

Absence of Protein Polymorphism in the *Ras* Genes of *Drosophila melanogaster*

Robert Gasperini, Greg Gibson*

Department of Biology, University of Michigan, Ann Arbor, MI 48109-1048, USA

Received: 5 November 1998 / Accepted: 26 March 1999

Abstract. Sequence analysis of 27 alleles of each of the three *Ras*-related genes in *Drosophila melanogaster* indicates that they all have low levels of polymorphism but may experience slightly different evolutionary pressures. No amino acid replacement substitutions were indicated in any of the sequences, or in the sibling species *D. simulans* and *D. mauritiana*. The *Dras1* gene, which is the major *ras* homologue in *Drosophila*, has less within-species variation in *D. melanogaster* relative to the amount of divergence from the sibling species than does *Dras2*, although the contrast was not significant by the HKA test. *Dras2* appears to be maintaining two classes of haplotype in *D. melanogaster*, one of which is closer to the alleles observed in the sibling species, suggesting that this is not likely to be a pseudogene despite the absence of a mutant phenotype. Although differences in level of expression may affect the function of the genes, it is concluded that genetic variation in the *Ras* signal transduction pathways cannot be attributed to catalytic variation in the *Ras* proteins.

Key words: *Drosophila melanogaster* — *Ras* — *Rap* — HKA test — Polymorphism — Oncogene

Introduction

The *Ras* family of protooncogenes are central regulators of cell proliferation and differentiation in animals (Paw-

son 1993). They encode cytoplasmic GTP/GDP binding proteins that transduce signals received by receptors at the cell surface (Katz and McCormick 1997). There are three classes of *Ras* genes in mammals (Valencia et al. 1991), the *H-Ras* and *K-Ras* protooncogenes, the *Ras*-related teratocarcinoma *TC21* genes, and the *Rap* *Ras*-related genes. Molecular genetic analysis indicates that *Ras* proteins are involved in an enormous array of developmental decisions, and there is increasing evidence that the functions of the different types diverged early in animal evolution.

In *Drosophila*, there are three *Ras* genes, all located in euchromatic portions of the third chromosome, which also appear to have distinct functions. Each of these genes shows between 71 and 93% sequence identity with one of the mammalian types over the full 550+ amino acids of the protein, which is significantly greater relatedness than among the three genes of a single species. *Dras1* is essential for development, as loss-of-function mutations disrupt processes as diverse as axial pattern formation, segmentation, and organogenesis (Wassarman et al. 1995). Like *Dras3*, otherwise known as *Roughened* (Hariharan et al. 1991), it also has a very well-characterized role in photoreceptor determination in the eye imaginal discs. *Dras2*, in contrast, is known solely as an open reading frame, despite extensive screens for associated mutant phenotypes (Harrison et al. 1995).

Given their involvement in both normal and uncontrolled cell proliferation, it is interesting to ask how genetic variation might affect *Ras* function, and in particular whether the *Ras* genes themselves are polymorphic. Two simple scenarios might be considered. On the one

* Present address: Department of Genetics, Gardner Hall, North Carolina State University, Raleigh, NC 27695-7614, USA
Correspondence to: Greg Gibson; e-mail: ggibson@unity.ncsu.edu

hand, it could be argued that since the *Ras* sequence is so highly conserved across divergent taxa, there must be very strong selective constraints acting on the gene. This would lead to an expectation of a relatively low sequence diversity within species in the genes and a low divergence between related species. On the other hand, by analogy with enzymes such as Pgi that play central roles in metabolism and yet can be unusually polymorphic (Watt 1994; Katz and Harrison 1997), it could be predicted that the *Ras* genes harbor a complex series of functionally distinct haplotypes. These would be maintained by balancing selection on isoforms with different optimal functions in different developmental processes. In this case, the genes might show a high ratio of within-to-between species polymorphism that could be detected by comparison of the sequences of multiple alleles.

One line of evidence in favor of the latter hypothesis is that despite the stability of Ras-mediated developmental processes, Ras activity is to some extent dosage dependent (Karim et al. 1996). There is also substantial genetic variation affecting the Ras/MAP-kinase pathway in photoreceptor determination, as shown by the divergent effects of wild-type genetic backgrounds on the sensitivity of eye development to constitutive activation of the signaling molecules that activate the pathway (Polaczyk et al. 1998). Similarly, different mouse strains respond quite distinctly to mutation of growth factor receptors that activate the Ras pathway (Threadgill et al. 1995). If *ras* genes themselves contribute to this variation, they would be expected to show evidence of functionally distinct polymorphisms. Alternatively, the *ras* genes could be monomorphic, with variation attributable solely to Ras cofactors.

Here we report an analysis of the sequences of approximately 1 kb (kilobase) of 27 alleles of each of the three *Drosophila Dras* genes. The genes appear to be experiencing slightly different evolutionary pressures as inferred from their patterns of nucleotide diversity. All three genes are among the most highly conserved genes yet documented in the fruitfly, showing essentially no variation in the coding sequence, and just a handful of synonymous substitutions relative to two sibling species. *Dras1* is further characterized by relatively low polymorphism in the introns and flanking noncoding sequences. In contrast, *Dras2* and *Dras3* have more normal levels of within species polymorphism, although a high level of shared polymorphism in *Dras2* with two sibling species suggests the maintenance of variation at this locus.

Materials and Methods

Source of Chromosomes

Sequences were obtained from third chromosomes that were isogenized by passage over a balancer chromosome (TM3 or TM6) using standard

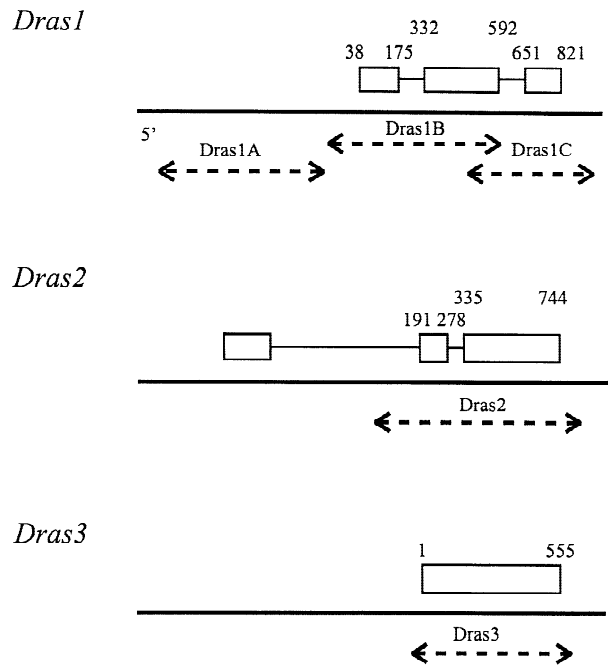


Fig. 1. Genomic structures of the three *Dras* genes. The genomic structures of *Dras1*, *Dras2*, and *Dras3* are shown, with boxes for exons, joined by solid lines, and positions of the first and last nucleotides of each exon relative to our consensus sequence, referred to in Fig. 2. The extent of 5' and 3' transcribed regions are not known with precision. Below each structure are the locations of the sequenced regions, indicated by dashed lines between primers. The 3' region of *Dras1* was amplified as two products.

Drosophila genetics. Each was extracted from a different inbred line. Six alleles originated in a collection of isofemales trapped in Ann Arbor, Michigan, in the summer of 1996; 20 alleles originated from the Wallace collection of wild-type lines trapped in various locations throughout the world 10–15 years ago, obtained from the Bowling Green Stock Center in February 1996 [see Gibson and van Helden (1997, Table 1) for a list of localities]; and 1 allele was from a lab stock containing a *Sevenless* transgene (Basler et al. 1992). The 94 lines used to study linkage disequilibrium between *Dras2* and *Dras3*, most of which originated in a Kenyan population, were obtained from the Bowling Green Stock Center in December 1997.

Primers and Sequencing Strategy

The structure of each locus and sequencing strategy is shown in Fig. 1. *Dras1* is located in 2 kb of cytological interval 85D on the right arm of chromosome 3, and the transcript includes two introns. Three pairs of primers were used to amplify the 5' flanking region and the coding region in two halves. *Dras2* is located in 2 kb of cytological interval 64B on the left arm of chromosome 3. The gene structures given by Neuman-Silberberg et al. (1984), Mozer et al. (1985), and Brock (1987) differ from one another with respect to the location of the intron/exon boundaries. This has an enormous effect on the nature of the polymorphisms (up to eight intron substitutions would affect coding sites, including insertion of a stop codon, if these structures were accepted). We used the analysis of Bishop and Corces (1988), which confirms aspects of the preceding studies and is definitive in terms of direct comparison of cDNA and genomic DNA and, also, produces almost-perfect alignment with the mammalian TC21 homolog. A single fragment covering exons 2 and 3 was amplified for sequencing of the

Drosophila alleles. *Dras3* is located in 0.7 kb of cytological interval 62C on the left arm of chromosome 3 and is encoded by a single exon. We report here only the sequence of the coding region, as the primers were designed to amplify immediately adjacent to the start and stop codons.

The primer sequences used both to amplify the fragments and to prime sequencing reactions were as follows: *Dras1A*-F, 5'-GCTAAGAAACGGTGATGCCAG-3'; *Dras1A*-R, 5'-GACTGTGCGTGTATGGGCTGC-3'; *Dras1B*-F, 5'-GCGTACGGAGAGAGAGACTG-3'; *Dras1B*-R, 5'-ACGAGGCCAGATCACATTTG-3'; *Dras1C*-F, 5'-GACCTACCGTGAGCAGATCAA-3'; *Dras1C*-R, 5'-GATATAGCAGCTGAACCAAGG-3'; *Dras2*-F, 5'-TTAGTCATTTGCGTTCATCTGC-3'; *Dras2*-R, 5'-TATATGTTGGCTCCTGCTTCC-3'; *Dras3*-F, 5'-AGAGATATACGAAGGATATAC-3'; and *Dras3*-R, 5'-GGTTTTTGAAGTCTTATAGCA-3'. Direct cycle sequencing was performed with an ABI 377 automated sequencer, following gel purification of the relevant PCR-generated fragments. Routinely, up to 700 nucleotides of readable sequence were obtained and analyzed using Sequence Navigator software. All alignments were unambiguous, as only a few small indels were observed in the total data set. All sequences were obtained in both directions, with the exception of 19 alleles of the *Dras3* locus, for which the forward strand sequence was both invariant and unambiguous. Representative sequences of each gene and species have been deposited in GenBank.

The ASOs used to test for linkage disequilibrium between *Dras2* and *Dras3* were 5'-ACTTGTACCCACCA-3', which recognizes the canonical *Dras2* sequence T516, but not C516, and 5'-CGCTGACGGTCCAGT-3', which recognizes the more rare G60 but not the A60 allele of *Dras3*. DNA was extracted from a single progeny of each of 94 crosses between a Kenyan wild-type fly and a *Sevenless* stock virgin female, and PCR amplification products for the two loci were checked on an agarose gel and then dot-blotted onto separate pieces of nylon membrane (Hybond N⁺; Amersham) and hybridized to ³²P-end-labeled ASOs following Saiki et al. (1986). Since the *Sevenless* chromosome alleles do not hybridize to either ASO, the genotypes and linkage phases of single wild-type chromosomes could be inferred from the presence or absence of a hybridization signal with each probe.

Statistics

Nucleotide diversity measures and the Tajima (1989) and Fu and Li (1993) test statistics were calculated using the SITES program obtained over the Internet (Hey and Wakeley, 1997). Indels were coded as single nucleotides in the analysis and, thus, carry the same weight as a single base substitution. HKA tests were as described by Hudson et al. (1987). The linkage disequilibrium parameter D' is equal to the ratio of D/D_{\max} , the statistical significance of which was confirmed by the standard χ^2 method (Hartl and Clark, 1997, pp. 101–103).

Results

Polymorphism Within *Drosophila melanogaster*

The three *Drosophila Ras* genes all have remarkably low levels of nucleotide diversity in the coding regions. In a global sample of 27 alleles of *D. melanogaster*, *Dras1* showed a single synonymous substitution present in a single allele, while *Dras2* and *Dras3* showed just four synonymous polymorphisms each. These ranged in frequency from 0.04 (singletons) up to 0.26. The corre-

sponding nucleotide diversity measures derived from the number of segregating sites (θ) and the average pairwise difference between alleles (π) are all less than half of the average for coding regions of autosomal genes in *Drosophila* [compare the value of 0.0044 per base pair for both of these parameters given by Moriyama and Powell (1996) with the values for 1cod, 2cod, and *ras3* in Table 1]. They lie in the same range as those of genes such as *tra* and *Adh(Dup)*, which have a significantly low nucleotide diversity by HKA comparisons (Moriyama and Powell 1996), although it should be recognized that the confidence limits on these measures are large.

The noncoding sequences of the three *Dras* genes by contrast have variable levels of polymorphism in *D. melanogaster*. Estimates of the neutral mutation parameters are provided on the left side of Table 1 for the 5' non-transcribed region of *Dras1*, the coding region including two introns of *Dras1*, most of the coding region derived from exons 2 and 3 of *Dras2*, and the uninterrupted coding region of *Dras3*. Breaking these regions down by structural criteria on the right side of Table 1 shows that there are three times fewer segregating sites in the non-coding sequences of *Dras1* than the introns of *Dras2*. The θ and π values for both the introns and the 5' flanking region of *Dras1* (but not of *Dras2*) are also low in comparison with the average θ for noncoding autosomal sequences in *D. melanogaster* [0.0116 per base pair, from Moriyama and Powell (1996)]. Note that the estimates shown are actually slightly inflated by the inclusion of three small indels in the analysis of *Dras1*. Also, differences between the θ and the π estimates approach significance for both regions of *Dras1*, as suggested by the large negative values of the Tajima, and Fu and Li, parameters as shown in Table 1 [compare with Table 1 of Tajima (1989) and Table 4 of Fu and Li (1993)]. This indicates that there are fewer average pairwise differences between alleles of *Dras1* than is expected from the number of segregating sites, possibly due to the presence of several divergent alleles in the sample: W3, W23, and W31 account for 10 of the 14 singleton nucleotide polymorphisms in the gene, as indicated in Fig. 2.

Divergence Between Species

Estimates of the genetic distances between species for each locus are shown in Table 1 as the modal number of differences between the *D. melanogaster* sequences and the sequences of two sibling species, *D. simulans* and *D. mauritiana*. The levels of between-species divergence and within-species polymorphism can be compared by examining the ratio of the divergence per base pair to π (D/π in Table 1). Although this number is sensitive to slight changes in either parameter, it is clear that divergence between the cosmopolitan species *D. melanogaster* and *D. simulans* is greater than expected relative to

Table 1. Parameters of *Dras* nucleotide diversity

	1 (5')	1 (3')	<i>ras2</i>	<i>ras3</i>	1cod	1nc	2cod	2nc
Length ^a	745	1046	780	555	570	1218	498	282
<i>E</i> (S) ^b	14	8	17	4	1	21	4	14
<i>E</i> (<i>D</i> _{mau}) ^b	38	31	19	5	7	62	2	17
<i>E</i> (<i>D</i> _{sim}) ^b	37	32	20	4	5	64	3	17
<i>E</i> (<i>D</i> _{m-s}) ^b	11	17	8	1	1	27	1	7
Shared ^c	0	0	7	1	0	0	1	6
θ (bp) ^d	0.0049	0.0020	0.0057	0.0019	0.0005	0.0045	0.0021	0.0121
π (bp) ^d	0.0031	0.0009	0.0051	0.0020	0.0001	0.0026	0.0018	0.0108
<i>D</i> / π ^e	15.5	34.4	4.5	3.5	90.0	20.0	3.3	4.7
Tajima ^f	-1.23	-1.76	-0.56	0.25	-1.15	-1.51	-0.33	-0.58
Fu & Li ^f	-1.70	-2.37	-0.03	0.08	-1.65	-2.11	0.09	-0.07

^a Lengths in nucleotides of the 5' and 3' halves of *Dras1*, *Dras2*, and *Dras3* and of the coding (cod) and noncoding (nc) regions of *Dras1* and *Dras2* (*Dras3* is all coding).

^b *E*(S) is the observed number of segregating sites, *E*(*D*) are estimates of the divergence between species calculated as the modal number of sites that differ between the *D. simulans* (sim) or *D. mauritiana* (mau) sequence and the *D. melanogaster* consensus sequence, or between the *D. simulans* and the *D. mauritiana* (m-s) sequences.

^c The number of polymorphic sites in *D. melanogaster* for which one of the nucleotides is the nucleotide in one or both sibling species.

^d θ is an estimate of the neutral mutation parameter per base pair, based on the number of segregating sites. π is an estimate of the neutral mutation parameter based on the average pairwise distance between alleles in the *D. melanogaster* sample.

^e The ratio of the modal divergence between the *D. simulans* and the *D. melanogaster* sequences, per base pair, to the estimate of π .

^f Values of the Tajima (1989) and Fu and Li (1993) statistics calculated by the SITES program.

the observed segregating polymorphism for both the coding and the noncoding sequences of *Dras1*. Closer inspection of the data indicates that *Dras1* and *Dras2* show similar overall divergences between species when introns and flanking sequences are included in the analysis, suggesting that there may be a deficiency of polymorphism in *Dras1*. All three predicted *Dras* proteins show about one-quarter of the typical divergence for *Drosophila* proteins. *D. simulans* and *D. mauritiana* also diverge from each other at all three loci to similar extents, and at about one-third of the level seen relative to *D. melanogaster*, as is typical. Thus, despite the differences in intraspecific variation, the *Dras* genes are diverging to similar extents between species.

The ratio of within-species to between-species polymorphism is expected to be constant across loci if they are evolving according to the infinite alleles model of neutral molecular evolution. This expectation has led to the development of a statistical test for neutrality known as the HKA test (Hudson, Kreitman, and Aguadé, 1987). None of the genes or gene regions in this data set showed significant departures from neutrality as assessed by the HKA test, compared with one another or with the 5' portion of the *Adh* gene (data not shown). This result may be an artifact of the low levels of nucleotide diversity in the sample, as the comparison of the coding regions of *Dras1* (one segregating site and seven fixed differences) and *Dras2* (four segregating sites and a mean of just two differences relative to *D. mauritiana*) would be significant if these values were all doubled. However, formally the results suggest that purifying selection and genetic drift are the major factors affecting nucleotide diversity in the *Dras* genes.

The pattern of polymorphism in *Dras2* shows some unusual features. Again, the level of polymorphism is too low to achieve statistical significance by standard tests, but the distribution of pairwise differences within *D. melanogaster* does not appear to be unimodal (Fig. 4A). This suggests that there are two, and possibly three, classes of haplotypes segregating in the sample, despite evidence for recombination between them. Thus, for example, the polymorphisms 516C, 633C, 751G, and 754G are grouped in several alleles and tend to be associated with 46C, 324T, and 327G.

Surprisingly, seven of the polymorphic sites in *Dras2*, including one of the synonymous substitutions, are found in the *D. simulans* and *D. mauritiana* alleles (Figs. 2 and 3). In contrast, none of the nonconsensus nucleotides in *Dras1* of *D. melanogaster* are present in either sibling species. Intriguingly, these "shared" sites are not randomly distributed throughout the sample, but are restricted to 9 of the 27 alleles, just one more than the number of the most frequent substitution (C751G), and in fact most of the sites belong to the rare haplotype grouping referred to above. It can be inferred that they were present in the ancestral population prior to the divergence of *D. simulans* and have been retained as a group despite their divergence from the majority of the alleles. The ancestral sites are now also associated with five other polymorphisms (160A, 174G, 288A, 330T, and 633C) that are only occasionally found in the remaining 18 alleles. W28 represents a third class of haplotype and also shares a single site with the sibling species.

Several of these "shared" polymorphisms at the 3' end of *Dras2* are in apparent significant linkage disequilib-

			C	A	A	A	A	A	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	S		
			o	1	3	7	1	1	2	1	2	3	7	8	9	1	1	1	1	1	2	2	2	2	2	3	
			n	4	8	0									0	2	3	4	7	8	1	3	4	5	7	8	
Dras1	5'	51	5' N 8	A	.	.	.	G	.	G	.	.	.	G	.	G	G	G	.	.	.	G	
		70	5' N 1	C	T	
		300	5' N 1	T	C	
		327	5' N 1	A	G	.	.	.	
		373	5' V 1	A	C	
		389	5' V 2	T	A	A	A	
		504	5' V 1	G	T	.	.	.	
		538	5' V 1	T	.	A	
		609	5' V 2	C	G	G	
		627	5' D 3	T	
		642	5' V 8	G	C	.	C	C	C	C	.	C	C	.	.	.	C	
		664	5' V 1	T	A	.	.	
		670	5' N 1	T	C	
		720	5' N 5	C	T	.	.	.	T	T	
Dras1	3'	225	I1 I 4	-	3	.	.	.	3	
		280	I1 D 1	2	
		292	I1 V 1	T	A	
		523	SY N 1	G	A	
		648	I2 N 1	C	T	
		846	3' N 2	T	C	C	
		889	3' V 1	G	T	
		1023	3' V 1	G	C	
Dras2		46	I1 N 4*	T	.	.	.	C	C	C	.	
		132	I1 N 1	A	G	.	.	
		146	I1 N 1	G	A	.	.	
		160	I1 * 5	C	.	.	.	A	.	.	.	A	T	.	.	T	A	.	
		161	I1 N 1*	G	A	.	
		174	I1 V 3	T	.	.	.	G	.	.	.	G	G	.	.	
		286	I2 V 1	T	A	.	
		288	I2 V 5	C	.	.	.	A	.	.	.	A	A	.	.	A	A	.	.	
		324	I2 V 3*	A	.	.	.	T	.	.	.	T	
		327	I2 V 6*	T	.	.	.	G	.	.	.	G	G	.	.	.	G	.	.	G	G	
		330	I2 N 4	C	.	.	.	T	.	.	.	T	T	T	.	
		462	SY N 1	G	A	
		516	SY N 6*	T	.	.	.	C	.	.	.	C	C	C	C	
		564	SY N 2	G	A	.	.	.	A	
		633	SY N 5	T	.	.	.	C	.	.	.	C	C	C	.	
		751	3' V 8*	C	.	G	.	G	.	.	.	G	G	G	G	G	G	
		754	3' N 6*	A	.	.	.	G	.	.	.	G	G	G	G	G	
Dras3		27	SY N 4	C	T	T	T	
		60	SY N 7*	A	.	G	.	.	.	G	G	G	G	G	
		174	SY N 1	G	A	
		318	SY N 7	G	A	A	A	.	A	.	A	.	A

Fig. 2. Distribution of polymorphic sites within *Dras* genes of *D. melanogaster*. The nucleotide types at each polymorphic position in a sample of 27 isogenic third chromosomes of *D. melanogaster* are shown for the upstream 5' noncoding region of *Dras1*, the coding region of *Dras1*, the coding region of *Dras2*, and the coding region of *Dras3*. Nucleotides are numbered in the left-hand column according to their position in the consensus sequence (available upon request). The next three columns indicate the type of polymorphism [5', 3', first or

second intron (I1 or I2), or coding (SY, synonymous)]; whether it is a transition (N), a transversion (V), or an indel (— represents a single-base deletion; 3, the insertion of the triplet TGA; and 2, the deletion of the doublet CT); and the number of alleles with the polymorphism. Site 160 in *Dras2* has two changes, one a transversion and the other a transition. An asterisk adjacent to the number of alleles indicates polymorphisms shared with the two sibling species.

rium with the single "shared" polymorphism in *Dras3* in the sequence data set (for example, $D' = 0.57$, $p < 0.005$, for T516C in *Dras2* with A60G in *Dras3*). This association was not confirmed in a larger sample of 94 individual flies picked at random from a collection of isofemale lines derived from sites in Kenya that were obtained from the Bowling Green Stock Center. Hybridization of allele-specific oligonucleotides (ASOs) that bind only to PCR products containing these two nucleotide changes indicated a frequency of 0.18 for the *Dras2* T516C polymorphism in the larger sample and 0.44 for the *Dras3* A60G polymorphism. The frequencies of the four two-allele combinations were not significantly different from random expectations (data not shown). There was no evidence that the two classes of *Dras2* haplotypes

are geographically restricted, as both were found in samples from three continents.

Discussion

Absence of Protein Variation in Dras1

The analysis of nucleotide variation reported here indicates that the major *Ras* protooncogene homologue, *Dras1*, is essentially monomorphic in *Drosophila melanogaster*, at least with respect to the protein sequence. There is no evidence for a balance of catalytically distinct alleles as is seen for some central metabolic en-

Dras1 5'

```

      1 1 1 1 1 1 1 2 2 3 3 3 3 4 4 5 5 5 6 6 6 6 6 6 6 6
      8 0 5 7 2 7 9 2 6'8 9'9 0 0 1 5 0 4 7 1 2 5 0 8 3 8'9 3 5 4 6 4 4 9 1 8 1' 1 2 0 3
Mel  T A T A A T G A - C - C I G A G C T T A G A G T T - A T T T A A C T T G -- T T T A
Mau  A T G G A T C A G G I A I - T A T A C G A T T G G G C C G A G G T G C T G C G C C G
Sim  T A T A C C C T G G I - - - T A T A T G A T T G G G C C G A G G T G C T G C G T C G

```

Dras1 coding

```

      intron                                intron                                3'UTR
-----                                -----                                -----
      1 1 2      2 2 2 2 2 2 2 2 273' 2 2 2 2 3 3 3 3 4 5 6 6 6 6 6 7 7 8 9 9 9
      9 3 5 0      1 1 2 2 2 4 5 5 5 1 22 3 4 7 8 9 9 0 1 1 1 1 0 2 2 3 3 3 2 7 3 1 2 3
      7 6 4 4      6 8 0 8 9 8 3 4 9 1 01 9 2 8 7 1 9 1 2 7 9 5 2 2 5 0 3 7 8 6 3 4 1 2
Mel  C C T AAGCT C C GG T A C C G T - - - - T C A G G G T A T A A C A T G G C T C G T
Mau  T T C TAAAA C G GG G C C T G A I TT G G C T G T A A T C C G C A C T A T A T A T C
Sim  C T C TAAAA A A -- T C T C T T I GG A T C T G T A G C C C G C A C G A G A A A T C

```

Dras2

Dras3

```

      1 1 1 1 1 1 1 1 2 2 2 3 3 3 3 4 5 7 7 7 7
      4 0 2 4 5 5 6 6 8 5 8 8 9 0 0 2 2 9 1 5 5 5 5
      6 9 2 2 2 9 1 5 6 3 4 8 1 0 9 4 7 8 6 1 2 4 6
mel  T A A T G G G C T C C C T T T A T C T C T A T
mau  C . T C . . A G C . T - A C G T G T C G C G G
sim  C G T C A A A . . A T - A C C T G T C G . G G
mel* C          A          A          T G          G          G
      1 1 2 3
      6 1 5 3 6
      0 7 0 8 3
mel  A T G C A
mau  G C A T G
sim  G C A . G
mel* G

```

Fig. 3. Divergence between *Dras* genes in three *Drosophila* species. The differences among the sequences of *D. melanogaster*, *D. simulans*, and *D. mauritiana* are shown relative to the consensus nucleotide positions in Fig. 2. Insertions present in the sibling species but not *D. melanogaster* are numbered with a prime (for example, 56' is the site of a G in the 5' region of *Dras1*). There is a 42-bp nucleotide insertion

at position 273' of the first *Dras1* intron that also shows four differences between the two sibling species. *Dras2* and *Dras3* have seven sites and one site, respectively, that segregate in *D. melanogaster* and are found in *D. simulans* and *D. mauritiana*, as indicated in the mel* line. Three fixed synonymous differences in *Dras2* are indicated by a +.

zymes. Further, the absence of any replacement substitutions relative to two sibling species indicates that *Dras1* is among the most highly constrained loci in *Drosophila*. There may of course be some segregating replacement polymorphisms not detected in our sample, but these would be very rare.

Changes in dosage of the *Dras* genes do not on their own have phenotypic effects, but loss of function alleles of *Dras1* and *Dras3* do modify the mutant phenotypes of other genes with which they interact, such as the receptors whose signals they transduce (Karim et al. 1996; Hariharan et al. 1991). Thus, levels of *Dras* gene expression may be just as tightly regulated as the catalytic efficiencies of the proteins. In the absence of extensive nucleotide variation, it is likely that variation in cofactors is the predominant source of the genetic variation in Ras activity in *Drosophila* that we have documented (Polaczyk et al. 1998). For example, Ayala and Hartl (1993) found several amino acid polymorphisms in a small sample of alleles of the *boss* gene of *D. melanogaster*, which encodes a ligand that activates the Ras pathway in R7 photoreceptor cells, though they concluded that the variation was evolving in accordance with neutral expectations.

Within- and Between-Species Variation in Dras1 and Dras2

Despite a qualitative difference in the ratio of within-species polymorphism to between-species divergence for

the coding regions of *Dras1* and *Dras2*, application of the HKA test failed to reject the null hypothesis of neutral evolution. Since the low levels of nucleotide diversity reduce the statistical power of this test, it might nevertheless be argued that the very low level of variation in the coding region of *Dras1* of *D. melanogaster* (a single variant site in the sample of 27 alleles) provides evidence for a recent selective sweep through the locus. This reduced diversity is, however, also consistent with the operation of deleterious background selection near the centromere, where recombination is generally low (reviewed by Aquadro 1997). Recombination parameters could not be reliably measured with this data set to confirm that *Dras1* does indeed have a lower effective population size than the other two *Dras* genes.

Alternatively, there may be some form of balancing selection acting on the *Dras2* locus. In the absence of any replacement variation, any such selection would by inference be acting on either noncoding, presumably regulatory, or synonymous sites. The two classes of *Dras2* haplotypes differ from those of the sibling species to similar extents, as indicated by the unrooted parsimonious gene network at the top of Fig. 4B. Without considerably more extensive sampling of *D. mauritiana*, *D. simulans*, and an outgroup species, it is not possible to resolve the location of the root, though this dramatically affects interpretation of the pattern of variation. Note that *D. mauritiana* is an island species that is thought to be derived from *D. simulans* (Hey and Kliman 1993), so it is reasonable to assume that the root does not lie between

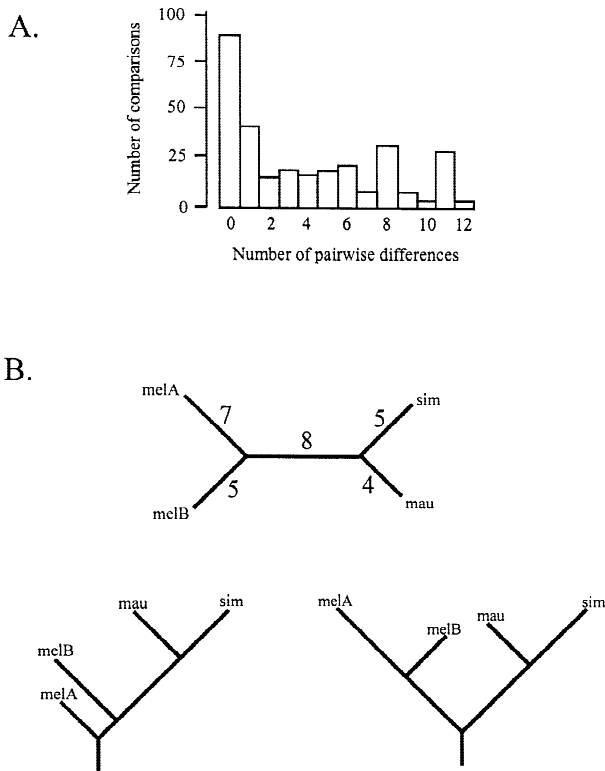


Fig. 4. Distribution of polymorphism in *Dras2* among *Drosophila* species. **A** Histogram showing the distribution of the 325 pairwise differences among the 26 alleles in the *Dras2* sample, ranging from 0 (for example, W8 vs W9) to 12 (A18 vs W13). **B Top:** Unrooted parsimony tree linking the rare and consensus *D. melanogaster*, *D. simulans*, and *D. mauritiana* haplotypes. The number of substitutions is indicated beside each branch. melA, consensus sequence (for example, the W10 allele); melB, rare haplotype represented by the A14 and W3 alleles that includes all of the shared sites plus five other polymorphisms. The tree was calculated by hand. **Bottom, left and right:** Two possible rootings, see text for discussion.

them. If the root is placed between the *D. melanogaster* haplotypes (bottom left, Fig. 4B), it implies that both the consensus and the rare haplotypes are ancestral and have reduced rates of evolution relative to the sibling species. If the root is placed in the middle branch (bottom right), the consensus haplotype would appear to be derivative. Although the relative rates of change would be similar, this model requires loss of all seven otherwise shared polymorphisms in the one haplotype, which seems unlikely by chance, unless the more common haplotype has recently entered the population.

It would be easier to begin to accept that selection is maintaining variation in *Dras2* if there were any phenotype associated with the gene. Despite an intensive screen for lethals that uncovered several mutations in the adjacent *Rop* gene, Harrison et al. (1994) were unable to find a mutation in *Dras2*. Bishop and Corces (1992) showed that expression of an activated "oncogenic" form of *Dras2* causes wing and eye defects but could not attribute any phenotype to overexpression of the wild-type gene product. Our demonstration of an absence of

nonsynonymous variation in *Dras2* implies that the gene is subject to strong purifying selection and confirms that it is not a pseudogene. It has recently been proposed based on gene expression studies and comparison with the most similar human Ras-related gene, *TC21*, that *Dras2* is actually involved in secretion and/or endocytosis in a restricted subset of tissues, including the central nervous system (Salzberg et al. 1993). If so, *Dras2* would not function in the same genetic pathways as the other two *Dras* genes. Consequently, epistatic interaction cannot account for the existence of linkage disequilibrium with *Dras3* in the sequenced sample, which is perhaps better regarded as an artifact of sampling and/or of isogenic chromosome extraction.

The fact that all three *Dras* proteins show similar and typical levels of divergence between *D. melanogaster* and *D. simulans* despite low and variable levels of intra-specific polymorphism remains to be explained. If we accept that all three genes have unusually low neutral mutation rates, due to strong selection in favor of an optimal haplotype, this problem can be dealt with by recognizing that there is inherent irregularity in the length of time between nucleotide substitutions. The absence of introns in *Dras3* means that there are few footprints of genic history to infer much about this locus. The lack of polymorphism in the transcribed portion of *Dras1* is at least consistent with the recent conclusion of a sweep through that region but, given the low rate of recombination, is more readily explained as a consequence of repetitive deleterious mutation preventing the maintenance of linked neutral polymorphism. The bifurcated distribution of polymorphism in *Dras2* suggests either the maintenance of two old alleles in one species and an accelerated rate of change in the other two or that a new allele is in the process of displacing the ancestral type in *D. melanogaster*. A more conservative interpretation is that the very low levels of within species polymorphism beget high variances, and hence all of the patterns are consistent with neutral evolution, allowing for variance in parameters such as recombination rates, migration, and fluctuating population size.

Acknowledgments. Many thanks go to Jody Hey, Michael Nachman, and Marta Wayne for their comments on an early version of the manuscript. This work was supported by grants to G.G. from the John and Suzanne Munn Endowed Research fund of the University of Michigan Comprehensive Cancer Research Center and from the David and Lucille Packard Foundation.

References

- Aquadro CF (1997) Insights into the evolutionary process from patterns of DNA sequence variability. *Curr Opin Genet Dev* 7:835–840
- Ayala FJ, Hartl DL (1993) Molecular drift of the *bride of sevenless* (*boss*) gene in *Drosophila*. *Mol Biol Evol* 10:1030–1040
- Basler K., Christen B, Hafen E (1991) Ligand-independent activation of the Sevenless receptor tyrosine kinase changes the fate of cells in the developing *Drosophila* eye. *Cell* 49:281–291

- Bishop JG, Corces V (1988) Expression of an activated *ras* gene causes developmental abnormalities in transgenic *Drosophila melanogaster*. *Genes Dev* 2:567–577
- Brock, HW (1987) Sequence and genomic structure of *ras* homologous *Dmras85D* and *Dmras64B* of *Drosophila melanogaster*. *Gene* 51:129–137
- Fu Y, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics* 133:693–709
- Gibson G, van Helden S (1997) Is function of the *Drosophila* homeotic gene *Ultrabithorax* canalized? *Genetics* 147:1155–1168
- Hariharan IK, Carthew RW, Rubin GM (1991) The *Drosophila* roughened mutation: Activation of a *rap* homolog disrupts eye development and interferes with cell determination. *Cell* 67:717–722
- Harrison SD, Solomon N, Rubin GM (1995) A genetic analysis of the 63E-64A genomic region of *Drosophila melanogaster*: Identification of mutations in a replication factor C subunit. *Genetics* 139:1701–1709
- Hartl D, Clark A (1997) *Principles of Population Genetics*, 3rd ed. Sinauer Associates, Sunderland MA
- Hey J, Kliman RM (1993) Population genetics and phylogenetics of DNA sequence variation at multiple loci within the *Drosophila melanogaster* species complex. *Mol Biol Evol* 10:804–822
- Hey J, Wakeley J (1997) A coalescent estimator of the population recombination rate. *Genetics* 145:833–846
- Hilton H, Kliman RM, Hey J (1994) Using hitchhiking genes to study adaptation and divergence during speciation within the *Drosophila melanogaster* species complex. *Evolution* 48:1900–1913
- Hudson R., Kreitman M, Aguadé M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics* 116:153–159
- Karim, FD, Chang H, Therrien M, Wassarman DA, Laverty T, and Rubin GM (1996) A screen for genes that function downstream of *Ras1* during *Drosophila* eye development. *Genetics* 143:315–329
- Katz LA, Harrison RG (1997) Balancing selection on electrophoretic variation of phosphoglucose isomerase in two species of field cricket: *Gryllus veletis* and *G. offnsylvanicus*. *Genetics* 147:609–621
- Katz ME, McCormick F (1997) Signal transduction from multiple Ras effectors. *Curr Opin Genet Dev* 7:75–79
- McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351:652–654
- Moriyama EN, Powell JR (1996) Intraspecific nuclear DNA variation in *Drosophila*. *Mol Biol Evol* 13:261–277
- Mozer B, Marlor R, Parkhurst S, Corces V (1985) Characterization and developmental expression of a *Drosophila ras* oncogene. *Mol Cell Biol* 5:885–889
- Neuman-Silberberg FS, Schjeter E, FM Hoffmann, Shilo BZ (1984) The *Drosophila ras* oncogenes: Structure and nucleotide sequence. *Cell* 37:1027–1033
- Pawson T (1993) Signal transduction—A conserved pathway from the membrane to the nucleus. *Dev Genet* 14:333–338
- Polaczyk P, Gasperini R, Gibson G (1998) Naturally occurring genetic variation affects *Drosophila* photoreceptor determination. *Dev Genes Evol* 207:462–470
- Saiki RK, Bugawan TL, Horn GT, Mullis K, Erlich HA (1986) Analysis of enzymatically amplified β -globin and HLA-DQ α DNA with allele specific oligonucleotide probes. *Nature* 324:163–166
- Salzberg A, Cohen N, Halachmi N, Kimchie Z, Lev Z (1993) The *Drosophila Ras2* and *Rop* gene pair: A dual homology with a yeast *Ras*-like gene and a suppressor of its loss-of-function phenotype. *Development* 117:1309–1319
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595
- Threadgill DW, et al. (1995) Targeted disruption of mouse EGF receptor: Effect of genetic background on mutant phenotype. *Science* 269:230–234
- Valencia A, Chardin P, Wittinghofer A, Sander C (1991) The ras protein family: Evolutionary tree and role of conserved amino acids. *Biochemistry* 30:4637–4648
- Wassarman DA, Therrien M, Rubin GM (1995) The Ras signaling pathway in *Drosophila*. *Curr Opin Gen Dev* 5:44–50
- Watt WB (1994) Allozymes in evolutionary genetics: Self-imposed burden or extraordinary tool? *Genetics* 136:11–16