

GENETICS

Supporting Information

<http://www.genetics.org/cgi/content/full/genetics.109.113522/DC1>

Convergently Recruited Nuclear Transport Retrogenes are Male Biased in Expression and Evolving Under Positive Selection in *Drosophila*

Charles Tracy, Javier Río, Mansi Motiwale, Shawn M. Christensen and Esther Betrán

Copyright © 2010 by the Genetics Society of America
DOI: 10.1534/genetics.109.113522

FILE S1

Alignments of *Ntf-2* and *ran* proteins and *Drosophila* retroduplicates highlighting amino acid residues of known function

A. Alignments of Dntf-2 retrogenes with their parental and with rat Ntf2. Da_Ntf-2r is shown as Dntf-2_Retana in the alignment and Dg_Ntf-2r is shown as Dntf-2_Retgrim in the alignment.

D. melanogaster

```
Dntf2_D.sim      MSLNPOYEETIGKGFVQQYYAIFDDPANRANRVNPFYSATDSFMTFFHGHQIQ 50
Dntf2_D.sec     MSLNPOYEETIGKGFVQQYYAIFDDPANRANRVNPFYSATDSFMTFFHGHQIQ 50
Dntf2_PA_D.mel  MSLNPOYEDIGKGFVQQYYAIFDDPANRANRVNPFYSATDSFMTFFHGHQIQ 50
Dntf2r_D.mel    MSLNLOYEDIGKEFVQQYYAIFDDPANRENVINEYNATDSFMTFFHGNQIQ 50
Dntf2r_D.sim    MSLNPOYEETIGKGFVQQYYAIFDDPVNRENVNPFYSATDSFMTFFHGRQIQ 50
Dntf2r_D.sec    MSLNPOYEETIGKGFVQQYYAILDDLANRENAVNPFYSVTSDFMTFFHGHQIQ 50
Ntf2_R.norvegicus MGDKPIWEQIGSSFIQHYYQLFDN--DRTQLGAIY-IDASCLTWFHGGQEQ 47
                * . : *:* . *:*:* :*: :* : * * :*:*:*:
```

```
Dntf2_D.sim      GAPKILEKVQSISEFQIITRVITVDSQPTFDGGVLINVLGRIQCCLIDPEH 100
Dntf2_D.sec     GAPKILEKVQSISEFQIITRVITVDSQPTFDGGVLINVLGRIQCCLIDPEH 100
Dntf2_PA_D.mel  GAPKILEKVQSISEFQIITRVITVDSQPTFDGGVLINVLGRIQCCLIDPEH 100
Dntf2r_D.mel    GAPKILEKVQSISEFQIITRVITVDSQPTSDGGVLIIVLGRICCLIDPEH 100
Dntf2r_D.sim    GAPKILEKVQSISEFQIITRVITVDSQPTFDGGVLIIVLGRICCLIDPEH 100
Dntf2r_D.sec    GAPKILEKVQSISEFQIITRVITVDSQPTFDGGVLIIVLGRICCLIDPEH 100
Ntf2_R.norvegicus GKAAIVEKLSISEFQIITRVITVDSQPTFDGGVLIIVLGRICCLIDPEH 97
                * . *:*:*:*:* *:* * * * * * . : : *:*:*:*:*:
```

```
Dntf2_D.sim      AFSQVFFLKANAGTEFFVAHDIEFRLNIHNSA 130
Dntf2_D.sec     AFSQVFLKANAGTEFFVAHDIEFRLNIHNSA 130
Dntf2_PA_D.mel  AFSQVFFLKANAGTEFFVAHDIEFRLNIHNSA 130
Dntf2r_D.mel    AFSQIFLLKPNNGSFLVAHDIEFRLNIHNSA 130
Dntf2r_D.sim    SFSQIFLLKPNNGSFLVAHDIEFRLNIHNSA 130
Dntf2r_D.sec    SFSQIFLLKPNNGSFLVAHDIEFRLNIHNSA 130
Ntf2_R.norvegicus GFHQMPELLKNINDAVVCTNDMEFLALHNF 127
                . * * : * * * * . : : . : * : * * * : * * .
```

Red box - Interacting interface with RanGDP.

Orange - Amino acids interacting with FxFG repeats of nucleoporins.

D. ananassae

```
Dntf2_Retana    MPLNPHYEPMGQEFVRCQYVIFDNPATRALTATFFSHNDSEMTFFHGHQVLYGKIFEKVK 60
Dntf2_D.ana     MSLNPOYEDIGKGFVQQYYAIFDDPANRANRVNPFYSATDSFMTFFHGHQIQGAPKILEKVQ 60
Ntf2_R.norvegicus MGDKPIWEQIGSSFIQHYYQLFDN--DRTQLGAIY-IDASCLTWFHGGQEQGKAAIVEKLS 57
                * . * : * : * . * : * * * * * * * : * : * : * : * * . * * * .
```

```
Dntf2_Retana    SISEFQIITRVITVDSQPTFDGGVLINVLGRIQCCLIDPEHSAFSQVFLKANAG--TEFVA 118
Dntf2_D.ana     SISEFQIITRVITVDSQPTFDGGVLINVLGRIQCCLIDPEHSAFSQVFLKANAG--TEFVA 118
Ntf2_R.norvegicus SISEFQIITRVITVDSQPTFDGGVLINVLGRIQCCLIDPEHSAFSQVFLKANAG--TEFVA 115
                ** : * * * * . : * * * * * . : : * : * : * : * * * * * * * * * * . : : :
```

```
Dntf2_Retana    HDIEFRLNIHDTE 132
Dntf2_D.ana     HDIEFRLNIHNSA 130
Ntf2_R.norvegicus NDMFRLALHNF 127
                * : * * * : * :
```

D. grimshawi

```
Dntf2_Retgrim   MAINPOYEA VGKGFVQQYYAIFDDPANRANRVNPFYSTDSFMTFFHGHQIQGAPKILEKVQ 60
Dntf2_D.grim    MALNPOYEDIGKGFVQQYYAIFDDPANRANRVNPFYSATDSFMTFFHGHQIQGAPKILEKVQ 60
Ntf2_R.norvegicus MGDKPIWEQIGSSFIQHYYQLFDN--DRTQLGAIY-IDASCLTWFHGGQEQGKAAIVEKLS 57
                * . * : * : * . * : * * * * * * * : * : * : * : * * . * * * .
```

```
Dntf2_Retgrim   SISEFQIITRVITVDSQPTFDGGVLINVLGRIQCCLIDPEHSAFSQVFLKANAGSFFVAHD 120
Dntf2_D.grim    SISEFQIITRVITVDSQPTFDGGVLINVLGRIQCCLIDPEHSAFSQVFLKANAGSFFVAHD 120
Ntf2_R.norvegicus SISEFQIITRVITVDSQPTFDGGVLINVLGRIQCCLIDPEHSAFSQVFLKANAGSFFVAHD 117
                ** : * * * * . : * * * * * . : : * : * : * : * * * * * * * * * * . : : :
```

```
Dntf2_Retgrim   IEFRLNIHNSA 130
Dntf2_D.grim    IEFRLNIHNSA 130
Ntf2_R.norvegicus MEFRLALHNF 127
                : * * * : * * .
```


FILE S2**Functional inferences from the analyses of *Dntf-2* and *ran* retrogene protein sequences**

Heterospecific complexes formed between rat NTF-2 protein and canine Ran protein have been crystallized, and their mutual interactions determined (BERMAN *et al.* 2002). Additional analysis and crystal structures reveal the interactions of NTF-2 and Ran with other nuclear transport proteins (BULLOCK *et al.* 1996; ISGRO and SCHULTEN 2007; MARCHLER-BAUER *et al.* 2007; MATSUURA and STEWART 2004; RENAULT *et al.* 2001; SEEWALD *et al.* 2002; STEWART *et al.* 1998; VETTER *et al.* 1999). We threaded the protein sequences of the *Dntf-2* and *ran* retrogenes onto the structures of the crystallized paralogs in an effort to infer structural conflicts that might prohibit known parental protein-protein interactions. These analyses were performed using PyMOL software (<http://www.pymol.org>).

Ntf-2 and *ran* proteins are well known and conserved proteins (QUIMBY *et al.* 2000) that physically interact with each other and play a central role in the transport of proteins to the nucleus (RIBBECK *et al.* 1998). Ran exists in GDP bound inactive form and

GTP bound active form. RanGDP predominantly localizes in the cytoplasm and RanGTP in the nucleus. NTF-2 is a dimer that interacts with nucleoporins and RanGDP during transport to the nucleus (QUIMBY *et al.* 2000). RanGDP also interacts with RanGEF in the nucleus during RanGDP to RanGTP conversion (ISGRO and SCHULTEN 2007), with exportins in the nucleus in order to transport complexes out of the nucleus (KUSANO *et al.* 2003; MATSUURA and STEWART 2004), with Importin β during Importin β 's return trip to the cytoplasm (ISGRO and SCHULTEN 2007), and with RanGAP in the cytoplasm during RanGTP hydrolysis to RanGDP (KUSANO *et al.* 2002). Together, these proteins maintain a gradient of RanGDP/RanGTP that is important for protein import and export. Details about the particular residues of NTF-2 and Ran that are known to be involved in protein-protein interactions are given in Figure 2 and File S1.

We have threaded proteins from *Dntf-2*, *Dntf-2* derived retrogenes, *ran*, and *ran* derived retrogenes onto rat NTF-2 and canine Ran, respectively, to identify possible changes in function in the newly duplicated proteins (see Materials and Methods). We know that *Dntf-2* is a gene under purifying selection in *Drosophila* ($K_A/K_S = 0.0188$; see above), but *ran* is under stronger purifying selection ($K_A/K_S = 0.0065$; see above), likely due to the fact that it carries multiple functions, as discussed above. The selection imposed on Ran relative to DNTF-2 is evident in the alignments of these *Drosophila* proteins with mammalian orthologs (Figure 2 and File S1).

By threading the *Dntf-2* retrogene proteins (and DNTF-2) from the three *Drosophila* lineages onto a known NTF-2 crystal structure, we observed that the amino acids involved in interacting with RanGDP and the FxFG repeats of the nucleoporins are conserved or have changes that do not seem to structurally exclude binding to RanGDP or to nucleoporins (File S2, panel A). Even for Da_NTF-2R, the most divergent retrogene protein, the residues important for function are conserved or do not seem to impose overt structural conflicts, suggesting that the retrogenes of *Dntf-2* encode transport competent proteins capable of carrying RanGDP across the nuclear membrane.

Similar threading of the *Drosophila ran* retrogene proteins (and Ran) onto the known Ran crystal structure (Figure 2 and File S1, panel B) showed that the proteins encoded by the retrogenes in *D. ananassae* and *D. grimshawi* lineages show little divergence from their parental genes. Most of the amino acids that are important for function are identical or underwent conservative changes that do not appear to impose structural conflicts, suggesting similar functions between Ran and its

retrogenes in these lineages. While *Da_ran-like* and *Dg_ran-like* proteins are similar to the parental proteins, this similarity is not due to a recent origin as the *KS* is saturated between parental and retrogenes. The *KS* is 3.3193 for *Da_ran-like* and is 1.5144 for *Dg_ran-like* calculated using PAML. Accordingly, the *KA/KS* ratios are 0.0408 for *Da_ran-like* and 0.0348 for *Dg_ran-like* (Table S4).

The Ran-like in the *D. melanogaster* subgroup is more diverged, and our threading analyses indicate that some changes are likely to abolish some protein-protein interaction interfaces while possibly retaining others (Figure 2). The interface with

DNTF-2 appears to be present, as most of the interacting amino acids are identical or show conservative changes. We also posit that Ran-like may still interact with RanGAP, although potentially at a reduced level, as several residues within—and proximal to—known RanGAP interacting residues have suffered non-conservative mutations (e.g., amino acids 91-94, which are likely under positive selection, and amino acid 128 which is not suggested to be under positive selection) (Figure 2). Mapping these mutations onto co-crystal structures of Ran/RanGAP indicate that there are no major steric or charge conflicts. Further, RanGAP has been shown to produce duplicate genes (e.g., *Sd* (KUSANO *et al.* 2003) and to be under positive selection itself (PRESGRAVES 2007)), and interactions with a changing RanGAP or its duplicates could also explain the observed changes in Ran-like's RanGAP interface.

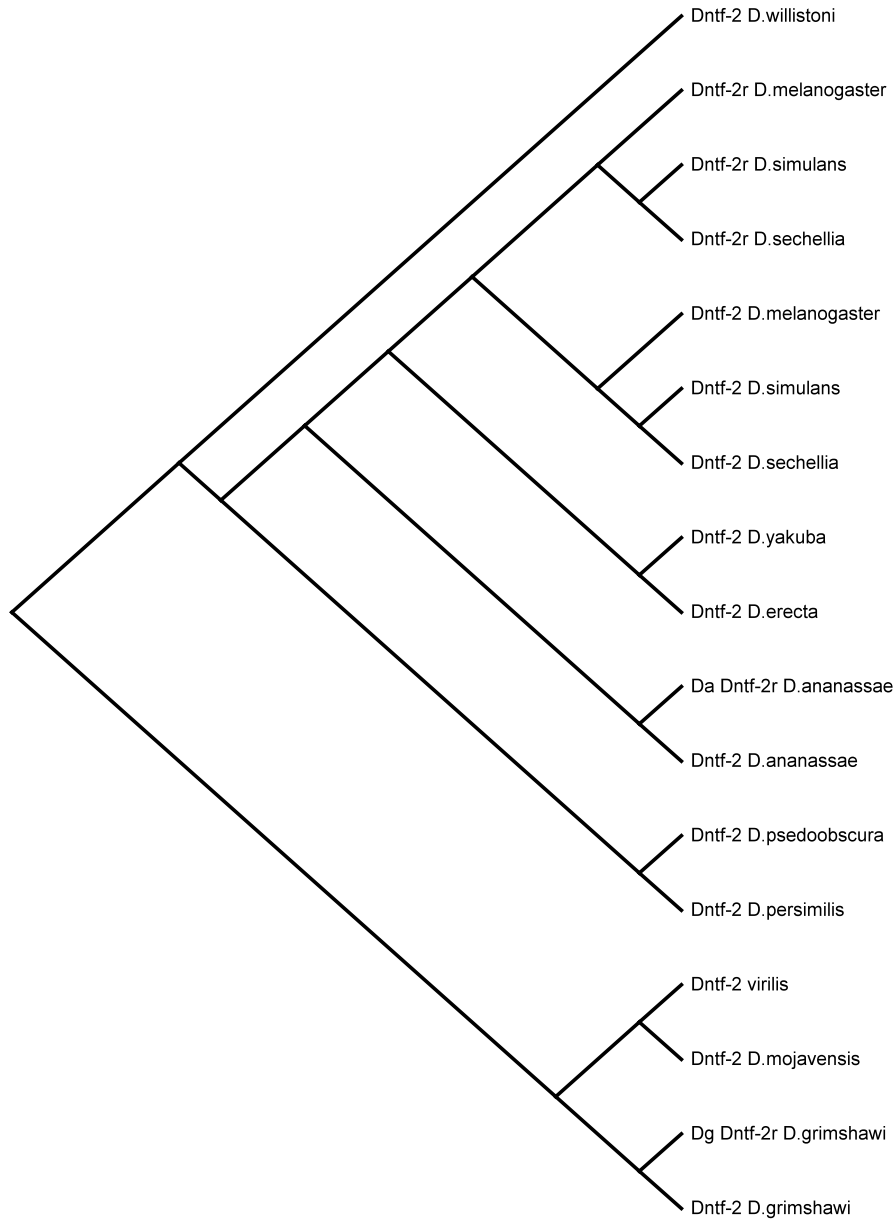
All other parental functional surfaces of Ran-like from the *D. melanogaster* subgroup seem to be even less conserved than the RanGAP interface. The analyses below focus on *D. melanogaster* Ran-like, but similar conclusions apply to the other Ran-like lineages analyzed. Regarding the interaction with Importin β , Ran-like proteins have likely reduced overall charge interactions through partial or complete changes in charge (E113G [disrupts hydrogen bond], K142T [in the basic patch], Y146L, and Y147I). In addition, one amino acid replacement (Q145Y) may possibly introduce a steric clash with position 163 (W163Y). The C-terminal end that is known to stabilize RanGDP (SEEWALD *et al.* 2002) may also have diverged. It is known that in the absence of this end, the RanGAP mediated hydrolysis of RanGTP to RanGDP is accelerated (SEEWALD *et al.*

2002), and the exchange of GDP to GTP catalyzed by RanGEF is also accelerated (RICHARDS *et al.* 1995). The C-terminal end is also required for the efficient binding of Ran to several of the Ran-binding proteins. Such binding is required for proper function of Ran (RICHARDS *et al.* 1995), but it is likely lost in Ran-like. RanGEF and exportins may also have a weaker interaction with Ran-like. Ran-like residues involved in RanGEF interaction have lost charge (partially or completely) or hydrophobicity (R95S, and V96N). Residue 95 is likely under positive selection (Table 2 and Figure 2). Ran-like residue 37 involved in exportin interaction has changed dramatically in charge and size (K37M).

The above analysis seems to indicate that the *D. melanogaster* subgroup *ran-like* protein has retained DNTF-2 and RanGAP interfaces while likely losing, or at least diminishing, all other protein-protein interfaces. The presence of DNTF-2 and RanGAP interfaces on Ran-like suggests that the Ran-like protein might exist in the RanGDP form and could be carried into the nucleus by DNTF-2r, where the Ran-like RanGDP could be converted to RanGTP. The Importin β interface on Ran-like, however, is likely diminished. As a result, Ran-GTP might not be transported out of the nucleus by Importin β , and Importin β might have a diminished capacity to release cargo upon nuclear entry. Similarly, export of RanGTP by the exportin complex may possibly be reduced. Additionally, the loss of Ran-like's C-terminal residues suggest that hydrolysis of RanGTP to RanGDP might possibly be accelerated in the presence of RanGAP. Other changes in Ran-like may affect the exchange of GDP to GTP by RanGEF. The binding of several lesser-known Ran binding proteins may also be affected. All these structural inferences remain to be experimentally tested. However, from these data it seems that Ran-like cannot completely replace Ran in testes in those *D. melanogaster* subgroup species where it is still functional.

REFERENCES

- BERMAN, H. M., T. BATTISTUZZI, T. N. BHAT, W. F. BLUHM, P. E. BOURNE *et al.*, 2002 The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr* **58**: 899-907.
- BULLOCK, T. L., W. D. CLARKSON, H. M. KENT and M. STEWART, 1996 The 1.6 angstroms resolution crystal structure of nuclear transport factor 2 (NTF2). *J Mol Biol* **260**: 422-431.
- ISGRO, T. A., and K. SCHULTEN, 2007 Association of nuclear pore FG-repeat domains to NTF2 import and export complexes. *J Mol Biol* **366**: 330-345.
- KUSANO, A., C. STABER, H. Y. CHAN and B. GANETZKY, 2003 Closing the (Ran)GAP on segregation distortion in *Drosophila*. *Bioessays* **25**: 108-115.
- KUSANO, A., C. STABER and B. GANETZKY, 2002 Segregation distortion induced by wild-type RanGAP in *Drosophila*. *Proc Natl Acad Sci U S A* **99**: 6866-6870.
- MARCHLER-BAUER, A., J. B. ANDERSON, M. K. DERBYSHIRE, C. DEWEESE-SCOTT, N. R. GONZALES *et al.*, 2007 CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res.* **35**: D237-240.
- MATSUURA, Y., and M. STEWART, 2004 Structural basis for the assembly of a nuclear export complex. *Nature* **432**: 872-877.
- PRESGRAVES, D. C., 2007 Does genetic conflict drive rapid molecular evolution of nuclear transport genes in *Drosophila*? *Bioessays* **29**: 386-391.
- QUIMBY, B. B., T. LAMITINA, S. W. L'HERNAULT and A. H. CORBETT, 2000 The mechanism of ran import into the nucleus by nuclear transport factor 2. *J Biol Chem* **275**: 28575-28582.
- RENAULT, L., J. KUHLMANN, A. HENKEL and A. WITTINGHOFER, 2001 Structural basis for guanine nucleotide exchange on Ran by the regulator of chromosome condensation (RCC1). *Cell* **105**: 245-255.
- RIBBECK, K., G. LIPOWSKY, H. M. KENT, M. STEWART and D. GORLICH, 1998 NTF2 mediates nuclear import of Ran. *EMBO Journal* **17**: 6587-6598.
- RICHARDS, S., K. LOUNSBURY and I. MACARA, 1995 The C terminus of the nuclear RAN/TC4 GTPase stabilizes the GDP-bound state and mediates interactions with RCC1, RAN-GAP, and HTF9A/RANBP1. *J Biol Chem* **270**: 14405-14411.
- SEEWALD, M. J., C. KORNER, A. WITTINGHOFER and I. R. VETTER, 2002 RanGAP mediates GTP hydrolysis without an arginine finger. *Nature* **415**: 662-666.
- STEWART, M., H. M. KENT and A. J. MCCOY, 1998 Structural basis for molecular recognition between nuclear transport factor 2 (NTF2) and the GDP-bound form of the Ras-family GTPase Ran. *J Mol Biol* **277**: 635-646.
- VETTER, I. R., A. ARNDT, U. KUTAY, D. GORLICH and A. WITTINGHOFER, 1999 Structural view of the Ran-Importin beta interaction at 2.3 Å resolution. *Cell* **97**: 635-646.

FIGURE S1.—Tree provided for the PAML branch analyses of *Dntf-2* and retrogenes

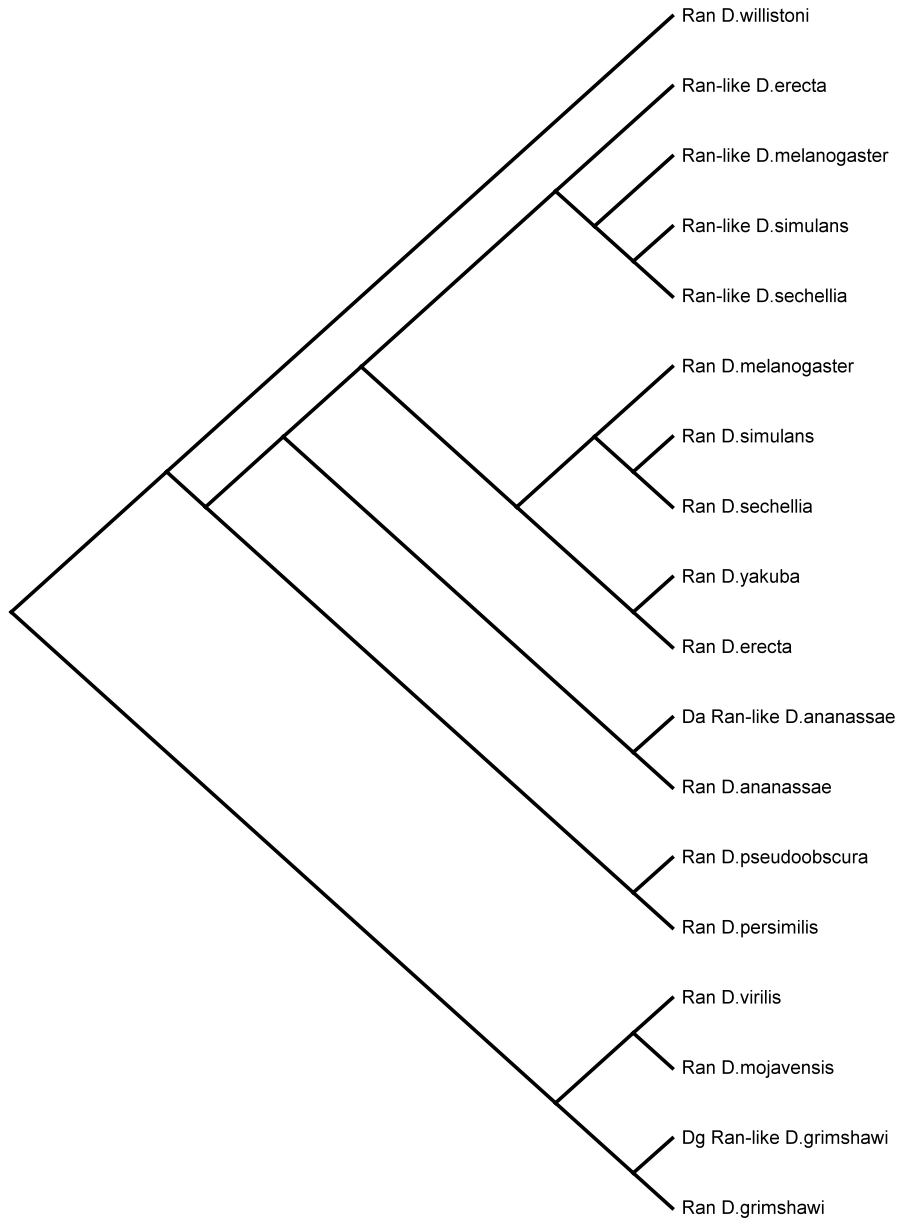


FIGURE S2.—Tree provided for the PAML branch analyses of *ran* and retrogenes

Tai21 ATGCAAGAGGTGACCTCATTCAAGCTGGTTCTTCTCGGAGACGGAGGAACTGGGAAAGCC 60
 Tai30 ATGCAAGAGGTGACCTCATTCAAGCTGGTTCTTCTCGGAGACGGAGGAACTGGGAAAGCC 60
 Tai15b ATGCAAGAGGTGACCTCATTCAAGCTGGTTCTTCTCGGAGACGGAGGAACTGGGAAAGCC 60
 Tai15a ATGCAAGAGGTGACCTCATTCAAGCTGGTTCTTCTCGGAGACGGAGGAACTGGGAAAGCC 60
 GE22850 ATGCAAGAGGTGACCTCATTCAAGCTGGTTCTTCTCGGAGACGGAGGAACTGGGAAAGCC 60
 GE19852 ATGCAAGAGGTGACCTCATTCAAG-TGGTTCTTCTCGGAGACGGAGGAACTGGGAAAGCC 59
 Tai6 ATGCAAGAGGTGACCTCATTCAAG-TGGTTCTTCTCGGAGACGGAGGAACTGGGAAAGCC 59
 Tai18 ATGCAAGAGGTGACCTCATTCAAG-TGGTTCTTCTCGGAGACGGAGGAACTGGGAAAGCC 59
 Tai59 ATGCAAGAGGTGACCTCATTCAAGCTGGTTCTTCTCGGAGACGGAGGAACTGGGAAAGCC 60
 Tai37 ATGCAAGAGGTGACCTCATTCAAGCTGGTTCTTCTCGGAGACGGAGGAACTGGGAAAGCC 60
 Tai26 ATGCAAGAGGTGACCTCATTCAAG-TGGTTCTTCTCGGAGACGGAGGAACTGGGAAAGCC 59
 Tai27 ATGCAAGAGGTGACCTCATTCAAGCTGGTTCTTCTCGGAGACGGAGGAACTGGGAAAGCC 60

Tai21 ACATTTATCAAGCGACACCTGACCGGCGAGTTCGAGAGGCGATACATTGCGACCCTGGGT 120
 Tai30 ACATTTATCAAGCGACACCTGACCGGCGAGTTCGAGAGGCGATACATTGCGACCCTGGGT 120
 Tai15b ACATTTATCAAGCGACACCTGACTGGCGAGTTCGATAGGCGATACATTGCGATCCTGGGT 120
 Tai15a ACATTTATCAAGCGACACCTGACCGGCGAGTTCGAGAGGCGATACATTGCGACCCTGGGT 120
 GE22850 ACATTTATCAAGCGACACCTGACCGGCGAGTTCGAGAGGCGATACATTGCGACCCTGGGT 120
 GE19852 ACATTTATCAAGCGACACCTGACCGGCGAGTTCGAGAGGCGATACATTGCGACCCTGGGT 119
 Tai6 ACATTTATCAAGCGACACCTGACCGGCGAGTTCGAGAGGCGATACATTGCGACCCTGGGT 119
 Tai18 ACATTTATCAAGCGACACCTGACCGGCGAGTTCGAGAGGCGATACATTGCGACCCTGGGT 119
 Tai59 ACATTTATCAAGCGACACCTGACCGGCGAGTTCGAGAGGCGATACATTGCGACCCTGGGT 120
 Tai37 ACATTTATCAAGCGACACCTGACCGGCGAGTTCGAGAGGCGATACATTGCGACCCTGGGT 120
 Tai26 ACATTTATCAAGCGACACCTGACCGGCGAGTTCGAGAGGCGATACATTGCGACCCTGGGT 119
 Tai27 ACATTTATCAAGCGACACCTGACCGGCGAGTTCGAGAGGCGATACATTGCGACCCTGGGT 120

Tai21 GTGGAGGTCCATCCAATACTCTTCCACACCAACCGAGGAGTGTACCGCTTCTATGTGTGG 180
 Tai30 GTGGAGGTCCATCCAATACTCTTCCACACCAACCGAGGAGTGTACCGCTTCTATGTGTGG 180
 Tai15b GTGGAGGTCCATCCAATACTCTTCCACACCAACCGAGGAGTGTACCGCTTCTATGTGTGG 180
 Tai15a GTGGAGGTCCATCCAATACTCTTCCACACCAACCGAGGAGTGTACCGCTTCTATGTGTGG 180
 GE22850 GTGGAGGTCCATCCAATACTCTTCCACACCAACCGAGGAGTGTACCGCTTCTATGTGTGG 180
 GE19852 GTGGAGGTCCATCCAATACTCTTCCACACCAACCGAGGAGTGTACCGCTTCTATGTGTGG 179

Tai6 GTGGAGGTCCATCCAATACTCTTCCACACCAACCGAGGAGTGTACCGCTTCTATGTGTGG 179
 Tai18 GTGGAGGTCCATCCAATACTCTTCCACACCAACCGAGGAGTGTACCGCTTCTATGTGTGG 179
 Tai59 GTGGAGGTCCATCCAATACTCTTCCACACCAACCGAGGAGTGTACCGCTTCTATGTGTGG 180
 Tai37 GTGGAGGTCCATCCAATACTCTTCCACACCAACCGAGGAGTGTACCGCTTCTATGTGTGG 180
 Tai26 GTGGAGGTCCATCCAATACTCTTCCACACCAACCGAGGAGTGTACCGCTTCTATGTGTGG 179
 Tai27 GTGGAGGTCCATCCAATACTCTTCCACACCAACCGAGGAGTGTACCGCTTCTATGTGTGG 180

Tai21 GACTGCGGTCAGGAGAAGTTCGGTGGCTACAAGATGGGTATTATGTCCAAGGTCAA 240
 Tai30 GACTGCGGTCAGGAGAAGTTCGGTGGCTACAAGATGGGTATTATGTCCAAGGTCAA 240
 Tai15b GACTGCGGTCAGGAGAAGTTCGGTAGCCTACAAGATGGGTATTATGTCCAAGGTCAA 240
 Tai15a GACTGCGGTCAGGAGAAGTTCGGTGGCTACAAGATGGGTATTATGTCCAAGGTCAA 240
 GE22850 GACTGCGGTCAGGAGAAGTTCGGTGGCTACAAGATGGGTATTATGTCCA----- 234
 GE19852 GACAC-----AAGATGGGTATTATGTCCAAGGTCAA 210
 Tai6 GACAC-----AAGATGGGTATTATGTCCAAGGTCAA 210
 Tai18 GACAC-----AAGATGGGTATTATGTCCAAGGTCAA 210
 Tai59 GACTGCGGTCAGGAGAAGTTCGGTGGCTACAAGATGGGTATTATGTCCAAGGTCAA 240
 Tai37 GACTGCGGTCAGGAGAAGTTCGGTGGCTACAAGATGGGTATTATGTCCAAGGTCAA 240
 Tai26 GACTGCGGTCAGGAGAAGTTCGGTGGCTACAAGATGGGTATTATGTCCAAGGTCAA 239
 Tai27 GACTGCGGTCAGGAGAAGTTCGGT-----AGACGGGTATTATGTCCAAGGTCAA 232

Tai21 TGTGCCATAATAATGTTTCGACGTGAGCTCGAGAATTACCTACAAGAATGTGGCACGTTGG 300
 Tai30 TGTGCCATAATAATGTTTCGACGTGAGCTCGAGAATTACCTACAAGAATGTGGCACGTTGG 300
 Tai15b TGTGCCATAATAATGTTTCGACGTGAGCTCGAGAATTACCTACAAGAATGTGGCACGTTGG 300
 Tai15a TGTGCCATAATAATGTTTCGACGTGAGCTCGAGAATTACCTACAAGAATGTGGCACGTTGG 300
 GE22850 TGTGCCATAATAATGTTTCGACGTGAGCTCGAGAATTACCTACAAGAATGTGGCACGTTGG 294
 GE19852 TGTGCCATAATAATGTTTCGACGTGAGCTCGAGAATTACCTACAAGAATGTGGCACGTTGG 270
 Tai6 TGTGCCATAATAATGTTTCGACGTGAGCTCGAGAATTACCTACAAGAATGTGGCACGTTGG 270
 Tai18 TGTGCCATAATAATGTTTCGACGTGAGCTCGAGAATTACCTACAAGAATGTGGCACGTTGG 270
 Tai59 TGTGCCATAATAATGTTTCGACGTGAGCTCGAGAATTACCTACAAGAATGTGGCACGTTGG 300
 Tai37 TGTGCCATAATAATGTTTCGACGTGAGCTCGAGAATTACCTACAAGAATGTGGCACGTTGG 300
 Tai26 TGTGCCATAAAATTGTTTCGACGTGAGCTCGAGAATTACCTACAAGAATGTGGCACGTTGG 299
 Tai27 TGTGCCATAATAATGTTTCGACGTGAGCTCGAGAATTACCTACAAGAATGTGGCACGTTGG 292

***** * *****

Tai21	CACCGCGACTTGGTGAGGGTATGCGGCAATATTCGATTGTTTTGTGTGGAAACAAGGTG	360
Tai30	CACCGCGACTTGGTGAGGGTATGCGGCAATATTCGATTGTTTTGTGTGGAAACAAGGTG	360
Tai15b	CACCGCGACTTGGTGAGGGTATGCGGCAATATTCGATTGTTTTGTGTGGAAACAAGGTG	360
Tai15a	CACCGCGACTTGGTGAGTGTATGCGGCAATATTCGATTGTTTTGTGTGGAAACAAGGTG	360
GE22850	CACCGCGACTTGGTGAGGGTATGCGGCAATATTCGATTGTTTTGTGTGGAAACAAGGTG	354
GE19852	CACCGCGACTTGGTGAGGGTATGCGGCAATATTCGATTGTTTTGTGTGGAAACAAGGTG	330
Tai6	CACCGCGACTTGGTGAGGGTATGCGGCAATATTCGATTGTTTTGTGTGGAAACAAGGTG	330
Tai18	CACCGCGACTTGGTGAGGGTATGCGGCAATATTCGATTGTTTTGTGTGGAAACAAGGTG	330
Tai59	CACCGCGACTTGGTGAGGGTATGCGGCAATATTCGATTGTTTTGTGTGGAAACAAGGTG	360
Tai37	CACCGCGACTTGGTGAGGGTATGCGGCAATATTCGATTGTTTTGTGTGGAAACAAGGTG	360
Tai26	CACCGCGACTTGGTGAGGGTATGCGGCAATATTCGATTGTTTTGTGTGGAAACAAGGTG	359
Tai27	CACCGCGACTTGGTGAGGGTATGCGGCAATATTCGATTGTTTTGTGTGGAAACAAGGTG	352

***** *****

Tai21	GATATCAAGCAACGGAAGGTTAGGCCAGGCGCTTTGACTTTCATCGTAAGAAAAACCTC	420
Tai30	GATATCAAGCAACGGAAGGTTAGGCCAGGCGCTTTGACTTTCATCGTAAGAAAAACCTC	420
Tai15b	GATATCAAGCAACGGAAGGTTAGGCCAGGCGCTTTGACTTTCATCGTAGGAAAAACCTC	420
Tai15a	GATATCAAGCAACGGAAGGTTAGGCCAGGCGCTTTGACTTTCATCGTAAGAAAAACCTC	420
GE22850	GATATCAAGCAACGGAAGGTTAGGCCAGGCGCTTTGACTTTCATCGTAAGAAAAACCTC	414
GE19852	GATATCAAGCAACGGAAGGTTAGGCCAGGCGCTTTGACTTTCATCGTAAGAAAAACCTC	390
Tai6	GATATCAAGCAACGGAAGGTTAGGCCAGGCGCTTTGACTTTCATCGTAAGAAAAACCTC	390
Tai18	GATATCAAGCAACGGAAGGTTAGGCCAGGCGCTTTGACTTTCATCGTAAGAAAAACCTC	390
Tai59	GATATCAAGCAACGGAAGGTTAGGCCAGGCGCTTTGACTTTCATCGTAAGAAAAACCTC	420
Tai37	GATATCAAGCAACGGAAGGTTAGGCCAGGCGCTTTGACTTTCATCGTAAGAAAAACCTC	420
Tai26	GATATCAAGCAACGGAAGGTTAGGCCAGGCGCTTTGACTTTCATCGTAAGAAAAACCTC	419
Tai27	GATATCAAGCAACGGAAGGTTAGGCCAGGCGCTTTGACTTTCATCGTAAGAAAAACCTC	412

***** *****

Tai21	CACTACATTGAAATGTCCGCCAAGTCAAACCTATA----ACATTGAGAGTCCCTTCGTCTA	476
Tai30	CACTACATTGAAATGTCCGCCAAGTCAAACCTATA----ACATTGAGAGTCCCTTCGTCTA	476
Tai15b	CACTACATTGAAATGTCCGCCAAGTCAAACCTATA----ACATTGAGAGTCCCTTCGTCTA	476
Tai15a	CACTACATTGAAATGTCCGCCAAGTCAAACCTATA----ACATTGAGAGTCCCTTCGTCTA	476

GE22850 CACTACATTG-----TCCGCCAAGTCAAACATAA----ATATTGAGAGTCCCTTCGTCTA 465
 GE19852 CACTACATTGAAATGTCCGCCAAGTCAAACATAA----ACATTGAGAGTCCCTTCGTCTA 446
 Tai6 CACTACATTGAAATGTCCGCCAAGTCAAACATAA----ACATTGAGAGTCCCTTCGTCTA 446
 Tai18 CACTACATTGAAATGTCCGCCAAGTCAAACATAA----ACATTGAGAGTCCCTTCGTCTA 446
 Tai59 CACTGCATTGAAATGTCCGCCAAGTCAAACATAA**CTTGA**CATTGATAGTCCCTTCGTCTA 480
 Tai37 CACTGCATTGAAATGTCCGCCAAGTCAAACATAA**CTTGA**CATTGATAGTCCCTTCGTCTA 480
 Tai26 CACTACATTGAAATGTCCGCCAAGTCAAACATAA----ACATTGAGAGTCCCTTCGTCTA 475
 Tai27 CACTACATTGAAATGTCCGCCAAGTCAAACATAA----ACATTGAGAGTCCCTTCGTCTA 468

Tai21 TCTGTTGCGGAAGTTGGTTGATGATCCCAACTTGCAATTGGTCAAGAACCCCGCTCTAAA 536
 Tai30 TCTGTTGCGGAAGTTGGTTGATGATCCCAACTTGCAATTGGTCAAGAACCCCGCTCTAAA 536
 Tai15b TCTGTTGCGGAAGTTGATTGATGATCCCAACTTGCAATTGGTCAAGAACCCCGCTCTAAA 536
 Tai15a TCTGTTGCGGAAGTTGGTTGATGATCCCAACTTGCAATTGGTTAAGAACCCCGCTCTAAA 536
 GE22850 TCTGTTGCGGAAGTTGGTTGATGATCCCAACTTGCAATTGGTCAAGAACCCCGCTCTAAA 525
 GE19852 TCTGTTGCGGAAGTTGGTTGATGATCCCAACTTGCAATTGGTCAAGGACCCCGCTCTAAA 506
 Tai6 TCTGTTGCGGAAGTTGGTTGATGATCCCAACTTGCAATTGGTCAAGGACCCCGCTCTAAA 506
 Tai18 TCTGTTGCGGAAGTTGGTTGATGATCCCAACTTGCAATTGGTCAAGGACCCCGCTCTAAA 506
 Tai59 TCTGTTGCGGAAGTTGGTTGATGATCCCAACTTGCAATTGGTCAAGAACCCCGCTCTAAA 540
 Tai37 TCTGTTGCGGAAGTTGGTTGATGATCCCAACTTGCAATTGGTCAAGAACCCCGCTCTAAA 540
 Tai26 TCTGTTGCGGAAGTTGGTTGATGATCCCAACTTGCAATTGGTCAAGAACCCCGCTCTAAA 535
 Tai27 TCTGTTGCGGAAGTTGGTTGATGATCCCAACTTGCAATTGGTCAAGAACCCCGCTCTAAA 528

Tai21 ACCCCCAGAAGTGGTTTTTACCGACGACG**CG**-----AATGGA 571
 Tai30 ACCCCCAGAAGTGGTTTTTACCGACGACG**CG**-----AATGGA 571
 Tai15b ACCCCCAGAAGTGGTTTTTACCGACGACG**CG**-----AATGGA 571
 Tai15a ACCCCCAGAAGTGTCTTTTACCGACGAGATGCGCCGTCAAGTGAACGCGGGTTAATGGA 596
 GE22850 ACCCCCAGATGTGGTTTTTACCGACGAGATGCGCCGTCAAGTGAACGCGGGTTAATGGA 585
 GE19852 ACCCCCAGAAGTGGTTTTTACCGACGAGATGCGCCGTCAAGTGAACGCGGGTTAATGGA 566
 Tai6 ACCCCCAGAAGTGGTTTTTACCGACGAGATGCGCCGTCAAGTGAACGCGGGTTAATGGA 566
 Tai18 ACCCCCAGAAGTGGTTTTTACCGACGAGATGCGCCGTCAAGTGAACGCGGGTTAATGGA 566
 Tai59 ACCCCCAGAAGTGGTTTTTACCGACGAGATGCGCCGTCAAGTGAACGCGGGTTAATGGA 600
 Tai37 ACCCCCAGAAGTGGTTTTTACCGACGAGATGCGCCGTCAAGTGAACGCGGGTTAATGGA 600

```

Tai26      ACTCCCAGAAGTTGTTTTTACCGACGAGATGCGCCGTCAAGTGGAAACGCGGGTTAATGGA 595
Tai27      ACCCCCAGAAGTGGTTTTTACCGACGAGATGCGCCGTCAAGTGGAAACGCGGGTTAATGGA 588
          ** ***** ** ** ********** *****

Tai21      GGCCAGCTTCTATCCTCTGCCCACTTATAACGATGATGATGATCTGTAA 620
Tai30      GGCCAGCTTCTATCCTCTGCCCACTTATAACGATGATGATGATCTGTAA 620
Tai15b     GGCCAGCTTCTATCCTCTGCCCACTTATAACGATGATGATGATCTGTAA 620
Tai15a     GGCCAGCTTCTATCCTCTGCCCACTTATAACGATGATGATGATCTGTAA 645
GE22850    GGCCAGCTTCTATCCTCTGCCCACTTATAACGATGATGATGATCTGTAA 634
GE19852    GGCCAGCTTCTATCCTCTGCCCACTTATAACGATGATGATGATCTGTAA 615
Tai6       GGCCAGCTTCTATCCTCTGCCCACTTATAACGATGATGATGATCTGTAA 615
Tai18      GGCCAGCTTCTATCCTCTGCCCACTTATAACGATGATGATGATCTGTAA 615
Tai59      GGCCAGCTTCTATCCTCTGCCCACTTATAACGATGATGATGATCTGTAA 649
Tai37      GGCCAGCTTCTATCCTCTGCCCACTTATAACGATGATGATGATCTGTAA 649
Tai26      GGCAAGCTTCTATCCTCTGCCCACTTATAACGATGATGATGATCTGTAA 644
Tai27      G-----ATGTTAATGATGATGATCTGTAA 612
          *                ** **********

```

FIGURE S3.—Alignment of *ran-like* disabled sequences from several strains of *D. yakuba*. See Materials and Methods and Results for more details.

Blue highlights deletions or insertions that change the frame (i.e. deletions that are not multiple of 3 base pairs). Deletions with insertion are marked in orange and they all change frame. The first premature stop codon in frame is shown in red.

TABLE S1
Oligonucleotide primers used in this work

Sequence	Type#	Gene	Species	Purpose
5'CTGGCAGGATAGGTTCAATAC 3'	F	<i>ran-like</i>	<i>D. melanogaster</i>	Genomic PCR
5'CAAAGATCATCGTTGCAC3'	R	<i>ran-like</i>	<i>D. melanogaster</i>	Genomic PCR
5'GCTGGCGGGATAAGTTC3'	F	<i>ran-like</i>	<i>D. simulans</i>	Genomic PCR
5'CCATGGGCACGAAGTAAG3'	R	<i>ran-like</i>	<i>D. simulans</i>	Genomic PCR
5'ATTACACAAGCCGCTCC3'	F	<i>ran-like</i>	<i>D. yakuba</i>	Genomic PCR
5'ACGCAGAAGGGGAAAAG3'	R	<i>ran-like</i>	<i>D. yakuba</i>	Genomic PCR
5'ATGCCCTCTCAATCCCCAC3'	F	<i>Da_Ntf-2r</i>	<i>D. ananassae</i>	Genomic PCR
5'TTATTCCGTGTCGTGGATATTC3'	R	<i>Da_Ntf-2r</i>	<i>D. ananassae</i>	Genomic PCR
5'ATGCCCTCTCAATCCCCAC3'	F	<i>Da_Ntf-2r</i>	<i>D. atripex</i>	Genomic PCR
5'TTATTCCGTGTCGTGGATATTC3'	R	<i>Da_Ntf-2r</i>	<i>D. atripex</i>	Genomic PCR
5'CAATCTCCTCGTGCAGACG3'	F	<i>Da_ran-like</i>	<i>D. ananassae</i>	Genomic PCR
5'CGGAGTGTCCAATTTGTGCG3'	R	<i>Da_ran-like</i>	<i>D. ananassae</i>	Genomic PCR
5'CAATCTCCTCGTGCAGACG3'	F	<i>Da_ran-like</i>	<i>D. atripex</i>	Genomic PCR
5'CGGAGTGTCCAATTTGTGCG3'	R	<i>Da_ran-like</i>	<i>D. atripex</i>	Genomic PCR
5'GATATTGGCAAGGGATTCGTC3'	F	<i>Dntf-2</i>	<i>D. ananassae</i>	RT-PCR
5'CGACCAAGAACGTTAATCAG3'	R	<i>Dntf-2</i>	<i>D. ananassae</i>	RT-PCR
5' CCAATGGGCCAGGAATTTGTG 3'	F	<i>Da_Ntf-2r</i>	<i>D. ananassae</i>	RT-PCR
5'GCAGTCTTCCCAGGACACTC3'	R	<i>Da_Ntf-2r</i>	<i>D. ananassae</i>	RT-PCR
5'CACATTCAAGTGCGTACTCGTC3'	F	<i>ran</i>	<i>D. ananassae</i>	RT-PCR
5'TGGCAACGAATTCCAGGTTGG3'	R	<i>ran</i>	<i>D. ananassae</i>	RT-PCR
5'GTGGCAGCGGTGATGGTATTC3'	F	<i>Da_ran-like</i>	<i>D. ananassae</i>	RT-PCR
5'CTTGCCGAAATGTCGTAG3'	R	<i>Da_ran-like</i>	<i>D. ananassae</i>	RT-PCR
5'ATTGTGACCACAGTCGGTTC 3'	F	<i>Gapdh2</i>	<i>D. ananassae</i>	RT-PCR
5'GTCGTACCAAGAGATCAGCTTCAC3'	R	<i>Gapdh2</i>	<i>D. ananassae</i>	RT-PCR
5'AGGACATTGGCAAGGGCTTTC3'	F	<i>Dntf-2</i>	<i>D. grimshawi</i>	RT-PCR
5'TGGGCTGAGAGTCAACTGTGG 3'	R	<i>Dntf-2</i>	<i>D. grimshawi</i>	RT-PCR
5'CCGTTGGCAAAGGTTTTGTCC3'	F	<i>Dg_Ntf-2r</i>	<i>D. grimshawi</i>	RT-PCR
5'CAGTTGCGAATAGGAGTGTGGTG3'	R	<i>Dg_Ntf-2r</i>	<i>D. grimshawi</i>	RT-PCR
5'CAAGGACCGCAAGGTCAAAG3'	F	<i>ran</i>	<i>D. grimshawi</i>	RT-PCR
5'TGCAGCTGCCAGTCTCTGTG3'	R	<i>ran</i>	<i>D. grimshawi</i>	RT-PCR
5'GCTGTGTGGCAACAAAGTCG3'	F	<i>Dg_ran-like</i>	<i>D. grimshawi</i>	RT-PCR
5'CCTCCTGCAAATCTCGTTCCG3'	R	<i>Dg_ran-like</i>	<i>D. grimshawi</i>	RT-PCR
5'GGTGCTGCCAAAACATCAT3'	F	<i>Gapdh</i>	<i>D. grimshawi</i>	RT-PCR
5'GCTGAGGAAATCGGTGGAGAC3'	R	<i>Gapdh</i>	<i>D. grimshawi</i>	RT-PCR

#F and R refer to forward and reverse, respectively

TABLE S2**PAML comparisons for *Dntf-2* and *Dntf-2r* retrogenes**

Model	l	p	\hat{W}_{Dntf-2}	$\hat{W}_{Dntf-2r}$	$\hat{W}_{Dntf-2r_dup}$	\hat{W}_{Dn_Ntf-2r}	\hat{W}_{Dg_Ntf-2r}
One-ratio	-2570.8689	34	0.0533	0.0533	0.0533	0.0533	0.0533
Two-ratio	-2531.7282	35	0.0243	0.2657	0.2657	0.2657	0.2657
Four-ratio	-2522.3422	37	0.0247	0.5311	0.5311	0.3309	0.0754
Five-ratio	-2522.0333	38	0.0247	0.6235	0.3655	0.3310	0.0754
Four-ratio $\hat{W}_{Dntf-2r} = \mathbf{1}$	-2524.2914	36	0.0248	1.0000	1.0000	0.3328	0.0758
Four-ratio $\hat{W}_{Da_Ntf-2r} = 1$	-2527.5541	36	0.0239	0.5362	0.5362	1.0000	0.0762

Log likelihood values and estimates of K_A/K_S ratios are shown. p is the number of parameters estimated in the model. \hat{W}_{Dntf-2} is K_A/K_S ratio for all *Dntf-2* genes. $\hat{W}_{Dntf-2r}$ is the K_A/K_S ratio for the melanogaster subgroup minus the branch immediately following duplication of *Dntf-2r*. $\hat{W}_{Dntf-2r_dup}$ is the K_A/K_S ratio for the branch immediately following duplication in the melanogaster subgroup. \hat{W}_{Da_Ntf-2r} and \hat{W}_{Dg_Ntf-2r} the K_A/K_S ratio for the retroposed sequence in *D. ananassae* and *D. grimshawi* respectively.

TABLE S4**PAML comparisons for *ran* and *ran* retrogenes**

Model	l	p	\hat{w}_{ran}	$\hat{w}_{ran-like}$	$\hat{w}_{ran-like_dup}$	$\hat{w}_{Da_ran-like}$	$\hat{w}_{Dg_ran-like}$
One-ratio	-4610.4156	36	0.0571	0.0571	0.0571	0.0571	0.0571
Two-ratio	-4492.8616	37	0.0042	0.1793	0.1793	0.1793	0.1793
Four-ratio	-4452.4538	39	0.0043	0.3593	0.3593	0.0395	0.0349
Five-ratio	-4420.8830	40	0.0044	0.7023	0.0249	0.0408	0.0348
Five-ratio $\hat{w}_{ran-like}=\mathbf{1}$	-4422.9256	39	0.0044	1.0000	0.0250	0.0409	0.0350

Log likelihood values and parameters estimated under differing models of selection. l refers to log likelihood values. p is the number of parameters estimated in the model. \hat{w}_{ran} is K_A/K_S ratio for all *ran* genes. $\hat{w}_{ran-like}$ is the K_A/K_S ratio for the melanogaster subgroup minus the branch immediately following duplication of *ran-like*. $\hat{w}_{ran-like_dup}$ is the K_A/K_S ratio for the branch immediately following duplication in the melanogaster subgroup. $\hat{w}_{Da_ran-like}$ and $\hat{w}_{Dg_ran-like}$ the K_A/K_S ratio for the retroposed sequence in *D. ananassae* and *D. grimshawi* respectively

