DOI: 10.1534/genetics.109.113522

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[1]

Department of Biology, University of Texas, Arlington, Texas 76019

Manuscript received December 22, 2009
Accepted for publication December 31, 2009

ABSTRACT

The analyses of gene duplications by retroposition have revealed an excess of male-biased duplicates generated from X chromosome to autosomes in flies and mammals. Investigating these genes is of primary importance in understanding sexual dimorphism and genome evolution. In a particular instance in Drosophila, X-linked nuclear transport genes (*Ntf-2* and *ran*) have given rise to autosomal retroposed copies three independent times (along the lineages leading to *Drosophila melanogaster*, *D. ananassae*, and *D. grimshawi*). Here we explore in further detail the expression and the mode of evolution of these Drosophila *Ntf-2*- and *ran*-derived retrogenes. Five of the six retrogenes show male-biased expression. The *ran-like* gene of *D. melanogaster* and *D. simulans* has undergone recurrent positive selection. Similarly, in *D. ananassae* and *D. atripex*, the *Ntf-2* and *ran* retrogenes show evidence of past positive selection. The data suggest that strong selection is acting on the origin and evolution of these retrogenes. Avoiding male meiotic X inactivation, increasing level of expression of X-linked genes in male testes, and/or sexual antagonism might explain the recurrent duplication of retrogenes from X to autosomes. Interestingly, the *ran-like* in *D. yakuba* has mostly pseudogenized alleles. Disablement of the *ran-like* gene in *D. yakuba* indicates turnover of these duplicates. We discuss the possibility that *Dntf-2r* and *ran-like* might be involved in genomic conflicts during spermatogenesis.

CONVERGENT duplications of two genes involved in nuclear transport, *Ntf-2* and *ran* in different lineages of Drosophila, have recently been described (Bai *et al.* 2007). The Drosophila *Ntf-2* (*Dntf-2*) appears to have generated retroposed copies—retrogenes (Brosius 1991)—three independent times. The retrogenes are observed in different genomic locations in different Drosophila species with respect to the six evolutionarily conserved chromosomal arms of Drosophila (*i.e.*, Muller elements A to F; Powell 1997): (1) *Dntf-2r* located in the 2L chromosomal arm (Muller element B) of the four species of the *Drosophila melanogaster* complex (*D. melanogaster*, *D. simulans*, *D. mauritiana*, and *D sechellia*; Betrán and Long 2003), (2) an unnamed retrogene in the *D. ananassae* lineage (*GF23973*) located in the chromosomal arm corresponding to Muller element D, and (3) a retrogene in the lineage leading to *D. grimshawi* (*GH16820*) located in the arm that corresponds to Muller element D. We call the *Dntf-2*-derived retrogenes of the *D. ananassae* and *D. grimshawi* lineages *Da_Ntf-2r* and *Dg_Ntf-2r*, respectively. The *ran* gene has also given rise to retrogenes three independent times—along some of the same lineages as the *Dntf-2*-derived retrogenes. It gave rise to *ran-like*, which is present in all the species of the *D. melanogaster* subgroup and is located in the 3L arm (Muller element D). *D. ananassae* and *D. grimshawi* also have independently derived *ran* retrogenes, *GF24088* and *GH16204*, located in Muller elements B and D, respectively. We call the *ran*-derived retrogenes of the *D. ananassae* and *D. grimshawi* lineages *Da_ran-like* and *Dg_ran-like*, respectively.

The parental *ran* and *Ntf-2* proteins physically interact with each other and play a central role in the transport of proteins into and out of the nucleus (Ribbeck *et al.* 1998). Both proteins are highly conserved and are required in all eukaryotes (Quimby *et al.* 2000). The presence of duplicates of both genes in several lineages may have an adaptive explanation, particularly if the expression of each duplicate overlaps the other. While the function of the duplicates is not known, all six independent retrogene events occurred from X-to-autosome locations—a direction of retroduplication known to be overrepresented in Drosophila, human, and mouse genomes (Betrán *et al.* 2002; Emerson *et al.* 2004). The X-to-autosome duplications may be the result of positive selection due to pressures relating to male meiotic X inactivation, sexual antagonism, and/or X chromosome dosage compensation (see below).

In the case of male meiotic X inactivation (Lifschytz and Lindsley 1972; Hense *et al.* 2007), it is thought to

Supporting information is available online at http://www.genetics.org/cgi/content/full/genetics.109.113522/DC1.

[1]*Corresponding author:* Biology Department, Box 19498, University of Texas, Arlington, TX 76019. E-mail: betran@uta.edu

be beneficial to have gene copies off the X and onto autosomes where transcription can occur during X chromosome inactivation (Betrán *et al.* 2002). It has previously been shown that the *D. melanogaster* retrogene *Dntf-2r* exhibits a strong testis-biased expression (Betrán and Long 2003), and the same is true for *ran-like* (Chintapalli *et al.* 2007). In addition, both *Dntf-2r* and *ran-like* represent X-to-autosome movements (Bai *et al.* 2007) . Thus, it is possible that the multiple parallel retropositions to autosomes of *Dntf-2* and *ran* have been driven by the selective advantage of providing a way of retaining nuclear transport function in males during meiotic X inactivation. Male meiotic X inactivation might be a consequence of the inactivation that unsynapsed meiotic chromosomes suffer (Turner *et al.* 2005; Turner 2007).

On the other hand, under the standard sexual antagonistic hypothesis, males and females need the same gene, but alleles that are good for males might be bad for females and vice versa (*i.e.*, some homologous traits are selected in different directions in the different sexes; Chippindale *et al.* 2001). In this case, if mutations are partially dominant, it might be beneficial to have the male-biased genes located on the autosomes, as opposed to the X chromosome. The X chromosome spends two-thirds of its time in females; thus, the female would be able to outcompete the male for alleles on the X chromosome (Rice 1984; Charlesworth *et al.* 1987; Ranz *et al.* 2003).

In the case of X chromosome dosage compensation, it is known that compensation in Drosophila occurs through hypertranscription of the X chromosome in males. The need for additional increased X-linked gene expression in males could be a pressure to duplicate X-linked genes onto the autosomes (Vicoso and Charlesworth 2009).

Additionally, the duplicates attaining novel male testes functions could also explain the multiple duplication events of the nuclear transport genes *Dntf-2* and *ran*. For example, it has been proposed that nuclear transport genes (including *Dntf-2r*) might play a role in segregation distortion and/or germline genomic conflicts involving transposable elements and viruses (Presgraves 2007; Presgraves and Stephan 2007). Interestingly, there is direct evidence that genes involved in nuclear transport play a role in the young *SD* segregation distortion system in spermatogenesis of *D. melanogaster* (see review, Kusano *et al.* 2003). The main distorter, *Sd*, is a truncated duplicate form of the nuclear transport gene *RanGAP* that (mis)localizes in the nucleus (Kusano *et al.* 2003). Nuclear transport genes (*e.g.*, *RanGAP* and six different nucleoporins) and their duplicates (*e.g.*, *Dntf-2r*) have been hypothesized to be involved in segregation distortion in Drosophila (Presgraves 2007). The arms race between distorters and suppressors of distortion could lead to the fixation of duplicate genes with testes expression and to fast gene evolution of these genes (Kusano *et al.* 2002; Burt and Trivers 2006; Presgraves 2007). Indeed, *RanGAP*, six nucleoporins, and *Dntf-2r* all have been shown to have evolved under positive selection (Betrán and Long 2003; Presgraves 2007; Presgraves and Stephan 2007). If this is the case, the selective advantage for the fixation, longevity, and fast evolution of duplicates like *Dntf-2r* will persist as long as segregation distortion conflicts involving nuclear transport keep emerging.

Transposable elements and viruses may also impose genomic conflicts involving nuclear transport since these elements often need to enter germline nuclei to replicate (Presgraves and Stephan 2007).

Previous work demonstrated that the *D. melanogaster Dntf-2r* has testis-biased expression and is under positive selection. *D. melanogaster ran-like* is similarly testes biased in expression. Here we explore the mode of evolution and the pattern of expression of other independently retroposed copies of *Ntf-2* and *ran*. In both *D. ananassae* and *D. grimshawi*, we show that *Dntf-2* and *ran* retrogenes are, for the most part, strongly male biased in their transcription. In addition, we reveal that all the retrogenes are evolving faster than the parental genes and are largely under positive selection. The *ran-like* gene is evolving under recurrent positive selection in several branches and for particular residues. In the lineages where *ran-like* is undergoing positive selection, some of the inferred selected amino acid changes appear to disturb specific protein–protein interaction surfaces. Polymorphism data for *ran-like* in the *D. yakuba* lineage indicate that *D. yakuba*'s *ran-like* is a pseudogene in most alleles (9 of 10) and is wrongly annotated as functional in FlyBase. The retrogenes in the *D. ananassae* lineage, *Da_Ntf-2r* and *Da_ran-like*, are also under positive selection. These data reveal strong selective pressures on the origin and evolution of *Dntf-2* and *ran* retrogenes.

MATERIALS AND METHODS

Strains used and DNA sequencing: Genomic DNA was extracted from single flies using a Puregene kit (Qiagen, Valencia, CA). The *ran-like* gene was PCR amplified from genomic DNA from 12 *D. melanogaster* flies from different Zimbabwe strains (ZH13, ZH18, ZH19, ZH20, ZH21, ZH23, ZH26, ZH27, ZH28, ZH29, ZH32, and ZH40; Hollocher *et al.* 1997), nine *D. simulans* flies from different Madagascar strains (M1, M4, M5, M24, M37, M50, M242, M252, and M258), and 10 *D. yakuba* strains (Taï6, Taï15a, Taï15b, Taï18, Taï21, Taï26, Taï27, Taï30, Taï37, and Taï59) from the Taï forest in Ivory Coast. These strains were kindly provided by the Wu, Aquadro, Long, and Llopart laboratories. Oligonucleotide primers used for these reactions are given in the supporting information, Table S1.

Da_Ntf-2r and *Da_Ran-like* were PCR amplified from genomic DNA from 10 *D. ananassae* flies from different strains (14024-0371.16, 14024-0371.17, 14024-0371.18, 14024-0371.25, 14024-0371.30, 14024-0371.31, 14024-0371.32, 14024-0371.33, 14024-0371.34, and 14024-0371.35), and four *D. atripex* flies from different strains (14024-0361.00, 14024-0361.01, 14024-0361.02, and 14024-0361.03) obtained from

the San Diego Stock Center. Oligonucleotide primers used in these reactions are given in Table S1. *Da_Ntf-2r* sequences of *D. atripex* are partial gene sequences (357 bp long).

PCR products were sequenced from both strands using an automated DNA sequencer and fluorescent BigDye terminators (Applied Biosystems, Foster City, CA). Internal primers were used in addition to the primers above to complete some of the sequencing. Heterozygotes were cloned using a TOPO cloning kit (Invitrogen, Carlsbad, CA), and one insert was sequenced to resolve haplotypes. Accession numbers for the sequences obtained are GU338155–GU338215.

Strains that were used for the whole genome sequence project of *D. ananassae* and *D. grimshawi* (Clark *et al.* 2007) were used to study pattern of expression of *Dntf-2*, *ran*, and their retrogenes in these species (see below). These strains were obtained from the former Tucson Drosophila Stock Center (currently the San Diego Stock Center).

Sequence analyses: The nucleotide sequences of the *Dntf-2* gene in the 12 sequenced Drosophila species (Clark *et al.* 2007) were retrieved from FlyBase along with the sequences of the paralog *Dntf-2r* in *D. melanogaster*, *D. simulans*, and *D. sechellia*. These 15 sequences were aligned with each other and with *Dntf-2* retroposed sequences from *D. ananassae* and *D. grimshawi* with Clustal W software (Thompson *et al.* 1994). Sequences were analyzed initially using the CODEML software package implemented in PAML 4 (Yang 2007). The tree provided is shown in Figure S1. Branch models were employed to determine selective pressures on the parental and retroposed sequences by calculation of K_A/K_S ratios for each branch (Yang 1998). The one-ratio model sets all branches to evolve at the same rate. Additional models allowed for the genes to evolve at different rates, for the branches to evolve at different rates after duplication, or for the parental gene to evolve at a different rate after duplication. Models where K_A/K_S is set at one in particular lineages were also compared to detect purifying selection. These were all *a priori* hypotheses that we wanted to test. These models were compared by calculating two times the log likelihood values and comparing this value to a χ^2 distribution with degrees of freedom equaling the difference in number of parameters estimated by each model.

Next, site models (NSsites) of CODEML software implemented in PAML were used to uncover the possibility of positive selection acting on a few sites. Site-specific likelihood models M7 and M8 were applied to the sequences with the appropriate tree topology (Nielsen and Yang 1998; Yang *et al.* 2000). Model M7, which does not allow for sites under positive selection, was compared to model M8, which does allow for sites under positive selection. M7 assumes a beta distribution for ω between 0 and 1 over all sites, while M8 adds an additional site class ($\omega \geq 1$), with ω estimated from the data. A likelihood ratio test was performed by calculating two times the log likelihood values and comparing this value to a χ^2 distribution with two degrees of freedom. Posterior probabilities of codons under positive selection were computed in model M8 using Bayes empirical Bayes when the LRT was significant. The tree provided for the site analyses was ((*Dntf-2r_D. simulans*, *Dntf-2r_D. sechellia*), *Dntf-2r_D. melanogaster*).

The same PAML methods and comparisons were employed for *ran* and its retroposed sequences. We used parental and retrogene sequences that are present in the same species, with the addition of a *ran-like* ortholog in *D. erecta*. The *ran-like* ortholog in *D. yakuba* was not added to the analyses because a careful look at the annotation and polymorphism data revealed that it is likely a pseudogene. The tree provided for the K_A/K_S ratio analyses in different branches is shown in Figure S2. The fact that the *ran-like* ortholog in *D. yakuba* is evolving differently, similar to a pseudogene, prompted us to investigate if the other *ran-like* lineages evolve at different rates. The tree provided for the site-specific analyses was (((*ran-like_D.simulans*, *ran-like_D.sechellia*), *ran-like_D.melanogaster*), *ran-like_D.erecta*).

Sequences of *Dntf-2r* and *ran-like* were also analyzed using the HyPhy package. *Dntf-2r* sequences and tree topology were ((*Dntf-2r_D. simulans*, *Dntf-2r_D. sechellia*), *Dntf-2r_D. melanogaster*). The *ran-like* sequences and tree topology were (((*ran-like_D. simulans*, *ran-like_D. sechellia*), *ran-like_D. melanogaster*), *ran-like_D. erecta*). These sequences were uploaded to the HyPhy package available at http://www.datamonkey.org (Pond and Frost 2005). Random effects likelihood (REL) and fixed effects likelihood (FEL) analyses were performed in an attempt to detect positively selected codons in the *Dntf-2r* and *ran-like* phylogenies. A Bayes factor threshold of 50, which corresponds to very low probablity (P∼1/Bayes factor), was used for the REL analysis, and $P < 0.10$ for the FEL analysis (Kosakovsky Pond and Frost 2005). These codon analyses are presumed to be more realistic than the PAML site models because they allow for synonymous rate variation across sites (Kosakovsky Pond and Frost 2005).

A McDonald–Kreitman test (McDonald and Kreitman 1991) was performed for the polymorphism data obtained for *ran-like* in *D. melanogaster* and *D. simulans*. Sequenced products were aligned using Clustal W (Thompson *et al.* 1994) and imported into DnaSP 5.0 (Librado and Rozas 2009) to perform the test. Lineage-specific McDonald–Kreitman tests were also performed for *ran-like* in *D. melanogaster* and *D. simulans* (Akashi 1995; Presgraves 2007). A modified McDonald–Kreitman test was also used. The latter test removes synonymous unpreferred changes from the calculations because they might contribute more to polymorphisms than to divergence (Akashi 1995; Schlenke and Begun 2003).

D. yakuba ran-like sequences obtained in this work were aligned with two genes of *ran-like* of *D. yakuba* annotated in FlyBase (*GE22850*; 3L random and *GE19852*; 3L) to reveal the disablements that we detected in 9 of the 10 alleles and all the genome sequences (see RESULTS and Figure S3). We infer that the two annotated genes of *ran-like* of *D. yakuba* in FlyBase are likely alleles of the same gene. Both genes are flanked by the same region but have different disablements that likely prevented initial assembly.

Standard McDonald–Kreitman tests (McDonald and Kreitman 1991) were also performed for the polymorphism data obtained for *Da_ran-like* and *Da_Ntf-2r* in *D. ananassae* and *D. atripex*. Sequenced products were aligned using Clustal W (Thompson *et al.* 1994) and imported into DnaSP 5.0 (Librado and Rozas 2009) to perform the tests. The lineage-specific and the modified McDonald–Kreitman tests were not performed in this case because of the absence of a close enough outgroup sequence and because of a lack of information on codon preferences in these species.

The action of recent positive selection can be addressed using proposed population genetics statistics that test for skew in the frequency spectrum of alleles when compared to the neutral model. Tajima's *D* (Tajima 1989) compares θ_π (the average number of nucleotide differences per site between two random sequences) and θ_W (Watterson's estimate of θ from the number of segregating sites; Watterson 1975). Differences between θ_π and θ_W (Tajima's *D*) reveal nonequilibrium conditions in the history of the gene. Negative significant values are consistent with recent events of positive selection. The *H* statistic, the difference between θ_π and θ_H estimates (Fay and Wu 2000), measures the excess of derived variants at high frequency. Again, negative significant values of this statistic are consistent with recent events of positive selection. Tajima's *D* and Fay and Wu's *H* values were tested using neutral coalescence simulations. They were computed and tested by

10,000 simulations using DNAsp 5.1 (Librado and Rozas 2009). Recombination rates for all those simulations were considered zero, to be conservative. When these tests were applied to the *ran-like* polymorphism data in *D. yakuba*, deletions were manually removed and coded as nucleotide changes.

Expression analyses: Tissues were homogenized, and total RNA was prepared as described by the Qiagen protocol (RNeasy kit, Qiagen). Total RNA was obtained from 30 virgin gonadectomized females, 30 gonadectomized males, 30 ovaries, and 100 testes plus accessory glands of *D. ananassae*. Gonadectomized males and females are flies from which we removed ovaries/testes and accessory glands by dissecting mature males and females in saline solution. After dissection, tissues were preserved in RNAlater solution (Applied Biosystems/Ambion, Austin, TX) at −20° after soaking them at 4° overnight until they were processed. Total RNA was also obtained from two virgin females and three males from *D. grimshawi*.

RT–PCR was conducted on total RNA for *Dntf-2, ran*, and the retrogenes derived from these genes in *D. ananassae* and *D. grimshawi*. *Gapdh2* and *Gapdh* were used as internal standards for the quantitative real-time RT–PCR (qRT–PCR) performed for the retrogenes in *D. ananassae* and *D. grimshawi*, respectively. Analysis of expression of intronless genes (such as *Dntf-2* and *ran* retrogenes) is challenging because genomic contamination can produce a band of the same size as that expected from the cDNA. Therefore, we digested possible contaminating DNA from the total RNA (DNase I amplification grade, Invitrogen) and ran controls including DNA digested total RNA without reverse transcriptase (RT−). Single strand complementary DNA (cDNA) was synthesized using Superscript and oligo-dT (Promega, Madison, WI). RT–PCRs were carried out using specific oligonucleotide primers given in Table S1. qRT–PCRs were performed using the ABI 7300 Real Time PCR system and the SYBR Green PCR Core reagents from Applied Biosystems. Two or three replicates of RT+ and RT− were performed for every retrogene, normalizing gene, and RNA extraction. The primers described above were used and produced similar amplification plots in the logarithmic scale that corresponds to similar efficiency (Schmittgen and Livak 2008). RT–PCR products were run in gels to control for any spurious amplification. Threshold cycle numbers (C_T values) were obtained with the default ABI software parameters. C_T values obtained for the retrogenes were normalized by subtracting the C_T value of the normalizing gene (ΔC_T). The mean difference in normalized threshold cycle number (ΔC_T) in different tissues was tested using ANOVA analyses. Post hoc Tukey tests were also performed. Fold changes in the expression were calculated using the expression $2^{-\Delta\Delta CT}$ (Schmittgen and Livak 2008).

RESULTS AND DISCUSSION

***Dntf-2* and *ran* retrogenes are mostly male biased in expression:** In *D. ananassae*, RT–PCR results reveal that both parental genes, *Dntf-2* and *ran*, are transcribed equally in both males and females across all tested tissue types (data not shown). The qRT–PCR results for the *Da_Nntf-2r* and *Da_ran-like* retrogenes are shown in Figure 1, A and B. The average of the retrogene C_T difference to *Gapdh2* (*i.e.*, ΔC_T) and standard error are shown for each tissue type. Small values of ΔC_T correspond to high values of transcription. An ANOVA comparing values for *Da_ ran-like* expression in different tissues (Figure 1A) reveals that these values are significantly different ($F_{(3,8)} = 317.5$; $P < 10^{-5}$). Post hoc Tukey tests show that all values are statistically significantly different from each other ($P < 0.05$ in all comparisons). *Da_ran-like* is highly expressed in testes and expressed at low or very low levels in gonadectomized male bodies, gonadectomized female bodies, and ovaries. Testes expression is 60-fold higher than in other tissues.

An ANOVA comparing values for *Da_ Ntf-2r* expression in different tissues (Figure 1B) shows that these values are significantly different ($F_{(3,7)} = 77.472$; $P < 10^{-4}$). Post hoc Tukey tests reveal that all values are statistically significantly different from each other ($P < 0.05$ in all comparisons), except the comparison between male and female gonadectomized body ($P = 0.9939$). *Da_ Ntf-2r* is highly expressed in testes and expressed at much lower levels in gonadectomized male bodies, gonadectomized female bodies, and ovaries. Expression in testes is 27-fold higher than in other tissues.

In *D. grimshawi* RT–PCR results reveal again that the parental genes *Dntf-2* and *ran* are transcribed equally in males and in females (data not shown). The retrogene *Dg_Nntf-2r* is highly expressed in males compared to females ($F_{(1,4)} = 103.96$; $P = 0.0005$; Figure 1C). We do not know whether this bias is due to testes expression as we were not able to obtain separate tissues for *D. grimshawi*. Unlike *Dg_Nntf-2r*, *Dg_ran-like* is expressed higher in females than males ($F_{(1,4)} = 43.40$; $P = 0.0028$; Figure 1C).

We conclude that *Dntf-2* and *ran* retrogenes are typically male biased in expression. The retrogenes are strongly testes biased in the *D. melanogaster* lineage (Betrán and Long 2003; Chintapalli *et al.* 2007) and the *D. ananassae* lineage (Figure 1). *Dg_Nntf-2r* is male biased in the *D. grimshawi* lineage (Figure 1), although it remains unknown whether *Dg_Nntf-2r* is testes biased.

Sequence evolution of *Dntf-2* and its retrogenes: To understand the mode of evolution of *Dntf-2* and its retrogenes, we performed sequence analyses using PAML software as described in materials and methods. Results of the branch PAML analyses for *Dntf-2* genes and *Dntf-2* retrogenes appear in Table S2. The free-ratio model (data not shown, $l = -2507.6789$, $P = 65$) was found to be significantly better ($P = 1.787 \times 10^{-13}$) than the one-ratio model in revealing rate differences. The two-ratio model, which estimates one K_A/K_S ratio for *Dntf-2* and one ratio for the *Dntf-2* retrogene branches, was found to fit the data better than the one-ratio model ($P < 1.110 \times 10^{-16}$). A four-ratio model was implemented to allow differing rates of evolution for the three retrogenes and *Dntf-2*. This model is significantly better than the two-ratio model ($P = 8.390 \times 10^{-5}$), indicating different rates of evolution among the recurrently recruited retrogenes. The retrogene in the *D. melanogaster*

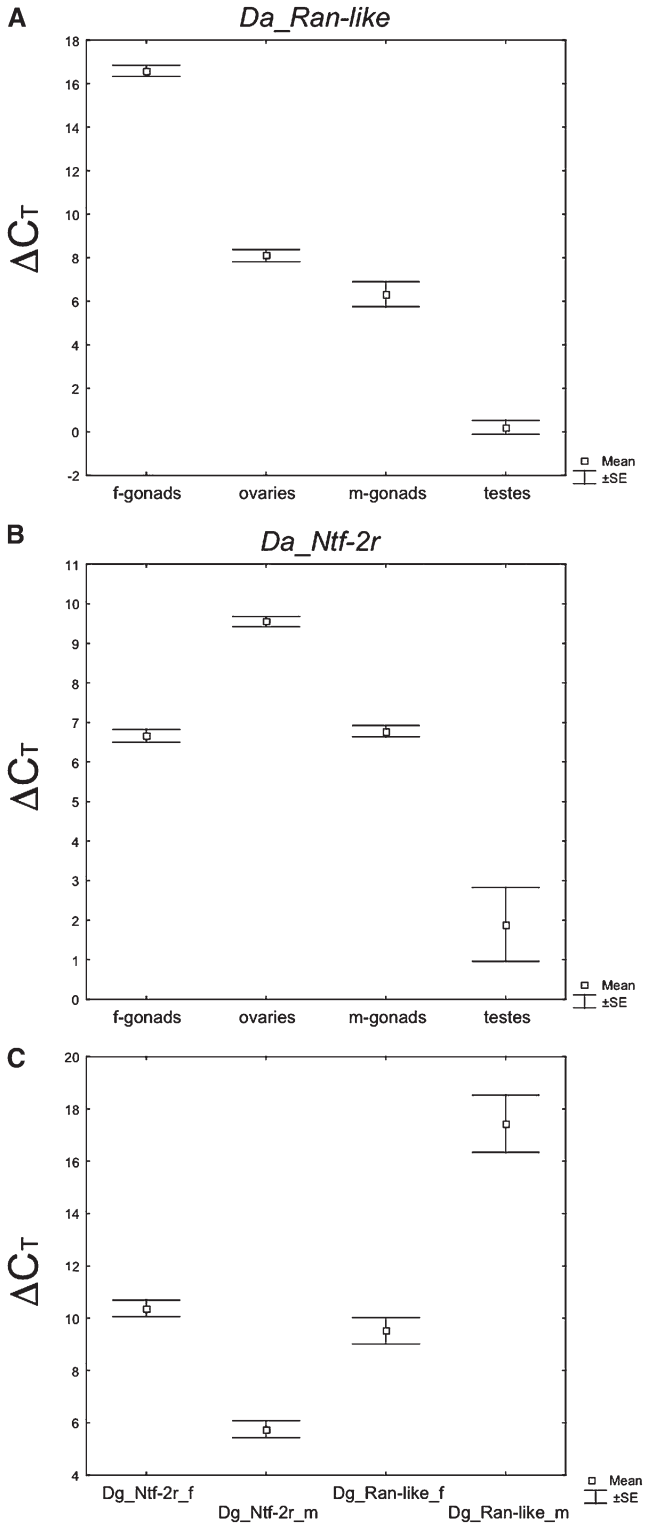


FIGURE 1.—qRT–PCR results for the *D. ananassae* and *D. grimshawi* retrogenes (*Da_ran-like*, *Da_Ntf-2r*, *Dg_ran-like*, and *Dg_ Ntf-2r*) are shown. The average of the retrogene C_T difference to the normalizing gene (*i.e.*, ΔC_T) and standard errors are shown for every tissue type. f, females; m, males.

complex is evolving the fastest ($K_A/K_S = 0.5311$). That said, all of the retrogene evolution rates are significantly higher than the parental rate—3- to 21-fold higher.

The retrogene rates can be explained either by positive selection acting on the retrogenes or by relaxation of constraint in the retrogene lineages. We have evidence of positive selection acting on *Dntf-2r* provided by the McDonald–Kreitman test (BETRÁN and LONG 2003) and on *Da_Ntf-2r* (see below). The results of the *Da_Ntf-2r* McDonald–Kreitman test is consistent with a high K_A/K_S ratio (0.3309) estimated in the *D. ananassae* lineage. However, no lines are available from *D. grimshawi* or close relatives to gather polymorphism data, and we have no evidence supporting selection in the *D. grimshawi* lineage at this point.

A five-ratio model is an extension of the four-ratio model. It allows a branch to evolve under a different K_A/K_S ratio postduplication. When the five-ratio model was applied to the retrogene in the *D. melanogaster* complex, it did not fit significantly better than the simpler four-ratio model. Models that considered a different rate for the parental genes postduplication were significantly worse at fitting the data than the four-ratio model (data not shown). Four additional models were constructed to determine whether the estimated K_A/K_S ratio for every gene was significantly less than one to ascertain whether purifying selection is acting. All four models were significantly worse at fitting the data than the four-ratio model, indicating that all of the K_A/K_S values in the phylogeny are significantly less than one and that purifying selection is acting in each of the lineages.

Although it has been pointed out that K_S will saturate at large evolutionary distances (*e.g.*, at distances larger than the *D. melanogaster* group that includes *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba*, *D. erecta*, and *D. ananassae*; LARRACUENTE *et al.* 2008), all of our *Dntf-2* retrogenes are younger than the distance represented by the *D. melanogaster* group. *Dntf-2r* is present only in *D. melanogaster*, *D. simulans*, and *D. sechellia*. *Da_Ntf-2r* is present only in *D. ananassae*. *Dg_Ntf-2r* is present only in *D. grimshawi*, whose lineage is estimated to be shorter than the *D. ananassae* lineage (CLARK *et al.* 2007). However, because the estimates of *Dntf-2* K_A/K_S listed in Table S2 suffer from saturation, we reestimated the K_A/K_S ratio for *Dntf-2* using only the *D. melanogaster* group of species. The PAML K_A/K_S ratio obtained was 0.0188. This K_A/K_S value confirms that the parental genes have evolved much slower than the retrogenes.

Site models (M7 and M8 and FEL) were also fitted to the data to test for positive selection acting on particular sites of *Dntf-2r* (see MATERIALS AND METHODS). From these results we did not infer positive selection acting on any sites of the retroposed *Dntf-2* sequences. However, REL analysis detected two codons [47 (Bayes factor = 52.9636), 87 (Bayes factor = 52.6307)] that are likely under positive selection. Nucleoporins are known to interact with amino acid number 47 of the parental NTF-2 (File S1).

Polymorphism data for part of *Da_Ntf-2r* in the *D. ananassae* and *D. atripex* lineages (Table S3) were

TABLE 1

McDonald–Kreitman tests for *Da_Ntf-2r* and *Da_ran-like* in the *D. ananassae* and *D. atripex* lineages

	Da_Ntf-2r		*Da_ran-like*	
	Fixed	Polymorphic	Fixed	Polymorphic
Replacement	36	2	15	0
Synonymous	27	16	33	21
	$G_{\text{Williams correction}} = 12.991$ $P = 0.0003$		Fisher's exact test $P_{(\text{two tailed})} = 0.0031$	

used to perform the McDonald–Kreitman and Tajima's *D* and Wu's *H* tests. The results of the McDonald–Kreitman test reveal a statistically significant excess of replacement substitutions (Table 1) that is usually interpreted as recurrent positive selection acting on the protein. Tajima's *D* and Fay and Wu's *H* tests were not significant for either sample set (data not shown). The above result, along with published results for *Dntf-2r* (Betrán and Long 2003), suggest that most retroposed *Dntf-2* sequences have been under strong selective pressures. We discuss the possible functional consequences of these protein changes in File S2.

Sequence evolution of *ran* and its retrogenes: To understand the mode of evolution of *ran* and its retrogenes, we performed sequence analyses similar to the ones performed for *Dntf-2* and its retrogenes. PAML-derived log likelihood values and maximum likelihood estimates of K_A/K_S ratios for the branches in *ran* and *ran-like*, as well as the other *ran* retrogenes, are given in Table S4. A free-ratio model (data not shown) was fitted ($l = -4410.17775$, $P = 69$) and compared to the one-ratio model. The free-ratio model gave a much better fit ($P = 0$) to the data due to differing K_A/K_S ratios along the branches of the tree. The one-ratio model was then compared to a two-ratio model ($P = 0$), showing a 43-fold increase in the rate of evolution in the *ran* retrogene lineages when compared to the *ran* branches (*e.g.*, 0.1793 for retrogene *vs.* 0.0042 for parental gene). Next, a four-ratio model was fitted to the data to allow differing rates of evolution for the branches that correspond to the three recurrent recruitments of *ran* retrogenes. This model shows accelerated evolution of *ran-like* in the *D. melanogaster* subgroup relative to all other branches in the tree when compared to the two-ratio model ($P = 0$). Other retrogene lineages are evolving much faster than the parental genes as well (9–10 times faster).

To determine the mode of evolution in the *D. melanogaster* subgroup immediately after duplication of *ran*, a five-ratio model was generated. Models that consider a different rate for the parental genes after duplication were significantly worse at fitting the data (data not shown), while models that consider a different rate for the retrogenes after duplication are significantly better than the previous four-ratio model ($P = 1.8873 \times 10^{-15}$). Our five-ratio model shows a marked increase in the K_A/K_S (0.7023) in *ran-like* of the *D. melanogaster* subgroup. Immediately after duplication of *ran* in the *D. melanogaster* subgroup, purifying selection ($K_A/K_S = 0.0249$) was acting on the newly retroposed gene, even though it was still evolving roughly six times faster than the parental gene. In fact, all ratios in the five-ratio model, including this last one, are significantly smaller than one (Table S4). Our conclusions (*i.e.*, that retrogenes evolve much faster than parental genes) remained unchanged when we reestimated the *ran* K_A/K_S ratio using only the species of the *D. melanogaster* group, where estimates should have higher accuracy (*i.e.*, K_S should not be saturated). The *ran* K_A/K_S ratio had a value of 0.0065 in this calculation, confirming that retrogenes evolve one to two orders of magnitude faster than the parental gene (Table S4). Again, either relaxation of constraint or positive selection could explain the large increase in evolutionary rate among the *ran-like* lineages, although additional analyses (see below) reveal that selection is more likely.

Site models (M7 *vs.* M8, REL, and FEL) were applied to test for positive selection acting on *ran-like*. The M8 site model allows for positively selected sites, while M7 does not. The likelihood ratio test between site models M7 and M8 was statistically significant ($2\Delta l = 8.885$, d.f. = 2; $P = 0.0118$), which is indicative of positive selection acting on *ran-like* because it reveals that model M8 fits the data significantly better than model M7. Site model M8 estimated that 36.4% of sites in the *ran-like* alignment experienced positive selection ($K_A/K_S = 2.52$). Codons that are most likely under positive selection as revealed by Bayes empirical Bayes analysis (posterior probabilities ≥0.95%) are shown in Table 2. REL analyses using the Hyphy package detected 14 sites (Table 2 and Figure 2) with a Bayes factor >50 ($P < 0.02$) that have likely been under positive selection. The more stringent FEL analyses detected only one codon (131) as being likely under positive selection ($P = 0.0973$).

Polymorphism data in *D. melanogaster*, *D. simulans*, and *D. yakuba* for *ran-like* were obtained. Table S5 shows the polymorphism data for *D. melanogaster* and *D. simulans*. McDonald–Kreitman and modified McDonald–Kreitman tests were performed for *ran-like* using polymorphism data for *D. simulans* and *D. melanogaster* (Table 3). Both McDonald–Kreitman tests showed a significantly higher ratio of replacements per substitu-

TABLE 2

Comparison of likely positively selected codons identified by site-specific model M8 BEB and REL analyses in *ran-like*

	Positively selected codons	
Codon	M8 BEB probability	REL Bayes factor
48 N	0.911	**572.1**
49 H	**0.955**	**558.1**
58 V	0.810	**52.2**
81 I	0.748	**53.1**
92 T	0.785	**54.2**
93 A	0.913	**598.3**
94 K	0.825	**53.1**
95 A	**0.958**	**469.5**
131 S	0.939	**583.9**
140 R	**0.960**	**467.6**
200 F	**0.985**	**3574.9**
202 D	0.888	**443.9**
207 Y	0.842	**55.3**
215 F	0.921	**389.1**

Codons that had Bayes empirical Bayes (BEB) posterior probabilities ≥0.95 or a Bayes factor ≥50 were included in the results. Boldface values represent sites identified with statistical significance by either method. Codon number and amino acids are relative to *D. simulans* full-length protein.

TABLE 3

McDonald–Kreitman and modified McDonald–Kreitman tests for *ran-like* in the *D. melanogaster* and *D. simulans* lineages

	Fixed	Polymorphic
Replacement	47	21
Synonymous	12 (2)	14 (5)

G (with Williams correction) = 4.062, $P = 0.0438$. Values in parentheses correspond to the changes to preferred codons used for the modified McDonald–Kreitman test: G (with Williams correction) = 4.028, $P = 0.0447$.

tion, as opposed to replacements per silent polymorphism. These data support recurrent positive selection acting on *ran-like* in *D. melanogaster* and *D. simulans*. Lineage-specific McDonald–Kreitman tests performed for *ran-like* in *D. melanogaster* and *ran-like* in *D. simulans* were not significant (data not shown). Likewise, Tajima's *D* and Fay and Wu's *H* tests were not significant, indicating that the positive selection revealed above was not recent.

D. yakuba polymorphism data were left out of the McDonald–Kreitman analysis because most *ran-like* alleles (9 of 10) were found to have deletions or insertions in the coding region (Figure S3). All of the deletions and insertions result in frameshifts and/or premature stop codons. The *ran-like* gene in *D. yakuba* is likely incorrectly annotated in FlyBase as two genes (one with an intron), but they are likely two disabled alleles (Figure S3). There are five kinds of disabling deletions and one kind of disabling insertion in our Taï *ran-like* data set. Some of the disablements overlap with the ones observed in the FlyBase sequences. These disablements are at frequencies that vary from 10 to 30% in our sample. This observation means that no single disablement has swept through the population under directional selection in agreement with the nonsignificant Tajima's *D* value (data not shown). We estimate the age of the alleles (Slatkin and Rannala 2000) to be less than $\sim 1.03 \times 10^6$ generations, assuming an effective population size of 10^6. If, in addition, we assume two generations per year in *D. yakuba*, this estimate leads to an age of 0.515 MY for the alleles present at 0.30 frequency. Since *D. santomea* and *D. yakuba* are estimated to have diverged ~400,000 years ago (Llopart *et al.* 2002), we conclude that some of the disablements may perhaps be shared between these species.

Polymorphism data were also obtained in *D. ananassae* and *D. atripex* for *Da_ran-like* (Table S6). The McDonald–Kreitman test was performed and revealed a significantly higher ratio of replacements per silent substitution when compared to the ratio of replacements per silent polymorphism. Fisher's exact test was applied in this case because some cells have no observations (Table 1). This result is compatible with recurrent positive selection occurring in these lineages. However, Tajima's *D* and Fay and Wu's *H* tests were not significant for either sample set (data not shown). Taken together, the site-specific analyses and McDonald–Kreitman tests for *ran-like* in *D. melanogaster* and *D. simulans* and *Da_ran-like* in *D. ananassae* and *D. atripex* strongly support the action of positive selection on *ran* retrogenes. We discuss the possible functional consequences of these protein changes in File S2.

Are X-to-autosome retrogenes convergently duplicated because of their function in genomic conflicts in Drosophila? A strong selective force, when acting across separate lineages, can often lead to convergent evolution. *Utp14* in mammals is an example of convergent evolution that involves the recurrent emergence and recruitment of retrogenes (Bradley *et al.* 2004). In this case, male meiotic X inactivation was suggested as the selective force driving the convergent recruitment of *Utp14* retrogenes.

In the present case, two interacting Drosophila nuclear transport genes, *Dntf-2* and *ran*, gave rise to three retrogenes each, in overlapping lineages. All these retrogenes represent X-to-autosome duplications. We further show that most of the retroposed genes have a male-biased transcription and have evolved under recurrent positive selection. We reveal that strong selection is associated with the origin and evolution of the duplicates of *Dntf-2* and *ran*. That said, *ran-like* appears to be pseudogenizing in *D. yakuba*. In further support to the X-to-autosome bias, we have identified another independent X-to-autosome retroposed copy of *Ntf-2* in *Anopheles gambiae* (data not shown).

In the Introduction, we outlined several selective hypotheses that could explain the recurrent duplication of *Ntf-2* and *ran*: male meiotic X inactivation, increasing

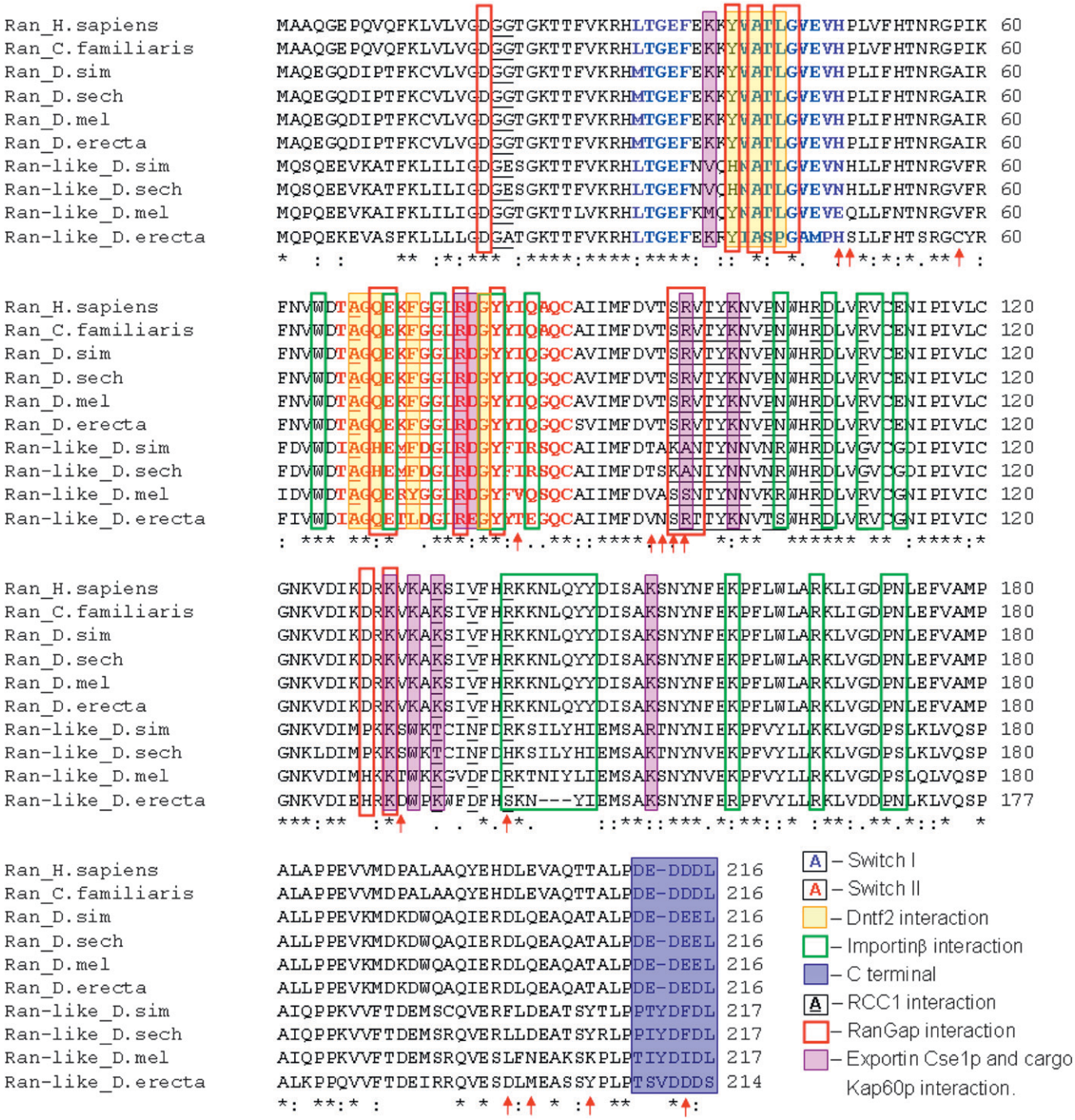


FIGURE 2.—Alignment of *ran* and Drosophila *ran-like* proteins highlighting amino acid residues of known function. Arrows point to sites that have likely been under positive selection in the retrocopies (Table 2).

level of expression of X-linked genes in male testes, sexual antagonism, and new function (*e.g.*, segregation distortion or other). In particular, meiotic genomic conflicts, be they meiotic drive driven or parasite driven, are a powerful force in shaping genes and genomes as these conflicts lead to an arms race and rapid evolution (BURT and TRIVERS 2006). Genes (or their duplicates) involved in conflicts evolve under recurrent positive selection (KINGAN *et al.* 2009) and may become fixed in the populations fairly quickly (BURT and TRIVERS 2006). These genes, particularly the duplicates, might be rendered of no use when the conflicts disappear or are dealt with in a different way. Indeed, examples exist of genes responding to strong selective pressures only to disappear after the pressures on them dissipate. Seminal fluid proteins have been observed to evolve under positive selection and have high turnover (BEGUN and LINDFORS 2005). In fact, this type of evolutionary

dynamic characterizes many male-biased genes (Zhang *et al.* 2007), possibly because male-biased genes are more likely to be engaged in male competition, sexual antagonism, segregation distortion, and/or parasite-related conflicts more often than non-male-biased genes. Very complex situations might emerge in meiotic drive systems as observed in the case of the Winters *sex ratio* genes in *D. simulans* (Kingan *et al.* 2009) where three genes of a gene family—one of them not even encoding for a protein anymore—are part of a meiotic drive system. Each of the Winters *sex ratio* genes are under positive selection in one population or another and are polymorphic for presence–absence of alleles.

In our case, if we hypothesize that nuclear transport is routinely involved in male germline-based genomic conflicts—segregation distortion or other—then this could explain recurrent duplication of *Ntf-2* and *ran* as well as the fast evolution of the their duplicates. We postulate that the *Ntf-2* and *ran* parental genes are (or have been) involved in male germline genomic conflicts. Different alleles of *Ntf-2* and *ran* in the population could act either as drivers or as suppressors of the conflict system. Alleles that increase in frequency in males, but are detrimental to females, would often end up being present at intermediate frequencies (Rice 1984; Patten and Haig 2009). Such an antagonism could promote the emergence of new genes under the model proposed by Proulx and Phillips (2006). Duplicates of *Ntf-2* and *ran* would then help alleviate the antagonism. At the same time, factors such as the male meiotic X inactivation, increasing level of expression of X-linked genes in male testes, and/or sexual antagonism may determine the location of the duplicated genes (*e.g.*, X to autosome), as described in the Introduction. The partial loss of function that we infer (File S2) might have occurred in some lineages (*e.g.*, *ran-like* in *D. melanogaster* subgroup species) would be explained by a segregation distortion or genomic conflict function followed by selective pressure to differentiate from the parental gene. The complete loss of *ran-like* in some lineages would occur if the conflicts become resolved through an alternative mechanism in that lineage (postrecruitment of *ran-like*) or vanish altogether. The retrogene would then become quickly pseudogenized or deleted (Petrov *et al.* 1996).

Finally, a knockout of *Dntf-2r* (*D. melanogaster P*-element insertion line EY05573) shows no obvious male fertility effects (M. Motiwale and E. Betrán, unpublished results), supporting a degree of dispensability of the gene. Experiments are being carried out to reveal the potential role of *Dntf-2r* and *ran-like* in segregation distortion in *D. melanogaster* using the *SD* system. Functions related to conflicts caused by transposable elements or viruses should also be tested.

We thank Elena de la Casa-Esperón, Daven Presgraves, Dave Begun, and several reviewers for comments on this work. We also thank Mao-Lien Wu, Chung-I Wu, Tessa Bauer DuMont, Chip Aquadro, Hongzheng Dai, Manyuan Long, Ana Llopart, and the Tucson and San Diego Drosophila Stock Centers for providing stocks. Research was supported by the University of Texas at Arlington startup funds and grant GM071813 from National Institutes of Health (to E.B.).

LITERATURE CITED

Akashi, H., 1995 Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in Drosophila DNA. Genetics **139:** 1067–1076.

Bai, Y., C. Casola, C. Feschotte and E. Betrán, 2007 Comparative genomics reveals a constant rate of origination and convergent acquisition of functional retrogenes in Drosophila. Genome Biol. **8:** R11.

Begun, D. J., and H. A. Lindfors, 2005 Rapid evolution of genomic Acp complement in the melanogaster subgroup of Drosophila. Mol. Biol. Evol. **22:** 2010–2021.

Betrán, E., and M. Long, 2003 *Dntf-2r:* a young Drosophila retroposed gene with specific male expression under positive Darwinian selection. Genetics **164:** 977–988.

Betrán, E., K. Thornton and M. Long, 2002 Retroposed new genes out of the X in Drosophila. Genome Res. **12:** 1854–1859.

Bradley, J., A. Baltus, H. Skaletsky, M. Royce-Tolland, K. Dewar *et al.*, 2004 An X-to-autosome retrogene is required for spermatogenesis in mice. Nat. Genet. **36:** 872–876.

Brosius, J., 1991 Retroposons–seeds of evolution. Science **251:** 753.

Burt, A., and R. Trivers, 2006 *Genes in Conflict. The Biology of Selfish Genetic Elements.* Harvard University Press, Cambridge, MA.

Charlesworth, B., J. A. Coyne and N. H. Barton, 1987 The relative rates of evolution of sex chromosomes and autosomes. Am. Nat. **130:** 113–146.

Chintapalli, V. R., J. Wang and J. A. Dow, 2007 Using FlyAtlas to identify better Drosophila melanogaster models of human disease. Nat. Genet. **39:** 715–720.

Chippindale, A. K., J. R. Gibson and W. R. Rice, 2001 Negative genetic correlation for adult fitness between sexes reveals ontogenetic conflict in Drosophila. Proc. Natl. Acad. Sci. USA **98:** 1671–1675.

Clark, A. G., M. B. Eisen, D. R. Smith, C. M. Bergman, B. Oliver *et al.*, 2007 Evolution of genes and genomes on the Drosophila phylogeny. Nature **450:** 203–218.

Emerson, J. J., H. Kaessmann, E. Betrán and M. Long, 2004 Extensive gene traffic on the mammalian X Chromosome. Science **303:** 537–540.

Fay, J. C., and C. I. Wu, 2000 Hitchhiking under positive Darwinian selection. Genetics **155:** 1405–1413.

Hense, W., J. F. Baines and J. Parsch, 2007 X chromosome inactivation during Drosophila spermatogenesis. PLoS Biol. **5:** e273.

Hollocher, H., C. T. Ting, M. L. Wu and C. I. Wu, 1997 Incipient speciation by sexual isolation in *Drosophila melanogaster*: variation in mating preference and correlation between sexes. Evolution **51:** 1175–1181.

Kingan, S. B., D. Garrigan and D. L. Hartl, 2009 Recurrent selection on the Winters *sex-ratio* genes in *Drosophila simulans*. Genetics **184:** 253–265.

Kosakovsky Pond, S. L., and S. D. Frost, 2005 Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol. Biol. Evol. **22:** 1208–1222.

Kusano, A., C. Staber and B. Ganetzky, 2002 Segregation distortion induced by wild-type RanGAP in Drosophila. Proc. Natl. Acad. Sci. USA **99:** 6866–6870.

Kusano, A., C. Staber, H. Y. Chan and B. Ganetzky, 2003 Closing the (Ran)GAP on segregation distortion in Drosophila. Bioessays **25:** 108–115.

Larracuente, A. M., T. B. Sackton, A. J. Greenberg, A. Wong, N. D. Singh *et al.*, 2008 Evolution of protein-coding genes in Drosophila. Trends Genet. **24:** 114–123.

Librado, P., and J. Rozas, 2009 DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics **25:** 1451–1452.

Lifschytz, E., and D. L. Lindsley, 1972 The role of X-chromosome inactivation during spermatogenesis. Proc. Natl. Acad. Sci. USA **69:** 182–186.

Llopart, A., S. Elwyn, D. Lachaise and J. A. Coyne, 2002 Genetics of a difference in pigmentation between *Drosophila yakuba* and *Drosophila santomea*. Evolution **56:** 2262–2277.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the *Adh* locus in Drosophila. Nature **351:** 652–654.

Nielsen, R., and Z. Yang, 1998 Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics **148:** 929–936.

Patten, M. M., and D. Haig, 2009 Maintenance or loss of genetic variation under sexual and parental antagonism at a sex-linked locus. Evolution **63:** 2888–2895.

Petrov, D. A., E. R. Lozovskaya and D. L. Hartl, 1996 High intrinsic rate of DNA loss in Drosophila. Nature **384:** 346–349.

Pond, S. L., and S. D. Frost, 2005 Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics **21:** 2531–2533.

Powell, J. R., 1997 *Progress and Prospects in Evolutionary Biology: The Drosophila Model*. Oxford University Press, New York.

Presgraves, D. C., 2007 Does genetic conflict drive rapid molecular evolution of nuclear transport genes in Drosophila? Bioessays **29:** 386–391.

Presgraves, D. C., and W. Stephan, 2007 Pervasive adaptive evolution among interactors of the Drosophila hybrid inviability gene, Nup96. Mol. Biol. Evol. **24:** 306–314.

Proulx, S. R., and P. C. Phillips, 2006 Allelic divergence precedes and promotes gene duplication. Evolution **60:** 881–892.

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.

Ranz, J. M., C. I. Castillo-Davis, C. D. Meiklejohn and D. L. Hartl, 2003 Sex-dependent gene expression and evolution of the Drosophila transcriptome. Science **300:** 1742–1745.

Ribbeck, K., G. Lipowsky, H. M. Kent, M. Stewart and D. Gorlich, 1998 NTF2 mediates nuclear import of Ran. EMBO J. **17:** 6587–6598.

Rice, W. R., 1984 Sex chromosomes and the evolution of sexual dimorphism. Evolution **38:** 735–742.

Schlenke, T. A., and D. J. Begun, 2003 Natural selection drives Drosophila immune system evolution. Genetics **164:** 1471–1480.

Schmittgen, T. D., and K. J. Livak, 2008 Analyzing real-time PCR data by the comparative C(T) method. Nat. Protoc. **3:** 1101–1108.

Slatkin, M., and B. Rannala, 2000 Estimating allele age. Annu. Rev. Genomics Hum. Genet. **1:** 225–249.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

Turner, J. M., 2007 Meiotic sex chromosome inactivation. Development **134:** 1823–1831.

Turner, J. M., S. K. Mahadevaiah, O. Fernandez-Capetillo, A. Nussenzweig, X. Xu *et al.*, 2005 Silencing of unsynapsed meiotic chromosomes in the mouse. Nat. Genet. **37:** 41–47.

Vicoso, B., and B. Charlesworth, 2009 The deficit of male-biased genes on the D. melanogaster X chromosome is expression-dependent: A consequence of dosage compensation? J. Mol. Evol. **68:** 576–583.

Watterson, G. A., 1975 On the number of segragating sites in genetical models without recombination. Theor. Popul. Biol. **7:** 256–276.

Yang, Z., 1998 Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. **15:** 568–573.

Yang, Z., 2007 PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. **24:** 1586–1591.

Yang, Z., R. Nielsen, N. Goldman and A. M. Pedersen, 2000 Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics **155:** 431–449.

Zhang, Y., D. Sturgill, M. Parisi, S. Kumar and B. Oliver, 2007 Constraint and turnover in sex-biased gene expression in the genus Drosophila. Nature **450:** 233–237.

Communicating editor: D. Begun

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

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 alingment and Dg_Ntf-2r is shown as Dntf-2_Retgrim in the alignment.

D.melanogaster

```
Dntf2_D.sim              MSLNPQYEEIGKGFVQQYYAIFDDPANRANVVNFYSATDSFMTFFGHQIQ 50
Dntf2_D.sec              MSLNPQYEEIGKGFVQQYYAIFDDPANRANVVNFYSATDSFMTFFGHQIQ 50
Dntf2_PA_D.mel           MSLNPQYEDIGKGFVQQYYAIFDDPANRANVVNFYSATDSFMTFFGHQIQ 50
Dntf2r_D.mel             MSLNLQYEDIGKEFVQQYYAIFDDPANRENVINFYNATDSFMTFFGNQIQ 50
Dntf2r_D.sim             MSLNPQYEEIGKGFVQQYYAIFDDPVNRENVVHFYSATDSFMTFFGRQIQ 50
Dntf2r_D.sec             MSLNPQYEEIGKGFVQQYYAILDDLANRENAVNFYSVTDSFMTFFGHQIQ 50
Ntf2_R.norvegicus        MGDKPIWEQIGSSFIQHYYQLFDN--DRTQLGAIY-IDASCLTWEGQQFQ 47
                         *. :  :*:**. *:*:** ::*:  :* :   :*    * :*:**.*:*

Dntf2_D.sim              GAPKILEKVQSLSFQFITRVITTVDSQPTFDGGVLINVLGRICCIDIPFH 100
Dntf2_D.sec              GATKILEKVQSLSFQFITRVITTVDSQPTFDGGVLINVLGRICCIDIPFH 100
Dntf2_PA_D.mel           GAPKILEKVQSLSFQFITRVITTVDSQPTFDGGVLINVLGRICCIDIPFH 100
Dntf2r_D.mel             GAPKILEKVQSLSFQFIARVITTVDSQPTSDGGVLIIVLGRIKCIDIPFH 100
Dntf2r_D.sim             GAPKILEKVQSLSFQFISIVITTVDSQPTFDGGVLISVLGRIKCIDIPFH 100
Dntf2r_D.sec             GAPKILEKVQSLSFQFISIVITTVDSQPTFDGGVLIFVLGRIKCIDIPFH 100
Ntf2_R.norvegicus        GKAAIVEKLSSLSFQFIQHSITAQDHQPTFDSCIISMVVGQIKADIPIM 97
                         * . *:**:.** ****   **: * *** *. ::  *:*:*:.*:**

Dntf2_D.sim              AFSQVFFLKANAGTFFVAHDIFRLNIHNSA 130
Dntf2_D.sec              AFSQVFVLKANAGTFFVAHDIFRLNIHNSA 130
Dntf2_PA_D.mel           AFSQVFFLKANAGTFFVAHDIFRLNIHNSA 130
Dntf2r_D.mel             AFSQIFLLKPNGGSLFVAHDIFRLNIHNSA 130
Dntf2r_D.sim             SFSQIFLLKPNGGSFLVAHDIFRLNIHNSA 130
Dntf2r_D.sec             SFSQIFLLKPNGGSFLVAHDIFRLNIHNSA 130
Ntf2_R.norvegicus        GFHQMFLLKNINDAWVCTNDMFRLALHNFG 127
                         .* *:*.**   .: . ::*:*** :** .
```

Red box - Interacting interface with RanGDP.

Orange – Amino acids interacting with FxFG repeats of nucleoporins.

D.ananassae

```
Dntf2_Retana             MPLNPHYEPMGQEFVKQYYVIFDNPATRALTATFFSHNDSFMTFFGEQVLGYYKIFEKVK 60
Dntf2_D.ana              MSLNPQYEDIGKGFVQQYYAIFDDPANRANVVNFYSATDSFMTFFGHQIQGAPKILEKVQ 60
Ntf2_R.norvegicus        MGDKPIWEQIGSSFIQHYYQLFDN--DRTQLGAIY-IDASCLTWEGQQFQGKAAIVEKLS 57
                         *  :* :* :*. *:::** :**:   *:    ::    * :*:**.*. *   *.**:.

Dntf2_Retana             SIFFQKVNRTLTNVDCQPTGDGGILMSVLGRICCIDIPSLSFSQIFVLKPDTSPNAYYLS 120
Dntf2_D.ana              SISFQFITRVITTVDSQPTFDGGVLINVLGRICCIDIPPHAFSQVFVLKANAG--TFFVA 118
Ntf2_R.norvegicus        SIFFQKIQHSITAQDHQPTPDSCIISMVVGQIKADIPIMGFHQMFLLKNIND--AWVCT 115
                         **.***: : :*  * *** *. ::  *:*:*:.*:**  .* *:*:**   .  ::  :

Dntf2_Retana             HDIFRLNIHDTE 132
Dntf2_D.ana              HDIFRLNIHNSA 130
Ntf2_R.norvegicus        NDMFRLALHNFG 127
                         :*:*** :*:
```

D.grimshawi

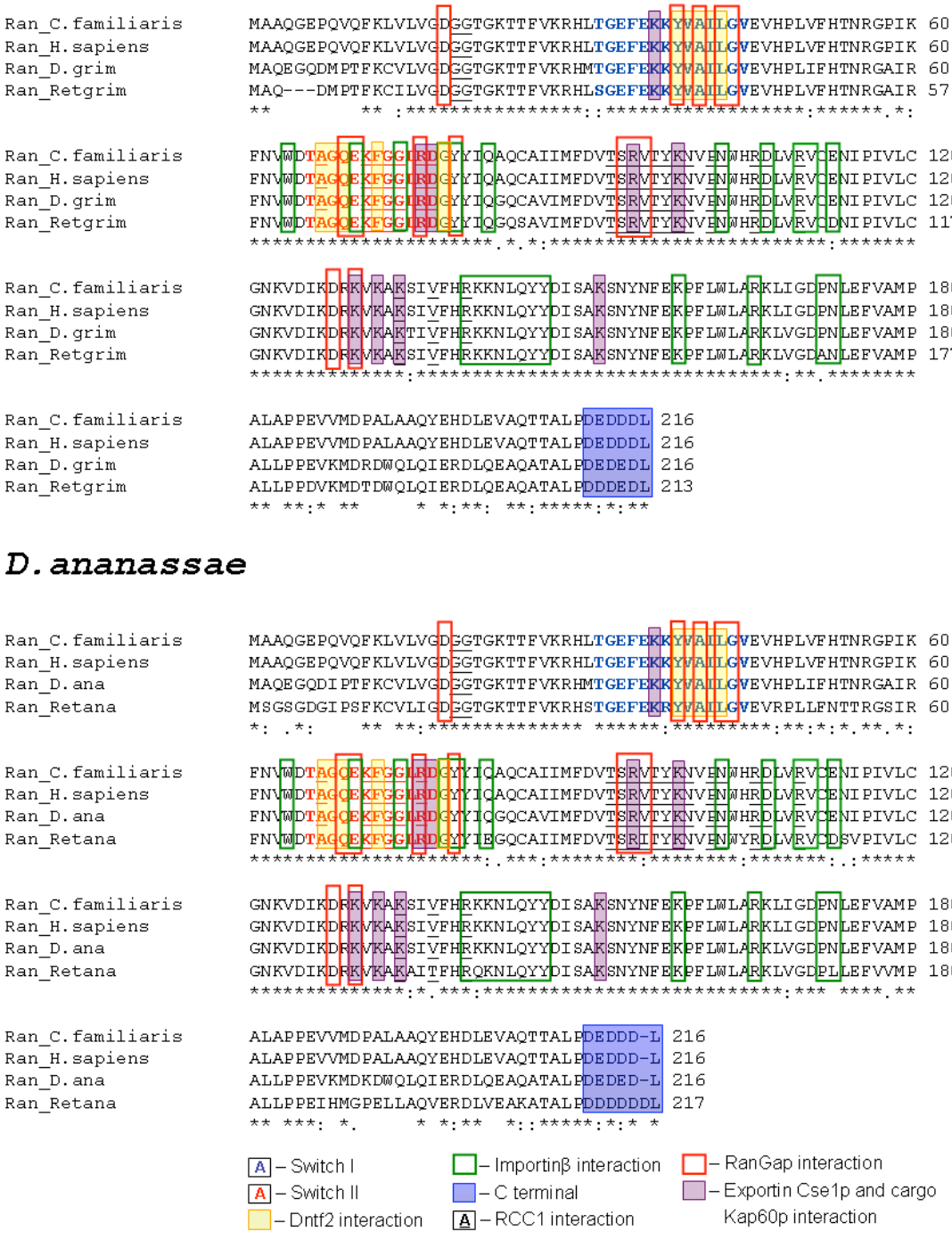
```
Dntf2_Retgrim            MAINPQYEAVGKGFVQQYYAIFDDPANRANVVNFYSTTDSFMTFFGHQIQGAPKILEKVQ 60
Dntf2_D.grim             MALNPQYEDIGKGFVQQYYAIFDDPANRANVVNFYSATDSFMTFFGHQIQGAPKILEKVQ 60
Ntf2_R.norvegicus        MGDKPIWEQIGSSFIQHYYQLFDN--DRTQLGAIY-IDASCLTWEGQQFQGKAAIVEKLS 57
                         *. :* :* :*..*:*:** :**:  :*:::  :*    * :*:**:*:** . *:**:.

Dntf2_Retgrim            SISFQKINRIITTIDSQPTFDGGVLINVLGRIQCIRIPPHSYSQVFLLKANAGSFFVAHD 120
Dntf2_D.grim             SISFQKISRVITTVDSQPTFDGGVLINVLGRVQCIDIPPLAFSQVFVLKANAGTYYVAHD 120
Ntf2_R.norvegicus        SIFFQKIQHSITAQDHQPTPDSCIISMVVGQIKADIPIMGFHQMFLLKNINDAWVCTND 117
                         **.****.: **: * *** *. ::  *:*:::.*:**  .: *:*:**   .::  ::*

Dntf2_Retgrim            IFRLNIHNSA 130
Dntf2_D.grim             IFRLNIHNSA 130
Ntf2_R.norvegicus        MFRLALHNFG 127
                         :*** :** .
```

B. Alignments of Da_Ran-like (Ran_Retana in the alingment) and Dg_Ran-like (RanRetgrim in the alingment) retrogenes with their parental and with dog Ran are shown below. Residues that were inferred to be of major functional importance from published crystallographic results (see text for references) are outlined.

D.grimshawi



D.ananassae

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 *K*S is saturated between parental and retrogenes. The *K*S is 3.3193 for *Da_ran-like* and is 1.5144 for *Dg_ran-like* calculated using PAML. Accordingly, the *K*A/*K*S 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. BATTISTUZ, 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 A resolution. Cell **97:** 635-646.

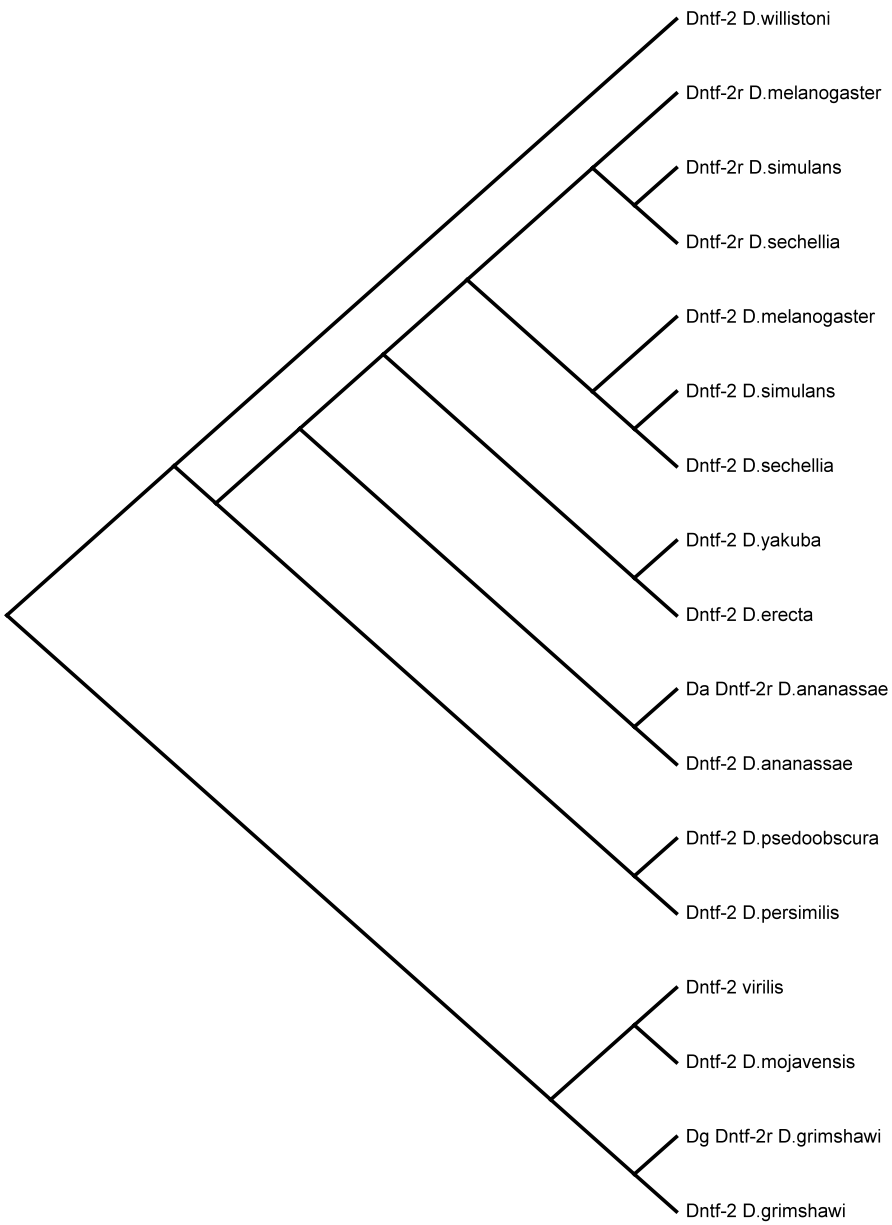


FIGURE S1.—Tree provided for the PAML branch analyses of *Dntf-2* and retrogenes

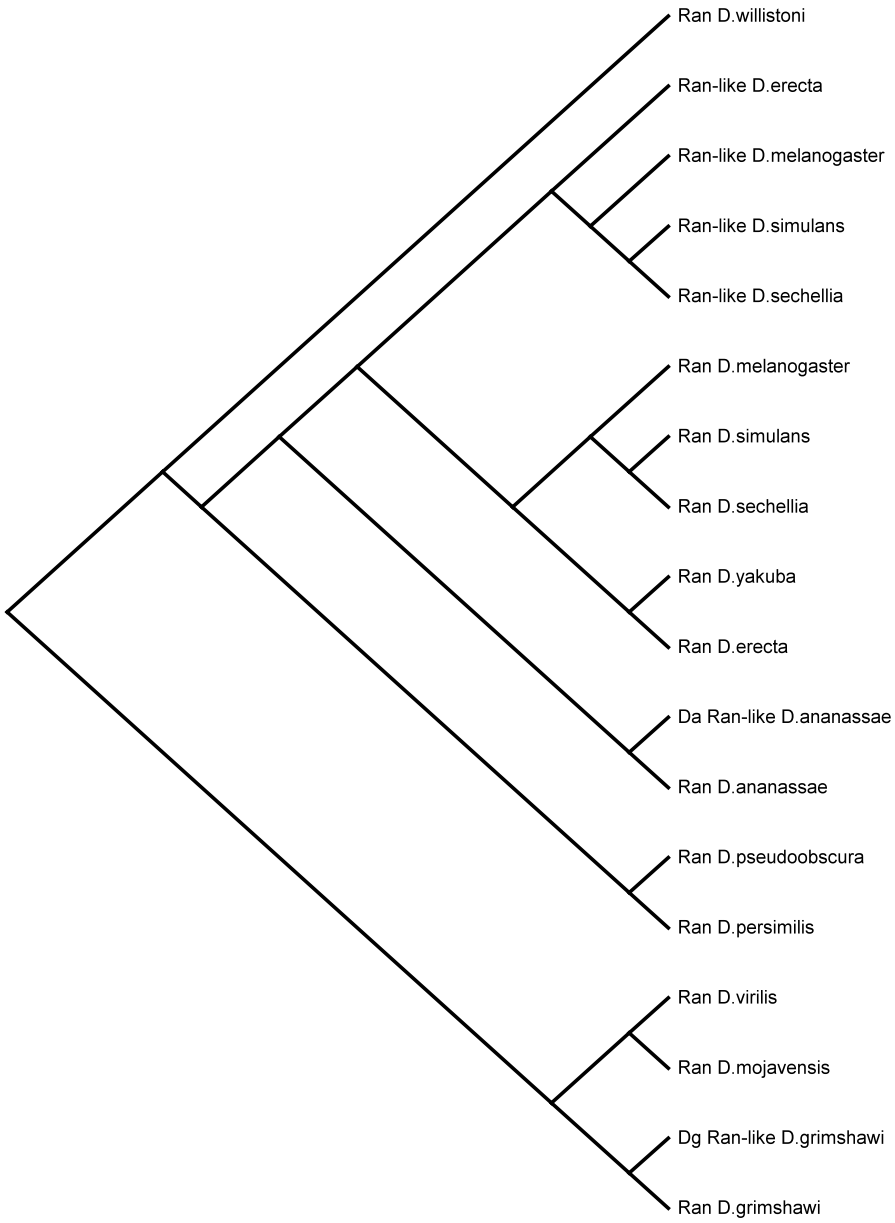


FIGURE S2.—Tree provided for the PAML branch analyses of *ran* and retrogenes

```
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           GACACTGCCGGTCAGGAGAAGTTCGGTGGCCTACAAGATGGGTATTATGTCCAAGGTCAA 240
Tai30           GACACTGCCGGTCAGGAGAAGTTCGGTGGCCTACAAGATGGGTATTATGTCCAAGGTCAA 240
Tai15b          GACACTGCCGGTCAGGAGAAGTTCGGTAGCCTACAAGATGGGTATTATGTCCAAGGTCAA 240
Tai15a          GACACTGCCGGTCAGGAGAAGTTCGGTGGCCTACAAGATGGGTATTATGTCCAAGGTCAA 240
GE22850         GACACTGCCGGTCAGGAGAAGTTCGGTGGCCTACAAGATGGGTATTATGTCCAA------ 234
GE19852         GACAC----------------------------AAGATGGGTATTATGTCCAAGGTCAA 210
Tai6            GACAC----------------------------AAGATGGGTATTATGTCCAAGGTCAA 210
Tai18           GACAC----------------------------AAGATGGGTATTATGTCCAAGGTCAA 210
Tai59           GACACTGCCGGTCAGGAGAAGTTCGGTGGCCTACAAGATGGGTATTATGTCCAAGGTCAA 240
Tai37           GACACTGCCGGTCAGGAGAAGTTCGGTGGCCTACAAGATGGGTATTATGTCCAAGGTCAA 240
Tai26           GACACTGCCGGTCAGGAGAAGTTCGGTGGCCTACAAGATGGGTATTATGTCCAAGGTCAA 239
Tai27           GACACTGCCGGTCAGGAGAAGTTCGGT--------AGACGGGTATTATGTCCAAGGTCAA 232
                *****                           *** ***************

Tai21           TGTGCCATAATAATGTTCGACGTGAGCTCGAGAATTACCTACAAGAATGTGGCACGTTGG 300
Tai30           TGTGCCATAATAATGTTCGACGTGAGCTCGAGAATTACCTACAAGAATGTGGCACGTTGG 300
Tai15b          TGTGCCATAATAATGTTCGACGTGAGCTCGAGAATTACCTACAAGAATGTGGCACGTTGG 300
Tai15a          TGTGCCATAATAATGTTCGACGTGAGCTCGAGAATTACCTACAAGAATGTGGCACGTTGG 300
GE22850         TGTGCCATAATAATGTTCGACGTGAGCTCGAGAATTACCTACAAGAATGTGGCACGTTGG 294
GE19852         TGTGCCATAATAATGTTCGACGTGAGCTCGAGAATTACCTACAAGAATGTGGCACGTTGG 270
Tai6            TGTGCCATAATAATGTTCGACGTGAGCTCGAGAATTACCTACAAGAATGTGGCACGTTGG 270
Tai18           TGTGCCATAATAATGTTCGACGTGAGCTCGAGAATTACCTACAAGAATGTGGCACGTTGG 270
Tai59           TGTGCCATAATAATGTTCGACGTGAGCTCGAGAATTACCTACAAGAATGTGGCACGTTGG 300
Tai37           TGTGCCATAATAATGTTCGACGTGAGCTCGAGAATTACCTACAAGAATGTGGCACGTTGG 300
Tai26           TGTGCCATAAAATTGTTCGACGTGAGCTCGAGAATTACCTACAAGAATGTGGCACGTTGG 299
Tai27           TGTGCCATAATAATGTTCGACGTGAGCTCGAGAATTACCTACAAGAATGTGGCACGTTGG 292
```

```
                ********** * **************************************************

Tai21           CACCGCGACTTGGTGAGGGTATGCGGCAATATTCCGATTGTTTTGTGTGGAAACAAGGTG 360
Tai30           CACCGCGACTTGGTGAGGGTATGCGGCAATATTCCGATTGTTTTGTGTGGAAACAAGGTG 360
Tai15b          CACCGCGACTTGGTGAGGGTATGCGGCAATATTCCGATTGTTTTGTGTGGAAACAAGGTG 360
Tai15a          CACCGCGACTTGGTGAGTGTATGCGGCAATATTCCGATTGTTTTGTGTGGAAACAAGGTG 360
GE22850         CACCGCGACTTGGTGAGGGTATGCGGCAATATTCCGATTGTTTTGTGTGGAAACAAGGTG 354
GE19852         CACCGCGACTTGGTGAGGGTATGCGGCAATATTCCGATTGTTTTGTGTGGAAACAAGGTG 330
Tai6            CACCGCGACTTGGTGAGGGTATGCGGCAATATTCCGATTGTTTTGTGTGGAAACAAGGTG 330
Tai18           CACCGCGACTTGGTGAGGGTATGCGGCAATATTCCGATTGTTTTGTGTGGAAACAAGGTG 330
Tai59           CACCGCGACTTGGTGAGGGTATGCGGCAATATTCCGATTGTTTTGTGTGGAAACAAGGTG 360
Tai37           CACCGCGACTTGGTGAGGGTATGCGGCAATATTCCGATTGTTTTGTGTGGAAACAAGGTG 360
Tai26           CACCGCGACTTGGTGAGGGTATGCGGCAATATTCCGATTGTTTTGTGTGGAAACAAGGTG 359
Tai27           CACCGCGACTTGGTGAGGGTATGCGGCAATATTCCGATTGTTTTGTGTGGAAACAAGGTG 352
                ***************** ****************************************

Tai21           GATATCAAGCAACGGAAGGTTAGGCCCAGGCGCTTTGACTTTCATCGTAAGAAAAACCTC 420
Tai30           GATATCAAGCAACGGAAGGTTAGGCCCAGGCGCTTTGACTTTCATCGTAAGAAAAACCTC 420
Tai15b          GATATCAAGCAACGGAAGGTTAGGCCCAGGCGCTTTGACTTTCATCGTAGGAAAAACCTC 420
Tai15a          GATATCAAGCAACGGAAGGTTAGGCCCAGGCGCTTTGACTTTCATCGTAAGAAAAACCTC 420
GE22850         GATATCAAGCAACGGAAGGTTAGGCCTAGGCGCTTTGACTTTCATCGTAAGAAAAACCTC 414
GE19852         GATATCAAGCAACGGAAGGTTAGGCCCAGGCGCTTTGACTTTCATCGTAAGAAAAACCTC 390
Tai6            GATATCAAGCAACGGAAGGTTAGGCCCAGGCGCTTTGACTTTCATCGTAAGAAAAACCTC 390
Tai18           GATATCAAGCAACGGAAGGTTAGGCCCAGGCGCTTTGACTTTCATCGTAAGAAAAACCTC 390
Tai59           GATATCAAGCAACGGAAGGTTAGGCCCAGGCGCTTTGACTTTCATCGTAAGAAAAACCTC 420
Tai37           GATATCAAGCAACGGAAGGTTAGGCCCAGGCGCTTTGACTTTCATCGTAAGAAAAACCTC 420
Tai26           GATATCAAGCAACGGAAGGTTAGGCCCAGGCGCTTTGACTTTCATCGTAAGAAAAACCTC 419
Tai27           GATATCAAGCAACGGAAGGTTAGGCCCAGGCGCTTTGACTTTCATCGTAAGAAAAACCTC 412
                ************************** ********************** **********

Tai21           CACTACATTGAAATGTCCGCCAAGTCAAACTATA----ACATTGAGAGTCCCTTCGTCTA 476
Tai30           CACTACATTGAAATGTCCGCCAAGTCAAACTATA----ACATTGAGAGTCCCTTCGTCTA 476
Tai15b          CACTACATTGAAATGTCCGCCAAGTCAAACTATA----ACATTGAGAGTCCCTTCGTCTA 476
Tai15a          CACTACATTGAAATGTCCGCCAAGTCAAACTATA----ACATTGAGAGTCCCTTCGTCTA 476
```

```
GE22850         CACTACATTG-----TCCGCCAAGTCAAACTATA----ATATTGAGAGTCCCTTCGTCTA 465
GE19852         CACTACATTGAAATGTCCGCCAAGTCAAACTATA----ACATTGAGAGTCCCTTCGTCTA 446
Tai6            CACTACATTGAAATGTCCGCCAAGTCAAACTATA----ACATTGAGAGTCCCTTCGTCTA 446
Tai18           CACTACATTGAAATGTCCGCCAAGTCAAACTATA----ACATTGAGAGTCCCTTCGTCTA 446
Tai59           CACTGCATTGAAATGTCCGCCAAGTCAAACTATACTTGACATTGATAGTCCCTTCGTCTA 480
Tai37           CACTGCATTGAAATGTCCGCCAAGTCAAACTATACTTGACATTGATAGTCCCTTCGTCTA 480
Tai26           CACTACATTGAAATGTCCGCCAAGTCAAACTATA----ACATTGAGAGTCCCTTCGTCTA 475
Tai27           CACTACATTGAAATGTCCGCCAAGTCAAACTATA----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-----------------------AATGGA 571
Tai30           ACCCCCAGAAGTGGTTTTTACCGACGACG-----------------------AATGGA 571
Tai15b          ACCCCCAGAAGTGGTTTTTACCGACGACG-----------------------AATGGA 571
Tai15a          ACCCCCAGAAGTTGTCTTTACCGACGAGATGCGCCGTCAAGTGGAACGCGGGTTAATGGA 596
GE22850         ACCCCCAGATGTGGTTTTTACCGACGAGATGCGCCGTCAAGTGGAACGCGGGTTAATGGA 585
GE19852         ACCCCCAGAAGTGGTTTTTACCGACGAGATGCGCCGTCAAGTGGAACGCGGGTTAATGGA 566
Tai6            ACCCCCAGAAGTGGTTTTTACCGACGAGATGCGCCGTCAAGTGGAACGCGGGTTAATGGA 566
Tai18           ACCCCCAGAAGTGGTTTTTACCGACGAGATGCGCCGTCAAGTGGAACGCGGGTTAATGGA 566
Tai59           ACCCCCAGAAGTGGTTTTTACCGACGAGATGCGCCGTCAAGTGGAACGCGGGTTAATGGA 600
Tai37           ACCCCCAGAAGTGGTTTTTACCGACGAGATGCGCCGTCAAGTGGAACGCGGGTTAATGGA 600
```

```
Tai26           ACTCCCAGAAGTTGTTTTTACCGACGAGATGCGCCGTCAAGTGGAACGCGGGTTAATGGA 595
Tai27           ACCCCCAGAAGTGGTTTTTACCGACGAGATGCGCCGTCAAGTGGAACGCGGGTTAATGGA 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	*ran-like*	*D. melanogaster*	Genomic PCR
5'CAAAGATCATCGTTGCAC3'	R	*ran-like*	*D. melanogaster*	Genomic PCR
5'GCTGGCGGGATAAGTTC3'	F	*ran-like*	*D. simulans*	Genomic PCR
5'CCATGGGCACGAAGTAAG3'	R	*ran-like*	*D. simulans*	Genomic PCR
5'ATTACACAAGCCGCTCC3'	F	*ran-like*	*D. yakuba*	Genomic PCR
5'ACGCAGAAGGGGAAAAG3'	R	*ran-like*	*D. yakuba*	Genomic PCR
5'ATGCCTCTCAATCCCCAC3'	F	*Da_Ntf-2r*	*D. ananassae*	Genomic PCR
5'TTATTCCGTGTCGTGGATATTC3'	R	*Da_Ntf-2r*	*D. ananassae*	Genomic PCR
5'ATGCCTCTCAATCCCCAC3'	F	*Da_Ntf-2r*	*D. atripex*	Genomic PCR
5'TTATTCCGTGTCGTGGATATTC3'	R	*Da_Ntf-2r*	*D. atripex*	Genomic PCR
5'CAATCTCCTCGTGCAGACG3'	F	*Da_ran-like*	*D. ananassae*	Genomic PCR
5'CGGAGTGTCCAATTTGTCG3'	R	*Da_ran-like*	*D. ananassae*	Genomic PCR
5'CAATCTCCTCGTGCAGACG3'	F	*Da_ran-like*	*D. atripex*	Genomic PCR
5'CGGAGTGTCCAATTTGTCG3'	R	*Da_ran-like*	*D. atripex*	Genomic PCR
5'GATATTGGCAAGGGATTCGTC3'	F	*Dntf-2*	*D. ananassae*	RT-PCR
5'CGACCAAGAACGTTAATCAG3'	R	*Dntf-2*	*D. ananassae*	RT-PCR
5' CCAATGGGCCAGGAATTTGTG 3'	F	*Da_Ntf-2r*	*D. ananassae*	RT-PCR
5'GCAGTCTTCCCAGGACACTC3'	R	*Da_Ntf-2r*	*D. ananassae*	RT-PCR
5'CACATTCAAGTGCGTACTCGTC3'	F	*ran*	*D. ananassae*	RT-PCR
5'TGGCAACGAATTCCAGGTTGG3'	R	*ran*	*D. ananassae*	RT-PCR
5'GTGGCAGCGGTGATGGTATTC3'	F	*Da_ran-like*	*D. ananassae*	RT-PCR
5'CTTGCCGGAAATGTCGTAG3'	R	*Da_ran-like*	*D. ananassae*	RT-PCR
5'ATTGTGACCACAGTCGGTTC 3'	F	*Gapdh2*	*D. ananassae*	RT-PCR
5'GTCGTACCAAGAGATCAGCTTCAC3'	R	*Gapdh2*	*D. ananassae*	RT-PCR
5'AGGACATTGGCAAGGGCTTC3'	F	*Dntf-2*	*D. grimshawi*	RT-PCR
5'TGGGCTGAGAGTCAACTGTGG 3'	R	*Dntf-2*	*D. grimshawi*	RT-PCR
5'CCGTTGGCAAAGGTTTTGTCC3'	F	*Dg_Ntf-2r*	*D. grimshawi*	RT-PCR
5'CAGTTGCGAATAGGAGTGTGGTG3'	R	*Dg_Ntf-2r*	*D. grimshawi*	RT-PCR
5'CAAGGACCGCAAGGTCAAAG3'	F	*ran*	*D. grimshawi*	RT-PCR
5'TGCAGCTGCCAGTCTCTGTG3'	R	*ran*	*D. grimshawi*	RT-PCR
5'GCTGTGTGGCAACAAAGTCG3'	F	*Dg_ran-like*	*D. grimshawi*	RT-PCR
5'CCTCCTGCAAATCTCGTTCG3'	R	*Dg_ran-like*	*D. grimshawi*	RT-PCR
5'GGTGCTGCCCAAAACATCAT3'	F	*Gapdh*	*D. grimshawi*	RT-PCR
5'GCTGAGGAAATCGGTGGAGAC3'	R	*Gapdh*	*D. grimshawi*	RT-PCR

#F and R refer to forward and reverse, respectively

TABLE S2

PAML comparisons for *Dntf-2* and *Dntf-2* retrogenes

Model	l	*p*	$\hat{w}_{Dntf\text{-}2}$	$\hat{w}_{Dntf\text{-}2r}$	$\hat{w}_{Dntf\text{-}2r\_dup}$	$\hat{w}_{Da\_Ntf\text{-}2r}$	$\hat{w}_{Dg\_Ntf\text{-}2r}$
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 $\hat{w}_{Dntf\text{-}2r} = 1$	-2524.2914	36	0.0248	1.0000	1.0000	0.3328	0.0758
Four-ratio $\hat{w}_{Da\_Ntf\text{-}2r} = 1$	-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. $\hat{w}_{Dntf\text{-}2}$ is K_A/K_S ratio for all *Dntf-2* genes. $\hat{w}_{Dntf\text{-}2r}$ is the K_A/K_S ratio for the melanogaster subgroup minus the branch immediately following duplication of *Dntf-2r*. $\hat{w}_{Dntf\text{-}2r\_dup}$ is the K_A/K_S ratio for the branch immediately following duplication in the melanogaster subgroup. $\hat{w}_{Da\_Ntf\text{-}2r}$ and $\hat{w}_{Dg\_Ntf\text{-}2r}$ the K_A/K_S ratio for the retroposed sequence in *D. ananassae* and *D. grimshawi* respectively.

TABLE S3

Polymorphism data for *Da_Ntf-2r* in *D. ananassae* and *D. atripex*

```
                                                    11111111111111111222222222222222222222233333333333333
                                  11111234455677778889999001223344677789990001222334455667888899000111222222345
                                 370678909082730125146036925437151790149015047001227024647612571203645656789302

D.atr 00                         CGTACAACTCCGACACCTTCACGTTCCCCCATTCAGCTGGTCCCCTTTCGGCATGTCGATGACGCCTGCATCAGCTGT
D.atr 01_A                       .....................................C...A........A..........................A
D.atr 01_B, 03_A                 .....................................C...A........A...........................
D.atr 02_A                       ............................T........C...A........A...........................
D.atr 02_B                       .....................................C............A........C..................
D.atr 03_B                       ...................T.......................................C..................
D.ana 16,17,18,25_A,30,35        .CACAGGA.GTATTCTTAA.T.C..ATTAGTCCTCCACAACAAGTCGGTA.GGGCCTTT.CTACTAGA..AATTGCT.
D.ana 25_B,33_A                  .CACAGGA.GTATTCTTAC.T.C.C..TAGTCCTCCACAACAAGTCGGTA.GGGCCTTT.CTACTAGA..AATTGCT.
D.ana 31                         .CACAGGA.GTATTCTTAA.T.C..ATTAGTCCTCCACAACAAGTCGGTA.GGGCCTTT.CTACTAGATCAATTGCT.
D.ana 32                         TCACAGGA.GTATTCTTAA.T.C...TTAGTCCTCCACAACTAGTCGGTA.GGGCCTTT.CTACTAGA..AATTGCT.
D.ana 33_B                       .CACAGGA.GTATTCTTAC.TTCC..TTAGTCCTCCACAACAAGTCGGTA.GGGCCTTT.CTACTAGATCAATTGCT.
D.ana 34                         .CACAGGACGTATTCTTAC.T.C.C..TAGTCCTCCACAACAAGTCGGTA.GGGCCTTT.CTACTAGA..AATTGCT.
```

TABLE S4

PAML comparisons for *ran* and *ran* retrogenes

Model	l	*p*	$\hat{w}_{ran}$	$\hat{w}_{ran\text{-}like}$	$\hat{w}_{ran\text{-}like\_dup}$	$\hat{w}_{Da\_ran\text{-}like}$	$\hat{w}_{Dg\_ran\text{-}like}$
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 $\hat{w}_{ran\text{-}like}$**=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. $\hat{w}_{ran\text{-}like}$ is the K_A/K_S ratio for the melanogaster subgroup minus the branch immediately following duplication of *ran-like*. $\hat{w}_{ran\text{-}like\_dup}$ is the K_A/K_S ratio for the branch immediately following duplication in the melanogaster subgroup. $\hat{w}_{Da\_ran\text{-}like}$ and $\hat{w}_{Dg\_ran\text{-}like}$ the K_A/K_S ratio for the retroposed sequence in *D. ananassae* and *D. grimshawi* respectively

TABLE S5

Polymorphism data for *ran-like* in *D. melanogaster* and *D. simulans*

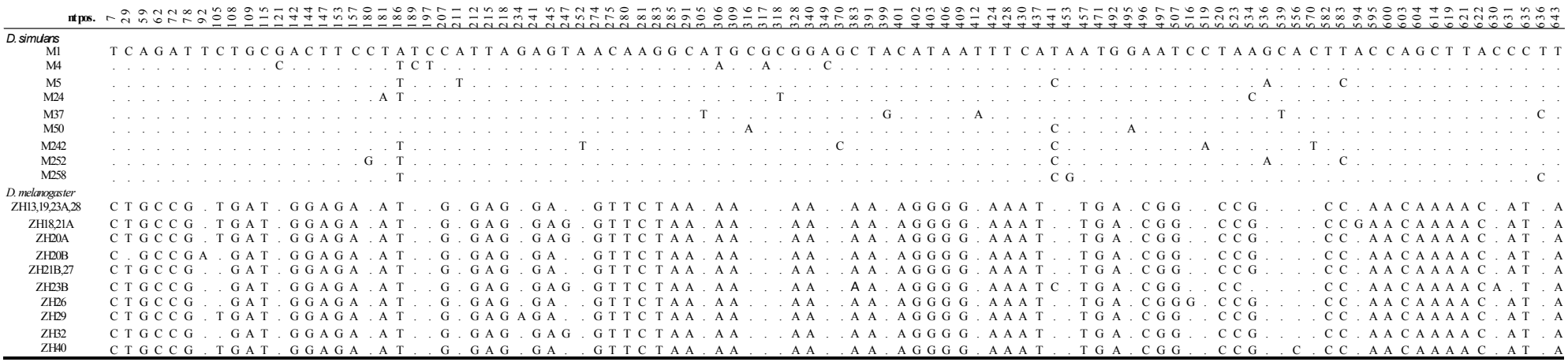


TABLE S6

Polymorphism data for *Da_ran-like* in *D. ananassae* and *D. atripex*

	nt.	8	19	20	57	69	85	93	96	105	108	111	114	120	126	135	153	162	174	189	234	244	246	247	249	264	297	303	318	336	341	343	354	363	390	421	423	430	432	438	441	447	453	459	465	490	498	502	513	516	517	531	546	552	558	564	571	573	578	581	591	597	605	612	618	625	627	636	648
D.atr 00_A		C	T	C	A	A	A	T	T	C	G	G	G	C	C	A	A	T	C	A	C	C	A	A	T	T	G	T	G	T	G	A	G	G	A	A	G	C	G	T	T	T	C	T	T	C	C	A	C	C	T	C	T	T	G	A	G	G	G	A	T	G	C	T	T	T	A	T	C
D.atr 00_B, 02		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.
D.atr 01 , 03		.	.	.	.	.	.	.	.	.	A	.	A	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	T
D.ana 16, 17, 35		G	G	G	T	G	C	C	.	T	.	.	.	G	T	C	T	C	.	G	A	G	G	G	G	A	.	G	C	C	C	G	A	A	G	C	.	T	.	C	C	.	.	C	C	T	A	T	T	T	C	T	C	C	A	C	C	C	T	T	G	.	T	.	A	C	G	.	.
D.ana 18_A		G	G	G	T	G	C	C	.	T	.	.	.	G	T	C	T	C	.	G	A	G	G	G	G	A	.	G	C	C	C	G	A	A	G	C	.	T	.	C	C	.	.	C	C	T	A	T	T	T	C	T	C	C	A	C	C	C	T	T	G	.	T	G	A	C	G	.	.
D.ana 18_B		G	G	G	T	G	C	C	.	T	.	.	.	G	T	C	T	C	.	G	A	G	G	G	G	A	A	G	C	C	C	G	A	A	G	C	.	T	.	C	C	.	.	C	C	T	A	T	T	T	C	T	C	C	A	C	C	C	T	T	G	.	T	.	A	C	G	.	.
D.ana 25		G	G	G	.	.	C	C	.	T	.	A	.	G	T	C	T	C	.	G	A	G	G	G	G	A	.	G	C	C	C	G	A	A	G	C	.	T	.	C	C	.	.	C	C	T	A	T	T	T	C	T	C	C	A	C	C	C	T	T	G	.	T	.	A	C	G	.	.
D.ana 30		G	G	G	T	G	C	C	.	T	.	.	.	G	T	C	T	C	.	G	A	G	G	G	G	A	.	G	C	C	C	G	A	A	G	C	.	T	.	C	C	.	.	C	C	T	A	T	T	T	C	T	C	C	A	C	C	C	T	T	G	.	T	.	A	C	G	C	.
D.ana 31		G	G	G	T	G	C	C	.	T	.	.	.	G	T	C	T	C	.	G	A	G	G	G	G	A	.	G	C	C	C	G	A	A	G	C	.	.	.	C	C	A	.	.	C	T	A	T	T	T	C	T	C	C	A	C	C	A	T	T	G	T	T	.	A	C	G	.	.
D.ana 32_A		G	G	G	.	.	C	C	.	T	.	A	.	G	T	C	T	C	.	G	A	G	G	G	G	A	.	G	C	C	C	G	A	A	G	C	A	.	.	C	C	A	.	.	C	T	A	T	T	T	C	T	C	C	A	C	C	A	T	T	G	T	T	.	A	C	G	.	.
D.ana 32_B		G	G	G	.	.	C	C	C	T	.	.	.	G	T	C	T	C	T	G	A	G	G	G	G	A	A	G	C	C	C	G	A	A	G	C	.	.	A	C	C	A	.	.	C	T	A	T	T	T	C	T	C	C	A	C	C	A	T	T	G	T	T	.	A	C	G	.	.
D.ana 33_A		G	G	G	T	.	C	C	.	T	.	.	.	G	T	C	T	C	.	G	A	G	G	G	G	A	A	G	C	C	C	G	A	A	G	C	.	.	.	C	C	A	T	.	C	T	A	T	T	T	C	T	C	C	A	C	C	A	T	T	G	T	T	.	A	C	G	.	.
D.ana 33_B		G	G	G	.	.	C	C	.	T	.	.	.	G	T	C	T	C	.	G	A	G	G	G	G	A	A	G	C	C	C	G	A	A	G	C	.	.	.	C	C	A	.	.	C	T	A	T	T	T	C	T	C	C	A	C	C	A	T	T	G	T	T	.	A	C	G	.	.
D.ana 34		G	G	G	.	.	C	C	C	T	.	.	.	G	T	C	T	C	T	G	A	G	G	G	G	A	A	G	C	C	C	G	A	A	G	C	.	.	.	C	C	A	.	.	C	T	A	T	T	T	C	T	C	C	A	C	C	A	T	T	G	T	T	.	A	C	G	.	.