

Figure S1. *Drosophila* phylogenetic tree drawn to scale illustrating the number of intron gains and losses. Pluses indicate the number of gained introns; minuses indicate the number of lost introns. Numbers at far right of the tree represent events identified in one species. Numbers at nodes represent events assumed to have occurred in ancestors. All branch lengths are drawn to scale except for the branch leading to *A. gam*, which likely radiated from the common ancestor of *Drosophila* and *A. gam* 250 million years ago [1].

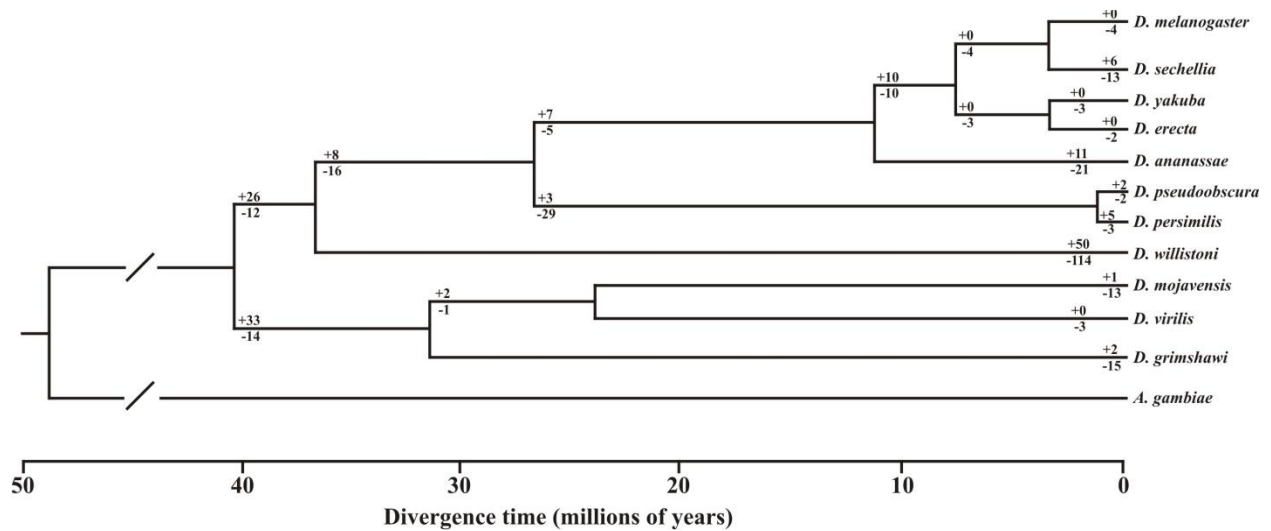


Figure S2. An example of intron sliding. (a) Global alignment of orthologous genes. All introns shown in this alignment are the fourth introns from their respective genes. Thirty shaded X's represent introns. (b) Alignment and score between the intronic sequence of the fourth intron in Dwil\GK22863 and the fourth intron in Dper\GL17458.

(a)

```

Dmel\CG8026      ATCAAACAGACTTGGAG-----GTACGAGCGCATG
Dsec\GM20618     ATCAAACAGACTTGGAG-----GTACGAGCGCATG
Dyak\GE22456     ATCAAACAAACTTGGAG-----GTACGAGCGCATG
Dere\GG10572     ATCAAACAGACTTGGAG-----GTACGAGCGCATG
Dana\GF12007     ATCCGACAGACTTGGAG-----GTACGAACGCATG
Dpse\GA20774     ATTAAACAGACTTGGAG-----GTACGAACGCATG
Dper\GL17458     ATTAAACAGACTTGGAG-----GTACGAACGCATG
Dwil\GK22863     ATCAAACAGACTTGGAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTTTGAGGGCATG
Dgri\GH19865     ATCAAACAGACTTGGAG-----GTACGAACGCATG
Dmoj\GI19815     ATCAAACAGACTTGGAG-----GTACGAACGCATG
Dvir\GJ18531     ATCAAACAGACTTGGAG-----GTACGAACGCATG

Dmel\CG8026      CGAGGTTTCTATAAGGGCCTG-GTGCCCTACCTGGTCCACGTCACGCCCAACATCTGCAT
Dsec\GM20618     CGAGGTTTCTATAAGGGCCTG-GTGCCCTACCTGGTCCACGTCACGCCCAACATCTGCAT
Dyak\GE22456     CGAGGTTTCTATAAGGGCCTG-GTGCCCTACCTGGTCCACGTCACGCCCAACATCTGCAT
Dere\GG10572     CGAGGTTTCTATAAGGGCCTG-GTGCCCTACCTGGTCCACGTCACGCCCAACATCTGCAT
Dana\GF12007     CGCGGATTCTATAAGGGCCTG-GTGCCCTACCTGGTTCACGTCACGCCCAATATCTGCAT
Dpse\GA20774     CGCGGTTTCTATAAGGGACTT-GTGCCGTACCTGGTCCACGTCACGCCCAACATCTGCAT
Dper\GL17458     CGCGGTTTCTATAAGGGACTT-GTGCCGTACCTGGTCCACGTCACGCCCAACATCTGCAT
Dwil\GK22863     CCCGGCTTCTATAAGGGATTGCAAGCCA-GTCTTGTGCGTGTGGT-CCCAGC--CTGTAT
Dgri\GH19865     CGTGGTTTCTATAAGGGACTC-GTGCCGTACTTGGTCCACGTAACGCCCAACATTTGCAT
Dmoj\GI19815     CGCGGTTTCTATAAGGGACTC-GTGCCGTACTTGGTCCACGTAACGCCCAACATTTGCAT
Dvir\GJ18531     CGCGGTTTCTATAAGGGACTA-GTGCCGTACCTGGTCCACGTAACGCCCAACATTTGCAT

Dmel\CG8026      G-----
Dsec\GM20618     GXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXCCTGCG-----TCTTTTCATTTG--
Dyak\GE22456     GXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXCCTGCG-----TCTTTTCATTTG--
Dere\GG10572     GXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXCCTGCG-----TCTTTTCATTTG--
Dana\GF12007     GXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXCGCGTCCAACGCGTCTTTTCATTTG--
Dpse\GA20774     GXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXCCAGAG-----TCTTTTAATTTG--
Dper\GL17458     GXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXCCAGAG-----TCTTTTAATTTG--
Dwil\GK22863     G-----
Dgri\GH19865     GXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXACGAGC-----TTTTTCATTTGCGC
Dmoj\GI19815     G-----
Dvir\GJ18531     GXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTAAGAG-----

```

(b)

E(10000): 1.5e-06; 57.6% identity (57.6% similar) in 774 nt overlap (144-849:10-695)

```

                110      120      130      140      150      160      170      180
Dwil\GK22863: TAGAGGAAGTCCAAATGCAGAATAGTATTAACCAGATGAATCTGCGGCAGAATTTAAGCACCTAGAAATACATGAATCTA
                ::::  ::  ::::  : :  : : ::::  :::  :
Dper\GL17458:      GTCATGCTGATCT--GGGAGAAGCTGACCAGCTAG-ATAGA----GTA

                        10      20      30      40

                190      200      210      220      230      240      250
Dwil\GK22863: TCTATCAATCAATTTATCTATAGTTTTCGACCCCTCTATTTTAAACATACAAAAAGAA-----AACATAGTA
                : : : :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :
Dper\GL17458: TTAACCA--GAACCTAGGATCTAGAATCTAGGA-----ATTATGATCTAGAATCTGGAACAGAATTTAAGCACCTAGAA
                50      60      70      80      90      100      110

```

```

                260          270          280          290          300          310          320
Dwil\GK22863: TA----AGTCTATC-AG----CAAGAAATG--ACAAAAGAAAACCAACAT--TATCTTATGAATTATTGTTACAAGTTC
                : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dper\GL17458: TACACGAATCTTTCTAGTTTCCTCCAATGTGCAGCAAAACAGCAACAAAATGACATGTTTGTATTATTGTTATAA-TTC
                120          130          140          150          160          170          180          190

                330          340          350          360          370          380          390
Dwil\GK22863: GATGGATTTGACTAGTTTGTGTTTAAATTATAATATTTAAATTAAGTGTATTCCATATATA-TGTGTAT-----
                : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dper\GL17458: GATTAA--GAACCACCT-----AAATACTAGCCCTAAGTTAAATCCATAACTAGTTTGTGTTTACGTTCTGCGC
                200          210          220          230          240          250

                400          410          420          430          440          450          460          470
Dwil\GK22863: ATCTAAAC-TTGAAAGTGATTGTTGTGATAGCTGCACTAGCAATTGCAAAAGTTATGAAATCTAATTTCAAAGATGCATT
                : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dper\GL17458: ATCTTCACGTTGAAAGTTATTGTTGTGAT-----GGTATG-----TAAATTCAAAGATGCATT
                260          270          280          290          300          310

                480          490          500          510          520          530          540
Dwil\GK22863: TCTT---CA-TTAACAAACGAAAACAACA-AAAACAAAAACACGCACACAACATTTAGTTTATAGTTTAGTTAAATTAT
                : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dper\GL17458: TCTTGCACACTTCAC---CG-AAGCAACACACAGCCACACACACACATTTAGTTCATAGTT--TAGTTAAATT-ATTTAT
                320          330          340          350          360          370          380

                550          560          570          580          590          600
Dwil\GK22863: TTATTTAAT--AATGAAATGT-TACACA----AAACAA-----TAAGAAAAGAAAAAATTTTCATTTTGGCAAAA
                : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dper\GL17458: TT-TGTAATGGATTGGAATGTAAACGCAGCTGAAACAGTTTCTAGTTTAGTTGAG-ACAAAATTTGTTTGGCAAAA
                390          400          410          420          430          440          450          460

                610          620          630          640          650          660          670          680
Dwil\GK22863: AGATTCTTCATAATTTTGCATATGGTTTTATAATTTT--TATAAACTTGACATACAATATTTATGTAAATAAGCATAAA
                : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dper\GL17458: AGATTCTTCATAGTTTCGCAAATGG-TTTATAATCTTAAGGATGAAATGAACCT-TGACACTTATGTAAATAAGCATAAA
                470          480          490          500          510          520          530          540

                690          700          710          720          730          740          750          760
Dwil\GK22863: AGTAAATCAAATCGCATTTGTGAA-CAAATACGAAAAACC--AAGCAAAAAA---CAAATGAAACAAGTGACAGCAAATTG
                : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dper\GL17458: A-----CGCATTTGTGAACCACCAACTAAGCTGCTGAAGCAGAACAATCCCAATCAAACCA--TACAACCAAGAG
                550          560          570          580          590          600

                770          780          790          800          810          820          830
Dwil\GK22863: AATCCTAACAAAAAGAAAAGAGAATTACTTA-ATTACAAT--TATGCTTAACTAT-TAAAATGCCAGGAAGGCAAAAAC
                : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dper\GL17458: AA---GAAAATACAAAAGTAGTTACTTACCTATATTAATGCCATG-ATGGCAATCCAAAAGGACACTAA--AAAAAAC
                610          620          630          640          650          660          670          680

                840          850          860          870          880          890          900          910
Dwil\GK22863: AAGAAA-GAAAGAAAAAATATCTTAAAGTCTCCGCTTGCAAGACGAGTCTTTTCATTACACCTTTACCAACAAAAAAC
                : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dper\GL17458: AACAACTGAAGAACGTCTCGGGTCCAA

```

Figure S3. An example of a global alignment of orthologous genes using artificial introns.
 Example of a global alignment of one group of orthologous genes. Thirty shaded X's represent an intron.

```

Dmel\CG7911      ATGGAGTCCGAGTCGTTTTATGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTGTCACA
Dsec\GM12196     ATGGAGTCCGAGTCGTTTTATGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTGTCACA
Dyak\GE23428     ATGGAGTCCGAGTCGTTTTATGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTGTCACA
Dere\GG11978     ATGGAGTCCGAGTCGTTTTATGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTGTCACA
Dana\GF16212     ATGGAATCCGAATCGTTTACGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTTTTACG
Dpse\GA20679     ATGGAGTCCGAATCTTTTTATGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTGTTACA
Dper\GL13489     ATGGAGTCCGAATCTTTTTATGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTGTTACA
Dwil\GK13375     ATGGAGACAGAATCTTTTTATGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTGTTACG
Dgri\GH22165     ATGGAGTCCGAATCGTTTACGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTGTTACG
Dmoj\GI21966     ATGGAGTCCGAGTCCTTTTACGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTGTTACC
Dvir\GJ14325     ATGGAGTCCGAATCTTTTTATGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTGTTACG

Dmel\CG7911      CTCAGCGAGAAGGAGGCCATCGCACAGTTCGAGGTCCCAGATGTACCCGAGGAGTACATC
Dsec\GM12196     CTCAGCGAGAAGGAGGCCATCGCACAGTTCGAGGTCCCAGATGTACCCGAGGAGTACATC
Dyak\GE23428     CTCAGCGAGAAGGAGGCCATCGCACAGTTCGAGGTCCCAGATGTACCCGAGGAGTATATC
Dere\GG11978     CTCAGCGAGAAGGAGGCCATCGCACAGTTCGAGGTCCCAGATGTACCCGAGGAGTATATC
Dana\GF16212     CTCAGCGAGAAGGAGGCCATCGCACAGTTCGAGGTCCCAGATGTACCCGAGGAGTATATC
Dpse\GA20679     CTCAGCGAGAAGGAGGCCATCGCACAGTTCGAGGTCCCAGATGTACCCGAGGAGTATATC
Dper\GL13489     CTCAGCGAGAAGGAGGCCATCGCACAGTTCGAGGTCCCAGATGTACCCGAGGAGTATATC
Dwil\GK13375     CTCAGCGAGAAGGAGGCCATCGCACAGTTCGAGGTCCCAGATGTACCCGAGGAGTATATC
Dgri\GH22165     CTCAGCGAGAAGGAGGCCATCGCACAGTTCGAGGTCCCAGATGTACCCGAGGAGTATATC
Dmoj\GI21966     CTCAGCGAGAAGGAGGCCATCGCACAGTTCGAGGTCCCAGATGTACCCGAGGAGTATATC
Dvir\GJ14325     CTCAGCGAGAAGGAGGCCATCGCACAGTTCGAGGTCCCAGATGTACCCGAGGAGTATATC

Dmel\CG7911      GTCCACTCGCACAAAGCTCATCATCAAACAGATTTCCTCGGCCAGAAAGCGAAGACCGGC
Dsec\GM12196     GTCCACTCGCACAAAGCTCATCATCAAACAGATTTCCTCGGCCAGAAAGCGAAGACCGGC
Dyak\GE23428     GTCCACTCGCACAAAGCTCATCATCAAACAGATTTCCTCGGCCAGAAAGCGAAGACCGGC
Dere\GG11978     GTCCACTCGCACAAAGCTCATCATCAAACAGATTTCCTCGGCCAGAAAGCGAAGACCGGC
Dana\GF16212     GTCCACTCGCACAAAGCTCATCATCAAACAGATTTCCTCGGCCAGAAAGCGAAGACCGGC
Dpse\GA20679     GTCCACTCGCACAAAGCTCATCATCAAACAGATTTCCTCGGCCAGAAAGCGAAGACCGGC
Dper\GL13489     GTCCACTCGCACAAAGCTCATCATCAAACAGATTTCCTCGGCCAGAAAGCGAAGACCGGC
Dwil\GK13375     GTCCACTCGCACAAAGCTCATCATCAAACAGATTTCCTCGGCCAGAAAGCGAAGACCGGC
Dgri\GH22165     GTCCACTCGCACAAAGCTCATCATCAAACAGATTTCCTCGGCCAGAAAGCGAAGACCGGC
Dmoj\GI21966     GTCCACTCGCACAAAGCTCATCATCAAACAGATTTCCTCGGCCAGAAAGCGAAGACCGGC
Dvir\GJ14325     GTCCACTCGCACAAAGCTCATCATCAAACAGATTTCCTCGGCCAGAAAGCGAAGACCGGC

Dmel\CG7911      GAATTCACGTTGTACAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGCGGAGACGAAC
Dsec\GM12196     GAATTCACGTTGTACAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGCGGAGACGAAC
Dyak\GE23428     GAATTCACGTTGTACAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGCGGAGACGAAC
Dere\GG11978     GAATTCACGTTGTACAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGCGGAGACGAAC
Dana\GF16212     GAATTCACGTTGTACAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGCGGAGACGAAC
Dpse\GA20679     GAATTTAATGTTGTACAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGCGGAGACGAAC
Dper\GL13489     GAATTTAATGTTGTACAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGCGGAGACGAAC
Dwil\GK13375     GAATTCATGTTGTACAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGCTGAAACGAAC
Dgri\GH22165     GAATTCAT-----XXXXXXXXXXXXXXXXXXXXXXXXXXXXGCGGAAACGAAC
Dmoj\GI21966     GAATTCACGTCGTTTCAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGCGGAAACGAAC
Dvir\GJ14325     GAATTCATGTTGTACAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGCGGAAACGAAC

Dmel\CG7911      ATAAACGACGATGGCGAGAAGAAACCTTGAAGATCCCCATTGCCGTGTTGAAGGTCGGC
Dsec\GM12196     ATAAACGACGATGGCGAGAAGAAACCTTGAAGATCCCCATTGCCGTGTTGAAGGTCGGC
Dyak\GE23428     ATAAACGACGATGGCGAGAAGAAACCTTGAAGATCCCCATTGCCGTGTTGAAGGTCGGC
Dere\GG11978     ATAAACGACGATGGCGAGAAGAAACCTTGAAGATCCCCATTGCCGTGTTGAAGGTCGGC
Dana\GF16212     ATAAACGACGATGG-----GGAGACGGTGAAGATTCCCATTGCCGTGTTGAAGGTCGGC
Dpse\GA20679     ATAAACGACGATGGTGAAGAAAGACTGTGAAGATTCCGATAGCAGTTCTGAAGGTCGGT
Dper\GL13489     GTGAACGATGATGGTGAAGAAAGACTGTGAAGATTCCGATAGCAGTTCTGAAGGTCGGT
Dwil\GK13375     GCTATTAACGAGGGTGAAGAAAGACTGTGAAGATTCCCATAGCCGTCTGAAGGTCGGT
Dgri\GH22165     GTGCACGATGACGGCGAGAAAGACAGTGAAGATTCCCATTGCCGTCTGAAGGTCGGT
Dmoj\GI21966     GTGAACGACGACGGCGAGAAGAAAGACAGTCAAAATTCCCATTGCTGTTCTGAAGGTCGGT
Dvir\GJ14325     GTGCACGACGACGGCGAGAAGAAAGACAGTCAAAATTCCCATTGCTGTTCTGAAGGTCGGT

```

```

Dmel\CG7911      GAGACCCGTAGCTTAAGACCAAAATGTCGAGTTCCCAATGGATCAGTGACCTTCAAACCTG
Dsec\GM12196     GAGACCCGTAGCTTAAGACCAAAATGTCGAGTTTCCCAACGGATCAGTGACCTTCAAACCTG
Dyak\GE23428     GAGACCCGTAGCTTAAGGCCAAATGTCGAGTTCCCAACGGATCAGTGACCTTCAAACCTG
Dere\GG11978     GAGACCCGTAGCTTAAGGCCAAATGTTGAATTTCCCAACGGATCCGTGACATTCAAGCTG
Dana\GF16212     GAAACCCGTAGCTTAAGGCCAAATGTTGAATTTCCCAACGGATCCGTGACATTCAAGCTG
Dpse\GA20679     GAAACGCGCAGCCTGAGACCAAAATGTGGAGTTCCCAATGGATCAGTCACTTTCAAACCTG
Dper\GL13489     GAAACGCGCAGCCTGAGACCAAAATGTGGAGTTCCCAATGGATCAGTCACTTTCAAACCTG
Dwil\GK13375     GAAACTCGTTGCTTGAAGCCAAATGTTGAGTTCCCAACGGATCCGTGACATTAAACTA
Dgri\GH22165     GAAACGCGCAGTTTGAGGCCAAATGTTGAGTTCCCAATGGCTCCGTGACATTCAAGCTC
Dmoj\GI21966     GAAACTCGCAGCTTGCGACCAAAATGTTGAATTTCCCAATGGATCCGTGACATTAAAGCTT
Dvir\GJ14325     GAAACTCGCAGTTTGAGGCCAAATGTTGAATTTCCCAATGGCTCCGTGACATTCAAACCTG

```

```

Dmel\CG7911      GTGCAGGGAAGTGGACCCGTACACGTATGCGGCAAGXXXXXXXXXXXXXXXXXXXXXXXXX
Dsec\GM12196     GTGCAGGGAAGTGGGCCCGTACACGTCTGCGGCAAGXXXXXXXXXXXXXXXXXXXXXXXXX
Dyak\GE23428     GTGCAGGGAAGTGGACCCGTCTACGTCTGCGGCAAGXXXXXXXXXXXXXXXXXXXXXXXXX
Dere\GG11978     GTGCAGGGAAGTGGACCCGTCTACGTCTGCGGCAAGXXXXXXXXXXXXXXXXXXXXXXXXX
Dana\GF16212     GTGCAGGGAAGTGGACCCGTCTATGTCTGTGGCAAGXXXXXXXXXXXXXXXXXXXXXXXXX
Dpse\GA20679     GTGCAGGGAAGTGGGCCCGTCTATGTCTGCGGCAAGXXXXXXXXXXXXXXXXXXXXXXXXX
Dper\GL13489     GTGCAGGGAAGTGGGCCCGTCTATGTCTGCGGCAAGXXXXXXXXXXXXXXXXXXXXXXXXX
Dwil\GK13375     ATCCAAGGCTCCGGTCCAGTTTATGTGTGCGGCAAG-----
Dgri\GH22165     GTCCAGGGCACAGGGCCCGTCTATGTTGTGGAAGXXXXXXXXXXXXXXXXXXXXXXXXX
Dmoj\GI21966     GTCCAGGGCACAGGGCCCGTCTATGTGTGCGGCAAGXXXXXXXXXXXXXXXXXXXXXXXXX
Dvir\GJ14325     GTCCAGGGCACAGGGCCCGTCTATGTCTGCGGCAAGXXXXXXXXXXXXXXXXXXXXXXXXX

```

```

Dmel\CG7911      XXXXXXGTGCAGATGAACTTTGGCGAGTTTGACGACGGT-----CAGATTTACGAGGAG
Dsec\GM12196     XXXXXXGTGCAGATGAACTTTGGCGAGTTTGACGACGGT-----CAGATTTACGAGGAG
Dyak\GE23428     XXXXXXGTGCAGATGAACTTTGGCGAGTTTGACGACGGT-----CAGATTTACGAGGAG
Dere\GG11978     XXXXXXGTGGAGATGAACTTTGGCGAGTTTGACGACGGT-----CAGATTTACGAGGAG
Dana\GF16212     XXXXXXACGGAGATGAACTTTCGGAGAGTTCAACGATGGT-----GAAATTTACGAAGAG
Dpse\GA20679     XXXXXXGCCGAGATGAACTTTGGCGAGTTTCGACGACGGG-----CAGATGTATGAGGAG
Dper\GL13489     XXXXXXGCCGAGATGAACTTTGGCGAGTTTCGACGACGGG-----CAGATGTATGAGGAG
Dwil\GK13375     -----GCGGAAATGAATTTTCGGCGAGTTTCGATGATGGTGATGGTCAAATGTATGAGGAT
Dgri\GH22165     XXXXXXGCTGAAATGAACTTGGGTGAATATGACGATGGT-----CAGATGTATGAGGAG
Dmoj\GI21966     XXXXXXGCGGAAATGAACTTGGGTGAATATGACGATGGT-----CAAATGTACGAGGAG
Dvir\GJ14325     XXXXXXGCGGAGATGAACTTGGGTGAATACGACGACGGG-----CAAATGTACGAGGAG

```

```

Dmel\CG7911      TATTTCGGACGAGGAGGAGGATAGCGAACTGGAATTC-GACGA-AGAAGCAGCTCCTCAGA
Dsec\GM12196     TATTTCGGACGAGGAGGAGGATAGCGAACTGGAATTC-GACGA-AGAAGCAGCTCCTCAGA
Dyak\GE23428     TATTCTGACGAGGAGGAGGATAGCGAACTGGAATTC-GACGA-AGAAGCAGCTCCTCAGA
Dere\GG11978     TATTTCGGACGAGGAGGAGGATAGCGAACTGGAATTC-GACGA-AGATGCAGCTCCTCAGA
Dana\GF16212     TATTCTGATGAGGAGGATGACAGCGAATTTGGAATTT-GACGA-GGATGCAGCTCCACAGA
Dpse\GA20679     TACTCCGACGACGAAGAGGACAGCGA-TGGGGACTTCGACGA-GGACGCTGTGCCGCAGA
Dper\GL13489     TACTCCGACGACGAAGAGGACAGCGA-TGGGGACTTCGACGA-GGACGCTGTGCCGCAGA
Dwil\GK13375     TATTTCAGATGAAGAGGAGGAGGA-----GGAGTAC-GACGA-GG-----TGCCGCAGA
Dgri\GH22165     TATTTCGGATGAGGACGATGACAGCGA-T--GAGCTC-GATGATGAGAGTT-CGCCACAGA
Dmoj\GI21966     TATTTCGGATGATGAAGATGACAGCGAGCTGGAGGCT-GAGGACGATTGCT-CTCCACAAA
Dvir\GJ14325     TATTTCGGACGACGAGGATGACAGTGAATTTGGAGCTT-GATGACGATTGTT-CACCACAGA

```

```

Dmel\CG7911      CGAACGGCAAGAGTAA---TAAGAAGAAGTAG
Dsec\GM12196     CGAACGGCAAGAGTAA---TAAGAAGAAGTAG
Dyak\GE23428     CGAACGGCAAGAGTAA---TAAGAAGAAGTAG
Dere\GG11978     CGAACGGCAAGAGTAA---TAAGAAGAAGTAG
Dana\GF16212     CGAATGGAAAGAGCAA---TAAAAAGAAGTAA
Dpse\GA20679     CAAACGGCAAGAGCAAAATTACAAGAAGTAA
Dper\GL13489     CAAACGGCAAGAGCAAAATTACAAGAAGTAA
Dwil\GK13375     CGAACGGCAAGAGCAA---TAAAAAGAAGTAA
Dgri\GH22165     CAAATGGCAAGAGCAT---AAAAAGAAGTAA
Dmoj\GI21966     CAAATGGAAAGAGCAA---TAAAAAGAAGTAA
Dvir\GJ14325     CAAATGGCAAGAGCAT---AAAAAGAAGTAA

```

Figure S4. An example of an intron excluded during manual inspection.

This intron was excluded from further analyses due to low sequence identity surrounding the intron-exon junctions in the global alignment. This was done as a precautionary measure to exclude potential false-positives. Thirty shaded X's represent an intron. This is a fragment from the global alignment of these orthologous genes; only the region surrounding the excluded intron is shown.

```

Dmel\CG11660      CCTGGCACAAG---TTATAGCAACCGCCAACGAAGAGGA---TGAACCGGAGGACGA--A
Dsec\GM24726      CCTGGCCAAG---TTATAGCAACCGCCAACGAAGAAGAGGATGAGCCGGAGGACGA--A
Dyak\GE20191      CCTGGCACAAG---TTATAGAAACCGCCAACGAAGAGGA---TGAACCGGAGGACGA--A
Dere\GG13901      CCTGGCACAAG---TTATAGAAACCAACCAACG-----CAGAGGATGA--A
Dana\GF24528      CCTCGTTCAAG---TCCTTGAAGCCAACAGCGATGGTGAGGATGCAGATGAGGACGAC-A
Dpse\GA11126      ----GC-----ACTGCCAACAGAAGGA----AAAGTAGAAGCTGGTGA
Dper\GL16346      ----GC-----AATXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-----
Dwil\GK10469      GATGGCTTCAAGTGCTTTAGAGGCTGTCAACGGACATGA-----GGAGGATGG---
Dgri\GH16627      -TCTCTGCCAA---C-----AACTACCGA-----TCCTGGCGAGGAATCTG---
Dmoj\GI11584      ACTTA-GCCAG---A-----AAGTTATAGAGAGTGCAAT-TTCCGGCGAAGAAACGG---
Dvir\GJ11263      -CTTGTGCCAG---C-----AAATTATTGATAACACCAA-TCCGGCCGAGGAAACTG---

Dmel\CG11660      GAGCCAGAG----GAATAC-GAT---GAAGACGATTACGAC-G-ATATCGGA-GATGACT
Dsec\GM24726      GAGCCAGAG----GAATAC-GAT---GAAGAGGACTACGAC-G-ATGTCGGA-GATGACT
Dyak\GE20191      GAGCCCGAG----GAATAC-GAT---GAAGAGGACTACGAC-G-ATGTCGGA-GATGACT
Dere\GG13901      GAGGTCGAG----GAATAC-GAT---GAAGAGGACTACGAC-G-ATGTCGGA-GATGACT
Dana\GF24528      ATGTCGAAG-----GCTAT-GACTACGAGGAGGATGACTATGA-GGAT--GGCGATGACT
Dpse\GA11126      GGATCTGGATTACGACTAT-GAT---GAAGAAGACGACGACTACGCGCTGGGCGATGAAT
Dper\GL16346      -----AT---GAAGAAGACCACGACTACGCACTGGGCGATGAAT
Dwil\GK10469      AGATGTGGA-----TTAT-GAT---GATGACGATGAAGACTACGATTTAAGCGATGGCT
Dgri\GH16627      --CCTCAGACT----ATGCAGATGATGATGAAGATGACGACTACGATCTGGGCGATAACT
Dmoj\GI11584      --ATTCAGACT----ATGCCGATGAAGATGAAGACGATGCCTTTGATGTTGGTGACGACT
Dvir\GJ11263      --ATTCGACT----ATGCGGATGAGTATGAAGACGACGATTACGACCAGGGCGACGACT

```

Figure S5. An example of an intron gain that was identified in a single species. An intron gain occurred in Dana\GF17707. This is the third intron in this gene. Thirty shaded X's represent an intron. This is a fragment from the global alignment of these orthologous genes; only the region surrounding the gained intron is shown.

Dmel\Hdac3	ACGATACTACAGCGTTAATGTGCCGCTAAAGGAGGGCATTGATGACCAGAGCTATTTTCA
Dsec\GM10631	ACGATACTACAGCGTTAATGTGCCGCTAAAGGAGGGCATTGATGATCAGAGCTATTTCCA
Dyak\GE25277	ACGATACTACAGCGTGAATGTCCCGCTAAAGGAGGGCATTGATGACCAGAGCTATTTCCA
Dere\GG10995	ACGATACTACAGCGTGAATGTGCCGCTAAAGGAGGGCATTGATGACCAGAGCTATTTCCA
Dana\GF17707	CCGCTACTACAGCGTGACGTCCCTCTTAAAGAAGGAATCGATGATCAGAGCTACTTTCCA
Dpse\GA27397	TCGATACTACAGCGTGAATGTGCCACTCAAGGAGGGCATTGACGATCAGAGCTACTTTCCA
Dper\GL22163	TCGATACTACAGCGTGAATGTGCCACTCAAGGAGGGCATTGACGATCAGAGCTACTTTCCA
Dwil\GK11850	TCGCTATTATAGTGTGAATGTTCGCTTAAAGAGGGAATCGATGATCAGAGCTACTTTCA
Dgri\GH18996	GCGATATTATAGCGTCAATGTGCCGCTTAAAGGAGGCATCGATGATCAGAGCTATTTTCA
Dmoj\GI24497	TCGTTATTACAGCGTTAATGTTCCTCTCAAAGAGGGAATTGATGATCAGAGCTATTTTCA
Dvir\GJ24563	CCGGTATTATAGCGTTAATGTGCCTCTCAAGGAGGTATTGACGACCAGAGCTACTTTCA
Dmel\Hdac3	G-----GTGTTCAAACCCATAATTTCTGCCATTAT
Dsec\GM10631	G-----GTGTTCAAACCCATCATTTTCGGCCATTAT
Dyak\GE25277	G-----GTGTTTAAACCAATCATATCGGCCATTAT
Dere\GG10995	G-----GTGTTCAAACCAATCATATCGGCCATTAT
Dana\GF17707	GXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTGTTTAAACCCATTATTTTCGGCCATCAT
Dpse\GA27397	G-----GTCTTCAAACCCATCATCTCGGCCATCAT
Dper\GL22163	G-----GTCTTCAAACCCATCATCTCGGCCATCAT
Dwil\GK11850	G-----GTTTTTAAGCCCATCATTTTCGGCCATAAT
Dgri\GH18996	A-----GTCTTCAAACCGATCGTTTCAGCAATTAT
Dmoj\GI24497	A-----GTGTTCAAACCAATAATTTTCAGCGATTAT
Dvir\GJ24563	A-----GTGTTCAAGCCGATCATTTCTGCGATTAT

Figure S6. An example of an intron gain that occurred at a node. An intron gain that likely occurred in the ancestor of *D. pse* and *D. per*. This is the first intron in both Dpse\GA19454 and Dper\GL25213. Thirty shaded X's represent an intron. This is a fragment from the global alignment of these orthologous genes; only the region surrounding the gained intron is shown.

```

Dmel\dbo          GTCCAGACCCCTCCTTCCCGCCGCCTGCCTGCTTCAACTGGTTGAGATTCAG-----
Dsec\GM25560      GTCCAGACCCCTCCTTACCCGCCGCCTGCCTGCTTCAACTGGTTGAGATTCAG-----
Dyak\GE22281      GTCCAGACCCCTCCTTCCCGCCGCCTGCCTGCTTCAACTGGTTGAGATTCAG-----
Dere\GG15931      GTCCAGACCCCTCCTTCCCGCCGCCTGCCTGCTTCAACTGGTTGAGATTCAG-----
Dana\GF24095      GTCCAGACTCTTCTGCCGCCGCCTGCCTCCTCCAGTTGGTCGAGATCCAG-----
Dpse\GA19454      GTGCAGACTCTGCTGCCAGCCGCCTGCCTGCTTCAGCTGGTGGAGATCCAGXXXXXXXXXX
Dper\GL25213      GTGCAGACTCTGCTGCCAGCCGCCTGCCTGCTTCAGCTGGTGGAGATCCAGXXXXXXXXXX
Dwil\GK15757      GTTCAGACTCTGTTGCCAGCCGCCTGCCTGCTACAGCTTGTGAGATTCAG-----
Dgri\GH17090      GTGCAAACTGCTGCCAGCCGCCTGCCTGCTGCAGTTGGTCGAAATTCAG-----
Dmoj\GI11691      GTACAGACGCTGCTGCCGCCGCCTGCCTGCTACAGCTCGTAGAGATTCAG-----
Dvir\GJ11367      GTCCAAACGCTGCTACCAGCCGCCTGCCTCCTGCAGCTCGTCGAAATACAG-----

Dmel\dbo          -----GACATCTGCTGCGAGTTCCTCAAACGGCAATTGGATCCC
Dsec\GM25560      -----GACATCTGCTGCGAGTTCCTCAAACGGCAGTTGGATCCC
Dyak\GE22281      -----GACATCTGCTGCGAGTTCCTCAAACGGCAGTTGGATCCC
Dere\GG15931      -----GACATCTGCTGCGAGTTCCTCAAACGGCAGTTGGATCCC
Dana\GF24095      -----GACATATGCTGTGAGTTCTCAAGCGCAACTGGATCCC
Dpse\GA19454      XXXXXXXXXXXXXXXXXXXXXXXGACATTTGCTGCGAGTTCCTCAAGCGACAACGGACCCG
Dper\GL25213      XXXXXXXXXXXXXXXXXXXXXXXGACATTTGCTGCGAGTTCCTCAAGCGCAACTGGACCCG
Dwil\GK15757      -----GACATTTGCTGTGAGTTCTCAAACGTCAGCTGGATCCG
Dgri\GH17090      -----GATATATGCTGTGAGTTCTCAAGCGCCAGCTCGACCCG
Dmoj\GI11691      -----GATATTTGCTGCGAGTTCTCAAGCGCAACTCGATCCG
Dvir\GJ11367      -----GACATATGCTGTGAGTTCTCAAGCGTCAACTCGATCCG

```


Figure S7. An example of an intron loss that was identified in a single species. An intron loss that occurred in Dwil\GK22366. Thirty shaded X's represent an intron. This is a fragment from the global alignment of these orthologous genes; only the region surrounding the lost intron is shown.

```

Dmel\CG4770      TCGGCAAGGCGAAATTGGGAAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXCTTGAGCA
Dsec\GM26874     TCGGGAAGGCGAAATTGGGAAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXCTTGAGCA
Dyak\GE25635     TCGGCAAGGCTAAATTAGGGAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXCTTAAGCA
Dere\GG23603     TCGGCAGGGCCAAATTGGGAAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXCTTGAGCA
Dana\GF17628     TCGGCAAGCAAAGTTAGGAAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXATTAAGCA
Dpse\GA18419     TCAACAAGGCAAAGTTGGGCAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXTCTGAGCA
Dper\GL12080     TCAACAAGGCAAAGTTGGGCAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXTCTGAGCA
Dwil\GK22366     TCCACAAGTCTCGATTGGGCAG-----CATTAGCA
Dgri\GH19617     TGCACAAGGCGCAGCTGGGCAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXCATCAGCA
Dmoj\GI22436     TGCACAAGGCGGAGCTGGGCAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXCATCAGCG
Dvir\GJ10999     TGCACAAGGCAGAGCTGGGCAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXCATCAGCA

```

Figure S8. An example of an intron loss that occurred at a node. An intron loss that likely occurred in the ancestor of *D. ere*, *D. yak*, *D. sec*, and *D. mel* but not *D. ana*. Thirty shaded X's represent an intron. This is a fragment of the global alignment of these orthologous genes; only the region surrounding the lost intron is shown.

Dmel\CstF-50	CTGATGATCAG-----CCAGCTGATGTACGACGGC
Dsec\GM16423	CTGATGATCAG-----CCAGCTGATGTACGATGGC
Dyak\GE23306	CTAATGATCAG-----CCAGCTGATGTACGACGGG
Dere\GG11859	CTGATGATCAG-----CCAGCTGATGTACGACGGG
Dana\GF23256	CTGATGATCAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXXCCAACTGATGTACGATGGC
Dpse\GA15331	CTGATGATAAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXXCCAGCTGATGTACGATGGC
Dper\GL23528	CTGATGATAAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXXCCAGCTGATGTACGATGGC
Dwil\GK10973	TTGATGATAAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXXTCAGTTGATGTACGATGGA
Dgri\GH18935	TTAATGATAAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXXCCAGTTGATGTACGATGGC
Dmoj\GI24451	TTAATGATAAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXXCCAATTAATGTACGATGGC
Dvir\GJ24120	TTAATGATAAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXXCCAGTTAATGTACGATGGC

Figure S9. Alignments that required the distant outlier *A. gambiae* to determine the ancestral state of the intron with Dollo parsimony. Thirty shaded X's represent an intron. These are fragments of the global alignment of these two groups of orthologous genes; only the region surrounding the intron in question is shown. (a) An example that required the ancestral state of the gene at this position, in this case intronless, to be known to infer that there was an intron gain in the ancestor of *D. gri*, *D. moj* and *D. vir*. (b) An example that required the ancestral state of the gene at this position, this case intron present, to be known to infer that there was an intron loss in the ancestor of *D. gri*, *D. moj* and *D. vir*.

(a)

```

Dmel\CG1332      TTCACCTTTGTGGACTTGAGCCCGCCAACAGAGGATGGTTCCAAATACGATTTTCAG----
Dsec\GM14636     TTCACCTTTGTGGACTTGAGTCCGCCAACGGAGGATGGTTCCAAATACGATTTTCAG----
Dyak\GE21425     TTCACGTTTGTGGACTTGAGTCCGCCAACGGAAGATGGTTCCAAATACGACTTCAG----
Dere\GG15207     TTCACCTTTGTGGACTTGAGTCCGCCAACGGAGGATGGTTCCAAATACGACTTCAG----
Dana\GF10565     TTCACCTTCGTGGACCTGAGTCTCCGACGGAGGATGGTTCCAAGCAGATTTCCG----
Dpse\GA12201     TTCACCTTCGTGGACCTGAGCCCGCCCACTGAGGATGGGTCCAAGCAGACTTCAG----
Dper\GL12837     TTCACCTTCGTGGACCTGAGCCCGCCCACTGAGGATGGGTCCAAGCAGACTTCAG----
Dwil\GK12734     TTCATTTTGTGTTGATCTAAGTCCACTCTCGGAGGACTCATCGAAGCATGATTTTCGT---
Dgri\GH15934     TTTACCTTTGTGGATCTGAGTCCAGCCTCCGAGGATGGCTCCAAGCATGATTTTCAGXXXX
Dmoj\GI16583     TTCACGTTTCGTGGACATGAGTCTCCAATGGAAGATGGCTCCAAGCATGATTTTAGXXXX
Dvir\GJ12835     TTCACCTTCGTGATCTGAGTCCGCCCTGGAGGATGGCTCCAAGCAGATTTTCAGXXXX
Agam\558841.2    CTCATGTTTCACGAAATAGCATCGCTGAAGGATGACGCCACGCGCAGACTTTAA----

Dmel\CG1332      -----GTTCTCCACCAACAAGATACGTGAGGTTTCATCTG
Dsec\GM14636     -----GTTCTCCACCAACAAGATACGTGAGGTTTCATCTG
Dyak\GE21425     -----GTTCTCCACCAACAAGATACGTGAGGTTTCATCTG
Dere\GG15207     -----GTTCTCCACCAACAAGATACGTGAGGTTTCATCTG
Dana\GF10565     -----ATTCTCCACGAACAAGATACGGGAAGTGCATCTC
Dpse\GA12201     -----GTTCTCCACCAACAAAATACGCGAGGTCCATCTG
Dper\GL12837     -----GTTCTCCACCAACAAAATACGCGAGGTCCATCTG
Dwil\GK12734     -----TTTCACATCAATAAAATTCGAGAAGTTCATTG
Dgri\GH15934     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXATTTTCAATCAACAAAATTCGGGAGGTTCATCTG
Dmoj\GI16583     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTTTTCATTAATAAAATACGTGAAGTTCATCTG
Dvir\GJ12835     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTTTTCATCAACAAAATCCGTGAGGTTCATCTG
Agam\558841.2    -----GTTTCCGATCGGTCAACTGCGCAGCTGCACCTG

```

(b)

```

Dmel\rtet        ACGGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTGTAATCATGACTTTGCTGCAGGGA
Dsec\GM15064     ACGGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTGTAATCATGACTTTGCTGCAGGGA
Dyak\GE25706     ACAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTGTAATAATGACTTTGCTGCAGGGA
Dere\GG24354     ACAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTGTAATAATGACGTTGCTGCAGGGA
Dana\GF16496     ACAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTATTATCATGACTCTGCTACAGGGA
Dpse\GA19110     ACAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTGTTATTATGACTTTGATTTCAGGGA
Dper\GL21785     ACAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTGTTATTATGACTTTGATTTCAGGGA
Dwil\GK14183     ACAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTATTATTATGACCATACTTCAAGGA
Dgri\GH18255     ACGG-----GAGTTGTGATGACTTTGCTGCAGGGT
Dmoj\GI23654     ACAG-----GGCTCGTTATGACTCTGCTACAGGGA
Dvir\GJ23498     ACGG-----GCCTCGTCATGACCTGCTGCAGGGT
Agam\321745.3    ACAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTGTACTGATGGCGCTACTGCAAGGT

```

References

1. Zdobnov EM, et al.: **Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*.** *Science* 2002, **298**:149-159.