J Mol Evol (1992) 34:130–140

Journal of Molecular Evolution

Complementary DNA-DNA Hybridization in *Drosophila*

Adalgisa Caccone,* Jennifer M. Gleason, and Jeffrey R. Powell

Department of Biology, PO Box 6666, Yale University, New Haven, CT 06511, USA

Summary. We have performed DNA–DNA hybridization experiments among several species of *Drosophila* using the evolutionarily conserved portion of the genome representing sequences coding for amino acids of proteins. This was done by using as tracer, radioactively labeled complementary DNA that was reverse transcribed from adult mRNA. We show that this procedure extends phylogenetically the distance over which the technique can be applied to fast-evolving groups such as *Drosophila.* The major phylogenetic conclusions are (1) the subgenus *Sophophora* is a monophyletic lineage; (2) within *Sophophora* the *melanogaster* subgroup is closer to the *obscura* group than either group is to the *willistoni* group; (3) the subgenus *Drosophila* is complex with most major lineages originating deep in the phylogeny; the subgenus may not be monophyletic; (4) as with most groups classically placed in *Drosophila,* the Hawaiian *Drosophila* originate early, supporting the notion that this lineage is older than the extant islands; and (5) the *virilis/repleta* lineage is monophyletic within *Drosophila.*

Key words: *Drosophila* — *Sophophora* — cDNA–DNA hybridization — Phylogenetics

Introduction

DNA–DNA hybridization of single-copy DNA (scDNA) sequences has been used to assess overall genetic relationships in several vertebrates and invertebrates (e.g., Hunt et al. 1981; Sibley and Ahlquist 1981, 1984, 1987; Caccone and Powell 1987, 1989; Sheldon 1987a; Caccone et al. 1988a; Catzeflis et al. 1989; Goddard et al. 1990; Kirsch et al. 1990). It has also been useful in revealing rates and patterns of genome evolution, which are strikingly different in different groups of organisms (e.g., Britten 1986; Sheldon 1987b; Catzeflis et al. 1987; Springer and Kirsch 1989; Caccone and Powell 1990; Werman et al. 1990). Insect genomes, in particular *Drosophila* and cave-crickets, are composed of sections of scDNA with extreme variation in rates of evolutionary change, variation more extreme than the known differences in rates between introns and exons (Powell and Caccone 1989; Caccone and Powell 1990). A similar, though less extreme situation holds for sea urchins (Werman et al. 1990). This finding contrasts markedly with what has been seen in other taxa, especially mammals and birds, where little evidence of such extreme heterogeneity has been found. Absolute rates of change also vary considerably among taxa (Britten 1986; Catzeflis et al. 1987). Insect DNA seems to evolve 5–10 times faster than many vertebrates. This comparison refers only to the more conservative part of the insect single-copy genome. If we consider the total single-copy genome, the rate difference is even more extreme (Caccone and Powell 1990).

Because of these dramatic differences in pattern and rate of evolution among vertebrates and invertebrates, the taxonomic level at which DNA–DNA hybridization provides reliable estimates of sequence divergence is very different in different groups of organisms. In vertebrates, DNA hybridization studies have been used to infer genetic relationships between species belonging to different families and orders. In insects, comparisons between different genera in the same subfamily of cave-crickets are at the resolution maximum of the technique (Caccone

* On leave from the Dipartimento di Biologia, II Università di Roma 'Tor Vergata,' Rome, Italy
Offprint requests to: J.R. Powell

Table 1. *Drosophila* species studied including strain used, source, and abbreviation

Drosophila species[a]	Strain	Source	Abbreviation
Subgenus *Sophophora*			
melanogaster group			
D. melanogaster	Oregon-R	D. Poulson	MEL
D. simulans	Arizona	D. Hartl	SIM
D. yakuba	02161.0; Ivory Coast, Africa	BG[b]	YAK
D. orena		D. Hartl	ORE
D. takahashii	0311.0; Nepal, Asia	BG[b]	TAK
obscura group			
D. pseudoobscura	DM3; Davis Mountain, Texas	J. R. Powell	PSE
willistoni group			
D. willistoni	St. Kitts, 32	J. R. Powell	WIL
Subgenus *Drosophila*			
repleta group			
D. mercatorum	S.αBr13; S. Salvador	A. Templeton	MER
D. hydei		S. Artavanis	HYD
virilis group			
D. virilis	1051.0; Pasadena, California	BG[b]	VIR
immigrans group			
D. immigrans	1731.5; Central Honshu, Japan	BG[b]	IMM
melanica group			
D. melanica	1141.0; Myakke Head, Florida	BG[b]	MELA
funebris group			
D. funebris	1911.0; Sturgis, Kentucky	BG[b]	FUN

[a] The species divisions follow Patterson and Stone (1952)
[b] Bowling Green *Drosophila* Stock Center

and Powell 1987). In *Drosophila,* it is difficult to reliably compare even species belonging to the same subgenera!

In this paper we expand the resolution power of the DNA–DNA hybridization technique by measuring sequence divergence of coding sequences only. The rationale behind this approach relies on the safe assumption that, in general, coding sequences evolve more slowly than noncoding sequences. By limiting our study to the subset of sequences coding for proteins, we should be able to assess genetic relationships between more distantly related taxa than by using overall single-copy sequences. Obviously, we are studying a much smaller fraction of the genome than when using standard DNA hybridization techniques. However, this small fraction of the genome contains millions of nucleotides, which is still much more than could be easily studied by sequencing. Thus, the major advantage of using an average divergence over a large number of genes is preserved such that the problem of confusing gene trees and species trees is minimized (Pamilo and Nei 1988).

The basic protocol for these experiments has been to isolate poly(A)$^+$ RNAs from adult flies and radioactively label DNA copied from this RNA. The resulting cDNAs were used as tracers in DNA–DNA hybridization experiments against total DNA preparations from its own species and others. Similar experiments have been performed with sea urchins (Roberts et al. 1985). Elsewhere we document that cDNA prepared from adult mRNA and embryo mRNA yield the same relative distances among species (Caccone et al., unpublished) so that the phylogenetic conclusions in the present report are not dependent upon developmental stage of expression of the mRNAs used.

We have chosen to use this approach to attempt to elucidate the genetic relationships of 11 species of *Drosophila* belonging to the two subgenera *Sophophora* and *Drosophila* (Table 1). These species and species groups represent the majority of species used in experimental work. Thus, in relating phylogenies to other work such as morphology, development, or behavior, it is the phylogenetic relationships of members of these two subgenera that are most important.

Materials and Methods

The strains of flies used in this study along with the abbreviations by which they will be subsequently referred and their traditional taxonomic classification are listed in Table 1. To our knowledge, all are isofemale lines, begun by single inseminated females from nature.

Total DNA was prepared for the 11 species studied using a method based on phenol/chloroform extractions and proteinase K digestion (Werman et al. 1990). The DNAs were then sheared by sonication to an average size of 500–2000 bp. They constitute the drivers in the DNA–DNA hybridization experiments.

Total RNA was obtained from adult flies of mixed sex for 6 of the 11 species (MEL, YAK, PSE, WIL, MELA, VIR) by the method of Chomczynski and Sacchi (1987). Poly(A)$^+$ RNA was prepared by two passes over oligo-dT columns. Integrity of the mRNA population was checked for some preparations by Northern blots hybridized to a cDNA clone of alcohol dehydrogenase, which yielded a single band of the appropriate size for intact message. The mRNA fraction was reverse transcribed (cDNA Synthesis System from BRL) with the addition of ^{3}H-dTTP and ^{3}H-dCTP and using poly-T as a primer. RNA was removed by denaturation with NaOH and boiling followed by neutralization (Roberts et al. 1985). The resulting cDNA was recovered and concentrated by use of Gene Clean (American Bioanalytical). Complementary DNA preparations were sized by alkaline gel electrophoresis (Hall et al. 1980; Hunt et al. 1981; Caccone and Powell, 1991). Their average sizes ranged from 300 to 600 bp. These preparations were the tracers in the DNA–DNA hybridization experiments.

We used the so-called TEACL method of DNA–DNA hybridization (Britten et al. 1978; Caccone and Powell 1989, 1991). The general protocol we used is very similar to our standard DNA–DNA hybridization method for single-copy DNA, which has been detailed elsewhere and will not be repeated here (Powell et al. 1986; Caccone and Powell 1987, 1990; Caccone et al. 1987; Powell and Caccone 1990). However, two significant differences should be noted. First, hybridization mixtures were incubated to an equivalent Cot >24,000 mol s/l, whereas in the standard DNA hybridization protocol, reassociations are stopped at a Cot >6000 mol s/l. Second, in standard DNA hybridization experiments with single-copy DNA, we mix tracer and driver in a controlled ratio (1:1000) to prevent the formation of tracer : tracer duplexes, which would bias results. However, this is not necessary when using cDNA tracers as they are single-stranded and do not have complements in the tracer preparation. This relaxation of the "ratio rule" allowed us to have more flexibility in the experimental design. However, it produces a major technical artifact: the percent reassociations both for homo- and heteroduplex comparisons are low due probably to large heterogeneity of copy number of the cDNA sequences because of differential mRNA concentrations in the initial preparations.

The statistical procedures used to summarize the data are detailed in Caccone et al. (1987) and Powell and Caccone (1990). J. Felsenstein's PHYLIP program FITCH version 3.2 (Felsenstein 1990) and J. Ferguson's neighbor-joining computer packages were used to construct trees based on Fitch and Margoliash (1967) and Saitou and Nei (1987) algorithms. To search for possible better trees we used the G and J options in the PHYLIP package. The G option allows for global pairwise branch swapping of taxa. The J option changes the species input order. To account for the missing entries we used the S option, which allowed us to specify the degree of replication of each distance. For missing entries we used a value of 999 with a degree of replication of 0.

Stability of tree branches was tested by jackknife and bootstrapping procedures. For the jackknife each species was removed in turn from the complete data set to determine whether or not other topologies could be found (Lanyon 1985). Each 10 taxa data set was then analyzed using FITCH. A bootstrap analysis was employed according to Marshall (1991). The data sets for the bootstrap were generated by defining a normal distribution using the ΔT_m means and standard errors. A pseudoreplicate data set was constructed by sampling just once for each cell in the data matrix. Sampling was done only once because the standard error already takes into account the number of replicates performed. This process was repeated until 100 pseudoreplicate data sets were produced. Each pseudoreplicate was then analyzed with the FITCH program and the resulting tree recorded (Marshall 1991).

Results

The results from all the experiments are summarized in Table 2. Complementary DNA tracers were obtained for 6 of the 11 species (MEL, YAK, WIL, PSE, VIR, MELA). Each comparison was replicated 3–6 times for a total of 271 median melting temperature determinations, t_m's. The standard errors of measurement of t_m ranged from 0.01 to 0.47 with an average of 0.13. For two of the six tracers we used three (MEL1, 2, and 3) and two (YAK1, 2) independently derived cDNA preparations. Although the Δt_m's obtained using different tracer for the same species varied considerably (up to 1.9°C), the ΔT_m's were very similar (within the two SEs). We use t_m to indicate the measured median melting temperature and T_m when it has been corrected for duplex length (Hall et al. 1980); all further discussion will refer to ΔT_m's. Hybrid duplexes varied in length from 47 to 302 bp.

Percent reassociation (%R) and normalized percent reassociation (NPR) are reported in the last two

columns of Table 2. These NPR estimates are not comparable to NPR when using the standard total scDNA procedures. Here, we have combined populations of sequences representing mRNA in the frequency of their expression at the time of isolation from the organism, or at least in their proportions resulting from reverse transcriptase reactions. This is then mixed with scDNA of the whole genome. We did not know the precise amount of tracer cDNA used in each reassociation; rather we used a constant number of cpms (300,000–400,000), which we estimated represented much less than 0.1 μg of DNA. Each tracer was mixed with a constant amount of driver total scDNA, usually 10 μg. Thus, the relatively low reassociation of tracer, even in reactions to scDNA from the same genome from which it was prepared, is due to the heterogeneity in copy number of cDNA sequences. Highly expressed genes, presumably highly represented in cDNA products, would not have the large molar excess of driver sequences that rare cDNA sequences would have. This also adds to the inaccuracy of measurement of NPR, and we do not use this parameter in the interpretation of the results presented here. However, because the ratio of cpm tracer to μg driver was the same in the homoduplex reaction and all heteroduplex reactions with the same tracer preparation, some information is present in the normalized values as we show elsewhere (Caccone et al., unpublished).

In the lower left part of Table 3, ΔT_m values for the data listed in Table 2, together with their standard errors, are reported. Reciprocal tests (each species used in turn as tracer and driver) were carried out for 10 out of the 36 pairwise comparisons. Reciprocity holds quite well in this data set. The mean difference between reciprocal ΔT_m's is 0.23. Because the standard errors associated with ΔT_m are usually 0.1–0.2, most reciprocal ΔT_m's are within 2 SEs. The overall ΔT_m's in cases of reciprocals are calculated as weighted averages as described in Caccone et al. (1987). These ΔT_m's were calculated on 6–12 replicate values. The matrix is not complete in the sense that not all pairwise comparisons were studied. However, all species were compared with at least one of the species belonging to the different species group. The percent base-pair mismatch shown in the upper right of Table 3 is based on the conversion that 1°C ΔT_m corresponds to 1.7% bp mismatch (Caccone et al. 1988b). This estimate was then corrected for multiple substitutions by the formula of Jukes and Cantor (1969). Corrected %bp mismatches range from 3.9% for the MEL–YAK comparison to up to 35–37% for comparison between the Hawaiian species (SIL) and MEL, PSE, MELA, or VIR. Because the experimental conditions allow sequences to form stable duplexes only if they have at least 70% bp matching, the measurement of a ΔT_m of 17°C (implying 29% bp mismatch without Jukes and Cantor correction, 37% with correction) indictes that we are at the limit of resolution for these techniques.

However, we must add two caveats in interpreting the ΔT_m's to base-pair change. First, the conversion of 1 ΔT_m to 1.7% mismatch is controversial. This estimate was made by use of the TEACL technique for sequences that differed only in base substitutions and virtually no deletions (Caccone et al. 1988b). Other estimates of this conversion range to as low as 0.7% change per 1°C ΔT_m. Recent measurements using hydroxylapatite and sequences differing in insertion/deletions as well as base substitutions (counting an insertion/deletion the same as a substitution) have indicated that a conversion of 1°C ΔT_m corresponds to a 1.2% base change (Springer et al., personal communication). Secondly, one may question whether the Jukes and Cantor correction is appropriate for coding sequences that have such selective constraints especially for replacement sites. However, at the distances considered (average substitutions per site much less than one), the Jukes and Cantor correction is about as accurate as the more complex corrections, taking into consideration complications of the code and transition/transversion biases (e.g., Li et al. 1985). However accurate or not the conversion, all subsequent phylogenetic analysis was done directly on ΔT_m's so that such questions are moot with respect to the phylogenies presented.

Figure 1 is a dendrogram based on the ΔT_m matrix in Table 3 using the neighbor-joining method (Saitou and Nei 1987). The algorithm will generate a tree that appears rooted using the most distant species as an outgroup, in this case the Hawaiian *Drosophila* are represented by SIL. The classically defined subgenera *Sophophora* and *Drosophila* do not appear as monophyletic taxa. However, the branch tips line up fairly evenly, implying relatively constant rates of accumulation of differences along lineages. Figure 2 is the neighbor-joining tree with the root forced between the subgenera. Now the tips do not line up, with the SIL being particularly out of line, implying a faster rate of accumulation of change along this branch.

Figure 3 shows a dendrogram obtained by running the Fitch program of Felsentein's PHYLIP package, which uses the Fitch and Margoliash (1967) method for fitting trees to distance matrices. This algorithm does not assume a constant rate of change along lineages. The tree presented was obtained after a global search, which examined 261 trees: the average percent standard deviation (APSD) was 2.091 and the sum of squares was 0.154. This tree had the smallest APSD out of 25 different trees obtained by

Table 2. Detailed data for the cDNA–DNA hybridization

Tracer	Driver	*n*	t_m	SE	Tracer length (bp)	T_{corr}	T_m	%R	NPR
MEL1	MEL	4	55.25	0.05	156	3.21	58.46	14.4	100.0
	YAK	5	53.83	0.11	209	2.39	56.22	17.5	121.7
	TAK	5	49.63	0.07	132	3.79	53.42	10.8	75.3
	PSE	4	47.84	0.08	115	4.35	52.19	23.7	164.7
	MELA	4	43.22	0.21	103	4.85	48.07	10.4	72.0
MEL2	MEL	4	57.18	0.09	127	3.93	61.11	9.6	100.0
	MAU	3	56.30	0.12	137	3.64	59.94	11.9	124.9
	SEC	3	55.61	0.17	133	3.77	59.38	5.8	60.7
	ERE	3	54.52	0.16	124	4.04	58.56	11.9	124.0
	TES	3	54.24	0.09	127	3.95	60.19	8.8	91.5
	SIL	4	39.46	0.38	108	4.65	44.11	10.6	111.3
MEL3	MEL	5	55.85	0.02	305	1.64	57.49	10.6	100.0
	SIM	5	54.11	0.11	235	2.13	56.24	16.5	156.0
	ORE	4	51.15	0.09	165	3.03	54.18	15.5	146.2
	TAK	4	49.84	0.15	215	2.33	52.17	7.5	70.8
	PSE	5	48.22	0.04	165	3.03	51.25	12.3	116.0
	WIL	4	45.34	0.03	140	3.57	48.91	5.0	47.2
	FUN	4	39.22	0.08	135	3.70	42.92	8.8	83.0
	HYD	4	44.58	0.19	302	1.66	46.24	4.0	37.7
	IMM	6	43.07	0.03	280	1.79	44.86	12.8	120.8
YAK1	YAK	4	57.24	0.08	161	3.11	60.35	16.3	100.0
	TAK	3	49.85	0.19	112	4.45	54.30	20.2	123.8
	ERE	3	53.94	0.07	106	4.72	58.65	12.8	78.6
	TES	3	55.34	0.05	107	4.66	60.00	11.8	72.4
YAK2	YAK	5	56.64	0.06	69	7.28	63.92	34.4	100.0
	MEL	5	51.01	0.15	47	10.65	61.66	14.5	42.2
	SIM	4	53.22	0.06	65	7.71	60.93	9.1	26.5
	WIL	4	48.36	0.12	69	7.24	55.60	42.1	122.4
	MELA	4	45.78	0.10	63	7.98	53.76	26.3	76.5
	MERC	4	42.73	0.22	60	8.37	51.10	17.4	50.6
	VIR	3	44.80	0.26	73	6.85	51.65	9.3	27.0
WIL	WIL	5	56.36	0.06	254	1.97	58.33	11.3	100.0
	PSE	3	47.70	0.06	173	2.89	50.59	8.2	72.7
	MEL	4	46.45	0.13	146	3.43	49.88	19.8	174.9
	YAK	4	46.68	0.08	166	3.00	49.68	11.3	99.7
	MERC	3	43.66	0.18	182	2.75	46.41	6.3	55.9
	HYD	3	43.75	0.40	187	2.67	46.42	2.1	18.4
	IMM	3	43.73	0.27	177	2.82	46.55	11.0	96.6
	MELA	4	43.38	0.13	162	3.09	46.47	13.4	117.8
	FUN	4	41.18	0.30	178	2.81	43.99	4.3	37.7
	VIR	3	43.81	0.01	236	2.13	45.94	6.3	55.9
PSE	PSE	4	57.54	0.05	135	3.69	61.23	15.0	100.0
	WIL	3	47.95	0.28	99	5.05	53.00	13.7	92.0
	VIR	3	45.02	0.05	99	5.03	50.05	10.7	71.4
	MERC	3	44.36	0.22	96	5.19	49.55	10.5	69.9
	SIL	3	39.54	0.47	96	5.23	44.77	7.6	51.0
	MEL	3	49.10	0.19	91	5.52	54.62	8.0	53.1
	IMM	4	44.74	0.07	111	4.51	49.25	7.6	50.3
VIR	VIR	4	55.93	0.06	128	3.90	59.83	24.0	100.0
	HYD	3	46.80	0.09	100	5.01	51.81	23.1	96.3
	MERC	3	46.51	0.13	94	5.33	51.84	27.8	116.1
	MEL	3	41.57	0.10	91	5.47	47.04	27.0	112.9
	PSE	4	43.42	0.11	104	4.79	48.21	8.8	36.9
	WIL	3	42.40	0.19	99	5.05	47.45	7.1	29.7
	FUN	4	38.72	0.39	91	5.52	44.24	9.6	40.0
	MELA	4	43.51	0.14	86	5.79	49.30	22.0	92.0
	IMM	4	42.52	0.05	92	5.46	47.98	10.5	44.0
	SIL	3	36.52	0.16	85	5.90	42.42	6.1	25.6
HYD	HYD	4	56.74	0.15	138	3.61	60.35	19.8	100.0
	VIR	4	47.05	0.18	99	5.03	52.08	11.9	60.4
	MERC	4	49.97	0.07	86	5.82	55.79	29.2	147.8
	YAK	4	43.31	0.15	102	4.93	48.24	14.1	71.5

Table 2. Continued

Tracer	Driver	*n*	t_m	SE	Tracer length (bp)	T_{corr}	T_m	%R	NPR
	FUN	4	41.05	0.02	106	4.72	45.77	14.5	73.5
MELA	MELA	3	54.83	0.01	163	3.08	57.91	18.9	100.0
	VIR	4	42.66	0.02	100	5.02	47.68	5.7	30.4
	SIL	4	35.36	0.33	90	5.55	40.91	17.5	92.6
	MEL	4	42.71	0.03	103	4.86	47.57	20.9	110.8
	PSE	4	40.45	0.09	96	5.20	45.65	19.0	100.9
	FUN	4	39.00	0.12	93	5.38	44.38	11.0	58.2
	IMM	4	41.03	0.06	102	4.91	45.95	18.7	98.9

Each pairwise species comparison is shown with the tracer first (first column) followed by the driver (second column). For each tracer the homo duplex is shown first, the heteroduplex comparisons follow. Abbreviations of the species are as in Table 1; other abbreviations: *n* = number of replicate t_m determinations; t_m = mean uncorrected median melting temperature; SE = standard error of t_m (and T_m); T_{corr} = temperature correction; T_m = corrected median melting temperature; %R = percentage reassociation; NPR = normalized percentage reassociation

Table 3. ΔT_m matrix for cDNA data

	MEL	YAK	PSE	WIL	MELA	IMM	VIR	HYD	MERC	FUN	SIL
MEL	—	3.93	11.60	16.18	20.03	25.29	25.68	22.08	—	30.06	36.50
YAK	2.25* 0.10	—	—	16.11	19.63	—	24.44	24.07	25.75	—	—
PSE	6.32* 0.08	—	—	14.47	—	23.76	22.17	—	23.07	—	35.02
WIL	8.56* 0.06	8.53* 0.08	7.74* 0.06	—	23.48	23.30	24.73	23.60	23.62	29.48	—
MELA	10.34* 0.03	10.16 0.11	—	11.86 0.15	—	23.72	19.79	—	—	27.47	36.50
IMM	12.63 0.06	—	11.98 0.09	11.78 0.24	11.96 0.06	—	23.46	—	—	—	—
VIR	12.79 0.11	12.27 0.26	11.29* 0.70	12.39* 0.10	10.23* 0.02	11.85 0.08	—	15.13	14.99	32.70	37.64
HYD	11.25 0.11	12.11 0.21	—	11.91 0.40	—	—	8.06* 0.10	—	8.18	30.09	—
MERC	—	12.82 0.23	11.68 0.23	11.92 0.19	—	—	7.99 0.14	4.56 0.17	—	—	—
FUN	14.57 0.10	—	—	14.34 0.31	13.53 0.01	—	15.59 0.39	14.58 0.15	—	—	—
SIL	17.00 0.39	—	16.46 0.47	—	17.00 0.33	—	17.41 0.17	—	—	—	—

Lower left display's ΔT_m's with 1 SE below. Asterisks indicate reciprocal tests were performed. Upper right indicates estimated base-pair mismatch calculated as explained in text

running the FITCH program on the same data matrix but with different species input orders. However, almost equally good trees (APSD = 2.092–2.095) were also produced in which either the position of MELA and IMM or the position of FUN and SIL was switched around.

The trees presented, based on different algorithms and assumptions, produce similarities and differences. In all cases, the *melanogaster* subgroup (MEL and YAK), PSE, and WIL cluster into a monophyletic lineage. VIR along with the two *repleta* group species, HYD and MERC, always cluster together. All other species join the trees relatively deeply and in different configurations depending upon the method used. To summarize our confidence in phylogenetic relationships deduced, we box the area in Fig. 3 that we do not consider defined.

An interesting aspect of these trees is the relatively long branches associated with the Hawaiian species, SIL. This could be an indication of an accelerated evolutionary rate, which would be expected on the basis of population genetic theory in species with small average population sizes (DeSalle and Templeton 1988). However, such an explana-

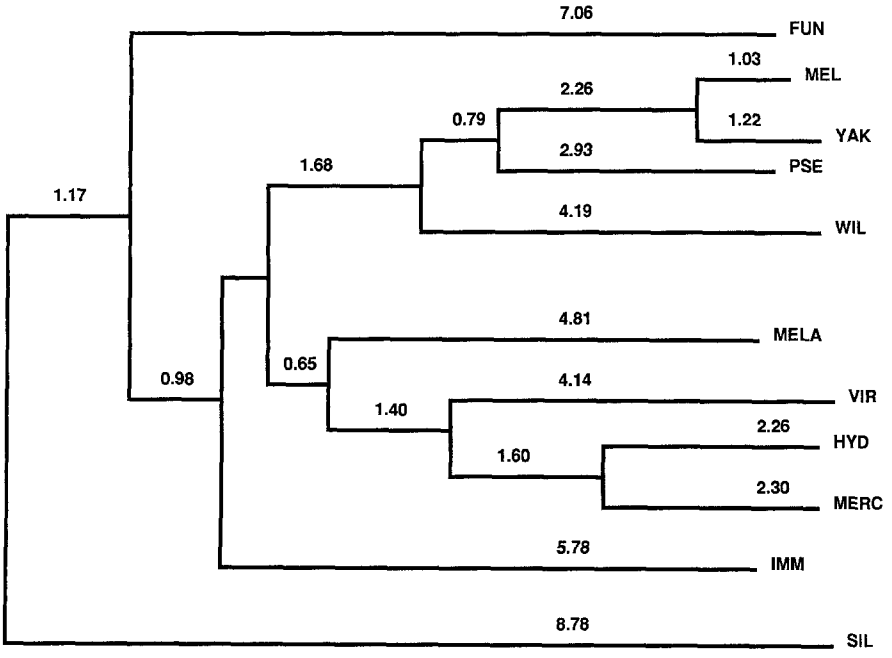

Fig. 1. Dendrogram constructed using the neighbor-joining method (Saitou and Nei 1987) on the adult cDNA ΔT_m's shown in Table 3. Numbers on branches are patristic distances. Species abbreviations are as in Table 1.

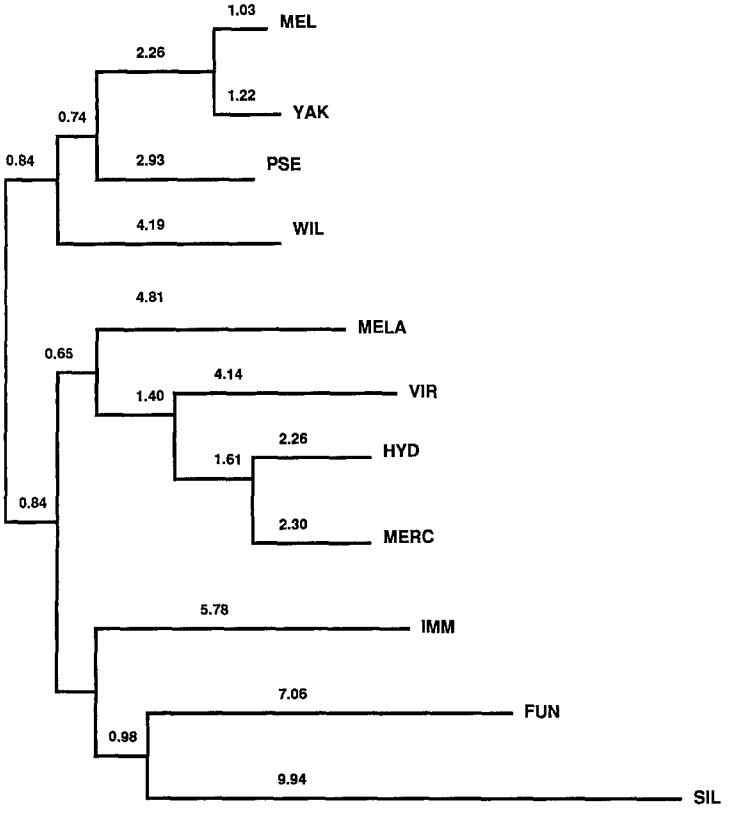

Fig. 2. Neighbor-joining tree based on the same data as the tree in Fig. 1 but constructed by forcing the root between the two traditionally recognized subgenera *Drosophila* and *Sophophora*

tion is difficult to adopt for *funebris* (FUN), which also has a relatively long branch; there is no evidence that this species (unlike Hawaiian species) has unusually small population sizes for a *Drosophila*. Whatever the exact position of the Hawaiian branch, our results are supportive of the notion that this lineage is quite old, certainly older than the extant islands (Beverley and Wilson 1985).

Figure 4 shows the results of the jackknife consensus tree (left side) and bootstrap consensus tree (right side) for the cDNA ΔT_m data matrix. Jackknifing has reduced to polytomies several of the branch-points of Fig. 3, leaving resolution limited to the species within the *Sophophora* subgenus and the *virilis–repleta* species group. The right side of Fig. 4 shows the bootstrap analysis. Nodal values

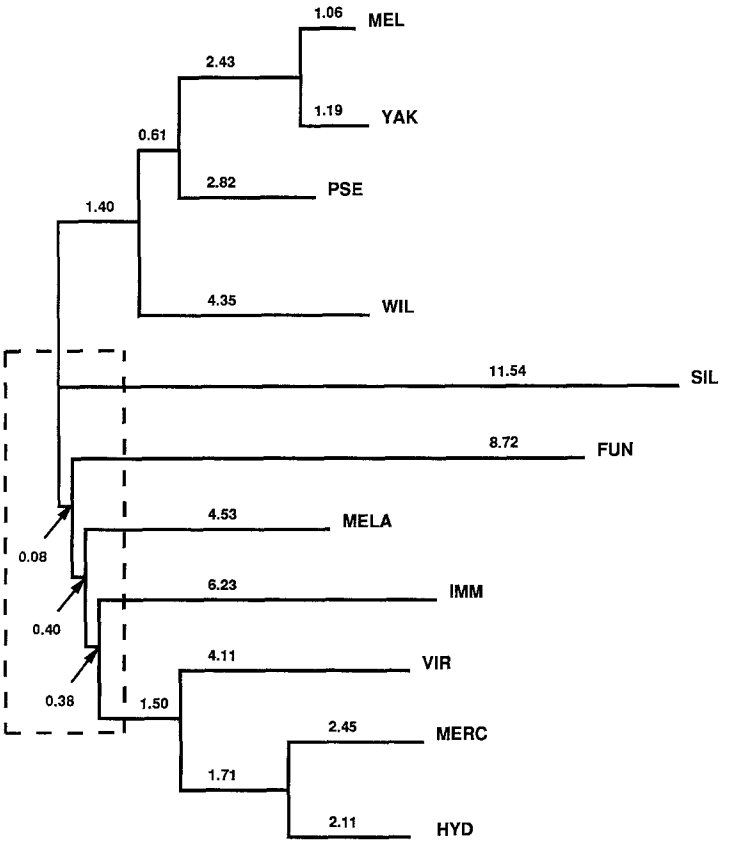

Fig. 3. Dendrogram obtained by running the FITCH program on Felsenstein's PHYLIP, which uses the Fitch and Margoliash (1967) method for fitting trees to distance matrices. This tree is unrooted. It was drawn with MEL at 0 and other species branches relative to this. Boxed area indicates area of ambiguity with little confidence in branch ordering.

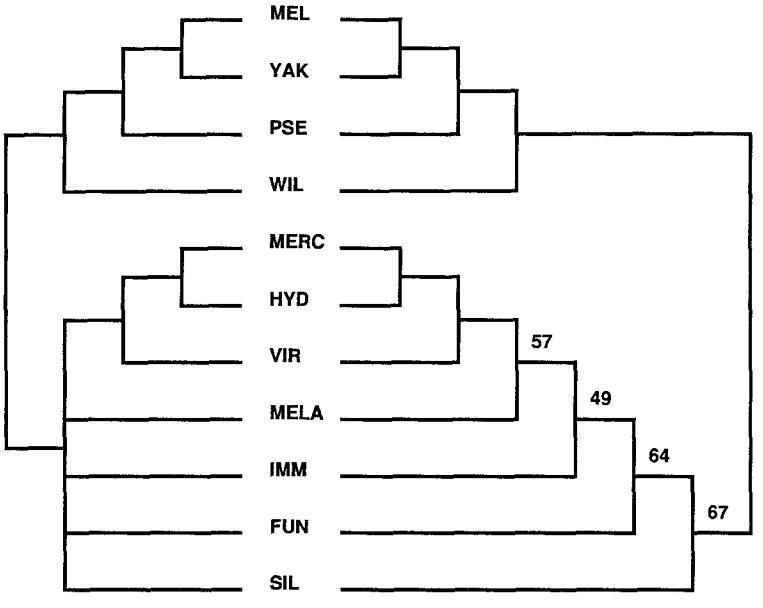

Fig. 4. Jackknife consensus tree (left) for the ΔT_m matrix in Table 3. Bootstrap consensus tree (right) based on 100 FITCH trees generated from pseudoreplicate matrices. Nodal values indicate the frequency out of 100 trials with which the indicated clade appeared in the pseudoreplicate trees. Nodes without numbers occurred with a frequency of 1.0.

indicate the proportion of times out of 100 pseudoreplicate analyses that each clade appeared. The bootstrap places low confidence on the same branches, which were collapsed by jackknifing. In all cases the nodes involved are the ones with small internodal branches and those that were already identified as suspect based on the dependence on algorithm or APSDs (boxed area Fig. 3).

Discussion

This study has two purposes. The first is to demonstrate the feasibility of performing DNA–DNA hybridization studies on the conserved subset of the genome represented by coding DNA. The usual manner of performing DNA–DNA hybridization, using total scDNA as tracer, has not proven reliable

at the higher distances in *Drosophila* primarily due to the fact that the percent of the genome that hybridizes becomes very low even at moderate distances, e.g., between the *melanogaster, willistoni,* and *obscura* groups all within the *Sophophora*; between *Sophophora* and *Drosophila* subgenera the problem is even more serious. Using total scDNA as tracer, Schulze and Lee (1986) found that between MEL and the *repleta* group only about 11% of the total scDNA hybridized and the ΔT_m's ranged from 8 to 13. With cDNA, we find 53% hybridization between MEL and the *repleta* group with ΔT_m's between 11.3 and 12.8. Thus, the measurements of the ΔT_m are more accurate and repeatable when a greater percentage of tracer hybridizes, perhaps due to the inaccuracies involved in removing nonreassociated sequences. This does not necessarily mean that the rate of evolution of cDNA is less than that of hybridizing scDNA; in fact it is about the same for these *Drosophila.* Presumably at the moderate and higher distances, much of the hybridizing scDNA is coding DNA so that one would expect the hybridizing scDNA and cDNA to evolve at about the same rate. Elsewhere we go into more detail about the implications of these and other studies for understanding patterns of genome evolution (Caccone et al., unpublished); our concern in this paper is primarily phylogenetic.

An obvious complication of this technique is that we have not used single-copy cDNA. Instead, the tracer population is heterogeneous in copy number because total mRNA populations were used to generate the cDNA. This results in, among other things, difficulty in interpreting percent reassociations. Using sea urchins, Roberts et al. (1985) devised a method for making such cDNA populations more nearly homogeneous in copy number by reassociation to total scDNA such that the tracer population would represent approximately the frequency of the expressed genes in the genome. We attempted to use their method on our *Drosophila* material without success; the losses of DNA at each step were so great that the final preparations did not contain enough radioactivity to perform experiments. However, in Roberts et al.'s experiments, the total cDNA preparations gave nearly identical results to those for cDNA made more homogeneous in copy number. Given the internal consistency of our results, their agreement with other data (see below), and the Roberts et al. results, using total cDNA as in this study yields reliable results for the purposes of phylogenetic analysis. Thus, this method would seem to extend the phylogenetic range over which DNA hybridization can reliably measure divergence of populations of sequences, a method particularly useful for fast-evolving groups such as *Drosophila.*

The second major concern is what the results indicate about *Drosophila* relationships. Despite the great attention bestowed upon *Drosophila* by experimental biologists, the understanding of their phylogenetic relationships is relatively poor. Relationships based primarily on morphological, and to a lesser extent chromosomal, data have been analyzed by Sturtevant (1942), Patterson and Stone (1952), Throckmorton (1975), Wheeler (1981), and Grimaldi (1987, 1990). Allozymic studies in general have not proven to be of particular utility with *Drosophila* as the rate of evolution is too rapid; Nei's Ds of one or more are quickly reached even between quite closely related species. The major molecular methods proven reliable at the higher systematic levels have been microcomplement fixation and two-dimensional electrophoresis of general proteins (Beverley and Wilson 1982, 1984, 1985; McIntyre and Collier 1986; Spicer 1988; Collier 1990). The DNA sequence of one protein-coding gene, alcohol dehydrogenase, is known from enough *Drosophila* species of diverse taxonomic affinities to become useful for understanding phylogenies (Sullivan et al. 1990). Ribosomal RNA-coding regions of both the mitochondrial DNA (DeSalle, unpublished) and the nuclear genome (Solignac, personal communication) are now also known for many diverse species of *Drosophila.*

The phylogenetic interpretation of the results of the present study are portrayed in Figs. 1–3. Together with the work just cited, some consistencies in the phylogenetic relationships of *Drosophila* groups are becoming clear. The following conclusions are generally supported by all the work, both morphological and molecular, or are at least consistent with all data sets. Some differences, generally minor and only for one data set, do exist, so the following represents a consensus.

1) The subgenus *Sophophora,* at least as represented by the *melanogaster* subgroup, the *obscura* group, and the *willistoni* group, is a monophyletic taxon.
2) Within *Sophophora,* the *melanogaster* subgroup and the *obscura* groups are more closely related to each other than either is to the *willistoni* group.
3) The *virilis/repleta* group is monophyletic.
4) The rest of the species studied here are not clearly defined in their phylogenetic affinities, but all have relatively deep branches, implying old divergences. These include *melanica, immigrans, funebris,* and the Hawaiians. Classically, all these species and their group members have been placed in the single subgenus *Drosophila.* This subgenus is not as clearly defined genetically as

is *Sophophora* and may not be a monophyletic taxon.

These conclusions are pictorially displayed in Fig. 3 where the boxed area should be considered undefined, and the branching patterns not shaded are well-supported by a variety of data, including that presented in this paper.

Acknowledgments. We are very grateful to Charles Marshall for the bootstrap analysis and Yale University for financial support. J.M.G. was supported by an NIH Genetics Training Grant and A.G. thanks her home institution for generous leave.

References

Beverley SM, Wilson AC (1982) Molecular evolution in *Drosophila* and higher Diptera. I. Micro-complement fixation studies of a larval hemolymph protein. J Mol Evol 18:251–264

Beverley SM, Wilson AC (1984) Molecular evolution in *Drosophila* and higher Diptera. II. A time scale for fly evolution. J Mol Evol 21:1–13

Beverley SM, Wilson AC (1985) Ancient origin for Hawaiian Drosophilinae inferred from protein comparisons. Proc Natl Acad Sci USA 82:4753–4757

Britten RJ (1986) Rates of DNA sequence evolution differ between taxonomic groups. Science 231:1393–1398

Britten RJ, Cetta A, Davidson EH (1978) The single-copy sequence polymorphism of the sea urchin *Strongylocentrotus purpuratus.* Cell 15:1175–1186

Caccone A, Powell JR (1987) Molecular evolutionary divergence among North American cave crickets. II. DNA–DNA hybridization. Evolution 41:1215–1238

Caccone A, Powell JR (1989) DNA divergence among hominoids. Evolution 43:925–942

Caccone A, Powell JR (1990) Extreme rates of heterogeneity in insect DNA evolution. J Mol Evol 30:273–280

Caccone A, Powell JR (1991) A protocol for the TEACL method of DNA–DNA hybridization. In: Hewitt G, Johnston A, Young J (eds) Molecular techniques in taxonomy. Springer-Verlag, Heidelberg, pp 385–407

Caccone A, Amato GD, Powell JR (1987) Intraspecific DNA divergence in *Drosophila*: a study on parthenogenetic *D. mercatorum.* Mol Biol Evol 4:343–350

Caccone A, Amato GD, Powell JR (1988a) Rates and patterns of scnDNA and mtDNA divergence within the *Drosophila melanogaster* subgroup. Genetics 118:671–683

Caccone A, DeSalle R, Powell JR (1988b) Calibration of the change in thermal stability of DNA duplexes and degree of base pair mismatch. J Mol Evol 27:212–217

Catzeflis FM, Sheldon FH, Ahlquist JE, Sibley CG (1987) DNA–DNA hybridization evidence of the rapid rate of muroid rodent DNA evolution. Mol Biol Evol 4:242–253

Catzeflis FM, Nevo E, Ahlquist JE, Sibley CG (1989) Relationships of the chromosomal species in the Eurasian mole-rats of the *Spalax ehrenbergi* group as indicated by DNA–DNA hybridization and an estimate of the spalacid–murid divergence time. J Mol Evol 29:223–232

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. Analyt Biochem 162:156–159

Collier GE (1990) Evolution of arginine kinase within the genus *Drosophila.* J Hered 81:177–182

DeSalle R, Templeton AR (1988) Founder effects and the rate of mitochondrial DNA evolution in Hawaiian *Drosophila.* Evolution 42:1076–1084

Felsenstein J (1990) PHYLIP manual version 3.2. University Herbarium, University of California, Berkeley

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. Science 155:279–284

Goddard K, Caccone A, Powell JR (1990) Evolutionary implications of DNA divergence in the *Drosophila* obscura group. Evolution 44:1656–1670

Grimaldi DA (1987) Phylogenetics and taxonomy of *Zygothrica* (Diptera: Drosophilidae). Bull Am Nat Hist 186:103–268

Grimaldi DA (1990) A phylogenetic revised classification of genera in the Drosophilidae (Diptera). Bull Am Nat Hist 197: 1–129

Hall TJ, Grula JW, Davidson EH, Britten RJ (1980) Evolution of sea urchin nonrepetitive DNA. J Mol Evol 16:95–110

Hunt JA, Hall TJ, Britten RJ (1981) Evolutionary distance in Hawaiian *Drosophila* measured by DNA reassociation. J Mol Evol 17:361–367

Jukes TH, Cantor C (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian Protein Metabolism. Academic Press, New York, pp 21–123

Kirsch JAW, Springer MS, Krajewski C, Archer M, Aplin K, Dickerman AW (1990) DNA/DNA hybridization studies of the carnivorous marsupials. I: The intergeneric relationships of bandicoots (Marsupialia: Perameloidea). J Mol Evol 30:434–448

Lanyon SM (1985) Detecting internal inconsistencies in distance data. Syst Zool 34:397–403

Li W-H, Lui C-C, Wu C-I (1985) Evolution of DNA sequences. In: MacIntyre RJ (ed) Molecular evolutionary genetics. Plenum Press, New York, pp 1–94

MacIntyre RJ, Collier GE (1986) Protein evolution in *Drosophila.* In: Ashburner M, Carson HL, Thompson JR (eds) The genetics and biology of *Drosophila,* vol 3e. Academic Press, New York, pp 39–146

Marshall CR (1991) Statistical test and bootstrapping: assessing the reliability of phylogenies derived from distance data. Mol Biol Evol 8:386–391

Pamilo P, Nei M (1988) Relationships between gene trees and species trees. Mol Biol Evol 5:568–583

Patterson JT, Stone WS (1952) Evolution in the genus *Drosophila.* Macmillan, New York

Powell JR, Caccone A (1989) Intra- and interspecific genetic variation in *Drosophila.* Genome 31:233–238

Powell JR, Caccone A (1990) The TEACL method of DNA–DNA hybridization: technical considerations. J Mol Evol 30: 267–272

Powell JR, Caccone A, Amato GD, Yoon C (1986) Rates of nucleotide substitution in *Drosophila* mitochondrial DNA and nuclear DNA are similar. Proc Natl Acad Sci USA 83:9090–9093

Roberts JW, Johnson SA, Kier P, Hall TJ, Davidson EH, Britten RJ (1985) Evolutionary conservation of DNA sequences expressed in sea urchin eggs and early embryos. J Mol Evol 22:99–107

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Schulze DH, Lee CS (1986) DNA sequence comparison among closely related *Drosophila* species of the *mulleri* complex. Genetics 113:287–303

Sheldon FH (1987a) Phylogeny of herons estimated from DNA–DNA hybridization data. Auk 104:97–108

Sheldon FH (1987b) Rates of single-copy DNA evolution in herons. Mol Biol Evol 4:56–69

Sibley CG, Ahlquist JE (1981) The phylogeny and relationships of the ratite birds as indicated by DNA–DNA hybridization. In: Scudder GGE, Reveal JL (eds) Evolution today. Carnegie-Mellon University, Pittsburgh, pp 301–335

Sibley CG, Ahlquist JE (1984) The phylogeny of the hominoid primates as indicated by DNA–DNA hybridization. J Mol Evol 20:2–15

Sibley CG, Ahlquist JE (1987) DNA hybridization evidence of homonoid phylogeny: results from an expanded data set. J Mol Evol 26:99–121

Spicer GS (1988) Molecular evolution among some *Drosophila* species groups as indicated by two-dimensional electrophoresis. J Mol Evol 27:250–260

Springer MS, Kirsch JA (1989) Rates of single-copy DNA evolution in phalangeriform marsupials. Mol Biol Evol 6:331–341

Sturtevant AH (1942) The classification of the genus *Drosophila* with the description of nine new species. Univ Tex Publ 4213: 5–51

Sullivan DT, Atkinson PW, Starmer WT (1990) Molecular evolution of the alcohol dehydrogenase genes in the genus *Drosophila*. Evol Biol 24:107–147

Throckmorton LH (1975) The phylogeny, ecology, and geography of *Drosophila*. In: King RC (ed) Handbook of genetics, vol 3. Invertebrates of genetic interest. Plenum Press, New York, pp 421–469

Werman SD, Davidson EH, Britten RJ (1990) Rapid evolution in a fraction of the *Drosophila* nuclear genome. J Mol Evol 30:281–289

Wheeler MR (1981) The Drosophilidae: a taxonomic overview. In: Ashburner M, Carson HL, Thompson JN (eds) The genetics and biology of *Drosophila*. Academic Press, New York, pp 61–140

Received May 17, 1991/Revised August 22, 1991/Accepted September 15, 1991