

P. papatasi dsx gene model

Scaffold7_Length=1359615

[illegible]

Figure S21. Manually-curated *P. papatasi* doublesex gene model. Exonic sequences are indicated by black cases; black upper cases indicate coding sequences. Intronic sequences are indicated by gray lower cases. Start and stop codons are highlighted in green and red, respectively. In yellow boxes, putative TRA/TRA-2 binding sites. Exon start/end positions in the scaffold/s are indicated. Intron boundaries were predicted using transcripts vs genome alignments and confirmed by *de novo* prediction with Berkeley BDGP Splice Site Prediction Tool with default parameters (http://www.fruitfly.org/seq_tools/splice.html); prediction scores are indicated in red.

L. longipalpis dsx gene model



Figure S22. Manually-curated *L. longipalpis doublesex* gene model. Exonic sequences are indicated by black cases; black upper cases indicate coding sequences. Intronic sequences are indicated by gray lower cases. Start and stop codons are highlighted in green and red, respectively. In yellow boxes, putative TRA/TRA-2 binding sites. Exon start/end positions in the scaffolds are indicated. Intron boundaries were predicted using transcripts vs genome alignments and confirmed by *de novo* prediction with Berkeley BDGP Splice Site Prediction Tool with default parameters (http://www.fruitfly.org/seq_tools/splice.html); prediction scores are indicated in red.

A)

827 EXON S1 1050 EXON S2

DNA: gcatagaATGATGACTACTCCGGACATCTTCGGCAGCCCCCTTCGCGGTATAGGTGTACGCCGTGGCTGACTTCCACCACGTCATGATAGTCGCGCAATCGCGGCACCTTGACCTCAAAAGCTATCCGACGTCAGGTTCTCCGGCATAACAACACCTGTCGCTTTACCTGAACGTGAAATGCCGATGCCCGACCTGCGTCATCTTCGGCAGCAACCAACCTAGTggcttctc 0.95 cacctagTtaggtc

+2fr: M·M·T·T·P·D·I·F·G·S·P·F·P·P·Y·R·C·Q·P·L·A·V·L·P·P·R·D·D·S·P·P·M·P·A·L·D·L·K·R·Y·P·S·T·E·V·P·P·A·S·T·T·P·V·R·L·P·E·R·E·M·P·M·P·H·L·R·P·S·S·A·Q·T·P·P·R·*

DNA: acctctggcgctagcacaattagcaggacacagtcgcagacaccacatcgcactctcaggagaaaaTCTGCAATCAACTgctcgcctgtgtcttatgtacagaatggcattgggggaagttagtgcaattgatcaattccctcgtagaacaaaaataaacctataccactgctttaaaggttttcatttcttcgggtgcactgttcaaccttctgtactcataaagatttcagtaacctc

DNA: cctagaaatcatcgagatttaattcaattatgcttttcagtaaatgcttctgcagagaaatttagaagaacatgatgaaatcggaagttaaaatagaggatctcaagatcacagtccttgatgaaagattcactcgcgagggaatctgcaggaagatccacatggaaaatgatattaaatgaatcagagatggtgatgtattatcagtggcacatgaagtacaagtcgaagttgg 0.76 caaggtTG

Tgaggtc ???

DNA: Tgaggtc caaaaacgttgaaacatacgggggtgttgtagctttgatcacctgcaaaaatacactcaagccctattgtatgattcttggcgcaactcttgccattttacatccatctcaagagagtcttttatgcttggcgcgattttcccttccctccactccatgaaatgttcaaaaattgcttcgtgaaaattctgcggaagctgcaattcaatatataaacctttggccat

DNA: tttctggagaatccttttgatatacagagacogtgaggacctgggcatcagggtgtgcttgcatctgtacaagtcaaacatgcccaagattgaaaagaattcctttttatgggaagtcccggaatacagatacaaatgtactacttttcagaaaatggaaaatgccttacttcaacaatgcacttaccaaattccatagtggtttcttctcagggacgctggccgttatgtatg

DNA: aatgttgattgtagattgtttttgttccaatgcatgttaatcgtgcaaaattgttgatgaaaataataatgattctttacgactattgcaataaataggatattactggcaaaaagcatttctatttgaattcttcaagtattctctgtttatcttaacttgcacttccattagcaaaactttttactctttactattatatttttgcctcgtgtgtctacattcttcttactattctg

DNA: gcggtgaaaattttctaatagatacacaagaattactatcgggcaaatatattgaaactccgcgattgcaatttttttgggtacttttttgcaaggaaaaatctttgtagattttccatactcggaaactctttatttttggtagtataacttgcattctcgtattcttttaaacagtgagaatgacgcgactgctataattgaaagaaacctctctg

DNA: gttagatttaattgtaaaagaatttgatgtagggtgtgtcgcacgaggatgcccgaattggaatcagggttttcacagtggatgaggaataaaggcaggggagatattgtaactctttaatttcaaaagtcaacaccttggccaatacaccagaaattttattctgtctccacattttcaactggtccttgaacatctttttttagtgtaaaacaaaaattcaaaaaatgaata

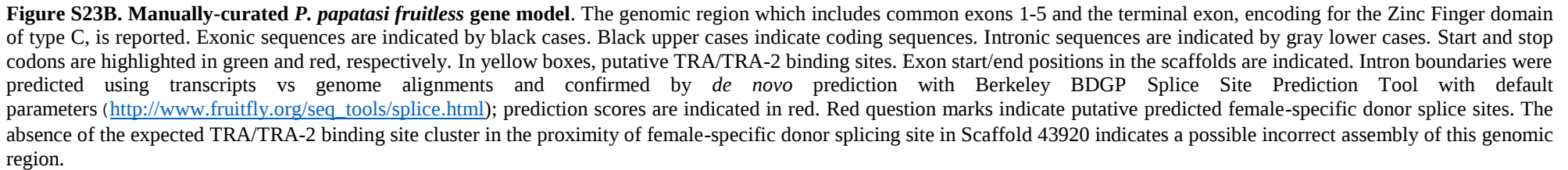
??? 0.75 cattttgTcagtaa ???

DNA: tacatttttggacagtaaatattctcattttgtcagtaataacgttaacattaacgcaaaattgctattaaactttaaacattctcgagatatacaaaatgatgattataaactctctattttccatttgcgaagaatttttattaattctatgcaattgagaattgtgtaatcccaattttaattacttgcgcaaaatttttttatatttttactttaaacctttaacgcagagacagt

DNA: ttttgcggcagcaaaaatcagcaataaaaaatgaaccgataaaaaaatttagtaaaagtcttagataccttagaaaaatctatcagaattcgtatcgaagtatttttgcctgtatgagcgatagggaacaaaaatgcaaaaaaatttagtaaaatttctgcacataccaatgacagataattttcactcacaagaaaaattattatgtttgagtattgattgttcttttctcaaacg

Figure S23A. Manually-curated *P. papatasi* fruitless gene model. The genomic region regulated by sex-specific alternative splicing, which includes the two sex-specific exons S1 and S2, is reported. Exonic sequences are indicated by black cases. Black upper cases indicate coding sequences. Intronic sequences are indicated by gray lower cases. Start codons are highlighted in green. In yellow boxes, putative TRA/TRA-2 binding sites. Exon start/end positions in the scaffolds are indicated. Intron boundaries were predicted using transcripts vs genome alignments and confirmed by *de novo* prediction with Berkeley BDGP Splice Site Prediction Tool with default parameters (http://www.fruitfly.org/seq_tools/splice.html); prediction scores are indicated in red. Red question marks indicate putative predicted female-specific donor splice sites. The absence of the expected TRA/TRA-2 binding site cluster in the proximity of female-specific donor splicing site in Scaffold 43920 indicates a possible incorrect assembly of this genomic region.

B)



L. longipalpis fru partial gene model

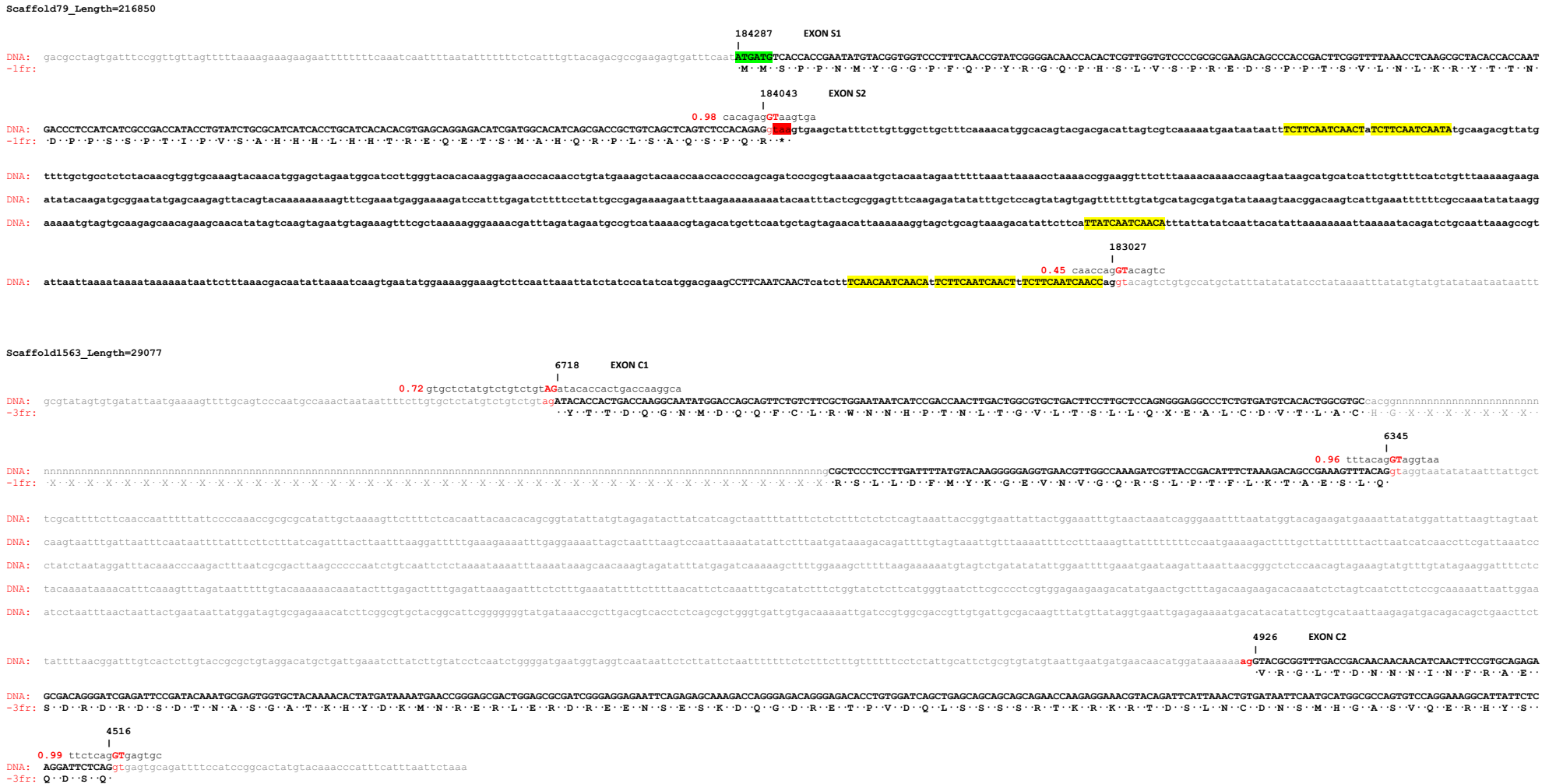


Figure S24. Manually-curated *L. longipalpis* fruitless partial gene model. The genomic region regulated by sex-specific alternative splicing is included. Exonic sequences are indicated by black cases. Black upper cases indicate coding sequences. Intronic sequences are indicated by gray lower cases. Start and stop codons are highlighted in green and red, respectively. In yellow boxes, putative TRA/TRA-2 binding sites. Exon start/end positions in the scaffolds are indicated. Intron boundaries were predicted using transcripts vs genome alignment and confirmed by *de novo* prediction with Berkeley BDGP Splice Site Prediction Tool with default parameters (http://www.fruitfly.org/seq_tools/splice.html); prediction scores are indicated in red.

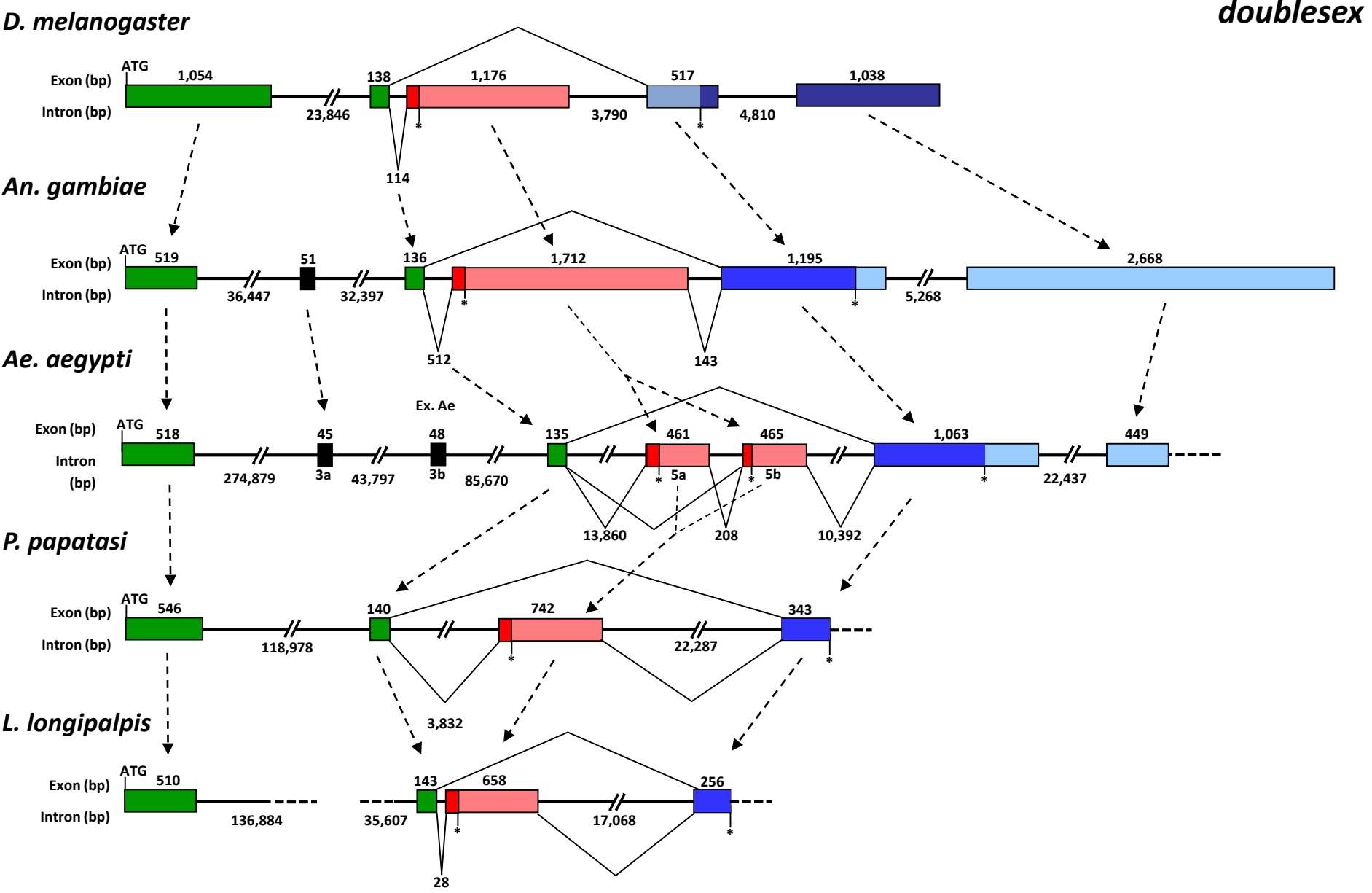


Figure S25. Comparison of genomic structures of diptera *dsx* genes. Green boxes represent the OD1 and OD2 domains encoding common exons. Black boxes represent exons encoding protein regions conserved in mosquitoes but not in fruit flies and sand flies. Alternative male-specific and female-specific exons are represented as blue boxes and pink boxes, respectively. Asterisks indicate the position of stop codons. In *Drosophila*, the *dsx* gene encompasses a 45 Kb-long region on chromosome 3R and is organized into 6 exons and 5 introns, with three common exons followed by a female-specific and two male-specific exons. DmDSXF translation initiates at the ATG within exon 2 and terminates within the female-specific exon 4, while in the case of DmDSXM, translation begins at the same ATG and terminates within the first male-specific exon 5.

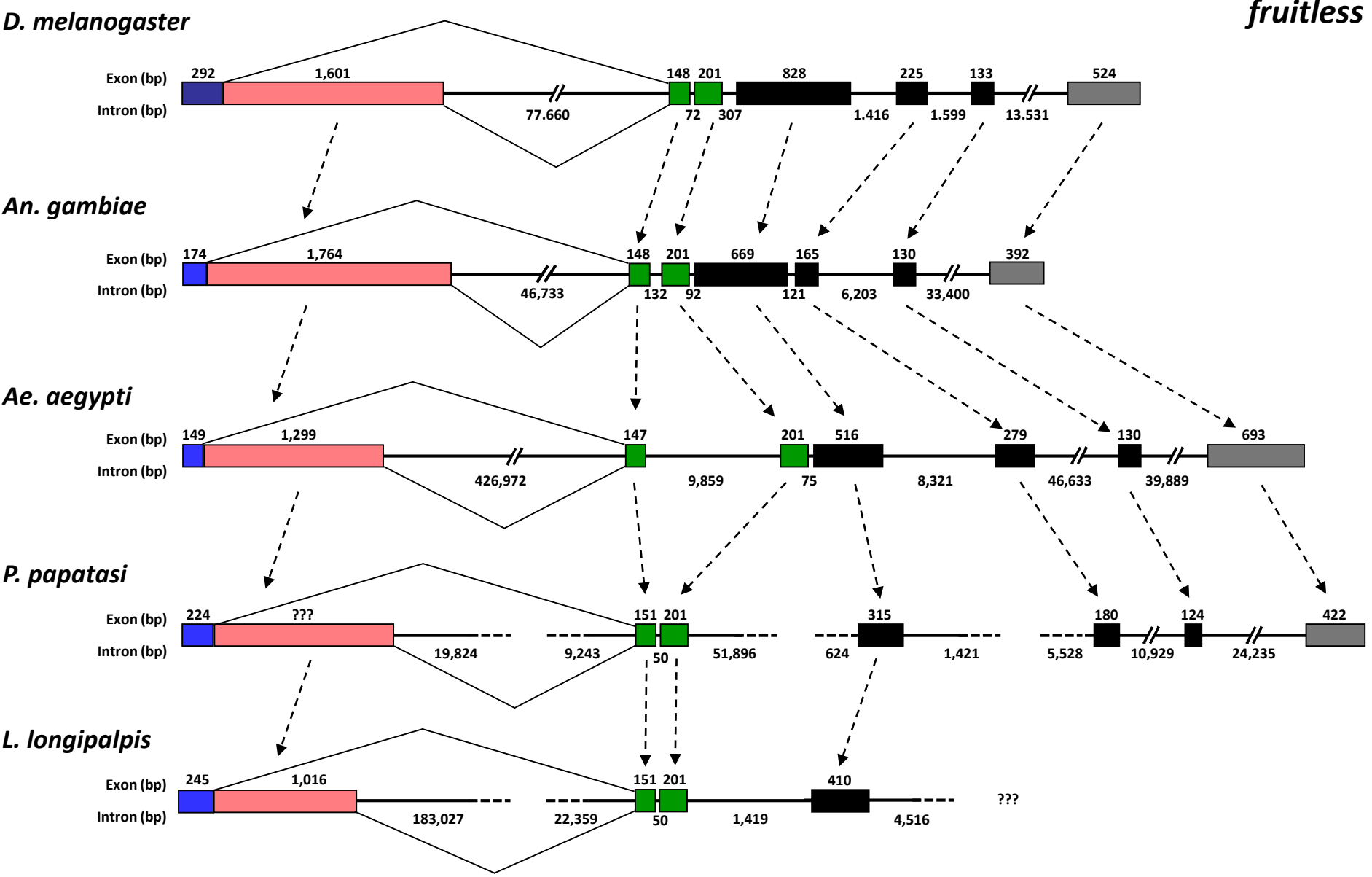


Figure S26. Comparison of genomic structures of diptera fru genes. Fru S1 common (but encoding the male-specific N-terminus) and S2 female-specific exons are represented as blue and pink boxes, respectively. Green boxes represent the non-sex-specific C1 and C2 exons, encoding the BTB domain. Black boxes represent the non-sex-specific C3-C5 exons encoding the Connector region of FRU proteins. The terminal grey boxes represent the ZnF-C domain encoding exons. The *Drosophila fru* gene encompasses a 98 Kb-long region and is organized in 7 exons and 6 introns with 6 common exons preceded by the sex-specific regulated region, with a common exon and a female-specific exon.