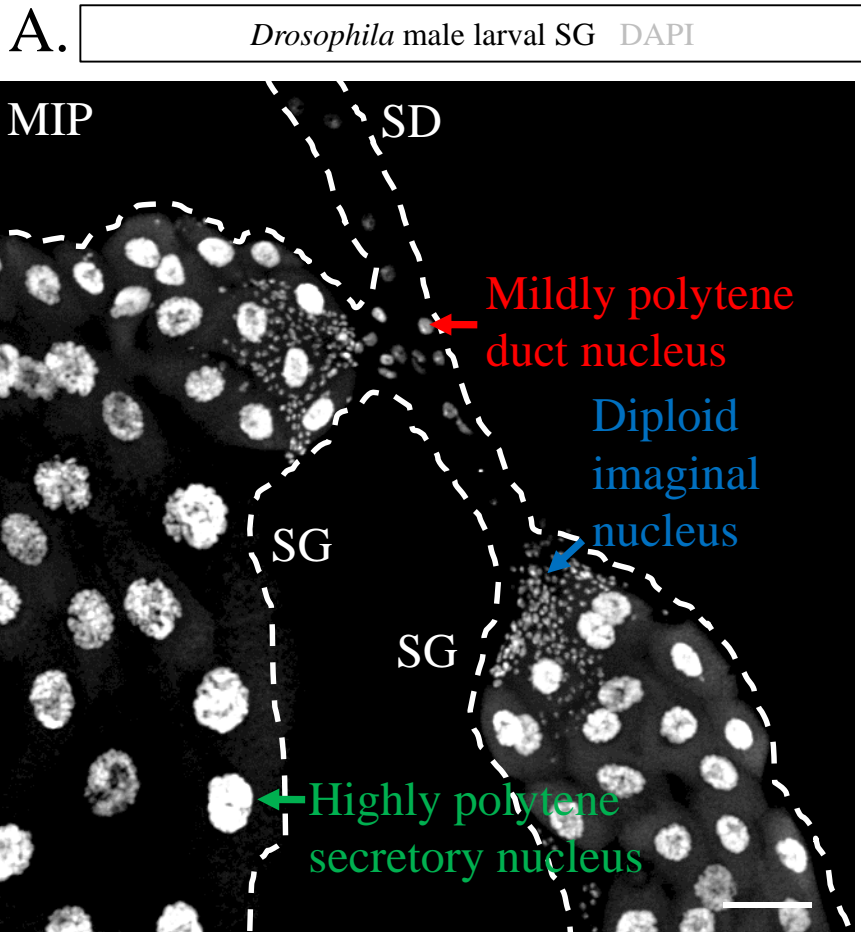
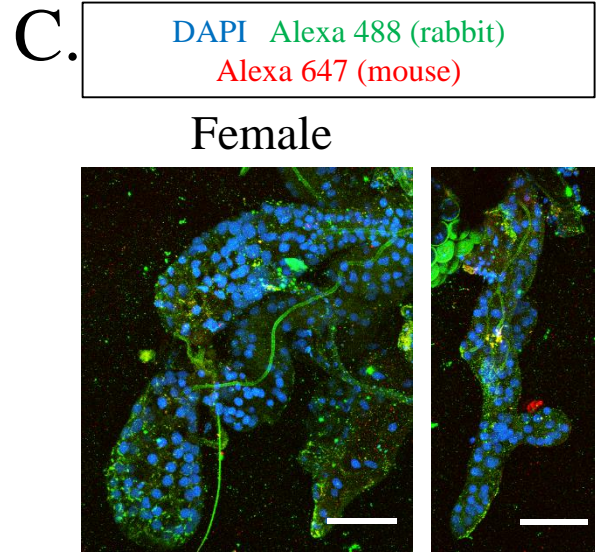
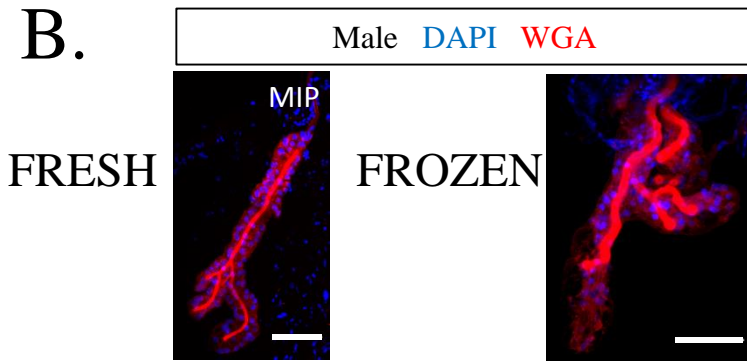
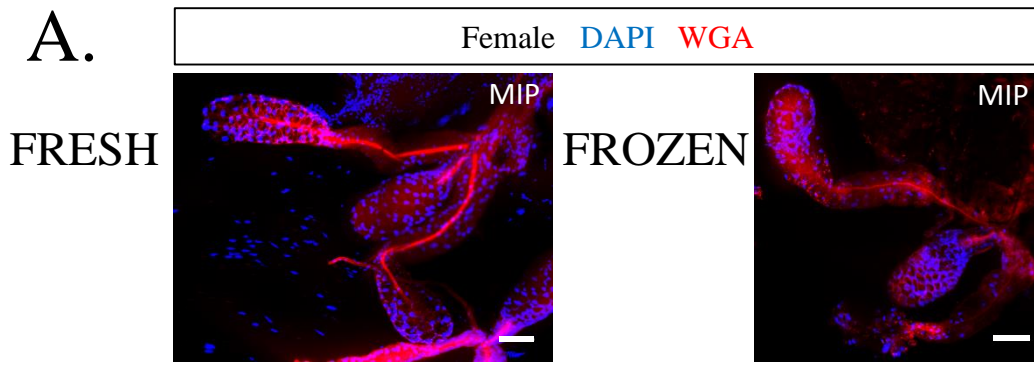


Figure S3









```

ADA0009605 MRECI SVHVGQAGVQIGNACWELYCLEHG IQ FD GQMP S DK TI GGGDD S FN TF FSE TGAG KHV PRAV FVD LE PT VVD EVRTG TY RQL FPH PEQL ITG KEDAANN YARGHYT IGKEI VDVVLD
AQUA001217 MRECI SVHVGQAGVQIGNACWELYCLEHG IQ FD GQMP S DK TI GGGDD S FN TF FSE TGAG KHV PRAV FVD LE PT VVD EVRTG TY RQL FPH PEQL ITG KEDAANN YARGHYT IGKEI VDVVLD
AAEL006642 MRECI SVHVGQAGVQIGNACWELYCLEHG IQ FD GQMP S DK TI GGGDD S FN TF FSE TGAG KHV PRAV FVD LE PT VVD EVRTG TY RQL FPH PEQL ITG KEDAANN YARGHYT IGKEI VDVVLD
AGA001219 MRECI SVHVGQAGVQIGNACWELYCLEHG IQ FD GQMP S DK TI GGGDD S FN TF FSE TGAG KHV PRAV FVD LE PT VVD EVRTG TY RQL FPH PEQL ITG KEDAANN YARGHYT IGKEI VDVVLD
ASTE006004 MRECI SVHVGQAGVQIGNACWELYCLEHG IQ FD GQMP S DK TI GGGDD S FN TF FSE TGAG KHV PRAV FVD LE PT VVD EVRTG TY RQL FPH PEQL ITG KEDAANN YARGHYT IGKEI VDVVLD
FBgn0003886 MRECI SVHVGQAGVQIGNACWELYCLEHG IQ FD GQMP S DK TI GGGDD S FN TF FSE TGAG KHV PRAV FVD LE PT VVD EVRTG TY RQL FPH PEQL ITG KEDAANN YARGHYT IGKEI VDVVLD
consensus MRECI SVHVGQAGVQIGNACWELYCLEHG IQ FD GQMP S DK TI GGGDD S FN TF FSE TGAG KHV PRAV FVD LE PT VVD EVRTG TY RQL FPH PEQL ITG KEDAANN YARGHYT IGKEI VDVVLD

ADA0009605 RIRKLABQCTGLQGFLI FHSFGGGTSGFTS LLMERLS VDYGKSKLEFAIY PAVQVSTAVVEPYNSILTTHTTLEHSDCAFMDVNEAIYDICHRRNLDIERPTY TNLNRLIQQIVS SI TA
AQUA001217 RIRKLABQCTGLQGFLI FHSFGGGTSGFTS LLMERLS VDYGKSKLEFAIY PAVQVSTAVVEPYNSILTTHTTLEHSDCAFMDVNEAIYDICHRRNLDIERPTY TNLNRLIQQIVS SI TA
AAEL006642 RIRKLABQCTGLQGFLI FHSFGGGTSGFTS LLMERLS VDYGKSKLEFAIY PAVQVSTAVVEPYNSILTTHTTLEHSDCAFMDVNEAIYDICHRRNLDIERPTY TNLNRLIQQIVS SI TA
AGA001219 RIRKLABQCTGLQGFLI FHSFGGGTSGFTS LLMERLS VDYGKSKLEFAIY PAVQVSTAVVEPYNSILTTHTTLEHSDCAFMDVNEAIYDICHRRNLDIERPTY TNLNRLIQQIVS SI TA
ASTE006004 RIRKLABQCTGLQGFLI FHSFGGGTSGFTS LLMERLS VDYGKSKLEFAIY PAVQVSTAVVEPYNSILTTHTTLEHSDCAFMDVNEAIYDICHRRNLDIERPTY TNLNRLIQQIVS SI TA
FBgn0003886 RIRKLABQCTGLQGFLI FHSFGGGTSGFTS LLMERLS VDYGKSKLEFAIY PAVQVSTAVVEPYNSILTTHTTLEHSDCAFMDVNEAIYDICHRRNLDIERPTY TNLNRLIQQIVS SI TA
consensus RIRKLABQCTGLQGFLI FHSFGGGTSGFTS LLMERLS VDYGKSKLEFAIY PAVQVSTAVVEPYNSILTTHTTLEHSDCAFMDVNEAIYDICHRRNLDIERPTY TNLNRLIQQIVS SI TA

ADA0009605 SLRFDGALNVDLTFQTNLVPYPR IHFPLVTYAPVISA EKAYHEQLS VAEITNACFE PANQMVKCD PRH GK YMACCML YRGDVPF KDVNAAIATI KTKR TIQ FVDWCPTGFVGIN YQPF
AQUA001217 SLRFDGALNVDLTFQTNLVPYPR IHFPLVTYAPVISA EKAYHEQLS VAEITNACFE PANQMVKCD PRH GK YMACCML YRGDVPF KDVNAAIATI KTKR TIQ FVDWCPTGFVGIN YQPF
AAEL006642 SLRFDGALNVDLTFQTNLVPYPR IHFPLVTYAPVISA EKAYHEQLS VAEITNACFE PANQMVKCD PRH GK YMACCML YRGDVPF KDVNAAIATI KTKR TIQ FVDWCPTGFVGIN YQPF
AGA001219 SLRFDGALNVDLTFQTNLVPYPR IHFPLVTYAPVISA EKAYHEQLS VAEITNACFE PANQMVKCD PRH GK YMACCML YRGDVPF KDVNAAIATI KTKR TIQ FVDWCPTGFVGIN YQPF
ASTE006004 SLRFDGALNVDLTFQTNLVPYPR IHFPLVTYAPVISA EKAYHEQLS VAEITNACFE PANQMVKCD PRH GK YMACCML YRGDVPF KDVNAAIATI KTKR TIQ FVDWCPTGFVGIN YQPF
FBgn0003886 SLRFDGALNVDLTFQTNLVPYPR IHFPLVTYAPVISA EKAYHEQLS VAEITNACFE PANQMVKCD PRH GK YMACCML YRGDVPF KDVNAAIATI KTKR TIQ FVDWCPTGFVGIN YQPF
consensus SLRFDGALNVDLTFQTNLVPYPR IHFPLVTYAPVISA EKAYHEQLS VAEITNACFE PANQMVKCD PRH GK YMACCML YRGDVPF KDVNAAIATI KTKR TIQ FVDWCPTGFVGIN YQPF

ADA0009605 TVVPGDLAKVQRAVCMLSNNTAI AEAWARLDHKFDLMYAKRA FVHWYVGEEMEEGEFS EAR ED LA ALE KD YE EVG MD SGE GE DE GAR EY
AQUA001217 TVVPGDLAKVQRAVCMLSNNTAI AEAWARLDHKFDLMYAKRA FVHWYVGEEMEEGEFS EAR ED LA ALE KD YE EVG MD SGE GE DE GAR EY
AAEL006642 TVVPGDLAKVQRAVCMLSNNTAI AEAWARLDHKFDLMYAKRA FVHWYVGEEMEEGEFS EAR ED LA ALE KD YE EVG MD SGE GE DE GAR EY
AGA001219 TVVPGDLAKVQRAVCMLSNNTAI AEAWARLDHKFDLMYAKRA FVHWYVGEEMEEGEFS EAR ED LA ALE KD YE EVG MD SGE GE DE GAR EY
ASTE006004 TVVPGDLAKVQRAVCMLSNNTAI AEAWARLDHKFDLMYAKRA FVHWYVGEEMEEGEFS EAR ED LA ALE KD YE EVG MD SGE GE DE GAR EY
FBgn0003886 TVVPGDLAKVQRAVCMLSNNTAI AEAWARLDHKFDLMYAKRA FVHWYVGEEMEEGEFS EAR ED LA ALE KD YE EVG MD SGE GE DE GAR EY
consensus TVVPGDLAKVQRAVCMLSNNTAI AEAWARLDHKFDLMYAKRA FVHWYVGEEMEEGEFS EAR ED LA ALE KD YE EVG MD Sge ge GE ga EY

```



```

AAEL008434 -MNIFRLAGLS HLLAIILLLIKIWKTRSCAIGSGKSQILFAIVVYS
AAEL002086 -MNIFRLAGLS HLLAIILLLIKIWKTRSCAIGSGKSQILFAIVVYS
ADAC005863 -MNIFRLSGLS HLLAIILLLIKIWKTRSCAIGSGKSQILFAIVVYS
AALB007873 -MNIFRLSGLS HLLAIILLLIKIWKTRSCAIGSGKSQILFAIVVYS
ASTE003703 -MNIFRLAGLS HLLAIILLLIKIWKTRSCAIGSGKSQILFAIVVYS
AGAP012756 -MNIFRLAGLS HLLAIILLLIKIWKTRSCAIGSGKSQILFAIVVYS
ADIR004038 -MNIFRLAGLS HLLAIILLLIKIWKTRSCAIGSGKSQILFAIVVYS
AFUN001915 -MLTGYRLLAGSFAFAIILYLFNINRWRKSCFVSGKQILVVTVDATRYADLVTFPEPTYSVYNVLMKTLFI SATLITVLLMHS IYRKT YDRENDTFYNEVLI LPCFV T ALFVNYRMEA FE
FBgn0022268 -MNIFRLAGLS HLLAIILLLIKIWKTRSCAIGSGKSQILFAIVVYS
consensus -mnifrlagDlsH11AIi1LLIkiWkTrSCaGiSGKsQILfaiVvys

```

```

AAEL008434 -----RYLDLVTFIFLYNTLKKLVFI
AAEL002086 -----RYLDLVTFIFLYNTLKKLVFI
ADAC005863 -----RYLDLVTFIFLYNTLKKLVFI
AALB007873 -----RYLDLVTFIFLYNTLKKLVFI
ASTE003703 -----RYLDLVTFIFLYNTLKKLVFI
AGAP012756 -----RYLDLVTFIFLYNTLKKLVFI
ADIR004038 -----RYLDLVTFIFLYNTLKKLVFI
AFUN001915 ILWSFSIFLEAVAILPQMDLICKTFRHVEPWFKCYLLLLGSYRALYVLHWMDRYSLYGLYDPLFAIAGGVQTVL FVLLALRIATLKHRRDRI VTVSTICYR
FBgn0022268 -----RYLDLVTFIFLYNTLKKLVFI
consensus -----RYLDLVTFIFLYNTLKKLVFI

```

```

AAEL008434 TSSFATLFLMYVFKKATYDHNHDSFRIEFLLLPCFVLA LLINSAFTPLVLTWTFIYLEAVA ILPQLFLVSKTGEAESITSHYLPALGSRALYLLNWIYRYYAEQHYDLIAIFAGAIQT
AAEL002086 TSSFATLFLMYVFKKATYDHNHDSFRIEFLLLPCFVLA LLINSAFTPLVLTWTFIYLEAVA ILPQLFLVSKTGEAESITSHYLPALGSRALYLLNWIYRYYAEQHYDLIAIFAGAIQT
ADAC005863 SAAIATI FLMYVFKKATYDHNHDSFRIEFLLLPCFVLA LLINSAFTPLVLTWTFIYLEAVA ILPQLFLVSKTGEAESITSHYLPALGSRALYLLNWIYRYYAEQHYDLIAIFAGAIQT
AALB007873 SAAIATI FLMYVFKKATYDHNHDSFRIEFLLLPCFVLA LLINSAFTPLVLTWTFIYLEAVA ILPQLFLVSKTGEAESITSHYLPALGSRALYLLNWIYRYYAEQHYDLIAIFAGAIQT
ASTE003703 STSVATI FLMYVFKKATYDHNHDSFRIEFLLVPCPLLA LLINNAFTPLVLTWTFIYLEAVA ILPQLFLVSKTGEAESITSHYLPALGSRALYLLNWIYRYYAEQHYDLIAIFAGAIQT
AGAP012756 STSVATI FLMYVFKKATYDHNHDSFRIEFLLVPCPLLA LLINNAFTPLVLTWTFIYLEAVA ILPQLFLVSKTGEAESITSHYLPALGSRALYLLNWIYRYYAEQHYDLIAIFAGAIQT
ADIR004038 STSVATI FLMYVFKKATYDHNHDSFRIEFLLVPCPLLA LLINNAFTPLVLTWTFIYLEAVA ILPQLFLVSKTGEAESITSHYLPALGSRALYLLNWIYRYYAEQHYDLIAIFAGAIQT
AFUN001915 STSVATI FLMYVFKKATYDHNHDSFRIEFLLVPCPLLA LLINNAFTPLVLTWTFIYLEAVA ILPQLFLVSKTGEAESITSHYLPALGSRALYLLNWIYRYYAEQHYDLIAIFAGAIQT
FBgn0022268 STSCATVFLMYVFKKATYDHNHDSFRIEFLLVPCPLLA LLINNAFTPLVLTWTFIYLEAVA ILPQLFLVSKTGEAESITSHYLPALGSRALYLLNWIYRYMVESYDLIAIFAGVVTQ
consensus stsiATI FLMYVFKKATYDHNHDSFRIEFLLVPCPLLA LLINNAFTPLVLTWTFIYLEAVA ILPQLFLVSKTGEAESITSHYLPALGSRALYLLNWIYRYYAEQHYDLIAIFAGAIQT

```

```

AAEL008434 ILYCDFFVLYITKVLKGGKLLQ LPA
AAEL002086 ILYCDFFVLYITKVLKGGKLLQ LPA
ADAC005863 ILYCDFFVLYITKVLKGGKLLQ LPA
AALB007873 ILYCDFFVLYITKVLKGGKLLQ LPA
ASTE003703 ILYCDFFVLYITKVLKGGKLLQ LPA
AGAP012756 ILYCDFFVLYITKVLKGGKLLQ LPA
ADIR004038 ILYCDFFVLYITKVLKGGKLLQ LPA
AFUN001915 ILYCDFFVLYITKVLKGGKLLQ LPA
FBgn0022268 ILYCDFFVLYITKVLKGGKLLQ LPA
consensus ILYCDFFVLYITKVLKGGKLLQ LPA

```



A.

```

AQUA002523 MAIRVELLAMVLEPLLLESVVPYAAAEKVVVDRDKVYCGHLDCTRVATFKGERFCTLCDTRHFCECKETRELPYMYACPGTEPCQSSDRLGSCSKTMHDVLCDRIDQAFLEQ
AGAP000150 MAIRVELLAMVLEPLLLESVVPYAAAEKVVVDRDKVYCGHLDCTRVATFKGERFCTLCDTRHFCECKETRELPYMYACPGTEPCQSSDRLGSCSKTMHDVLCDRIDQAFLEQ
AARA001017 MAIRVELLMVLEPLLLESVVPYAAAEKVVVDRDKVYCGHLDCTRVATFKGERFCTLCDTRHFCECKETRELPYMYACPGTEPCQSSDRLGSCSKTMHDVLCDRIDQAFLEQ
ASTE000264 NATFWALLTMVLEPVLQTSG-PYAAAEKVVVDRDKVYCGHLDCTRVATFKGERFCSPCDTRHFCECKETRESLPYMYACPGTEPCQSSDRRGSQQTMSDELCSRIDQAFLEA
AFUN003883 MAIRVELLAMVLEPVLQVGG-PYATAEKVVVDRDKVYCGHLDCTRVATFKGERFCSPCDTRHFCECKETRESLPYMYACPGTEPCQSSDRRSTQCKMHDVLCSLIKPFLEQ
ADIR000828 MKAFWRLLMVAAAVLLCLLSP---VIAGQTHIDRDKTYCEHIDCTKLAKYKGERFCSPCDTRHFCECKEVRRESLPLVGTSPGSGEQRKTSRSRQKCKTLDHNNLCSLILKPYM--
consensus Mai-VeLlIamVlLp1l1LlEsVvPyaaAekVwVDRdkvYcGh1DcTrvAtfKGeRfCt-CDTRHfCECKeTrE-LPYmyaCPGtePCQsSDr-GsCsktmhdvLC-zIDqafleq
    
```

LEGEND

Collapsed Alignments

- 0 - 33% Aligned AA
- 33 - 66% Aligned AA
- 66 - 100% Aligned AA

Genes

- Gene ID gene of interest
- Gene ID within-sp. paralog

Nodes

- gene node
- speciation node
- duplication node
- ambiguous node
- gene split event

B.

