Copyright © 1994 by the Genetics Society of America

Molecular Population Genetics of Sex Determination Genes: The *transformer* Gene of *Drosophila melanogaster*

C. Scott Walthour and Stephen W. Schaeffer

Department of Biology and Institute of Molecular Evolutionary Genetics, The Pennsylvania State University, University Park, Pennsylvania 16802

Manuscript received July 16, 1993
Accepted for publication December 17, 1993

ABSTRACT

The *transformer locus* (*tra*) produces an RNA processing protein that alternatively splices the *doublesex* pre-mRNA in the sex determination hierarchy of *Drosophila melanogaster.* Comparisons of the *tra* coding region among Drosophila species have revealed an unusually high degree of divergence in synonymous and nonsynonymous sites. In this study, we tested the hypothesis that the *tra* gene will be polymorphic in synonymous and nonsynonymous sites within species by investigating nucleotide sequence variation in eleven *tra* alleles within *D. melanogaster.* Of the 1063 nucleotides examined, two synonymous sites were polymorphic and no amino acid variation was detected. Three statistical tests were used to detect departures from an equilibrium neutral model. Two tests failed to reject a neutral model of molecular evolution because of low statisitical power associated with low levels of genetic variation (Tajima/Fu and Li). The Hudson, Kreitman, and Aguade test rejected a neutral model when the *tra* region was compared to the 5′-flanking region of alcohol dehydrogenase (*Adh*). The lack of variability in the *tra* gene is consistent with a recent selective sweep of a beneficial allele in or near the *tra* locus.

THE *transformer* (*tra*) gene in *Drosophila melanogaster* encodes one of the proteins necessary for correct sexual differentiation of somatic cells in females (BAKER and BELOTE 1983; STEINMANN-ZWICKY *et al.* 1990). *tra* regulates the expression of *doublesex*, a gene whose male-specific form represses female-specific differentiation genes and whose female-specific form represses male-specific differentiation genes (BAKER 1989). Loss-of-function mutations in *tra* have no effect in chromosomal males (*XY*), but chromosomal females (*XX*) fail to express the female form of *doublesex* and are transformed into sterile adult flies with male morphological characters such as sex combs, male-colored abdomen, external genitalia, and reduced gonads (STURTEVANT 1945).

The sex-specific regulation of *tra* is mediated by alternative splicing of its pre-mRNA (BOGGS *et al.* 1987). The default splicing pathway of *tra* used by both sexes encodes a truncated protein that has no effect on the *doublesex* gene. A female-specific protein, *Sex-lethal*, alternatively splices the *tra* pre-mRNA to yield a message that encodes an RNA-binding protein (Figure 1) (NAGOSHI *et al.* 1988; SOSNOWSKI *et al.* 1989; INOUE *et al.* 1990). The transformer protein and a second RNA-binding protein, transformer-2, alternatively splice the *doublesex* pre-mRNA in females (AMREIN *et al.* 1988; BAKER and WOLFNER 1988; BURTIS and BAKER 1989; GORALSKI *et al.* 1989; INOUE *et al.* 1992). The default *doublesex* transcript produced in males encodes the protein that represses female development while the alternatively spliced mRNA of females encodes the protein that represses male development.

The importance of the transformer protein in sexual development suggests that its amino acid sequence should be highly conserved throughout the evolution of Drosophila. This assumption was shown to be invalid by O'NEIL and BELOTE (1992) who have demonstrated that the transformer protein has an unusually high degree of divergence when the coding regions were compared among five Drosophila species. The data of O'NEIL and BELOTE suggest that the transformer protein will be highly polymorphic within species because mutations in populations are the source of diversity between species.

The high level of protein divergence observed among *transformer* sequences may be a result of neutral forces where amino acid substitutions accumulate by mutation and random genetic drift (KIMURA 1983). Alternatively, the adaptive fixation of beneficial protein alleles over time may contribute to the accumulation of amino acid differences among species (MCDONALD and KREITMAN 1991; EANES *et al.* 1993). These two hypotheses may be distinguished by an examination of diversity in synonymous and nonsynonymous nucleotides within and between species of Drosophila. The neutral fixation of new amino acid variants will have no effect on the fixation rate of linked neutral variation such as synonymous sites (BIRKY and WALSH 1988), while fixations of protein alleles by directional selection will remove linked synonymous variation by genetic hitchhiking (KAPLAN *et al.* 1989; STEPHAN *et al.* 1992). We present here a nucleotide

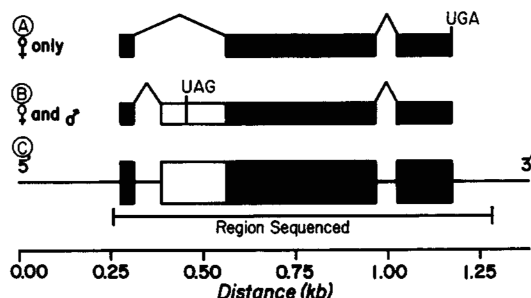

FIGURE 1.—Fine structure of the *transformer* gene of *D. melanogaster.* (A) The female-specific transcript of the *transformer* gene. The position of the termination codon UGA is indicated above the transcript. (B) The non-sex-specific transcript of the *transformer* gene. The position of the termination codon UAG is indicated above the transcript. (C) The *transformer* genomic region and the fragment of DNA that was sequenced. Sequences that are translated are represented by black boxes while sequences that are transcribed but not translated are shown as open boxes.

sequence analysis of 12 *tra* genes within and between species of the melanogaster subgroup of Drosophila to determine the level of amino acid diversity in the transformer protein and to suggest the evolutionary forces that may explain the rapid divergence of the transformer protein between species.

MATERIALS AND METHODS

Fly strains and genomic DNA preparation: Ten strains of *D. melanogaster* were collected from localities near State College, Pennsylvania. Two strains with the designation "F" were collected during August of 1989 by ANDREW G. CLARK and his colleagues and eight strains with the notation "ST" were obtained in August of 1991. The 10 strains were made homozygous for the third chromosome with the balancer stock TM3Sb/TM6 with the methods described in ASHBURNER (1989). Genomic DNA was isolated from each isochromosomal strain with the procedure of BINGHAM *et al.* (1981).

Polymerase Chain Reaction (PCR) and nucleotide sequencing: The nucleotide positions of the *tra* sequence are those of the previously published *D. melanogaster* sequence (BOGGS *et al.* 1987). A 1,111-base pair fragment was amplified from each of the 10 strains of *D. melanogaster* (SAIKI *et al.* 1988) (Figure 1), with the following oligonucleotide primers: the 5′ primer begins at nucleotide -11 (5′GAGTTGGCGGGCACATTGCAAGGT3′); and the 3′ primer begins at nucleotide 1100 (5′TAATCAACGTTCTCCGTTTGCAAT3′). Single-stranded templates for nucleotide sequencing were generated with the λ exonuclease procedure (HIGUCHI and OCHMAN 1989) and sequenced with eight oligonucleotide primers by the chain-termination method of SANGER *et al.* (1977). The eight oligonucleotide primers were separated by an average of 250 nucleotides and sequenced the *tra* regions on both strands. The overlapping sequences for each strain were assembled with the SEQMAN and SEQMANED programs (DNASTAR, Madison, Wisconsin).

The 10 *tra* sequences of *D. melanogaster* determined in this study were deposited in the GenBank/EMBL Data Libraries under the accession numbers L19464–L19470 and L19618–L19620. The sequences of previously published *tra* sequences from *D. melanogaster* and *Drosophila simulans* (GenBank/EMBL Data Library accession numbers M17478 and X66930, respectively) were also used in this study (BOGGS *et al.* 1987; O'NEIL and BELOTE 1992).

Nucleotide sequence alignments and statistical analyses: The 10 *tra* sequences generated in this study were aligned based on the published comparison of the *D. melanogaster* and *D. simulans* genes (O'NEIL and BELOTE 1992). The aligned sequences were assembled with the Eyeball Sequence Editor [ESEE, version: 1.09; CABOT and BECKENBACH (1989)]. Nucleotide sequences within the PCR primers were excluded from all analyses leaving a total of 1063 aligned bases. Any nucleotide site with two nucleotides present was defined as a segregating or polymorphic site. Nucleotide sites found in insertions or deletions assumed in aligned sequences were excluded from further analyses.

We estimated heterozygosity per nucleotide site for *tra* sequence domains with two methods. NEI (1987; equations 10.6 and 10.9) estimates nucleotide diversity, π, by determining the average number of differences between all pairs of sequences. An alternative measure of nucleotide variability, Θ or L, is estimated from the number of segregating sites, S [(WATTERSON (1975); equation 1.4a; NEI (1987) uses the notation, L, for the Watterson estimator Θ]. The quantities π and L are both estimates of the neutral mutation parameter $4N\mu$, where N is the effective population size and μ is the neutral mutation rate per nucleotide per generation. We also estimate a lower (L_L) and upper limit (L_U) on L that is a 95% confidence interval based on the probability distribution of S [HUDSON (1990; equations 10, 11, and 12); KREITMAN and HUDSON (1991; equation 3)].

The *tra* region was tested for departures from an equilibrium neutral model with three statistical tests. The three tests were designed to determine if the within species nucleotide diversity in *tra* is consistent with the expectations of a neutral model. The TAJIMA (1989) test computes a test statistic, D, which is the difference of two estimates of heterozygosity per locus, $M(S)$ and $M(k)$. D should not differ significantly from zero under a strictly neutral model. The TAJIMA test statistic was estimated for the synonymous variation in the *tra* region within *D. melanogaster.*

We also used the statistic suggested by HUDSON *et al.* (1987) (HKA test) to test for departures from neutral expectations. The neutral theory predicts that intraspecific polymorphism will be correlated to interspecific divergence (KIMURA 1983). The HKA test rejects a neutral model if the ratio of polymorphism to divergence differs significantly among independent loci (HUDSON *et al.* 1987), provided that one genetic locus evolves according to predictions of the neutral theory. The numbers of variable synonymous sites in the *tra* region were compared to the 5′-flanking region of alcohol dehydrogenase (*Adh*) in *D. melanogaster,* (KREITMAN and HUDSON 1991). Tests were performed on synonymous and total nucleotide variation. The between species divergence data is determined from comparisons of each locus with its homolog in *D. simulans.* The comparison of the *tra* data with the 5′-flanking region of *Adh* probably reflects the most reasonable test of neutrality because this segment has the least functional constraint on its sequence and probably has not been influenced by positive Darwinian selection in its recent history. Comparisons of the *Adh* or *Adh-Dup* coding sequences with the *tra* region are likely to be inappropriate for the HKA test because levels of nucleotide diversity have been altered in each gene by genetic hitchhiking events associated with some form of positive Darwinian selection (KREITMAN and HUDSON 1991).

FU and LI (1993, equation 32) have recently derived a test of selective neutrality that examines the number of mutations that occur on external *vs.* internal branches of a genealogy. The test uses within and between species data to estimate the

```
      a    b  c d           e                      f    g              h
            1 1 22223333333 3344444444555555566677 7778 88888888899999 9
      2233 80 1 12581112345 7801268899456666813812 6891 23335789901225 7
      2389 21 7 97332693605 8502627956990246828813 5634 61676081301167 6
DMEL  GTGA TC A GGAGCACAAGA TCTTCATGATAATAAACCACTC TCCT TTAGATCCACCTTT T
ST03  .... .. . ........... ...................... .... .............. .
ST05  .... .. . ....A...... ...................... .... .............. .
ST06  .... .. . ........... ...................... .... .............. .
ST08  .... .. . ........... ...................... .... .C............ .
ST09  .... .. . ........... ...................... .... .............. .
ST12  .... .. . ........... ...................... .... .C............ .
ST15  .... .. . ........... ...................... .... .............. .
ST16  .... .. . ........... ...................... .... .C............ .
F283  .... .. . ........... ...................... .... .............. .
F296  .... .. . ........... ...................... .... .............. .
DSIM  AAAG CG G ACGA.GTGTCT GTCATGAACCCGACTGTTGAAT AATA CCGACGTATTAGCC G
            R                RRR R R  RR R  R           R R    R RRR
```

FIGURE 2.—Polymorphic and divergent sites in the *transformer* region. The letters a through h are used to designate the eight sequence domains of the transformer region: a, 5′-flanking sequence; b, Exon 1; c, Intron 1; d, Non-Sex-Specific Exon 2; e, Female-Specific Exon 2; f, Intron 2; g, Exon 3; and h, 3′-flanking sequence. The vertical numbers above the sequence indicate the position of the variable site in the published *D. melanogaster* sequence (DMEL). The 11 *D. melanogaster* sequences are designated: DMEL, ST03, ST05, ST06, ST08, ST09, ST12, ST15, ST16, F283, and F296. The *D. simulans* sequence is labeled DSIM. The dots indicate identical bases to the DMEL sequence. Amino acid replacements are labeled by an R underneath the DSIM sequence.

number of internal and external mutations, η_i and η_e. An external mutation is a nucleotide site where the frequency of the rare base is $1/n$ within *D. melanogaster* and is not shared with the outgroup species, *D. simulans*, while an internal mutation is any other polymorphic nucleotide site. The test statistic D is used to determine if there is an excess or deficiency of external mutations. A significant negative value of D indicates directional selection due to an excess of external mutations, while a significant positive value of D denotes balancing selection due to a deficiency of external mutations.

Fisher's exact test (SOKAL and ROHLF 1981) was used to determine if the ratio of nonsynonymous to synonymous differences was the same for polymorphisms within species *vs.* fixed differences between species (MCDONALD and KREITMAN 1991). These two ratios should be equivalent for polymorphisms and fixed differences if protein evolution between species is neutral. Fisher's exact test was used rather than the *G*-test suggested by MCDONALD and KREITMAN (1991) because one of the cells of the 2 × 2 contingency table is zero such that the *G*-test is inappropriate.

RESULTS

Nucleotide variation: Two of the 1063 nucleotide sites sequenced were polymorphic within *D. melanogaster* (Figure 2). Both segregating sites, nucleotide positions 312 and 831 in the DMEL sequence, occurred in synonymous positions. No amino acid polymorphism was observed within *D. melanogaster*. Estimates of L and π in each of the sequence domains can be found in Table 1. The lower and upper bounds on L overlap for all the eight sequence domains in the *tra* region. Estimates of nucleotide diversity were quite low for all sequence domains compared to other genes that have been examined in *D. melanogaster* (Table 1) (AQUADRO 1993). The nucleotide divergence in the *tra* region was much greater when the *D. melanogaster* and *Drosophila simulans* sequences were compared. Forty-two synonymous substitutions and 16 amino acid replacements have accumulated between the two species (Table 1, Figure 2).

Sequence length variation: One length variant was found among the 11 *D. melanogaster tra* sequences studied. The strain ST15 had a 5 base-pair insertion in the 5′-flanking region of *tra* relative to the 10 other sequences. The inserted sequence, CATTT, is identical to the five bases immediately 5′ to the site of the insertion (C. S. WALTHOUR and S. W. SCHAEFFER, unpublished data). Four deletions of sequence were observed in the *D. simulans* sequence relative to the *D. melanogaster* sequence. Deletions of 39 and 3 base pairs were found in the Female-Specific Exon 2, and Intron 2, respectively. Two deletions of one nucleotide occurred in the 3′-Flanking sequence (O'NEIL and BELOTE 1992).

Tests of the neutral mutation hypothesis: The nucleotide data in the *tra* region fails to reject a neutral model of molecular evolution with either the TAJIMA (1989) or the FU and LI (1993) tests (TAJIMA: $M(S) = 0.68$, $M(k) = 0.62$, $D = -0.29$, $P > 0.05$; FU and LI: and $\eta_e = 1$, $\eta_i = 1$, $D = -0.45$, $P > 0.05$). The two test statistics in each case above were estimated from the information in two segregating sites observed within species. Thus, these results should be viewed with caution because of the low statistical power of these tests based on so few polymorphic sites. The results of the HKA test (1987) test stand in sharp contrast to the two tests discussed above. The comparison of the *tra* region with the 5′-flanking region of *Adh* rejects an equilibrium neutral model (Table 2). The test results are significant because *tra* nucleotide diversity in synonymous and nonsynonymous sites within *D. melanogaster* was significantly lower than that expected given the genetic divergence between *D. melanogaster* and *D. simulans*. The MCDONALD-KREITMAN (1991) test failed to reject a neutral model of protein evolution between species ($P = 0.99$) (Table 3).

DISCUSSION

The prediction that the *tra* gene would be polymorphic at the amino acid level within species was not supported by our data. The high level of conservation in amino acid sequence observed within species could re-

TABLE 1

Estimates of nucleotide diversity within and between species in the *transformer* region

Region	m	S	D	L_L	L (SE)	L_U	π (SE)
5′-Flanking	58	0	4	0.000	0.000 (0.000)	0.029	0.000 (0.000)
Exon 1[a]	10	0	1	0.000	0.000 (0.000)	0.170	0.000 (0.000)
Intron 1	73	0	1	0.000	0.000 (0.000)	0.023	0.000 (0.000)
Exon 2 (*m* and *f*)	175	1	10	0.000	0.002 (0.002)	0.016	0.001 (0.002)
Exon 2 (*f*)	101	0	13	0.000	0.000 (0.000)	0.017	0.000 (0.000)
Intron 2	57	0	4	0.000	0.000 (0.000)	0.030	0.000 (0.000)
Exon 3[a]	37	1	8	0.000	0.009 (0.009)	0.074	0.012 (0.013)
3′-Flanking	109	0	1	0.000	0.000 (0.000)	0.016	0.000 (0.000)
Summary							
Exons	148	1	22	0.000	0.002 (0.002)	0.018	0.003 (0.003)
Noncoding	472	1	20	0.000	0.001 (0.001)	0.006	0.0004 (0.001)

m, number of effectively synonymous sites; *S*, number of segregating sites; *D*, number of divergent sites between the *D. melanogaster* and *D. simulans* sequences, DMEL and DSIM; L_L and L_U, lower and upper bounds on *L* for which there is a 2.5% probability of observing the same number or more extreme values of *S*; π and *L*, nucleotide diversity determined from the number of pairwise differences (Nei, 1987) and the number of segregating sites (WATTERSON 1975), respectively; and SE, standard error. Exon 2 is shown in two parts, the non-sex-specific portion (m/f) and the female-specific coding region (f).

[a] The probability of observing this number of segregating sites or larger values in this region is equal to 0.05 given an average *L* value of 0.001 among all regions of the *transformer* locus.

TABLE 2

HKA test of selective neutrality in the *transformer* region of *D. melanogaster*

	Synonymous nucleotides		Total nucleotides	
	Polymorphism	Divergence	Polymorphism	Divergence
transformer	2 (9.3)	42 (34.7)	2 (11.4)	58 (48.6)
Adh 5′-flanking	30 (22.7)	78 (85.3)	30 (20.6)	78 (87.4)
Test	4.13	$P < 0.05$	5.50	$P < 0.05$

Note: Synonymous refers to synonymous substitutions in coding and noncoding regions. Total includes both synonymous and nonsynonymous substitutions. Numbers in parentheses are expected values for the HUDSON, KREITMAN and AGUADE (HKA) (1987) test. The row designated Test gives the results of the HKA (1987) test whose statistic approximates a chi-square distribution with one degree of freedom and the probability of the chi-square value, *P*.

TABLE 3

Numbers of fixed differences and polymorphisms in nonsynonymous and synonymous sites between *D. melanogaster* and *D. simulans*

Sites	Fixed differences	Polymorphisms
Nonsynonymous	16	0
Synonymous	22	1

sult from strong functional constraints on the *tra* protein product or from a recent directional selection event in or near the *tra* locus. An important developmental protein such as *tra* might be expected to be highly conserved at the amino acid level because of its fundamental role in sex determination. If strong constraint on amino acid sequence was the main force restricting variation in the protein, then amino acid sequence should be highly conserved between species and diversity in synonymous sites should be unaffected by changes in nonsynonymous positions (BIRKY and WALSH 1988). Sixteen amino acid replacements have accumulated in the *tra* protein between *D. melanogaster* and *D. simulans*. This is much less constrained than ADH, which has only two amino acid differences between the two species (BODMER and ASHBURNER 1984). The most striking feature of the *tra* data set is the virtual lack of nucleotide diversity within *D. melanogaster* compared to the numbers of synonymous and nonsynonymous changes that have accumulated since *D. melanogaster* and *D. simulans* diverged (O'NEIL and BELOTE 1992). Therefore, strong selective constraints on *tra* may be ruled out as an explanation for the low levels of *tra* genetic variation within *D. melanogaster* because nucleotide substitutions have accumulated among *tra* sequences during long term evolution.

A recent selective sweep of a new beneficial allele in the region is the most likely explanation for the within species nucleotide diversity pattern in *tra* (MCDONALD and KREITMAN 1991; BEGUN and AQUADRO 1992). A selective sweep will decrease variation in linked sites due to the hitchhiking effect (MAYNARD SMITH and HAIGH 1974; KAPLAN *et al.* 1989; STEPHAN *et al.* 1992). The *tra* gene region was shown to have a significant deficiency of synonymous and nonsynonymous variation when the HKA (1987) test compared *tra* with the 5′ flanking region of *Adh*.

Is the *tra* locus the site of the selective sweep? The size of the region affected by a selective sweep depends on the strength of selection and the recombination rate (STEPHAN *et al.* 1992). Selective sweeps in regions of low recombination would result in a reduction of nucleotide

variation over a large region (MAYNARD SMITH and HAIGH 1974; KAPLAN *et al.* 1989), which would make it difficult to know what gene was being selected. Alternatively, selective sweeps in regions with free recombination would have more localized reductions in heterozygosity.

E. C. KINDAHL and C. F. AQUADRO have estimated relative rates of recombination or coefficients of exchange (COE) on the third chromosome of *D. melanogaster* with methods described previously (BEGUN and AQUADRO 1992). The COE estimates they have obtained range from 0.001 to approximately 0.070, where a region with a COE of zero experiences little or no recombination while a region with a COE greater than zero experience higher levels of recombination. They have estimated the COE for the *tra* gene region to be 0.015 (E. C. KINDAHL and C. F. AQUADRO, Cornell University, personal communication). The estimate of COE for *tra* suggests that recombination rates are low to moderate in the region. Thus, a selective sweep in or near the *tra* locus might be expected to effect a large to moderate sized chromosomal segment. We sequenced approximately 1 kilobase of the *tra* gene, but we do not know to what extent variation is reduced in the local region around the gene. Additional sequence from the 5′ and 3′ regions may delimit the region affected by the selective sweep in the *tra* region.

Are multiple adaptive fixations of beneficial alleles responsible for the protein evolution observed between species? The MCDONALD-KREITMAN (1991) test failed to reject a neutral model of molecular evolution, which suggests that the between species divergence from mutation and random genetic drift. The MCDONALD-KREITMAN test lacked sufficient power to reject the neutral hypothesis because we had one polymorphic site among the *D. melanogaster tra* sequences. It is possible that multiple selective sweeps have been responsible for the rapid evolution of the transformer protein, however, the recent directional selection event in or near *tra* prevent a strong inference from being drawn.

CHARLESWORTH *et al.* (1993) have considered the effect that slightly deleterious alleles have on levels of variation in the genome, the so called background selection model. They have shown that selection against slightly deleterious alleles maintained by mutation can lower genetic diversity at linked neutral sites. The effect of background selection is most pronounced in regions where recombination rates are low such as the centromeric or proximal regions of autosomes. Background selection is a less potent force in regions with free recombination because the few deleterious genes contribute to reductions in fitness levels. Background selection may contribute to reduced levels of heterozygosity in *tra* because the estimate of recombination rate in the region is low to moderate.

The HKA test assumes that the sampled populations are at equilibrium. If the equilibrium assumption is not valid, then the significant test results could suggest that the our local *D. melanogaster* population is not at equilibrium. BEGUN and AQUADRO (1993) found that populations of *D. melanogaster* collected from the United States and Zimbabwe differ significantly in genetic composition in a four-cutter survey of eight genes. Their data suggest that U.S. populations have lost variation compared to the Zimbabwe population, which is consistent with various population bottleneck scenarios. The reduced heterozygosity in *tra* may have resulted from a recent bottleneck that established U.S. populations. A population bottleneck would reduce diversity in all loci within the genome, but other loci in *D. melanogaster* have not had the severe loss of variation as that observed in *tra* (BERRY and KREITMAN 1993). An examination of nucleotide diversity in African populations may clarify whether the *tra* gene has low diversity because of a selective sweep or a recent population bottleneck.

Why would the two other statistical tests (TAJIMA 1989; FU and LI 1993) of selective neutrality fail to reject a neutral model? These two tests have little power to detect departures from a neutral model when the number of segregating sites is small. Without interspecies data, we would have concluded that transformer protein lacks diversity because of high selective constraint on the *tra* sequence. This highlights the importance of collecting between species data to control for differences in selective constraints among genes.

The data presented in this report suggest that a selective fixation of a beneficial allele occurred in or near the *tra* locus. Further data are necessary to show that *tra* was the source of the selected variation. These data are intriguing because they suggest that an important developmental gene may have been the target of adaptive evolution.

We would like to thank ELLEN L. MILLER for her technical assistance and ANDREW G. CLARK for two genomic DNAs and balancer stocks. We would also like to thank CHARLES F. AQUADRO, DAVID BEGUN, MARTHA HAMBLIN, MARTIN KREITMAN and an anonymous reviewer for helpful comments and discussions used in the preparation of this manuscript. This work was supported in part by grant GM42472 to S.W.S. from the National Institutes of Health.

LITERATURE CITED

AMREIN, H., M. GORMAN and R. NOTHIGER, 1988 The sex-determining gene *tra-2* of Drosophila encodes a putative RNA binding protein. Cell **55:** 1025–1035.

AQUADRO, C. F., 1993 Molecular population genetics of *Drosophila*, pp. 222–266 *Molecular Approaches to Fundamental and Applied Entomology*, edited by J. OAKESHOTT and M. J. WHITTEN. Springer-Verlag, New York.

ASHBURNER, M., 1989 *Drosophila: A Laboratory Handbook*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

BAKER, B. S., 1989 Sex in flies: the splice of life. Nature **340:** 521–524.

BAKER, B. S., and J. M. BELOTE, 1983 Sex determination and dosage compensation in *Drosophila melanogaster*. Ann. Rev. Genet. **17:** 345–393.

BAKER, B. S., and M. F. WOLFNER, 1988 A molecular analysis of *doublesex*, a bifunctional gene that controls both male and female sexual differentiation in *Drosophila melanogaster*. Genes Dev. **2:** 477–489.

BEGUN, D. J., and C. F. AQUADRO, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster.* Nature **356:** 519–520.

BEGUN, D. J., and C. F. AQUADRO, 1993 African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature **365:** 548–550.

BERRY, A., and M. KREITMAN, 1993 Molecular analysis of an allozyme cline: alcohol dehydrogenase in *Drosophila melanogaster* on the east coast of North America. Genetics **134:** 869–893.

BINGHAM, P. M., R. LEVIS and G. M. RUBIN, 1981 Cloning of DNA sequences from the white locus of D. melanogaster by a novel and general method. Cell **25:** 693–704.

BIRKY, C. W., and J. B. WALSH, 1988 Effects of linkage on rates of molecular evolution. Proc. Natl. Acad. Sci. USA **85:** 6414–6418.

BODMER, M., and M. ASHBURNER, 1984 Conservation and change in the DNA sequences coding for alcohol dehydrogenase in sibling species of *Drosophila.* Nature **309:** 425–430.

BOGGS, R. T., P. GREGOR, S. IDRISS, J. M. BELOTE and M. MCKEOWN, 1987 Regulation of sexual differentiation in D. melanogaster via alternative splicing of RNA from the *transformer* gene. Cell **50:** 739–747.

BURTIS, K. C., and B. S. BAKER, 1989 Drosophila *doublesex* gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. Cell **56:** 997–1010.

CABOT, E. L., and A. T. BECKENBACH, 1989 Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. Comput. Appl. Biosci. **5:** 233–234.

CHARLESWORTH, B., M. T. MORGAN and D. CHARLESWORTH, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics **134:** 1289–1303.

EANES, W. F., M. KIRCHNER and J. YOON, 1993 Evidence for adaptive evolution of the G6pd gene in the *Drosophila melanogaster* and *Drosophila simulans* lineages. Proc. Natl. Acad. Sci. USA **90:** 7475–7479.

FU, Y.-X., and W.-H. LI, 1993 Statistical tests of neutrality of mutations. Genetics **133:** 693–709.

GORALSKI, T. J., J.-E. EDSTROM and B. S. BAKER, 1989 The sex determination locus *transformer-2* of Drosophila encodes a polypeptide with similarity to RNA binding proteins. Cell **56:** 1011–1018.

HIGUCHI, R. G., and H. OCHMAN, 1989 Production of single-stranded DNA templates by exonuclease digestion following the polymerase chain reaction. Nucleic Acids Research **17:** 5865.

HUDSON, R. R., 1990 Gene genealogies and the coalescent process. Oxf. Surv. Evol. Biol. **7:** 1–44.

HUDSON, R. R., M. KREITMAN and M. AGUADE, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153–159.

INOUE, K., K. HOSHIJIMA, H. SAKAMOTO and Y. SHIMURA, 1990 Binding of the *Drosophila Sex-lethal* gene product to the alternative splice site of *transformer* primary transcript. Nature **344:** 461–463.

INOUE, K., K. HOSHIJIMA, I. HIGUCHI, H. SAKAMOTO and Y. SHIMURA, 1992 Binding of the *Drosophila* transformer and transformer-2 proteins to the regulatory elements of doublesex primary transcript for sex-specific RNA processing. Proc. Natl. Acad. Sci. USA **89:** 8092–8096.

KAPLAN, N. L., R. R. HUDSON and C. H. LANGLEY, 1989 The "hitchhiking effect" revisited. Genetics **123:** 887–899.

KIMURA, M., 1983 *The Neutral Theory of Molecular Evolution.* Cambridge University Press, New York.

KREITMAN, M., and R. R. HUDSON, 1991 Inferring the evolutionary histories of the *Adh* and *Adh-Dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. Genetics **127:** 565–582.

MAYNARD SMITH, J., and J. HAIGH, 1974 The hitch-hiking effect of a favorable gene. Genet. Res. **23:** 23–35.

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

NAGOSHI, R. N., M. MCKEOWN, K. C. BURTIS, J. M. BELOTE and B. S. BAKER, 1988 The control of alternative splicing at genes regulating sexual differentiation in D. melanogaster. Cell **53:** 229–236.

NEI, M., 1987 *Molecular Evolutionary Genetics.* Columbia University Press, New York.

O'NEIL, M. T., and J. M. BELOTE, 1992 Interspecific comparison of the *transformer* gene of Drosophila reveals an unusually high degree of evolutionary divergence. Genetics **131:** 113–128.

SAIKI, R. K., D. H. GELFAND, S. STOFFEL, S. J. SCHARF, R. HIGUCHI *et al.*, 1988 Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239:** 487–491.

SANGER, F., S. NICKEN and A. R. COULSON, 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SOKAL, R. R., and F. J. ROHLF, 1981 *Biometry.* W. H. Freeman and Co., New York.

SOSNOWSKI, B. A., J. M. BELOTE and M. MCKEOWN, 1989 Sex-specific alternative splicing of RNA from the *transformer* gene results from sequence-dependent splice site blockage. Cell **58:** 449–459.

STEINMANN-ZWICKY, M., H. AMREIN and R. NOTHIGER, 1990 Genetic control of sex determination in *Drosophila.* Adv. Genet. **27:** 189–237.

STEPHAN, W., T. H. E. WIEHE and M. W. LENZ, 1992 The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. Theor. Popul. Biol. **41:** 237–254.

STURTEVANT, A. H., 1945 A gene in Drosophila melanogaster that transforms females into males. Genetics **30:** 297–299.

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

WATTERSON, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Pop. Biol. **7:** 256–276.

Communicating editor: W. F. EANES