Genet. Res., Camb. (1997), **70**, *pp.* 185–194. *With* 4 *figures.* Printed in the United Kingdom © 1997 Cambridge University Press

A multilocus view of speciation in the *Drosophila virilis* species group reveals complex histories and taxonomic conflicts

HOLLY HILTON[1] AND JODY HEY*

Department of Microbiology and Genetics, Rutgers University, Nelson Hall, PO Box 1059, Piscataway, NJ 08855-1059, USA

(*Received 6 May 1997 and in revised form 9 September 1997*)

Summary

The historical population genetic processes associated with the divergence of members of the *Drosophila virilis* species group were examined using DNA sequence variation from two loci. New data on DNA sequence variation from the *oskar* locus, taken from within and among all five closely related taxa in the *virilis* phylad of the *D. virilis* species group, were examined and compared with similar data previously collected from the *period* locus. Overall, the *oskar* and *period* data sets reveal similar patterns of variation. Both loci support the conclusion that the two subspecies of *D. americana* have had a large historical population size and are exchanging genes in nature. From these data there is little reason to consider them as distinct taxa. In the case of *D. novamexicana*, from which six lines were sequenced at each locus, there is an intriguing difference in the pattern seen at the two loci. Both loci reveal two distinct groups that are considerably divergent from each other, with very little evidence of gene flow between them. However, the grouping of lines into distinct subgroups based on *oskar* is different from the grouping based on *period*. The simplest explanation seems to be that *D. novamexicana* includes two distinct species, and that the sample of six lines happens to include cases of recent gene exchange. Alternatively, both *oskar* and *period* could be linked to sites of strong balancing selection and limited recombination.

1. Introduction

One way to explore the genetic changes that occur in populations around the time of speciation is to examine the patterns of DNA sequence variation in species that recently shared a common ancestor. The levels of divergence and of shared variation can be interpreted in light of speciation models, particularly when the data come from more than one portion of the genome (Hey, 1994). Inclusion of multiple loci is necessary so that forces that have affected all loci (e.g. population subdivision or changes in population size) can be distinguished from those that act on individual loci (e.g. natural selection).

The *Drosophila virilis* group, which includes several well-studied species that are very closely related, is traditionally split into two phylads: the *virilis* phylad and the *montana* phylad (Throckmorton, 1982). Our research has focused on the species of the *virilis* phylad, which contains five taxa: *D. virilis*, *D. lummei*, *D. novamexicana*, *D. americana americana* and *D. americana texana*. Previously we examined DNA sequence diversity at the X-linked *period* locus, within and among the members of the *virilis* phylad (Hilton & Hey, 1996). In this report we add a second locus, *oskar*, to our population genetic study of speciation.

In *D. melanogaster*, the *oskar* gene is a maternal effect gene required for both posterior body patterning and germline formation in the early embryo, the *D. virilis oskar* homologue (also called *virosk*) is required for body patterning but may not be as important in pole cell formation (Webster *et al.*, 1994). In *D. melanogaster*, *oskar* is located on the third chromosome (3R). Element 3R corresponds to chromosome 2 of the *D. virilis* phylad (Sturtevant & Novitski, 1941). A 2nd chromosome location of *oskar* in the *D. virilis* phylad is expected because of the high degree of conservation of chromosomal elements between *D. melanogaster* and *D. virilis* (Sturtevant & Novitski, 1941; Alexander, 1976). This conservation of linkage groups between *D. melanogaster* and *D. virilis* has also been confirmed for many individual loci (Whiting *et al.*, 1989; Tonzetich *et al.*, 1990; Neufeld *et al.*, 1991)

* Corresponding author. Tel: +1(908)-445-5272. Fax: +1(908)-445-5870. e-mail: hey@mbcl.rutgers.edu. www: http://heylab.rutgers.edu/.

[1] Present address: Department of Plant Sciences, Rutgers University, New Brunswick, NJ 08903, USA.

including *period* (Kress, 1993). In both *D. a. americana* and *D. a. texana*, chromosome 2 is fused with chromosome 3.

The *oskar* locus was chosen for this analysis because it had been sequenced in *D. virilis* previously; because it showed a moderate amount of divergence between *D. virilis* and *D. melanogaster* (Webster *et al.*, 1994); and because of its location on an autosome, which makes it unlinked to the *period* locus.

D. virilis is found in wild habitats in Japan and China, while in North America it is restricted to domestic habitats (Patterson & Stone, 1952). *D. lummei* is found in northeastern Europe. The North American taxa, *D. a. americana*, *D. a. texana* and *D. novamexicana*, are closely related and are referred to as the american complex. *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 (Patterson, 1942; Carson & Blight, 1952; Patterson & Stone, 1952; Throckmorton, 1982) with *D. a. american* found to the north and *D. a. texana* found to the south. These two subspecies can only be distinguished on the basis of a chromosomal fusion of elements X-4 found exclusively in *D. a. americana*, and the hybrid zone has been described on the basis of the geography of chromosomal variation. 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 colour than *D. a. americana* or *D. a. texana*, which have a dark-brown body colour. This change in mesothorax colour is presumably for desiccation resistance. It has been suggested that the change accompanied speciation and the ability to live in the drier habitat (Spicer, 1991*a*; Orr & Coyne, 1992).

2. Materials and methods

(i) *The flies*

All strains were obtained from the National Drosophila Species Resource Center (NDSRC) and, with the exception of a *D. a. texana* line number 1041.24 from New Orleans and *D. virilis* line number 1051.47 from China, they are the same lines used in the study of the *period* locus (Hilton & Hey, 1996). All extension numbers used here correspond to those from the stock center (Table 1). The strains are sometimes referred to by species name and extension number. For example, 'virilis.0' corresponds to NDSRC #1051.0 from Pasadena, CA, and 'americana.1' corresponds to NDSRC #0951.1 from Poplar, MT (Table 1).

(ii) *Inbreeding the lines*

To produce strains of flies that had very low heterozygosity levels, so that a single individual fly could be expected to carry identical copies of the *oskar* locus, all lines were inbred using full brother–sister mating for eight generations. This is expected to create an inbreeding coefficient of 0·82 (Falconer, 1989, p. 93) or a reduction of heterozygosity to an average of 0·18 of the original heterozygosity. In addition, most of these lines have been kept in captivity at low population size for at least 10 years. This long and slow inbreeding should have reduced heterozygosity considerably, and it may also have helped to eliminate deleterious alleles that can limit the effectiveness of rapid inbreeding. One inbred line (americana.0) was not used in this study, due to suspected heterozygosity. This line appeared to be polymorphic for a small deletion found within some of the *D. americana* lines.

Table 1. *List of lines sequenced*

Species name	Line number	Location
D. virilis	1051.0	Pasadena, CA
	1051.8	Truckee, CA
	1051.47	Hangchow, China
	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	0951.1	Poplar, MT
	0951.3	Millersburg, PA
	0951.4	Keelers Bay, VE
	0951.5	Jackson, MI
	0951.6	Chadson, NE
	0951.9	Myrtle Beach, SC
D. a. texana	1041.0	St Francisville, LA
	1041.23	Morrilton, AR
	1041.24	New Orleans, LA
	1041.27	Goldenhead Branch, FL
	1041.31	Hollandale, MS
D. novamexicana	1031.0[1]	Grand Junction, CO
	1031.4[1]	Moab, UT
	1031.7[1]	Patagonia, AZ
	1031.8[2]	San Antonio, NM
	1031.11[2]	Gila, NM
	1031.12[2]	Antlers, CO

All lines are from the National Drosophila Species Resource Center. [1] Indicates a member of group Nova-1, [2] indicates a member of group Nova-2.

(iii) *DNA preparation and sequencing*

DNA preparations were made from single male flies (protocol 48 in Ashburner, 1989). A 1·2 kb region of the *oskar* gene was PCR amplified, using 20-mer oligonucleotide primers, starting at positions 2285 ('+' primer 5′ base) and 3448 ('−' primer 5′ base) of Webster *et al.* (1994), GenBank accession L22556 (Fig. 1). PCR and DNA sequencing methods were identical to those of Kliman & Hey (1993). Sequence alignment was done by eye.

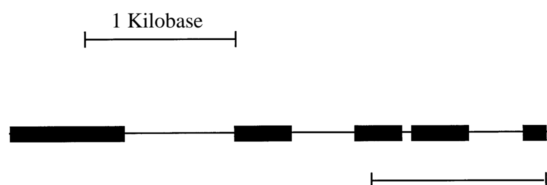

Fig. 1. Diagram of the exon (black box) and intron (black line) structure of the *oskar* locus (Webster *et al.*, 1994). The 5′ end is to the left. The specific region sequenced is marked with the lower black line.

(iv) *Estimating population genetic parameters*

The computer program SITES was used to estimate various population genetic parameters (Hey & Wakeley, 1997). Both π, the average number of pairwise nucleotide differences, and $\hat{\theta}$, calculated from the number of segregating sites, were used to estimate the neutral population mutation rate $4Nu$, with N defined as the effective population size and u as the neutral mutation rate (Tajima, 1993). The population migration rate Nm, where m is the migration rate per generation, was estimated from the Fst estimate of Hudson *et al.* (1992). The estimator γ of the population recombination rate $4Nc$ (Hey & Wakeley, 1997) was also calculated using the SITES program.

(v) *Statistical tests of population subdivision and natural selection*

The HKA test of natural selection (Hudson *et al.*, 1987) was applied to the data from both loci. This test is based on neutral theory predictions that regions of the genome that evolve at high rates between species will also have high levels of polymorphism within species (Hudson *et al.*, 1987). The test compares within- and between-species variation for two species at two or more loci. The test is similar to a χ^2 test comparing the observed and expected values for the number of polymorphic sites within each of two species and the between-species divergence. A Mantel test (Mantel, 1967) was used to check for evidence of population subdivision, as described by Hilton & Hey (1996).

(vi) *Double checking the* D. novamexicana *lines*

At the *period* locus, the *D. novamexicana* lines fell into two very distinct groups of three lines each (Hilton & Hey, 1996). At *oskar* there were also two very distinct groups, each with three lines of *D. novamexicana*, but there was a switch of a pair of the lines relative to the *period* locus groupings (see Section 3). In order to be sure that these observations truly represented our sample of *D. novamexicana* lines, and were not caused by a mixup within our laboratory, we conducted three experiments. In principle, there were three possible occasions for mixing: the single fly DNA preparations used for the sequencing of the *period* locus could have been mixed; the single fly DNA preparations used for sequencing *oskar* could have been mixed; or the actual fly strains could have been mixed, perhaps during inbreeding, or at some time after we received them. To test the first possibility, we resequenced portions of the *oskar* locus from the single fly DNA preparations that had previously been used to obtain the *period* sequences (and which had been stored in the freezer in Parafilmed tubes). Second, to test whether the DNA preparations used for *oskar* sequencing had been mixed up, we resequenced portions of *period* using the single fly DNA preparations that had previously been used for *oskar* sequencing. The third possible scenario, the mixing of fly lines, was checked by reordering strains from the stock centre; obtaining DNA from single flies; and resequencing portions of both *oskar* and *period* from those samples. No evidence of mixing was found in any of these investigations, and all three experiments support the finding that the *oskar* and *period* groupings of *D. novamexicana* lines are indeed different.

3. Results

(i) *Grouping the taxa and patterns of polymorphism*

Fig. 2 shows all the variable sites and Table 2 lists the types of variation found in each taxon for *oskar*. The sequenced region is rich in variation, and reveals several synonymous and amino acid replacement variants as well as considerable indel variation and many base pair changes in the introns. Because of taxonomic uncertainties, most analyses have been applied to two different groupings in both the case of *D. a. americana* and *D. a. texana* and the case of *D. novamexicana*.

D. a. americana and *D. a. texana* have been distinguished on the basis of a chromosomal fusion that defines a hybrid zone. *D. a. texana* lacks the fusion, which is found exclusively in *D. a. americana*. However, their biological and taxonomic relationships remain unresolved (Throckmorton, 1982; Spicer, 1992; Hilton & Hey, 1996). Consequently, most analyses were done on *D. a. americana* and *D. a. texana* together as one species, under the name *D. americana*, and as two separate species.

Several related analyses on the *oskar* data find little reason to distinguish between *D. a. americana* and *D. a. texana*. First, there were no fixed differences between the two taxa and they shared nine polymorphisms. Second, the Fst estimate of the population migration rate, Nm, was 1·944 (Hudson *et al.*, 1992). This is above 1·00, the value above which there is expected to be fairly little heterogeneity among populations (Wright, 1940). Lastly, a comparison of the pairwise differences within *D. a. americana* and *D. a. texana* versus values of pairwise differences between the groups did not find a significant difference between the two groups of differences (Mantel test, $P = 0{\cdot}08$). This specific test examines the null hypothesis that

```
Base
position                1222222   334   455   566666666666677777777777777777777777777788888888888888888888888888888
                      193124559   225   878   925666677777800011122234556666666677889900000000001111111111222222233
            7         5926656733  485   663   752458901239045926846871272345678969385901234567890123456789012345727
Comment     r         sssiiiddr   ssr   ssr   ssiiiddddddddiiiiiiiiiiiiiiiddddddiiiiidddddddddddddddddddddddddddidi
virilis-0   G(V)ACTAAG**T(I)ATA(N)GGA(I)ACGGG********CATACTAGCCCAAAG******TCTCTACAACAAGCGTGATTAGATAAATGCTAG*T
virilis-8   -(-)---------(-)---(-)---(-)--------------C--T-T-------C---------T-------------------------------
virilis-47  -(-)----G----(-)---(-)---(-)--AT----------C----T-----------------T-T-----------------------------
virilis-48  -(-)--C-G----(-)---(-)---(-)--A-----------C----T----G--C---------T-------------------------------
lummei-1    C(L)G--------(-)G-G(S)-C-(-)----C---------C-C--T-TTT-C-CTCGAACCGC-C***************************A--
lummei-2    C(L)---------(-)G-G(S)-C-(-)----C---------C-C--T-TTT-C-CTCGAACCGC-C***************************A--
lummei-4    C(L)G--G-----(-)G-G(S)-C-(-)----C---------C-C--T-TTT-C-CTCGAACCGC-C***************************A--
lummei-8    C(L)G--------(-)G-G(S)-C-(-)----C---------C-C--T-TTT-C-CTCGAACCGC-C***************************A--
Amer-1      C(L)GG-T-ATC-(-)G-G(S)-C-(-)----CTATATATAGC-C--T-TTT-C-CTCGAACCGC--***************************AT-
Amer-3      C(L)G--T-ATC-(-)G-G(S)-C-(-)----CTATATATA-C-C-CT-TTT-C-CTCGAACAGC-C***************************A--
Amer-4      C(L)GG-T-ATCC(T)G-G(S)-C-(-)T---CTATATATA-C-C--T-TTT-C-CTCGAACCGC--***************************AT-
Amer-5      C(L)GG-T-ATT-(-)G-G(S)-C-(-)T---CTATATATA-C-C--T-TTT-C-CTCGAACCGC--***************************AT-
Amer-6      C(L)GG-T-ATT-(-)GCG(S)-C-(-)----CTATATATA-C-C--T-TTT-C-CTCGAACCGC-C***************************A--
Amer-9      C(L)GG-T-ATC-(-)G-G(S)-C-(-)----C---------C-C--T-TTT-C-CTCGAACCGC-C***************************A--
Texana-27   C(L)GG-T-ATC-(-)G-G(S)-T-(-)----CTATATATA-CGC--T-TTT-CGCTCGAACCGC-C***************************A--
Texana-24   C(L)GG-T-ATC-(-)G-G(S)-T-(-)----CTATATATA-C-C--T-TTT-CGCTCGAACCGC-C***************************A--
Texana-31   C(L)GG-T-ATC-(-)G-G(S)-CC(L)----CTATATATA-C-C--T-TTT-C-CTCGAACCGC-C***************************A--
Texana-23   C(L)GG-T-ATC-(-)NCG(S)-C-(-)-A--CTATATATA-C-C--T-TTT-C-CTCGAACCGC-C***************************A-C
Texana-0    C(L)GG-T-ATT-(-)G-G(S)-C-(-)-A--CTATATATA-C-C--T-TTT-C-CTCGAACCGC-C***************************A-C
Novamex-0   C(L)GG-T-ATC-(-)G-G(S)AC-(-)----CTATATATAGC-C--T-TTT-C-CTCGAACCGC-C***************************A-C
Novamex-4   C(L)GG-T-ATC-(-)G-G(S)-C-(-)----CTATATATAGC-C--T-TTT-C-CTCGAACCGC-C***************************A-C
Novamex-7   C(L)GG-T-ATC-(-)G-G(S)-C-(-)----CTATATATAGC-C--T-TTT-C-CTCGAACCGC-C***************************A-C
Novamex-8   C(L)GG-T-ATC-(-)G-G(S)AC-(-)----CTATATATAGC-C--TCTTT-C-CTCGAACCGC-C***************************A--
Novamex-11  C(L)GG-T-ATC-(-)G-G(S)AC-(-)----CTATATATAGC-C--TCTTT-C-CTCGAACCGC-C***************************A--
Novamex-12  C(L)GG-NNNNNN(-)G-G(S)AC-(-)----CTATATATAGC-C--TCTTT-C-CTCGAACCGC-C***************************A--

Base                                                                                                     1111111
position    888888888888888888888888888888888888888899999999999999999999999999999999999999999999999999999990000001
            344444444455555555556666677777777888888900011111111112222222222333333334444445558888999990000113
            801234567901234567890134901456789012357978901234567890123456789012345780134672362348013683689567
Comment     iiiiiiiiidddddddddddddiiiiiiidddddddiiiiiddddddddddddddddddddddddddddiiiidiiiiiiiiiiiiiiiiiiiiii
virilis-0   CAGCTTAAGAGCAACGTATATTTTCCGCCATTTTATATTTT****************************TAGGTCGCGGAAAAAATACACATTATT
virilis-8   ---------------------------------------C--------------------------------C-----------------T-----
virilis-47  -----------------T---------------C-----C-------------------------------------------T------------
vililis-48  ---------*************-C---T-----------C-------------------------------------------T------------
lummei-1    GG--C---------A--T----A--T-------C-----CGTTCTGAAAGCAGCAGGTTAAAGTTCAGCAC---A--------G--GT--TACTA-
lummei-2    GG--C---------A--T----A--T--G----C-----CGTTCTGAAAGCAGCAGGTTAAAGTTCAGCAC---A--------G--GT--TACTA-
lummei-4    GG--C---------A--T----A-TT-------C-----CGTTCTGAAAGCAG-------------CAGG--A----------G--GT--TACTA-
lummei-8    GG--C---------A--T----A--T-------C-----CGTTCTGAAAGCAG-------------CAGG--A----------G--GT--TACTA-
Amer-1      -G---------A--A--T----C----A-----C----GCGTTCTGAAAGCAG-------------CAGG-----------TG---GT--TAC---
Amer-3      -G--C-GG------A--TC---C------*******---CGTTCTGAAAGCAG-------------CAGG---------T------GTCTTAC---
Amer-4      -G---------A--A--T----C---A------C-----CGTTCTGAAAGCAG-------------CAGG-------A-T------GT--TAC---
Amer-5      -G---------A--A--T----C---A------C-----CGTTCTGAAAGCAG-------------CAGG-C-C**A----G-----GT--TAC---
Amer-6      -G--C-G-A-----A--T----C----------C----GCGTTCTGAAAGCAG-------------CAGG---------T------GT--TAC---
Amer-9      -G--C-GGA-----A--T----C----------C-----CGTTCTGAAAGCAG-------------CAGG---------T------GT--TAC---
Texana-27   -G--C-GG------A--T----C----------C----GCGTTCTGAAANCAG-------------CAGG------A--TG---TGGT-TTAC---
Texana-24   -G--C-GG------A--T----C----------C----GCGTTCTGAAAGCAG-------------CAGG------A--TG---TGGT-TTAC---
Texana-31   -G---A-----T--A--T----G----------C-----CGTTCTGAAAGCAG-------------CAGG---------T------GTC-TAC---
exana-23    -GCG-A-----T--A--T----C----------C--GAGCGTTCTGAAAGCAG-------------CAGG---------T------GT--TAC---
Texana-0    -GCG-A-----T--A--T----C----------C--GAGCGTTCTGAAAGCAG-------------CAGG---------T------GT--TAC---
Novmaex-0   -G---------T--A--T----C----------C-----CGTTCTGAAAGTAG-------------CAGG---------T------GT--TAC--A
Novamex-4   -G---------T--A--T----C----------C-----CGTTCTGAAAGTAG-------------CAGG---------T------GT--TAC--A
Novamex-7   --A--------T--A--T----C----------C-----CGTTCTGAAAGTAG-------------CAGG---------T------GTG-TAC--A
Novamex-8   -G--C-G-A-----A--T----C----------C---G-CGTTCTGAAAACAG-------------CAGG------A-TT------GT--TAC---
Novamex-11  -G--C-G-A-----A--T----C----------C---G-CGTTCTGAAAACAG-------------CAGG------A-TT------GT--TAC---
Novamex-12  -G--C-G-A-----A--T----C----------C---G-CGTTCTGAAAACAG-------------CAGG------A-TT------GT--TAC--N
```

Fig. 2. Variable sites in *oskar*. 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 two non-coding intron regions that correspond to base positions 203–262 and 645–1133, and a non-coding region after the last exon corresponding to base positions 1134–1148. 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 (L, Leu; S, Ser; I, Ile; T, Thr; V, Val; N, Asn). Length variation is indicated by an asterisk (∗) in sequences shortened relative to others.

pairwise difference values between taxa are not different from those within taxa. The null hypothesis could not be rejected, consistent with no population subdivision between the groups. The Mantel test and the estimated migration rate are suggestive of limited population structure between *D. a. americana* and *D. a. texana*. This is because the *D. a. texana* samples have less variation than those from *D. a. american* (Table 3).

Among the six lines of *D. novamexicana* there was considerable variation (16 polymorphic sites, Table 3); however, there appear to be two distinct groups in a pattern quite similar (but not identical) to that found at the *period* locus (Hilton & Hey, 1996). When taken as a group of six sequences, there was little to distinguish them from *D. americana*. There were no fixed differences distinguishing *D. novamexicana* from *D. americana* and there were six shared polymorphisms. However, the *D. novamexicana* sequences fell into two divergent sets, each of three lines. Nearly all the variation found within *D. novamexicana* exists as differences between the two groups. There were 12 fixed differences between the groups and they shared no polymorphisms. In Tables 2 and 3, *D. novamexicana* is treated as one group (of six lines) and also as two groups, named 'Nova-1' and 'Nova-2' (Note: these are not identical to the *D. novamexicana* grouping based on the *period* data from Hilton & Hey,

Table 2. *The number of polymorphic sites within species at* oskar

		Exons		Introns	
	n	Synonymous	Replacement	Base	Length
D. virilis	4	1	0	18	1
D. lummei	4	1	0	9	1
D. americana	11	5	2	36	5
D. a. americana	6	3	1	24	5
D. a. texana	5	3	1	21	0
D. novamexicana	6	1	0	15	0
Nova-1	3	1	0	3	0
Nova-2	3	0	0	0	1

n is the number of sequences in the sample. The average number of synonymous sites among the sequences is 142 and the average number of replacement sites is 497. The total is 639, corresponding to 213 codons. Under 'Introns', 'Base' refers to the number of polymorphic base positions and 'Length' refers to the number of distinct indel polymorphisms. The average number of intron bases was 512. The data set also includes 14 bases just 3′ of the termination codon, but this region contained no variation. *D. americana* includes both *D. a. americana* and *D. a texana*, and *D. novamexicana* includes both Nova-1 and Nova-2 (see text).

Table 3. *DNA sequence variation summary*

	n	*S*	$\hat{\theta}$	π	Tajima's *D*
D. virilis	4	19	0·0102	0·0098	−0·359
D. lummei	4	9	0·0046	0·0050	0·860
D. americana	11	43	0·0140	0·0128	−0·474
D. a. americana	6	28	0·0117	0·0113	−0·239
D. a. texana	5	25	0·0113	0·0128	0·992
D. novamexicana	6	16	0·0071	0·0086	1·346
Nova-1[a]	3	4	0·0026	0·0026	na
Nova-2[a]	3	0	0·0	0·0	na

S is the number of polymorphic sites within groups, π is the average number of pairwise difference, and $\hat{\theta}$ is Watterson's estimator (Watterson, 1975; Tajima, 1993). Both π and $\hat{\theta}$ are estimates of $4Nu$ and both are shown divided by the number of base pairs. Tajima's *D* (Tajima, 1989) compares the similarity of measures of π and $\hat{\theta}$; it requires at least four sequences to calculate. None of the *D* values is significantly different from zero.

1996). Nova-1, consists of three lines: nova.0, nova.4 and nova.7. Within that group, nova.0 and nova.4 differ by one site, nova.0 and nova.7 differ at four sites, and nova.7 and nova.4 differ by three sites. In Nova-2, lines nova.8, nova.11 and nova.12 are all identical. There is one unique change that is found in all three of the Nova-2 lines and nova.0. Six of the 12 fixed differences found between Nova-1 and Nova-2 are polymorphic within *D. americana*. Of the six remaining fixed differences between the two groups, two are found to be unique to Nova-1 and four to Nova-2. These six unique changes may have arisen since each group became isolated from the species that was ancestral to *D. americana* and *D. novamexicana*. Alternatively, these unique changes could be polymorphisms within *D. americana* that did not appear in our sample or that were once polymorphic within *D. americana* but are now monomorphic.

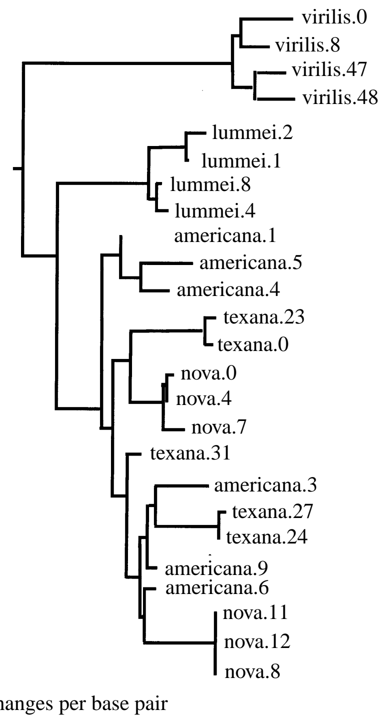

Fig. 3. A neighbour-joining tree (Saitou & Nei, 1987) of the *oskar* data generated with the PHYLIP programs DNADIST and NEIGHBOR (PHYLIP 3.5; Felsenstein, 1989). The distance matrix was generating using the Kimura 2 parameter method (Kimura, 1980) with a transition–transversion ratio of 1·1, as suggested by the polymorphism data (Fig. 2).

(ii) *Double checking the* D. novamexicana *lines*

To be sure that the *D. novamexicana* pattern of different groupings found for *oskar* and *period* was not caused by a laboratory mixup, we conducted three additional experiments (see Section 2). The original DNA preparations used for the *period* locus sequencing were used to obtain partial *oskar* sequence

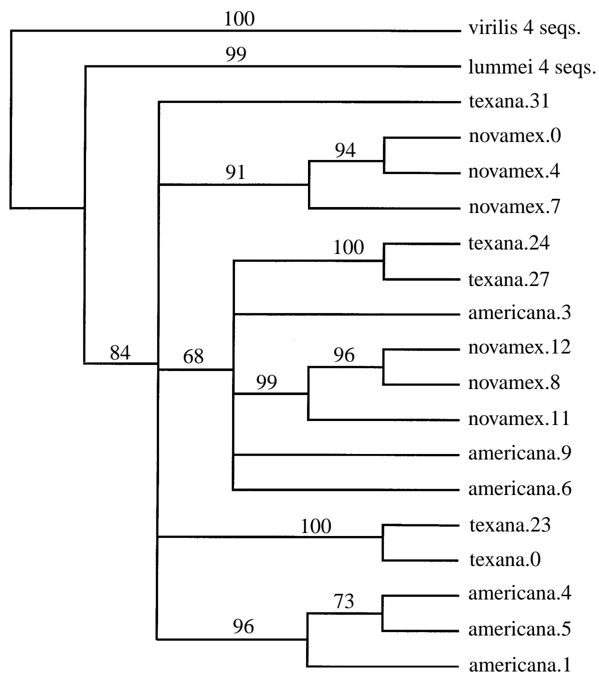

Fig. 4. A consensus distance tree of *oskar* based on bootstrapping; branches that appeared in less than 60% of trees were collapsed. Neighbour-joining bootstrap trees were produced by using NEIGHBOR in conjunction with the programs SEQBOOT, DNADIST and CONSENSE (PHYLIP 3.5; Felsenstein, 1989).

and the DNA preparations used for the *oskar* sequencing were used to obtain partial *period* sequence. Also new lines of flies were ordered from the NDSRC and partial sequences of both loci were obtained from common single fly DNA preparations. The *period* DNAs produced the correct *oskar* sequences and the *oskar* DNAs produced the correct *period* sequences, and the new lines checked out as well. Thus our conclusions on the structure of *D. novamexicana* subgroups at *period* and at *oskar* accurately reflect our *D. novamexicana* sample.

(iii) *Genealogical inference and recombination*

A gene tree estimate is shown in Fig. 3, and a consensus of bootstrap sampling is shown in Fig. 4. Both trees reveal that the relationships of *D. virilis*, *D. lummei* and the americana complex (*D. a. americana*, *D. a. texana* and *D. novamexicana*) are consistent with other phylogenetic analyses (Throckmorton, 1982; Spicer, 1992; Hilton & Hey, 1996). The tree figures also show two distinct *D. novamexicana* groups. These analyses did not use an outgroup sequence, however, the large divergence between *D. virilis* and *D. lummei*, as well as other information (Throckmorton, 1982; Spicer, 1991*b*, 1992), strongly suggests the root is along this branch.

If recombination has occurred in the history of a sample of sequences, then the true genealogy is an intercalated network and not a bifurcating tree. From the pattern of variation, it is possible to assess how much recombination has occurred (Table 4). Three groups (*D. a. texana*, *D. lummei* and *D. novamexicana*) revealed no evidence of recombination. In these cases all pairs of informative polymorphisms were consistent with one another so that both the estimate of the minimal number of recombination events (Hudson & Kaplan, 1985) and γ are zero. It seems likely that small sample sizes, in terms of numbers of sequences (e.g. *D. lummei* had the smallest sample size, four, that can be used for these analyses) and in numbers of polymorphisms have contributed to these zero estimates. In *D. a. americana* (and thus also in *D. americana*), where there were more sequences and more polymorphisms, considerable recombination was detected. The values of γ suggest that there have been about six recombination events for each mutation event in these taxa at this locus, and that there have been hundreds of recombination events since the time of the common ancestor of the samples from these taxa (Table 4). These estimates are obtained by dividing the estimate of $4Nc$, γ, by an estimate of $4Nu$, $\hat{\theta}$, to generate an estimate of the number of recombination events per neutral mutation events (i.e. c/u). Then this quantity is multiplied by the observed number of mutations in the history of the sequences since the time of the common ancestor (i.e. the observed number of polymorphic sites, S; Table 3). The difference in the minimum number of recombination events (Table 4; Hudson & Kaplan, 1985) and the estimated number of events reflects the fact that most recombination events in the history of a sample of sequences leave no trace of their occurrence (Hudson & Kaplan, 1985).

Table 4. *Recombination estimates*

	Minimum no. of events	γ	$\gamma/\hat{\theta}$	Estimated no. of events
D. virilis	2	na	—	—
D. lummei	0	na	—	—
D. americana	6	0·082	5·88	253
D. a. americana	2	0·074	6·31	177
D. a. texana	0	na	—	—
D. novamexicana	0	na	—	—

The minimum number of recombination events were determined using the method of Hudson & Kaplan (1985). γ is an estimate of $4Nc$ (Hey & Wakeley, 1997). $\gamma/\hat{\theta}$ is an estimate of the number of recombination events per mutation event (i.e. $4Nc/4Nu = c/u$). The estimated number of recombination events is calculated by multiplying the estimated number of recombination events per mutation event ($\gamma/\hat{\theta}$) by the estimated number of mutation events, S (Table 3).

(iv) *Polymorphism levels and tests of natural selection*

The values for measures of variation, π and $\hat{\theta}$, are given in Table 3. Both quantities are estimates of the

Table 5. *HKA tests*

Species pair	χ^2
virilis–lummei	1·23
virilis–americana	0·77
lummei–americana	0·34
a. americana–a. texana	0·12
Nova-A–Nova-B	5·96

Tests for natural selection between the *period* and *oskar* loci in several species pairs (Hudson *et al*., 1987). All tests have 2 degrees of freedom, in which case a significant χ^2 value would be above 5·99 at the $P < 0{\cdot}05$ level.

population neutral mutation rate $4Nu$, under assumptions of a Wright–Fisher population (Ewens, 1979) and an infinite sites neutral mutation model (Kimura, 1969). Comparisons of these two values can be used to calculate D (Tajima, 1989), which is proportional to the difference between the two values. A value of Tajima's D that is significantly different from zero can be considered as evidence of natural selection or a recently changing population size (Tajima, 1989). Tajima's D is slightly negative in *D. virilis*, *D. americana* and *D. a. americana*, and slightly positive in *D. lummei* (Table 3). In *D. a. texana* and *D. novamexicana* it is moderately positive, but none of the values is significantly different from zero.

The HKA test (Hudson *et al*., 1987) of selection was used on data from *oskar* and *period*. Table 5 shows the results of HKA tests on several of the species pairs. There is no evidence of selection between any of the pairs, including *D. virilis*, *D. lummei* and *D. americana*. Therefore, within these species, the different levels of polymorphism between *period* and *oskar* are consistent with variation in neutral mutation rates, and the differences in levels of polymorphism among the species may credibly reflect different population sizes (see below). When *D. novamexicana* was grouped for both loci as the two divergent groups found at the *period* locus (Nova-A and Nova-B), the HKA test statistic was nearly statistically significant, because there was little within-group variation at *period* but a large amount of between-group divergence, while at *oskar*, these '*period*' groupings created quite a bit of within-group variation and little between-group divergence.

Differences among taxa in the estimates of $4Nu$ (π and $\hat{\theta}$) can reflect differences in population size if the population size assumptions and mutation model assumptions of these estimators roughly hold. We see that the rank order of levels of variation has *D. americana* with the most, followed by *D. virilis* and *D. lummei*. The subgroups of *D. novamexicana* had the least variation (Nova-2 was invariant). This ordering of levels of variation is identical to that seen at the *period* locus (Hilton & Hey, 1996), as expected if polymorphism levels reflect relative population sizes. This parity of polymorphism levels among loci is also reflected in the low χ^2 values of the HKA tests (Table 5).

4. Discussion

When species are only recently diverged it is possible that they may still segregate genetic variation that arose prior to the onset of divergence. The amount and pattern of ancient variation, in comparison with more recent variation, can be analysed in light of historical models of speciation. In a sense, comparative DNA sequence studies that draw on variation within and between closely related populations or taxa permit population genetic analyses on relatively ancient events (Avise, 1989; Hey, 1994).

Our study of the divergence of the taxa in the virilis phylad of the *D. virilis* species group began with the X-linked *period* locus (Hilton & Hey, 1996). Like all such studies that focus on one region of the genome with limited recombination, questions persist from the *period* locus study as to whether the patterns of variation that were observed are representative of the genome, and thus reflect evolutionary forces that affect all portions of the genome. In this paper we extended the study to include the autosomal locus *oskar*.

Most of the patterns found within and among taxa at the *period* locus are repeated at the *oskar* locus. Regarding the two most distinct taxa in this group – *D. virilis* and *D. lummei* – both have considerable variation with more in the former. The apparent phylogenetic relationships in relation to the other taxa also are consistent with the *period* locus data and with the relationships proposed by Throckmorton (1982) and Spicer (1992).

(i) Drosophila americana

The pattern of variation found in *D. a. americana* and *D. a. texana* at the *oskar* locus shows only limited differentiation between these groups. The Mantel test, which compares pairwise differences within and between the two subspecies, failed to reject the null hypothesis that these samples came from a single population. Similarly, the estimate of Nm, also based on pairwise differences within and between the two groups, is greater than 1, the value at which mixing between the two is thought to be common. Finally, the two samples share nine polymorphisms and there are no fixed differences between them. This polymorphism pattern is also reflected in the mixing of lineages in the gene tree estimates (Figs. 3, 4). All these indications of gene flow between *D. a. americana* and *D. a. texana* at *oskar* were also seen at *period* (Hilton & Hey, 1996). Indeed at *period* estimated migration rates were even larger.

The subspecies designations of *D. a. americana* and *D. a. texana* are based on karyotype differences (Throckmorton, 1982), and they are supported to some extent by a small amount of prezygotic isolation

(Patterson *et al.*, 1947; Coyne & Orr, 1989). However, on the basis of *oskar* and *period* data, the subspecies designations do not appear to be useful. We found that strains labelled on the basis of karyotype differences (i.e. the chromosomal fusion difference that also defines the hybrid zone) have engaged in large amounts of gene flow. It follows that the geographic pattern of the chromosomal fusion variation does not reflect gene flow or differentiation for other parts of the genome. Our findings are consistent with those of Blight (1955) who studied the karyotypic frequencies in several populations that contained hybridizing populations of *D. a. americana* and *D. a. texana* near St Louis, Mo. He found hybrids and pure types existed in Hardy–Weinberg equilibrium, and concluded that the subspecies distinction was not useful for his populations. An explanation for the geographic pattern of chromosomal variation must then fall to natural selection, perhaps clinal selection causing the fused chromosome to be more advantageous in the north than in the south. There is no evidence of meiotic difficulty in hybrids, so recombination is probably normal in hybrids (Patterson & Stone, 1952). This means that the chromosomal region that is the site of selection must be tightly linked to the centromere (i.e. the site of the chromosomal fusion). If not, recombination would lead to linkage equilibrium between the chromosomal fusion and the site of selection and the sharp geographic pattern of karyotype variation would not be observed. Although *period* is on the X chromosome (one of the chromosomes fused in *D. a. americana*) it is located near the distal end, far from the fusion point (Kress, 1993).

Another noteworthy aspect that has emerged for *D. americana* is that it is one of the most polymorphic of *Drosophila* species – more variable at this locus than *D. pseudoobscura* or close relatives (Wang & Hey, 1996) and more variable than the species of the *D. melanogaster* complex (Hey & Kliman, 1993). The *oskar* numbers are even higher than those for *period* and are among the highest reported for any locus in an intraspecific DNA sequence study in *Drosophila* (Moriyama & Powell, 1996). The high level of variation probably indicates a very large historical population size. Although *D. americana* may currently not be abundant in nature, it may well have been very abundant in the past (Speith, 1979). The flies breed on the exposed bark of willow trees and are thought to be partially commensal with the American beaver, *Castor* (Speith, 1979). Beavers, beaver meadows and subsequently many willow trees were very common in North America before the 1700s (Speith, 1979).

(ii) Drosophila novamexicana

Drosophila novamexicana has revealed perhaps the most unexpected genealogies of the phylad. There had been no previous reports of population subdivision within the species (Patterson & Stone, 1952; Throckmorton, 1982). However, at both loci, the species contains two groups that are divergent at the DNA level but which have not diverged morphologically or chromosomally. Each locus has two groups that have little within-group polymorphism but contain quite a bit of divergence between them. Also for each locus, when the two groups are pooled there is very little divergence from *D. americana* – many of the differences between the two *D. novamexicana* groups occur as polymorphisms within *D. americana*. Thus at each locus the age of the two groups appears coincident with the time of separation of *D. novamexicana* from *D. americana* (Hilton & Hey, 1996). Surprisingly, the two subgroups are not identical at the two loci. At *period* the groups were named: Nova-A, consisting of nova.0, nova.7 and nova.8; and Nova-B, consisting of nova.4, nova.11 and nova.12. At *oskar*, group Nova-1 consists of nova.0, nova.7 and nova.7, while group Nova-2 consists of nova.8, nova.11 and nova.12. In essence, there is a switch between the two loci of lines nova.4 and nova.8. The possibility of nova.4 and nova.8 being mixed up in our laboratory was definitively ruled out by extensive resequencing.

There appear to be two explanations for this pattern. The first is that the pattern of divergence is due to a population-level process of non-interbreeding and divergence between groups that each have relatively small effective population sizes. The evidence for this model is that nearly identical levels of divergence and of polymorphism have been observed at two independent loci. Both loci suggest that the onset of divergence occurred at about the same time as did separation from the *D. americana*-like ancestor. The evidence against this model is that the two groups do not involve the same subsets of lines. Also when the geographic locations of the sample origins are considered, neither the *period* nor the *oskar* groupings easily fit a simple geographic partitioning of the variation. This model is essentially one of two cryptic species within *D. novamexicana*. To reconcile the model with the observations requires invoking some process whereby at least two of the sampled lines (nova.4 and nova.8) came to have recent ancestors from both cryptic species. Both lines do not fit the cryptic species hypothesis, and so two events need to be invoked. One explanation is that they simply represent gene flow that occurs between the cryptic species. However, this would imply that gene flow is fairly regular, which is in turn very much at odds within the high levels of divergence observed at both loci. An alternative explanation is that some reciprocal mixture occurred between lines after they were collected from the wild and before they came into our laboratory. If some samples from different cryptic species were mixed and then separated, it is possible that subsequent segregation and loss of variation with inbreeding could lead to the patterns observed.

The second explanation is that the pattern of variation at each locus is due to natural selection and that neither locus is representative of *D. novamexicana* population structure. Balancing selection acting on functional differences between two alleles could produce a pattern of two distinct groups of sequence as occurs at HLA loci (Klein *et al.*, 1990). Under this model for one locus, the age of the polymorphism is approximately the age of the common ancestor of both groups of alleles and the two groups continue to diverge as long as the polymorphism persists. The evidence against this model is that the polymorphism would have to be fairly old (to account for the considerable divergence) and yet no recombination is seen within the pooled *D. novamexicana* at either locus (Table 4; Hilton & Hey, 1996). The longer the polymorphism persists, the more recombination will limit the region of tight linkage and the shorter the portion of the genome that supports a deep divergence (Hudson & Kaplan, 1988). For example, balancing selection has been invoked to explain high polymorphism within the *Alcohol dehydrogenase* locus of *D. melanogaster* (Hudson *et al.*, 1987). However, that locus experiences considerable recombination and the region that seems to be affected by balancing selection is just a couple of hundred base pairs, much smaller than the *period* and *oskar* data sets analysed here. Given that both loci revealed considerable recombination in the other virilis phylad taxa, this model does not seem very tenable. One thing that might restrict recombination within *D. novamexicana*, despite the recombination within *D. americana*, would be if a locus resided in a within-species inversion polymorphism. There are some inversions found between species on both the X chromosome and chromosome 2; however, none is known to be polymorphic within *D. novamexicana* (Patterson & Stone, 1952). In addition, no inversions were seen on the X chromosome between four of the *D. novamexicana* lines sequenced and any *D. americana* line (G. Allison, personal communication). One additional piece of information that helps to rule out double balancing selection would be that there was no evidence of selection at these loci in any of the other virilis phylad taxa. Finally, if the balancing selection hypothesis is true, then it must be argued that both loci support polymorphisms of very similar ages and that both polymorphisms divide the population into two allelic groups of apparently very similar frequencies (though samples are small).

On balance we consider the first hypothesis – two cryptic species within *D. novamexicana* and mixing of samples – to be the simplest. Cryptic species identified on the basis of genetic evidence are a common phenomenon in *Drosophila*. The original distinction between *D. a. americana* and *D. a. texana* is an example (Patterson, 1941), as is the splitting of *D. pseudoobscura* and *D. persimilis* (Dobzhansky & Epling, 1944).

This research was supported by a grant from the National Science Foundation (DEB-9306625). Thanks to Bryant McAllister, who confirmed the 2nd chromosomal location of *oskar*.

References

Alexander, M. L. (1976). The genetics of *Drosophila virilis*. In *The Genetics and Biology of* Drosophila, vol. 1C (ed. M. Ashburner & E. Novitsky), pp. 1365–1419. New York: Academic Press.

Avise, J. C. (1989). Gene trees and organismal histories: a phylogenetic approach to population biology. *Evolution* **43**, 1192–1208.

Ashburner, M. (1989). Drosophila: *A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Blight, W. C. (1955). A cytological study of linear populations of *Drosophila americana* near St Louis, Missouri. PhD dissertation, Washington University, St Louis, Missouri.

Carson, H. L. & Blight, W. C. (1952). Sex chromosome polymorphism in a population of *Drosophila americana*. *Genetics* **37**, 572.

Coyne, J. A. & Orr, H. A. (1989). Patterns of speciation in *Drosophila*. *Evolution* **43**, 362–381.

Dobzhansky, T. & Epling, T. (1944). Taxonomy, geographic distribution and ecology of *Drosophila pseudoobscura* and its relatives. In *Contributions to the Genetics*, *Taxonomy*, *and Ecology of* Drosophila pseudoobscura *and its Relatives*, vol. 554 (ed. T. Dobzhansky & T. Epling), pp. 1–46. Washington, DC: Carnegie Institute of Washington.

Ewens, W. J. (1979). *Mathematical Population Genetics*. New York: Springer.

Falconer, D. S. (1989). *Introduction to Quantitative Genetics*. Harlow, UK: Longman.

Felsenstein, J. (1993). PHYLIP (Phylogeny Inference Package), version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.

Hey, J. (1994). Bridging phylogenetics and population genetics with gene tree models. In *Molecular Approaches to Ecology and Evolution*. (ed. B. Schierwater, B. Streit, G. Wagner & R. DeSalle). Basel: Birkhäuser.

Hey, J. & Kliman, R. M. (1993). Population genetics and phylogenetics of DNA sequence variation at multiple loci within and among species of the *D. melanogaster* complex. *Molecular Biology and Evolution* **10**, 804–822.

Hey, J. & Wakeley, J. (1997). A coalescent estimator of the population recombination rate. *Genetics* **145**, 833–846.

Hilton, H. & Hey, J. (1996). DNA sequence variation at the *period* locus reveals the history of species and speciation events in the *Drosophila virilis* group. *Genetics* **144**, 1015–1025.

Hudson, R. R. & Kaplan, N. L. (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. & Kaplan, N. L. (1988). The coalescent process in models with selection and recombination. *Genetics* **120**, 819–829.

Hudson, R. R., Kreitman, M. & Aguadé, M. (1987). A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159.

Hudson, R. R., Slatkin, M. & Maddison, W. P. (1992). Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**, 583–589.

Kimura, M. (1969). The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* **61**, 893–903.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through com-

parative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**, 111–120.

Klein, J., Gutknecht, J. and Fischer, N. (1990). The major histocompatibility complex and human evolution. *Trends in Genetics* **6**, 7–11.

Kliman, R. M. & Hey, J. (1993). DNA sequence variation at the *period* locus within and among species of the *Drosophila melanogaster* complex. *Genetics* **133**, 375–387.

Kress, H. (1993). The salivary gland chromosomes of *Drosophila virilis*: a cytological map, pattern of transcription and aspects of chromosome evolution. *Chromosoma* **102**, 734–742.

Mantel, N. (1967). The detection of disease clustering in a generalized regression approach. *Cancer Research* **27**, 209–220.

Moriyama, E. N. & Powell, J. R. (1996). Intraspecific nuclear DNA variation in *Drosophila*. *Molecular Biology and Evolution* **13**, 261–277.

Neufeld, T. P., Cathew, R. W. & Rubin, G. M. (1991). Evolution of gene position: chromosomal arrangement and sequence comparison of the *Drosophila melanogaster* and *Drosophila virilis sina* and *Rh4* genes. *Proceedings of the National Academy of Sciences of the USA* **88**, 10203–10207.

Orr, H. A. & Coyne, J. A. (1992). The genetics of adaptations: a reassessment. *American Naturalist* **140**, 725–742.

Patterson, J. T. (1941). The *virilis* group of *Drosophila* in Texas. *American Naturalist* **75**, 523–539.

Patterson, J. T. (1942). Distribution of the *virilis* group in the United States. In *Studies in the Genetics of* Drosophila, University of Texas publications no. 4228, pp. 53–161.

Patterson, J. T. & Stone, W. S. (1952). Evolution in the genus *Drosophila*. New York: Macmillan.

Patterson, J. T., McDonald, L. W. & Stone, W. S. (1947). *Sexual Isolation Between Members of the* Virilis *Group of Species*, University of Texas Publications no. 4720, pp. 7–31.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406–425.

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

Spicer, G. S. (1991 *b*). Molecular evolution and phylogeny of the *Drosophila virilis* species group as inferred by two-dimensional electrophoresis. *Journal of Molecular Evolution* **33**, 379–394.

Spicer, G. S. (1992). Reevaluation of the phylogeny of the *Drosophila virilis* species group (Diptera: Drosophilidae). *Annals of the Entomological Society of America* **85**, 11–25.

Spieth, H. L. (1979). The *virilis* group of *Drosophila* and the beaver *Castor*. *American Naturalist* **114**, 312–316.

Sturtevant, A. H. & Novitski, E. (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. In *Mechanisms of Molecular Evolution* (ed. N. Takahata & A. G. Clarke), pp. 37–59. Sunderland, Mass.: Sinauer Associates.

Throckmorton, L. (1982). The *virilis* species group. In *The Genetics and Biology of* Drosophila, vol. 3b (ed. M. Ashburner, H. L. Carson & J. N. Thompson, pp. 227–296. New York: Academic Press.

Tonzetich, J., Hayashi, S. & Grigliatti, T. A. (1990). Conservation of sites of tRNA loci among the linkage groups of several *Drosophila* species. *Journal of Molecular Evolution* **30**, 182–188.

Wang, R. L. & Hey, J. (1996). The speciation history of *Drosophila pseudoobscura* and close relatives: inferences from DNA sequence variation at the period locus. *Genetics* **144**, 1113–1126.

Watterson, G A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* **7**, 256–275.

Webster, P. J., Suen, J. & Macdonald, P. M. (1994). *Drosophila virilis oskar* transgenes direct body patterning but not pole cell formation or maintenance of mRNA localization in *D. melanogaster*. *Development* **120**, 2027–2037.

Whiting, J. H. Jr, Piley, M. D., Farmer, J. L. & Jeffery, D. E. (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. *American Naturalist* **74**, 232–248.