DNA Sequence Variation at the *Period* Locus Reveals the History of Species and Speciation Events in the *Drosophila virilis* Group

Holly Hilton and Jody Hey

Department of Biological Sciences, Rutgers University, Piscataway, NJ 08855-1059

Manuscript received February 27, 1996
Accepted for publication June 21, 1996

ABSTRACT

The *virilis* phylad of the *Drosophila virilis* group consists of five closely related taxa: *D. virilis, D. lummei, D. novamexicana, D. americana americana* and *D. americana texana.* DNA sequences from a 2.1-kb pair portion of the *period* locus were generated in four to eight individuals from each of the five taxa. We found evidence of recombination and high levels of variation within species. We found no evidence of recent natural selection. Surprisingly there was no evidence of divergence between *D. a. americana* and *D. a. texana,* and they collectively appear to have had a large historical effective population size. The ranges of these two taxa overlap in a large hybrid zone that has been delineated in the eastern U.S. on the basis of the geographic pattern of a chromosomal fusion. Also surprisingly, *D. novamexicana* appears to consist of two distinct groups each with low population size and no gene flow between them.

THE basic process of evolution that has led to the diversity of life is the formation of new species. Different genetical theories of speciation have been proposed (DOBZHANSKY 1937; MAYR 1942; CARSON 1968, 1975; WHITE 1978; TEMPLETON 1980), yet there is still much debate over the details of the process (BARTON and CHARLESWORTH 1984; TEMPLETON 1989; COYNE and ORR 1989; COYNE 1992). In general, speciation events cannot be observed directly because they occur on an evolutionary time scale. However, measurements of the pattern of genetic variation within and among closely related species can provide information suitable for exploring speciation processes and examining the role of population size, gene flow and population subdivision.

This research compares DNA sequence data from within and among the five closely related taxa in the *virilis* phylad of the *Drosophila virilis* species group (THROCKMORTON 1982). These five taxa are *D. virilis, D. lummei, D. novamexicana, D. americana americana* and *D. americana texana. D. virilis* is found in wild habitats in Japan and China, while in North America it is restricted to domestic habitats (PATTERSON and STONE 1952). In contrast, the other *virilis* phylad species are found exclusively in woodland settings (PATTERSON 1942a). *D. lummei* is found in northeastern Europe. The North American taxa, *D. a. americana, D. a. texana,* and *D. novamexicana* are closely related (PATTERSON and STONE 1952; THROCKMORTON 1982; SPICER 1992) and are collectively referred to as the americana complex (PATTERSON and STONE 1952). *D. a. americana* and *D. a. texana* reside in the eastern United States, and their ranges overlap in a hybrid zone running through North Carolina, Tennessee and Arkansas (Figure 1). *D. a. americana* is found north of the hybrid zone, and *D. a. texana* is found to the south (PATTERSON 1942b; CARSON and BLIGHT 1952; PATTERSON and STONE 1952; THROCKMORTON 1982). The third North American species, *D. novamexicana,* is found in the drier habitat of lower river valleys of New Mexico and the surrounding states. *D. novamexicana* has a lighter mesothorax color than both *D. a. americana* and *D. a. texana,* which have a dark body color and are virtually indistinguishable morphologically from each other. The lighter mesothorax of *D. novamexicana* may be an adaptation for desiccation resistance (SPICER 1991a). It has been suggested that the change accompanied speciation and the ability to live in the drier habitat (SPICER 1991a).

The five virilis phylad taxa exhibit three different metaphase karyotypes. *D. virilis* possesses what is considered the primitive karyotype, with five pairs of rods and one pair of dots (HSU 1952; THROCKMORTON 1982). This pattern is shared by *D. novamexicana* and *D. lummei. D. a. americana* and *D. a. texana* both have a fusion of the second and third chromosomes, making a large V-shaped chromosome. *D. a. americana* has a unique fusion of the *X* and a fourth chromosome, also a large V-shaped chromosome, which causes both the *Y* and a free fourth chromosome (both rods) to be male limited (STURTEVANT and NOVITSKI 1941; HSU 1952; PATTERSON and STONE 1952; EVGEN'EV 1971; THROCKMORTON 1982; GUBENKO and EVGEN'EV 1984). It is this X-4 chromosomal fusion that distinguishes *D. a. americana* from *D. a. texana,* and the hybrid zone between these species has been described on the basis of measurements of the frequency of X-4 fusion karyotypes.

We report DNA sequence data from the *period* (*per*)

Corresponding author: Holly Hilton, Department of Biological Sciences, Rutgers University, Nelson Labs, P.O. Box 1059, Piscataway, NJ 08855-1059. E-mail:hilton@mbcl.rutgers.edu

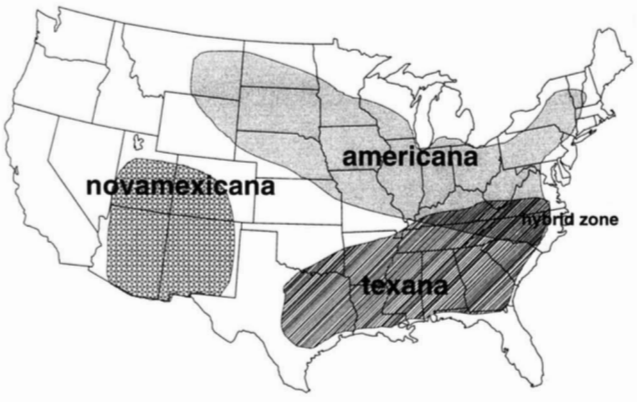


FIGURE 1.—Generalized range of *D. a. americana, D. a. texana,* and *D. novamexicana* including *D. a. americana* and *D. a. texana* hybrid zone. Compiled from data of PATTERSON and STONE (1952) and information on collection locations from The National Drosophila Species Resource Center.

locus. Mutations at the *per* locus have been found to affect circadian rhythms as well as courtship songs (KRYRIACOU and HALL 1984). In *D. melanogaster, per* is found on the *X* chromosome. An *X* chromosome location for *per* is also expected in species of the *virilis* phylad because of the high degree of conservation of chromosomal elements between *D. melanogaster* and *D. virilis* (STURTEVANT and NOVITSKI 1941; ALEXANDER 1976). This conservation of linkage groups between *D. melanogaster* and *D. virilis* has also been confirmed for many individual loci (TONZETICH *et al.* 1990; WHITING *et al.* 1989; NEUFELD *et al.* 1991; NURMINSKY *et al.* 1996). We generated data from a 2.1-kb region from four lines of *D. virilis* and *D. lummei,* seven lines of *D. a. americana,* eight lines of *D. a. texana* and six lines of *D. novamexicana* (Figure 2). The *per* locus was chosen for this analysis for a variety of reasons. KLIMAN and HEY (1993) studied variation in a 1.9-kb region of the *per* locus from six individuals of each of the four species of the *D. melanogaster* group and found it was a good choice for their study of speciation (KLIMAN and HEY 1993; HEY and KLIMAN 1993). Also, *per* evolves quickly, so that even when examining closely related species, there is ample variation for a variety of analyses (COLOT *et al.* 1988; KLIMAN and HEY 1993). Lastly, the expected *X* chromosome location of the *per* gene simplifies the procedure for generating single copy genomic DNA. DNA prepared from a single male contains sequences from *X*-linked genes in hemizygous, rather than diploid, proportion.

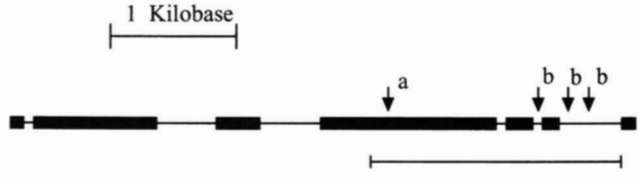


FIGURE 2.—Diagram of the *period* locus and the region sequenced. The specific region sequenced is marked with the lower black line that corresponds to bases 2870–4864 of COLOT *et al.* (1988). a marks the spot of the Thr-Gly repeat; b marks the three locations where sequence was not used in the analysis, see text for details.

TABLE 1

List of lines sequenced

Species name	Line no.	Location
D. virilis	1051.0	Pasadena, CA
	1051.8	Truckee, CA
	1051.9	Sendai, Japan
	1051.48	Texmelucan, Mexico
D. lummei	1011.1	Moscow, Russia
	1011.2	Overhalix, Sweden
	1011.4	Kukkola, Finland
	1011.8	Sakata, Japan
D. A. americana	0951.0[c]	Anderson, IN
	0951.1[c]	Poplar, MT
	0951.3[c]	Millersburg, PA
	0951.4[c]	Keelers Bay, VE
	0951.5[c]	Jackson, MI
	0951.6[c]	Chadson, NE
	0951.9[c]	Myrtle Beach, SC
D. A. texana	1041.0[c]	St. Francisville, LA
	1041.22[c]	New Orleans, LA
	1041.23[c]	Morrilton, AR
	1041.25	So. Richmond, VA
	1041.26	Tallahassee, FL
	1041.27[c]	Goldenhead Branch, FL
	1041.29	Jamestown, SC
	1041.31	Hollandale, MS
D. novamexicana	1031.0[a]	Grand Junction, CO
	1031.4[b]	Moab, UT
	1031.7[a]	Patagonia, AZ
	1031.8[a]	San Antonio, NM
	1031.11[b]	Gila, NM
	1031.12[b]	Antlers, CO

All lines are from the National Drosophila Species Resource center.

[a] A member of group Nova-A.

[b] A member of group Nova-B.

[c] These lines were checked for the appropriate metaphase chromosome compliment; see text for details.

MATERIALS AND METHODS

The flies: All strains were obtained from the National Drosophila Species Resource Center (NDSRC) (Table 1). In this paper, strains are referred to by species name and the NDSRC extension number, for example, "virilis.0," corresponds to NDSRC #1051.0 (Table 1). Confirmation of chromosomal karyotype in some strains was done using mitotic squash protocol #1 in ASHBURNER (1989). In addition, JEANNE HNILICKA and B. CHARLESWORTH (personal communication) found the following lines obtained from the stock center had the expected chromosomal patterns (THROCKMORTON 1982): americana.0, americana.1, americana.3, americana.4, americana.6, americana.9, and texana.22 (Table 1).

DNA preparation and sequencing: DNA preparations were made from single male flies (protocol 48 in ASHBURNER 1989). A 2.1-kb region of the *per* gene was PCR amplified using 20-mer oligonucleotide primers starting at positions 2803 ("+" primer 5′ base) and 4911 ("−" primer 5′ base) of COLOT *et al.* (1988; GenBank accession X13877). PCR and DNA sequencing methods were identical to those of KLIMAN and

HEY (1993). Sequences have been submitted to GenBank, acquisition numbers (L81296–L81324).

Alignments: Sequence alignment was first done by eye and then with the Genetics Computer Group program PILEUP (DEVEREUX and HAEBERLI 1991). Three small areas within introns revealed large amounts of insertion-deletion (indel) variation and were not included in the study because of alignment uncertainty. One area that was removed (corresponding to position 4197–4227 of COLOT *et al.* 1988) contained variations on a CT repeat, ranging from four CT pairs in all *D. virilis* lines to 21 pairs in a *D. novamexicana* line. The other regions that were not included correspond to positions 4436–4456 and 4585–4635 of COLOT *et al.* (1988).

Estimating Nm and Mantel test: Nm, the product of effective population size and migration rate, was estimated using the Fst estimate of HUDSON *et al.* (1992). To test whether the divergence between subspecies is greater than expected by chance, the nonparametric Mantel test (MANTEL 1967) was used to compare the similarity between two matrixes. The first matrix contained, for a set of DNA sequences from two subspecies, the number of differences between all pairs of sequences. The second matrix represented the hypothesis that sequences were more similar within subspecies and was made up of zeros and ones. A zero was placed in the matrix at those positions that corresponded to positions in the first matrix that contained the pairwise difference between sequences drawn from the same subspecies. A one corresponded to a difference between sequences from different subspecies. The test of association between the two matrices is straightforward: a coefficient of association, z, is calculated as the sum of all pairwise matrix cell projects (*i.e.*, the product of matrix 1 cell i,j with matrix 2 cell i,j summed over all i and j); an empirical distribution of this statistic is determined by a repeated process of 1000 random permutation of rows and columns of one matrix and recalculation of z for each permutation, and the probability of getting an equal or more extreme value than the observed value of z is assessed by comparing the observed value with the random distribution. The Mantel test was carried out using the NTSYS (ROHLF 1985) computer program package.

Measuring variation: The average number of pairwise nucleotide differences, π, is calculated from

$$\pi = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{k_{ij}}{\binom{n}{2}}, \quad (1)$$

where n is the number of sequences sampled and k_{ij} is the number of differences between sequences i and j. A second measure of sequence variation, θ, is a simple function of the number of polymorphic sites, S, and the sample size (WATTERSON 1975)

$$\theta = S/\sum_{i=1}^{n-1} \frac{1}{i}. \quad (2)$$

Both π and θ have expected values of $4Nu$, where N is the effective population size and u is the neutral mutation for the locus per generation. For a sex-linked gene, in the case where the effective population size is similar for males and females, the expected value for π and θ is $3Nu$.

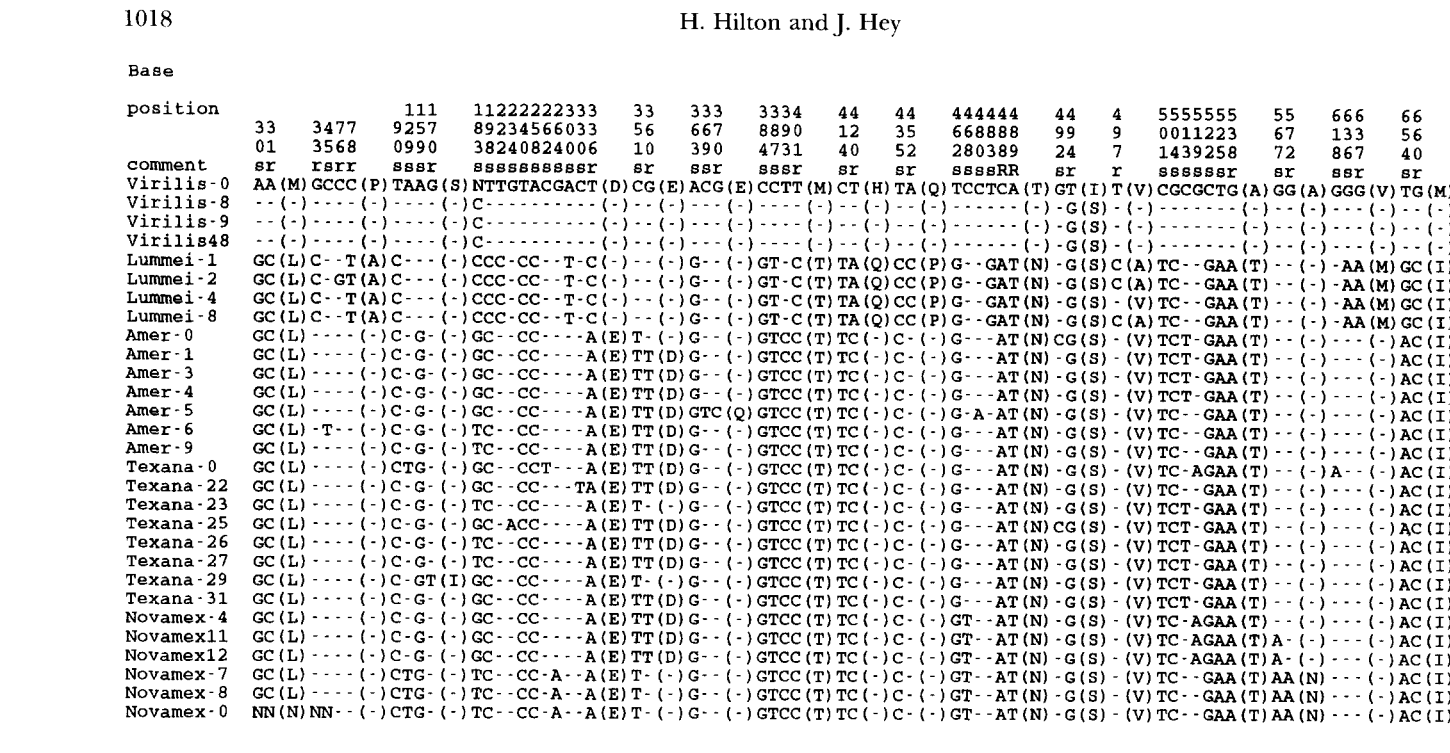
RESULTS

Grouping the lines: The three taxa in the americana complex, *D. a. americana*, *D. a. texana*, and *D. novamexicana*, have been distinguished on the basis of morphological and chromosomal comparisons. Typically, one individual has been chosen to represent each taxa in later studies of the *D. virilis* group (*e.g.*, REINBOLD and COLLIER 1990). However, these *a priori* hypotheses of species and subspecies status, based on limited genetic data, are not supported by our genealogical study. We have two examples where comparative sequence data do not support the prior species designations. First, *D. a. americana* and *D. a. texana* appear to be indistinguishable on the basis of the *per* data. Second, our *D. novamexicana* sequences seem to have come from two groups that have not recently exchanged genes.

Three analyses of the *per* data failed to reveal a pattern of divergence between the *D. a. americana* and the *D. a. texana* samples. First, these two groups had 33 shared polymorphisms, which are base pair positions where both *D. a. americana* and *D. a. texana* were segregating the same two bases. We also found no fixed differences (positions where all of the *D. a. americana* lines had one base and all of the *D. a. texana* had a different base) between these two groups. Second, the Fst estimate of Nm was 55.166 (HUDSON *et al.* 1991). Typically an Nm value greater than or equal to one leads to considerable homogeneity among populations (WRIGHT 1940). Third, a comparison of pairwise differences within and between *D. a. americana* and *D. a. texana* was no different from random contrasts, as determined by a Mantel test (the observed divergence was not different than zero; $P = 0.176$; see MATERIALS AND METHODS). For the rest of the analyses we have treated *D. a. americana* and *D. a. texana* as one group under the name, *D. americana*.

A second question about how to group individuals arose in *D. novamexicana*. When taken as a group of six sequences, there was little to distinguish them from *D. americana*. There was just one fixed difference between the *D. novamexicana* and the *D. americana* sequences (a synonymous change at position 468). The *D. novamexicana* samples, when grouped together, where highly variable; however this was misleading. The *D. novamexicana* sequences include two divergent sets, each of three very similar sequences. Nearly all of the variation within the *D. novamexicana* sample occurs as differences between these two groups, which we have named "Nova-A" and "Nova-B". There were 23 fixed differences between Nova-A and Nova-B, and no shared polymorphisms. The Fst estimate of Nm between the two groups was 0.01, revealing little evidence of intermixing at *per*. For the remainder of the analysis, *D. novamexicana* was treated both as one group, as well as treating Nova-A and Nova-B separately.

Both Nova-A and Nova-B revealed little within-group polymorphism (Table 3, Figure 3). In Nova-A, lines nova.7 and nova.8 were identical, sharing a unique 3-bp insertion in the coding sequence. The third member of Nova-A, nova.0, was also different from the other two, in having a unique intron base change. In group Nova-B, each of the three lines, nova.4, nova.11, and nova.12, differed by one or two changes. This is quite different from *D. americana* where after lines texana.26 and texana.27, which were identical, the next most closely related lines differed by 14 changes.



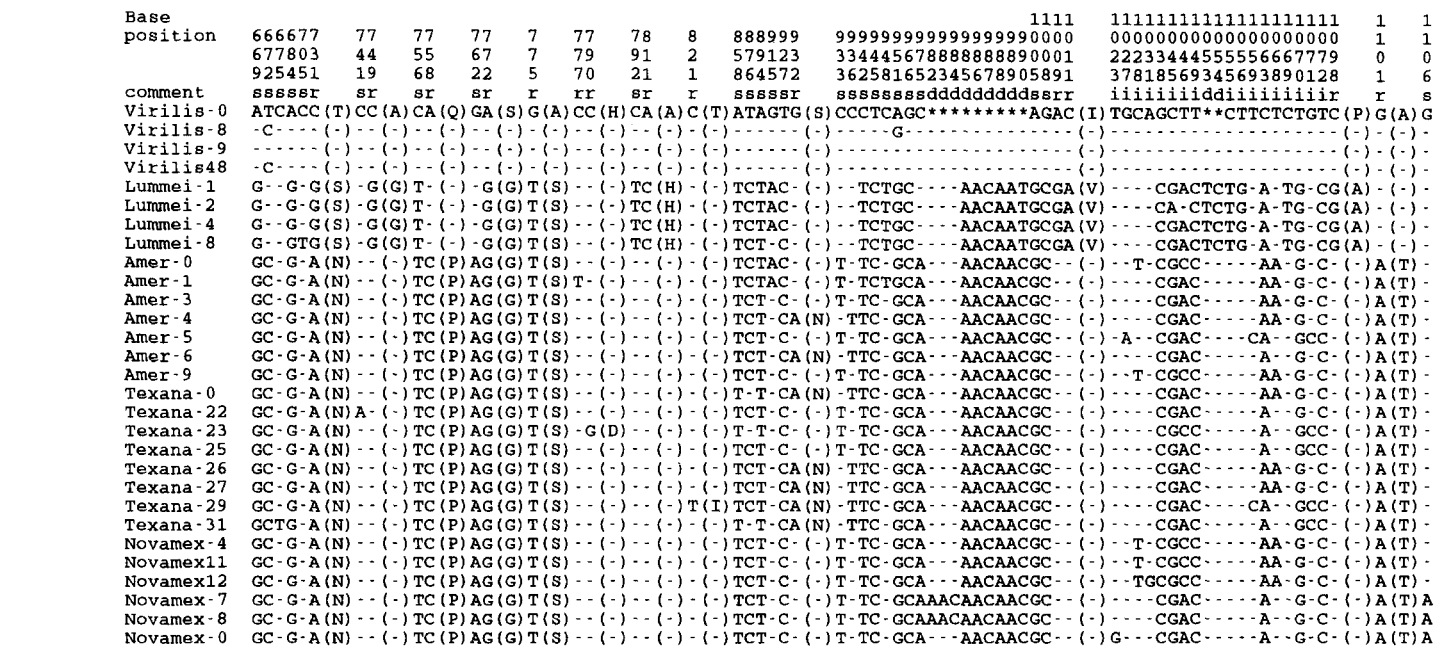


FIGURE 3.—Variable sites in *period*. The first rows indicate the base position of variable sites within the sequenced region. In the comment row, s, synonymous substitution; r, amino acid replacement substitution; I, intron change; d, deletion change. There are three noncoding intron regions that correspond to base positions 1015–1088, 1311–1432, and 1574–2088. The virilis.0 (1051.0) sequence is used as a reference. Nucleotides identical to the reference are indicated by a dash. At amino acid replacement sites, the nucleotide is followed in parentheses by the one letter code for the resulting amino acid (M, Met; L, Leu; P, Pro; A, Ala; S, Ser; I, Ile; D, Asp; E, Glu; Q, Gln; T, Thr; N, Asn; V, Val; G, Gly; H, His; K, Lys; W, Trp; R, Arg). Length variation is indicated by * in sequences shortened relative to others.

Eleven of the 23 fixed differences between the two *D. novamexicana* groups are polymorphic within *D. americana*. In other words, at 11 base positions where *D. americana* was found to segregate two bases, it was also found that the three Nova-A sequences all possessed one of the bases found in *D. americana* while the three Nova-B sequences all possessed the other base that was found in *D. americana*. Of the 12 remaining fixed differences between Nova-A and Nova-B, five base changes are unique to Nova-A (*i.e.*, Nova-A sequences are different from Nova-B and *D. americana*) and seven are unique to Nova-B. These 12 base changes may have arisen since each group has become isolated from the species that was ancestral to *D. americana* and *D. novamexicana*. It is also possible that these changes are polymorphisms in *D. americana* that did not appear in our

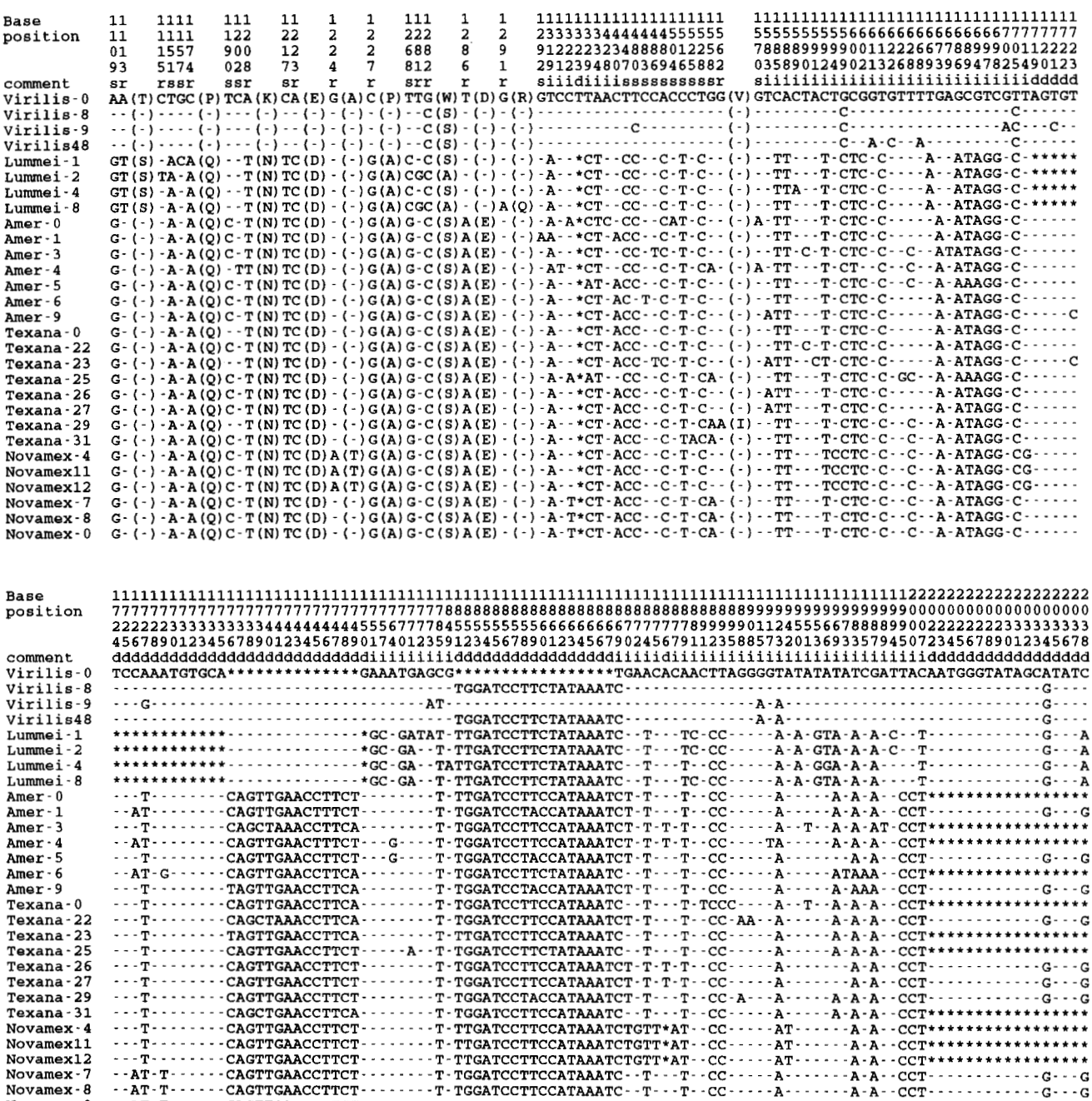


FIGURE 3.—*Continued*

sample; or they may have once been polymorphic in *D. americana* but are not now. Interestingly, the collection sites of the six *D. novamexicana* lines do not show a geographic pattern of separation between groups A and B (Table 1, Figure 1).

DNA sequence variation summary: If DNA sequence variation is neutral, then the patterns of DNA sequence variation can be used to estimate relative historical population sizes. Figure 3 shows all of the variable sites, and Table 2 lists the types of variation found in each group. Two estimates of variation are shown in Table 3: π, the average pairwise difference, and θ, calculated from the number of segregating sites (see MATERIALS AND METHODS, Measuring variation). Both *D. virilis* and *D. lummei* have levels of variation similar to the range observed at the *per* locus in *D. melanogaster* (KLIMAN and HEY 1993). In contrast, *D. americana* shows a level of variation twice that of *D. virilis* and *D. lummei,* (this is true whether the *americana* and *texana* subspecies are grouped together or not), suggesting a large historical effective population size. *D. novamexicana,* when divided into groups A and B, contains very low levels of variation, which is consistent with small population size (Table 3).

These measures of variation can also be used to examine the history of natural selection. Both π and θ have

TABLE 1

List of lines sequenced

Species name	Line no.	Location
D. virilis	1051.0	Pasadena, CA
	1051.8	Truckee, CA
	1051.9	Sendai, Japan
	1051.48	Texmelucan, Mexico
D. lummei	1011.1	Moscow, Russia
	1011.2	Overhalix, Sweden
	1011.4	Kukkola, Finland
	1011.8	Sakata, Japan
D. A. americana	0951.0[c]	Anderson, IN
	0951.1[c]	Poplar, MT
	0951.3[c]	Millersburg, PA
	0951.4[c]	Keelers Bay, VE
	0951.5[c]	Jackson, MI
	0951.6[c]	Chadson, NE
	0951.9[c]	Myrtle Beach, SC
D. A. texana	1041.0[c]	St. Francisville, LA
	1041.22[c]	New Orleans, LA
	1041.23[c]	Morrilton, AR
	1041.25	So. Richmond, VA
	1041.26	Tallahassee, FL
	1041.27[c]	Goldenhead Branch, FL
	1041.29	Jamestown, SC
	1041.31	Hollandale, MS
D. novamexicana	1031.0[a]	Grand Junction, CO
	1031.4[b]	Moab, UT
	1031.7[a]	Patagonia, AZ
	1031.8[a]	San Antonio, NM
	1031.11[b]	Gila, NM
	1031.12[b]	Antlers, CO

All lines are from the National Drosophila Species Resource center.

[a] A member of group Nova-A.

[b] A member of group Nova-B.

[c] These lines were checked for the appropriate metaphase chromosome compliment; see text for details.

the same expected value, however θ is more influenced by low frequency polymorphic bases. This is because a single rare segregating base contributes little to the average pairwise differences (π), but it is counted as an additional segregating site in the calculation of θ. A measure of the discrepancy between π and θ, TAJIMA's D, is proportional to the difference between these two measures of variation (TAJIMA 1989). When there is an excess of low frequency polymorphisms (as expected with purifying selection or selective sweeps), θ will be bigger than π, and TAJIMA's D will have a negative value. A positive value is expected with balancing selection or population subdivision (TAJIMA 1989). Tajima's D is slightly negative in *D. virilis, D. lummei,* and *D. americana,* but these values are not significant and neutrality cannot be rejected (Table 3). Also, the power of Tajima's D is low with the small sample sizes used here (SIMONSEN *et al.* 1995). In *D. novamexicana* Tajima's D is significant and positive, suggesting that the subdivision into two distinct groups is appropriate.

The Fu and Li test (FU and LI 1993) is similar to that of TAJIMA (1989) and can be used to explore the same selective forces as Tajima's D. This test compares the numbers of mutations that occur in external branches of a genealogy to those that occur on internal branches. Under some types of selection, the number of external mutations deviate from the expectation based on numbers of internal mutations. Fu and Li's D will be negative when there is an excess of external mutations (suggestive of purifying selection or selective sweeps) and positive when there is an abundance (suggestive of balancing selection or population subdivision). Fu and Li's D is slightly negative in *D. lummei* and *D. americana,* and slightly positive in *D. virilis* (Table 3). These values are not significant and neutrality cannot be rejected. In *D. novamexicana* the value of Fu and Li's D is significant and positive (Table 3), in accordance with the findings of Tajima's D for the group.

A third way to look for evidence of natural selection is to compare the numbers of substitutions that result in amino acid replacements with those that do not. If natural selection is acting to fix amino acid replacement mutations within species, we may expect a higher proportion of replacement differences in interspecific contrasts than in intraspecific contrasts. Alternatively, natural selection may be preventing the fixation of

TABLE 2

The number of polymorphic sites within species

		Exons			Introns		
	n	Synonymous	Replacement	No. bases	Base	Length	No. bases[a]
D. virilis	4	3	2	1367	14	1	681
D. lummei	4	3	5	1367	9	0	678
D. americana	15	29	8	1367	44	1	700
D. novamexicana	6	7	3	1367	16	3	690
Nova-A[b]	3	0	0	1367	1	1	711
Nova-B[b]	3	1	0	1367	1	0	700

n is the number of DNA sequences in the sample. Under introns, base refers to base substitutions at the sequence level and length refers to differences in sequence length.

[a] Intron lengths are an average because of length polymorphisms.

[b] Nova-A and Nova-B are two subdivisions of *D. novamexicana.*

TABLE 3

DNA sequence variation summary

	n	S	π	θ	Tajima's D	Fu and Li's D
D. virilis	4	19	0.0057 (0.0033)	0.0058 (0.0033)	−0.195	0.322
D. lummei	4	17	0.0049 (0.0029)	0.0051 (0.0030)	−0.484	−0.189
D. americana	15	81	0.0109 (0.0051)	0.0136 (0.0051)	−0.894	−0.928
D. novamexicana	6	26	0.0077 (0.0041)	0.0059 (0.0030)	1.891*	1.656*
Nova-A	3	1	0.0004 (0.0004)	0.0004 (0.0004)	NA	NA
Nova-B	3	2	0.0007 (0.0006)	0.0007 (0.0006)	NA	NA

n is the number of DNA sequences. S is the number of polymorphic sites within groups, π and θ were calculated using expressions (1) and (2), respectively, and then these quantities were divided by the number of base pairs in the DNA sequences. The standard errors of the estimates, per base pair, are in parentheses. To calculate these, first the variances were determined using expressions (4) and (13) in TAJIMA (1993) for θ and π, respectively. For each variance, the square root was taken and then this quantity was divided by the number of base pairs sequenced. Tajima's D (TAJIMA 1989) compares the similarity of measures of π and θ; it requires at least four sequences to perform the test. Fu and Li's D (FU and LI 1993) also requires four sequences to perform the test. The D values of *D. novamexicana* are significant at the 0.05 level. NA, not available.

replacement polymorphisms. In this case the proportion of replacement polymorphisms, relative to synonymous polymorphisms within species, may be higher than expected on the basis of interspecific fixed differences. MCDONALD and KREITMAN (1991) formulated a test that compares the numbers of sites that are polymorphic within species to those fixed between species for replacement *vs.* synonymous sites. We tested several different species pairs and found no evidence of selection (Table 4).

Recombination and genealogical inference: HUDSON and KAPLAN (1985) described a way to estimate the minimum number of recombination events that are consistent with the polymorphism patterns in a sample of four or more DNA sequences. In general, this estimate is expected to be larger with larger sample sizes, and to be far lower than the actual number of recombination events (HUDSON and KAPLAN 1985). We found that *D. virilis* and *D. lummei,* each with four sequences, must have had recombination occur at least once. *D. americana,* with 15 sequences, has experienced recombination at least 13 times. There is no recombination seen within *D. novamexicana,* however this is not surprising, given the pattern of variation of two distinct types with no intermediate forms.

TABLE 4

MCDONALD-KREITMAN tests

Species pair	χ^2_{1df}
virilis-lummei	0.733
virilis-americana	0.248
lummei-americana	0.970
americana-Nova-A	0.094
americana-Nova-B	1.625

χ^2 tests are for differences between replacements and synonymous sites, within and between species (all contrasts not significant, 1 df). MCDONALD and KREITMAN (1991). See text for details.

Recombination makes the process of gene tree estimation not only problematic but it also causes any particular estimate to be "not real." When there has been recombination within a gene, each piece of the nonrecombined DNA has its own gene tree (HUDSON 1990). Although these different gene trees are not independent, the history of multiple trees means that there is no true bifurcating tree for the gene as a whole. Despite these limitations on their usefulness in the face of recombination, gene tree estimates can still be informative in the case of the presence of deep branches that separate widely divergent taxa. In addition, tree estimates for sequences with a history of recombination share certain structural characteristics. For example, when there has been a lot of recombination scrambling the relationships among different sequences, a tree estimate is expected to have short internal branches relative to the terminal branch tips.

In our tree estimates an outgroup was not used, although the large divergence between *D. virilis* and *D. lummei,* as well as other information (THROCKMORTON 1982; SPICER 1991b, 1992), strongly suggest the root is along this branch. Distance matrices were created using the program DNADIST (PHYLIP 3.5; FELSENSTEIN 1989). A neighbor-joining tree (SAITOU and NEI 1987) was produced by using the PHYLIP program NEIGHBOR (Figure 4A). Neighbor-joining bootstrap trees were produced by using NEIGHBOR in conjunction with the programs SEQBOOT, DNADIST, and CONSENSE. The majority

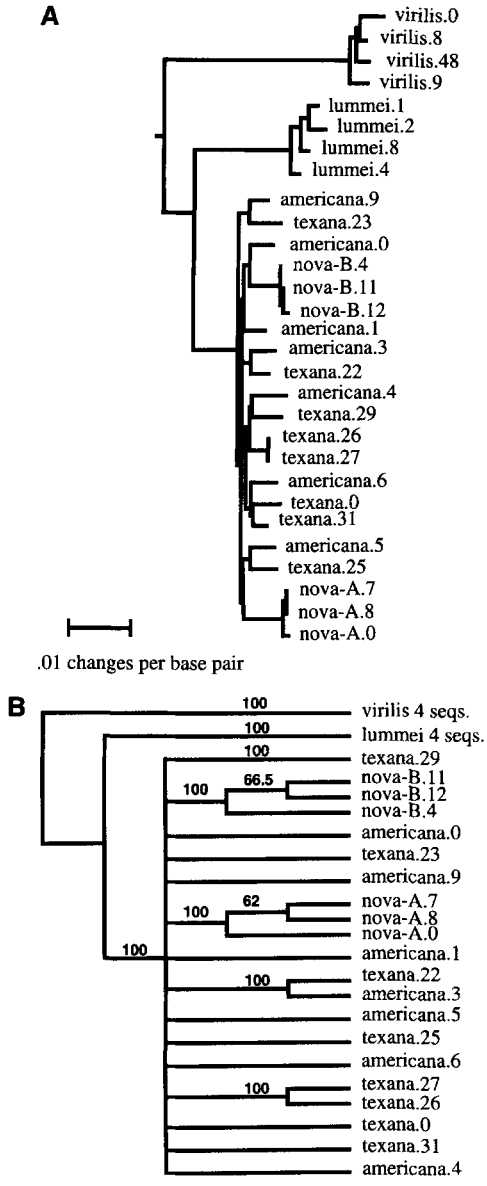


FIGURE 4.—Neighbor joining trees. (A) A standard distance tree. (B) A consensus tree based on bootstrapping; branches that appeared in <60% of trees were collapsed.

rule consensus tree based on 100 replicates is shown in Figure 4B. Branches with bootstrap values of <60% were collapsed. Figure 4 reveals that the relationships of *D. virilis, D. lummei* and the americana complex flies (*D. a. americana, D. a. texana* and *D. novamexicana*) are consistent with other phylogenetic analyses (THROCKMORTON 1982; SPICER 1992).

The effect of recombination can be seen in the two trees in Figure 4. Figure 4A shows that the branches connecting the sequences sampled from the americana complex are joined by short internal branches. This pattern is reminiscent of a star phylogeny and could be taken as evidence of recent population bottleneck and expansion. However, the method of HUDSON and KAPLAN (1985) has revealed multiple recombination events in the history of these sequences. The effect of this scrambling is to distribute the variation among sequences uniformly so that all sequences are about equally different from all others. Most of the short internal branches collapse in the consensus tree (Figure 4B), revealing that various *D. americana* lines are all related to about the same degree and that Nova-A and Nova-B both arise out of this.

DISCUSSION

Selection at *period*: It does not appear that the pattern of variation in this 2.1-kb region of the *per* locus has been strongly affected by natural selection. First, a McDonald-Kreitman test showed no evidence of selection (Table 4, MCDONALD and KREITMAN 1991). Second, Tajima's D and Fu and Li's D were not significantly different from zero in *D. virilis, D. lummei,* and *D. americana* (Table 3, TAJIMA 1989; FU and LI 1993). These tests were significant in *D. novamexicana,* and this result will be discussed below. Third, recombination has occurred at *per,* reducing the length of tight linkage groups, which in turn reduces the probability that any particular portion of the sequence is tightly linked to a site under selection (MAYNARD-SMITH and HAIGH 1974). Similar observations were made by KLIMAN and HEY (1993) for a 1.9-kb portion of the *per* locus studied in the four species of the *D. melanogaster* complex. The region sequenced by KLIMAN and HEY (1993) in the *D. melanogaster* group ends ~150 bases upstream to our *D. virilis* sequence.

Within the region we sequenced in *D. virilis,* the homologous corresponding region of *D. melanogaster* group has a large Thr-Gly repeat of variable length (PEIXOTO *et al.* 1992). ROSATO *et al.* (1994) examined the Thr-Gly region in eight populations of *D. simulans* (a member of the *D. melanogaster* group) and found significant departures from neutrality based on Tajima's D, suggesting balancing selection in *D. simulans.* In the *D. virilis* sequence, there are just two pairs of Thr-Gly repeats (position 3044 of COLOT *et al.* 1988; position 185 of the region sequenced for this paper). We found no variation in this pattern among the lines we sequenced in any of the species. Our results reveal no evidence of selection acting on this very short Thr-Gly region in the *D. virilis* phylad.

***D. a. americana*/*D. a. texana* divergence:** These two subspecies have been differentiated on the basis of a chromosomal fusion of the *X* and the fourth. Both subspecies share the fusion of chromosomes *2* and *3,* which is not seen in the three other species of the phylad, so this fusion is presumably the derived state (PATERSON and STONE 1952; ALEXANDER 1976). However, we show that at the *X*-linked *per* gene there is no divergence between the two subspecies. A trivial explanation, that can be ruled out, is that the stocks were cross contaminated or misidentified with regard to subspecies. First, the subspecies designations were confirmed with mitotic chromosome squashes (see Table 1 and MATERIALS AND METHODS). Second, cross contamination of ameri-

cana/texana stocks is expected to lead to the appearance of identical *per* sequences among different stocks. However, at the sequence level, all the lines were very divergent with the exception of texana.26 and texana.27, which were both collected in Florida.

Chromosomal changes may contribute to speciation in many groups (WHITE 1978). For example, two chromosomal types that differ by inversions or fusions may have different selective advantages in separate environments, and this could lead to speciation if the hybrids between the two are at a selective disadvantage. Low fitness in hybrids could be expected because recombination within the germ line of hybrid individuals that are heterozygous for different karyotypes will generate inviable gametes. However, the X-4 fusion in *D. a. americana* does not seem to lead to an increase in inviable gametes when mixed with *D. a. texana*. BLIGHT (1955) studied the karyotypic frequencies in several populations that contained hybridizing populations of *D. a. americana* and *D. a. texana* near St. Louis. He found hybrids and pure types existed in Hardy-Weinberg equilibrium and concluded that the subspecies distinction was not useful for his populations.

The dual observation of the presence of an X-4 fusion hybrid zone and a lack of divergence at the *X*-linked *per* locus may be the result of a combination of selection, gene flow and recombination. There may be some selection acting, with the X-4 fusion being advantageous in the north, and that advantage diminishes as one moves south. Within the hybrid zone, recombination within hybrids would lead to gene flow between types, which could swamp any effect of selection seen at *per*. Under this scenario, the site or sites of selection that maintain the X-4 fusion hybrid zone are not expected to be in tight linkage with the *per* locus.

The *per* locus data are consistent with a large historical effective population size in *D. americana*. This conclusion is based on the combination of two pieces of evidence. First, the level of variation is high (Table 3). Second, a large portion of the variation looks old, because it is well scrambled by recombination.

The divergence of *D. novamexicana* from *D. americana* may give some insight into the history of chromosomal evolution. Considered together, the two groups of *D. novamexicana* have little *per* locus divergence from *D. americana*, yet they do have a distinct karyotype. *D. novamexicana* has the "ancestral" chromosomal type of no chromosomal fusions, while *D. a. texana* has the 2-3 fusion and *D. a. americana* has both the 2-3 fusion and the X-4 fusion. One explanation is that after *D. novamexicana* split off from ancestral *D. americana*, both the X-4 and 2-3 fusion occurred. Yet from the pattern of sequence data, in which *D. americana* is segregating variation that separates the two *D. novamexicana* groups (see below), it appears that ancestral *D. americana*'s population size was large before the split of *D. novamexicana*. Therefore it seems likely that multiple chromosomal types, including the 2-3 fusion and possibly the X-4 fusion, were segregating before the origin of *D. novamexicana*. Alternatively, the origin of the 2-3 fusion may have been directly associated with the origin of *D. novamexicana*. An additional piece of evidence suggesting that a variety of chromosomal types existed in the ancestor to the americana complex is that *D. novamexicana*, although it has an "ancestral" chromosomal type, contains many of the inversions found in the americana complex relative to *D. virilis* and *D. lummei* (PATTERSON and STONE 1952; THROCKMORTON 1982).

***D. novamexicana*:** The history of *D. novamexicana* seems to have been different than for *D. americana*. Based on *per* locus sequences, the species contains two groups that are divergent at the DNA level, but which have not diverged morphologically or chromosomally. The divergence in the sequence variation is confirmed by the significant results of Tajima's D and Fu and Li's D (Table 3). One explanation for this pattern is that balancing selection is maintaining two distinct "alleles" at high frequency within *D. novamexicana*. Under this model, our designations of Nova-A and Nova-B reflect the divergence of functional *per* locus alleles and are not representative of variation elsewhere in the genome. The most appropriate test of this hypothesis is to examine a second unlinked locus. The alternative explanation is that the *per* locus pattern reflects population level processes and not balancing selection. If this is so, similar patterns of variation are expected elsewhere in the genome. One piece of evidence that argues against balancing selection is that at *D. americana*, the *per* locus sequences reveal a history with considerable recombination. If *D. novamexicana* is a single species with balancing selection maintaining two functionally distinct *per* alleles, then a history of *per* locus recombination is also expected here. Thus the balancing selection model also requires an additional component to explain the absence of recombination in *D. novamexicana*.

Regardless of whether the pattern of *per* locus variation has been due to natural selection, or whether it is because *D. novamexicana* consists of two populations that are not exchanging genes (*e.g.*, "cryptic" species), the variation does show that *D. novamexicana* is very closely related to *D. americana*. Furthermore, it does appear that *D. novamexicana* probably arose from an ancestral species that had a large population size. This can be inferred from the 11 fixed differences between Nova-A and Nova-B that were found still segregating in *D. americana*.

If the pattern of variation at *per* is taken to be representative of the genome, then we can consider the kinds of processes that might have given rise to two groups within *D. novamexicana*. Both groups have in common a light mesothorax color, a chromosomal karyotype, and geographic range that is separate from that for *D. americana*. The two groups also share a single fixed difference, with respect to *D. americana*, in the *per* locus sequence. Though the two groups could have arisen

independently, these shared characteristics suggest that *D. novamexicana* arose once and then split into two groups.

If a single origin of *D. novamexicana* is taken as a working hypothesis, then some other aspects of this speciation event and initial divergence can be explored. During the time between the origin of *D. novamexicana* and the divergence of Nova-A and Nova-B, *D. novamexicana* (1) acquired its lighter mesothorax color, (2) began living in a drier habitat, (3) did not lose much *per* locus variation, and (4) may have acquired one new substitution at *per*. However, the very low level of divergence between *D. novamexicana* (including both Nova-A and Nova-B) and *D. americana* suggests that it was soon after the origin of *D. novamexicana* that this new species split into two groups. If this model of two splitting events in rapid succession is correct, it follows that the evolution of the lighter mesothorax color was fairly rapid.

KLIMAN and HEY (1993) examined four closely related taxa in the *D. melanogaster* group for DNA sequence variation at *per*. They concluded that *D. simulans*, a large population still segregating very old lineages, gave rise to two island species, *D. mauritiana* and *D. sechellia*. They found that *D. simulans* was still segregating polymorphism fixed between *D. mauritiana* and *D. sechellia*. They felt it was appropriate to consider *D. simulans* a parent species to the two island species. In our analysis, we found that *D. americana* is a large population that is still segregating variation that predates the split of Nova-A and Nova-B. These two findings of large populations that still segregate old variation suggest that speciation may often proceed by the formation of daughter species that bud off of larger species.

This work was supported by National Science Foundation grant DEB-9306625 to J.H. and an Anne B. and James H. Leathem Scholarship fund for dissertation improvement to H.H.

LITERATURE CITED

ALEXANDER, M. L., 1976 The genetics of *Drosophila virilis*, pp. 1365–1419 in *The Genetics and Biology of Drosophila*, Vol. 1C, edited by M. ASHBURNER and E. NOVITSKY. Academic Press, New York.

ASHBURNER, M., 1989 *Drosophila: A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

BARTON, N. H., and B. CHARLESWORTH, 1984 Genetic revolutions, founder effects, and speciation. Ann. Rev. Ecol. Syst. **15:** 133–164.

BLIGHT, W. C., 1955 A cytological study of linear populations *Drosophila americana* near St. Louis, Missouri. Ph.D. dissertation, Washington University, St. Louis, MO.

CARSON, H. L., 1968 The population flush and its genetic consequences, pp. 123–137 in *Population Biology and Evolution*, edited by R. C. LEWONTIN. Syracuse University Press, Syracuse, NY.

CARSON, H. L., 1975 The genetics of speciation at the diploid level. Am. Nat. **109:** 83–92.

CARSON, H. L., and W. C. BLIGHT, 1952 Sex chromosome polymorphism in a population of *Drosophila americana*. Genetics **37:** 572.

COLOT, H. V., J. C. HALL and M. ROSBASH, 1988 Interspecific comparison of the period gene of drosophila reveals large blocks of non conserved coding DNA. EMBO J. **7:** 3929–3937.

COYNE, J. A., 1992 Genetics and speciation. Nature **355:** 511–515.

COYNE, J. A., and H. A. ORR, 1989 Patterns of speciation in *Drosophila*. Evolution **43:** 362–381.

DEVEREUX, J., and P. HAEBERLI, 1991 Genetics computer group: program manual for the GCG package. version 7. Genetics Computer Group, Madison, WI.

DOBZHANSKY, T., 1937 *Genetics and the Origin of Species.* Columbia University Press, New York.

EVGEN'EV, M. B., 1971 The pattern of polytene chromosome conjugation and crossing-over in interspecific hybrids of *Drosophila*. Theoret. Appl. Genet. **41:** 249–254.

FELSENSTEIN, J., 1989 PHYLIP. Phylogeny Inference Package, version 3.2. Cladistics **5:** 164–166.

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

GUBENKO, S., and M. B. EVGEN'EV, 1984 Cytological and linkage maps of *Drosophila virilis* chromosomes. Genetica **65:** 127–139.

HEY, J., and R. M. KLIMAN, 1993 Population genetics and phylogenetics of DNA sequence variation at multiple loci within and among species of the *D. melanogaster* complex. Mol. Biol. Evol. **10:** 804–822.

HSU, T. C., 1952 III. Chromosomal variation and evolution in the virilis group of Drosophila, pp. 35–72 in University of Texas Publications #5204 *Studies in the Genetics of Drosophila VII.*

HUDSON, R. R., 1990 Gene genealogies and the coalescent process, pp. 1–44 in *Oxford Surveys in Evolutionary Biology*, Vol. 7, edited by D. FUTUYMA and J. ANTONOVICS. Oxford University Press, New York.

HUDSON R. R., and N. L. KAPLAN, 1985 Statistical Properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111:** 147–164.

HUDSON, R. R., M. SLATKIN and W. P. MADDISON, 1992 Estimation of levels of gene flow from DNA sequence data. Genetics **132:** 583–589.

KLIMAN, R. M., and HEY, J. 1993 DNA sequence variation at the period locus within and between species of the *D. melanogaster* complex. Genetics **133:** 375–387.

KYRIACOU, C. P., and J. C. HALL, 1984 Learning and memory mutations impair acoustic priming of mating behavior in Drosophila. Nature **308:** 62–65.

MANTEL, N., 1967 The detection of disease clustering in a generalized regression approach. Cancer Res. **27:** 209–220.

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

MAYR, E., 1942 *Systematics and the Origin of Species.* Columbia University Press, New York.

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

NEUFELD, T. P., R. W. CATHEW and G. M. RUBIN, 1991 Evolution of gene position: chromosomal arrangement and sequence comparison of the *Drosophila melanogaster* and *Drosophila virilis sina* and *Rh4* genes. Proc. Natl. Acad. Sci. USA **88:** 10203–10207.

NURMINSKY, D. I., E. N. MORIYAMA, E. R. LOZOVSKAYA and D. L. HARTL, 1995 Molecular phylogeny and genome evolution in the *Drosophila virilis* species group: duplications of the alcohol dehydrogenase gene. Mol. Biol. Evol. **13:** 132–149.

PATTERSON, J. T., 1942a Drosophila and speciation. Science **95:** 153–159.

PATTERSON, J. T., 1942b IX. Distribution of the virilis group in the United States, pp.153–161 in University of Texas Publications #4228 *Studies in the Genetics of Drosophila.*

PATTERSON, J. T., and W. S. STONE, 1952 *Evolution in the genus Drosophila.* MacMillian, New York.

PEIXOTO, A. A., R. COSTA, D. A. WHEELER, J. C. HALL and C. P. KYRIACOU, 1992 Evolution of the threonine-glycine repeat region of the *period* gene in the *melanogaster* species subgroup. J. Mol. Evol. **35:** 411–419.

REINBOLD, S. L., and G. E. COLLIER, 1990 Molecular systematics of the *Drosophila virilis* species group (Diptera:Drosophilidae). Ent. Soc. Amer. **83:** 468–474.

ROHLF, F. J., 1985 NTSYS. Numerical Taxonomy System of Multivariate Statistical Programs. State University of New York, Stony Brook.

ROSATO, E., A. A. PEIXOTO, G. BARBUJANI, R. COSTA and C. P. KYRIACOU, 1994 Molecular polymorphism in the period gene of *Drosophila simulans*. Genetics **138:** 693–707.

SAITOU, N., and M. NEI, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

SIMONSEN, K. L., G. A. CHURCHILL and C. F. AQUADRO, 1995 Proper-

ties of statistical tests of neutrality for DNA polymorphism data. Genetics **141:** 413–429.

SPICER, G. S., 1991a The genetic basis of a species specific character in the *Drosophila virilis* group. Genetics **128:** 331–337.

SPICER, G. S., 1991b Molecular evolution and phylogeny of the *Drosophila virilis* species group as inferred by two-dimensional electrophoresis. J. Mol. Evol. **33:** 379–394.

SPICER, G. S., 1992 Reevaluation of the phylogeny of the *Drosophila virilis* species group (Diptera: Drosophilidae). Ann. Entomol. Soc. Am. **85:** 11–25.

STURTEVANT, A. H., and E. NOVITSKI, 1941 The homologies of the chromosome elements in the genus *Drosophila*. Genetics **26:** 517–541.

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

TAJIMA, F., 1993 Measurement of DNA polymorphism, pp. 37–59 in *Mechanisms of Molecular Evolution,* edited by N. and A. G. CLARKE. Sinauer Associates, Sunderland, MA.

THROCKMORTON, L., 1982 The virilis species group, pp. 227–296 in *The Genetics and Biology of Drosophila,* Vol. 3b, edited by M. ASHBURNER and E. NOVITSKY. Academic Press, New York.

TEMPLETON, A. R., 1980 The theory of speciation via the founder principle. Genetics **94:** 1011–1038.

TEMPLETON, A. R., 1989 The meaning of species and speciation: a genetic perspective, pp. 3–27 in *Speciation and Its Consequences,* edited by D. OTTE and J. ENDLER. Sinaur Press, Sunderland, MA.

TONZETICH, J., S. HAYASHI and T. A. GRIGLIATTI, 1990 Conservation of sites of tRNA loci among the linkage groups of several Drosophila species. J. Mol. Evol. **30:** 182–188.

WATTERSON, G. A., 1975 On the number of segregating sites in genetical models with recombination. Theoret. Pop. Biol. **7:** 256–276.

WHITE, M. J. D., 1978 *Modes of Speciation.* Freeman and Co., San Francisco.

WHITING, J. H. JR., M. D. PILEY, J. L. FARMER and D. E. JEFFERY, 1989 *In situ* hybridization analysis of chromosomal homologies in *Drosophila melanogaster* and *Drosophila virilis.* Genetics **122:** 99–109.

WRIGHT, S., 1940 Breeding structure of populations in relation to speciation. Am. Nat. **74:** 232–248.

Communicating editor: A. G. CLARK