Rates and Patterns of scnDNA and mtDNA Divergence Within the Drosophila melanogaster Subgroup

Adalgisa Caccone,[1] George D. Amato and Jeffrey R. Powell

Department of Biology, Yale University, New Haven, Connecticut 06511

ABSTRACT

Levels of DNA divergence among the eight species of the *Drosophila melanogaster* subgroup and *D. takahashii* have been determined using the technique of DNA-DNA hybridization. Two types of DNA were used: single-copy nuclear DNA (scnDNA) and mitochondrial DNA (mtDNA). The major findings are: (1) A phylogeny has been derived for the group based on scnDNA which is congruent with chromosomal data, morphology, and behavior. The three homosequential species, *simulans*, *sechellia*, and *mauritiana*, are very closely related; the scnDNA divergence indicate the two island species are a monophyletic group. (2) The rates of change of scnDNA and mtDNA are not greatly different; if anything scnDNA evolves faster than mtDNA. (3) The rates of scnDNA evolution are not closely correlated to chromosomal (inversion) evolution. (4) The *Drosophila* genome appears to consist of two distinct classes of scnDNA with respect to rate of evolutionary change, a very rapidly evolving fraction and a relatively conservative fraction. (5) The absolute rate of change was estimated to be at least 1.7% nucleotide substitution per one million years. (6) DNA distance estimates based on restriction site variation are correlated with distances based on DNA-DNA hybridization, although the correlation is not very strong.

D ROSOPHILA have played a central role in experimental genetics, evolution, and, more recently, molecular biology. Among *Drosophila, D. melanogaster* has been most extensively studied. For a long time, *D. melanogaster* was thought to have only one close relative, *D. simulans*. Relatively recently several other close relatives have been discovered in Africa and neighboring islands so that presently the *melanogaster* subgroup is comprised of eight known species [see LEMEUNIER *et al.* (1986) for a review]. These are two cosmopolitan species, *D. melanogaster* and *D. simulans*; four mainland Africa species, *D. yakuba, D. erecta, D. teissieri,* and *D. orena*; and two insular species, *D. sechellia* and *D. mauritiana*. Because of the importance of this group, we chose to study levels of DNA divergence among species. The purposes of this study are to compare the relative rates of evolution of two kinds of DNA, single-copy nuclear DNA (scnDNA) and mitochondrial DNA (mtDNA), and to compare the phylogenetic relationships indicated by the two sets of data.

The technique we have used to quantitatively assess genetic divergence is DNA-DNA hybridization. The technique relies on the fact that the thermal stability of DNA duplexes is determined by the fidelity of base-pair (bp) matching, A with T and G with C. Recently we (CACCONE, DESALLE and POWELL 1988) have determined empirically the relationship between

percent bp mismatch and change in median melting temperature, ΔTm. The conversion is that $1° \Delta Tm$ corresponds to 1.7% bp mismatch; the relationship is linear, at least over the range of values in the present study. Thus ΔTm is a valid overall measure of DNA divergence.

MATERIALS AND METHODS

The strains of flies used in this study, along with the abbreviations by which they will be referred subsequently, are listed in Table 1. To our knowledge, all are isofemale lines begun by single inseminated females from nature. For three species we studied four different strains to either assess intraspecific variability or to more accurately determine a branching pattern by independent measures based on different strains. In addition to the *melanogaster* subgroup species, we included *D. takahashii* as an outgroup species to "root" phylogenetic trees.

The methods of DNA extraction, DNA-DNA hybridization, and analysis have been reported elsewhere (POWELL *et al.* 1986; CACCONE, AMATO and POWELL 1987; CACCONE and POWELL 1987) and will not be repeated here. However, three aspects of our methods are worth emphasizing. First, we have used the so-called TEACL method of DNA-DNA hybridization (BRITTEN, CETTA and DAVIDSON 1978). With this method the DNA melting is done in a compound, tetraethylammonium chloride, which at a certain concentration (2.4 M) negates the base composition effect on the thermal stability of DNA duplexes. This is important because the A + T/G + C ratios are quite different in mtDNA and scnDNA and may also vary among species. The percentage of single-stranded DNA at each temperature is determined by S1 nuclease digestion. Second, with this method, one can make a correction for the dependence

---

[1] On leave from II Università di Roma "Tor Vergata," Rome, Italy.

TABLE 1

Strains of *Drosophila* used in this study, abbreviations used in text, and sources of the strains

Drosophila species	Strain	Abbreviation	Source
D. melanogaster	Oregon-R	MEL	D. POULSON
	France 9-1	MEL-A	R. SINGH
	West Africa 25A	MEL-B	R. SINGH
	Australia 13	MEL-C	R. SINGH
D. simulans	Arizona	SIM	D. HARTL
	Capetown 20	SIM-D	R. SINGH
	S. France 45	SIM-E	R. SINGH
	Congo Brazzeville 32	SIM-F	R. SINGH
D. mauritiana	—[a]	MAU	D. HARTL
	David 207	MAU-G	J. COYNE
	David 75	MAU-H	J. COYNE
	David 102	MAU-I	J. COYNE
D. erecta	—	ERE	D. HARTL
D. teissieri	—	TES	D. HARTL
D. yakuba	—	YAK	D. HARTL
D. sechellia	—	SEC	D. HARTL
D. orena	—	ORE	D. HARTL
D. takahashii	0311.0	TAK	Bowling Green Drosophila Stock Center

[a] Dashes indicate no strain designation.

TABLE 2

Intraspecific scnDNA distances for *D. melanogaster* and *D. simulans*

Tracer		Driver[a]	
D. melanogaster	MEL-A	MEL-B	MEL-C
Oregon-R	0.11 ± 0.06	0.04 ± 0.09	0.10 ± 0.09
D. simulans	SIM-D	SIM-E	SIM-F
Arizona	0.15 ± 0.12	0.20 ± 0.10	0.25 ± 0.20

[a] Driver strains represent a wide geographic range; see Table 1 for origin. Figures are ΔTm ± 1 standard error.

of melting on the size of the duplexes (HALL *et al.* 1980; HUNT, HALL and BRITTEN 1981). Thus in our data tables we will present tracer length measurements determined by alkaline agarose electrophoresis and the temperature correction for the Tm. Third, the condition under which reassociation was carried out (1 M TEACL, 45°) requires a bp match of about 75% or greater to form stable duplexes. Reassociation reactions are incubated to equilibrium ($C_o t > 6,000$ mol sec/liter) after which no further reassociation occurs. Percent reassociation is determined by the fraction of the reaction which remains S1 nuclease resistant.

For the mtDNA studies, tracer mtDNA was isolated by two CsCl gradients to be certain to exclude nuclear DNA. The label for mtDNA was ^{3}H-CTP so that the effect of the A + T-rich region was minimized.

We have presented elsewhere (CACCONE, AMATO and POWELL 1987) the formulas we used to calculate overall average ΔTm's, with and without reciprocals, and their standard errors. We also used the PHYLIP package of programs developed by FELSENSTEIN (1985) to obtain phylogenetic trees based on the assumption of branches pro-

ceeding synchronously (the KITSCH program) and trees allowing asynchrony (the FITCH program).

RESULTS

Details of the results are presented in the APPENDIX. Not all pairwise comparisons were performed for either scnDNA or mtDNA. However, in the case of scnDNA, all species were used as tracers and drivers in multiple comparisons and all nodes of trees were multiply determined. The mtDNA data are not so complete due to the difficulty of purifying mtDNA from the different species. However, all possible nodes have been tested for mtDNA divergence.

One test of the accuracy of this technique is to test for reciprocity. In theory, reciprocals in which each species is used as a tracer and driver should yield the same result. Also if the temperature correction for tracer length makes the measurements more accurate, then reciprocals based on ΔTm (corrected Tm's) should be closer than reciprocals based on Δtm (uncorrected). This is the case in the present data. For the twelve reciprocal tests using scnDNA the mean difference between reciprocal Δtm's is 0.62° while for ΔTm it is 0.32°. For the four reciprocals using mtDNA the mean difference between Δtm's is 0.62° and between ΔTm's is 0.30°. Since the standard errors associated with Δtm and ΔTm are usually 0.1° to 0.2°, reciprocal Δtm's are often significantly different while reciprocal ΔTm's are very seldom significantly different. This both validates the accuracy of the technique and justifies using the temperature

TABLE 3

ΔTm matrix for scnDNA[a]

	MEL	SIM	MAU	SEC	YAK	TES	ERE	ORE	TAK
MEL	0	2.17*	2.51*	2.38*	4.02*	3.97	3.86	5.11	6.16*
		0.09	0.10	0.07	0.08	0.17	0.09	0.08	0.11
SIM	3.78	0	0.80*	1.11*	4.40	4.34	4.49	5.12	6.85
			0.09	0.11	0.19	0.27	0.16	0.18	0.13
MAU	4.39	1.37	0	0.64*	4.46	4.50	4.79*	5.28	6.58
				0.07	0.12	0.36	0.14	0.24	0.10
SEC	4.16	1.91	1.10	0	4.05	4.12	4.10	5.27	6.64
					0.12	0.25	0.06	0.17	0.11
YAK	7.17	7.88	7.99	7.22	0	2.20	2.83*	5.53	6.37*
						0.08	0.12	0.25	0.11
TES	7.07	7.77	8.07	7.35	3.84	0	2.74	5.50	6.09
							0.11	0.17	0.10
ERE	6.87	8.05	8.62	7.32	4.97	4.81	0	5.40*	6.38
								0.15	0.15
ORE	9.23	9.25	9.56	9.54	10.04	9.99	9.79	0	6.21
									0.21
TAK	11.28	12.66	12.11	12.23	11.70	11.14	11.72	11.38	0

[a] Upper right displays ΔTm's with one standard error below. Asterisks indicate reciprocal tests were performed. Lower left indicates estimated bp mismatch calculated as explained in text.

correction. All analyses and discussion will concern ΔTm's.

In order to assess the significance of interspecific divergence, it is useful to know levels of intraspecific variation. We studied four strains each of *D. melanogaster* and *D. simulans*, the geographically most widely dispersed species in the group. The origin of the strains is indicated in Table 1. Table 2 presents the ΔTm's between strains. While all the mean ΔTm's are positive, only one is significantly different from zero. Note also in APPENDIX Tables 6 and 7, the normalized percent reassociation (NPR) is virtually 100% for the intraspecific tests which also indicates lack of significant intraspecific differentiation. These results contrast to our previous results with *D. mercatorum* (CACCONE, AMATO and POWELL 1987); this will be discussed in the DISCUSSION.

Table 3 presents the matrix of ΔTm values and estimated percent bp mismatch. Reciprocals refer to cases where both species were used as tracer and driver in separate experiments. Note that the standard errors are not dependent upon the magnitude of ΔTm, a fact that validates some of the analytical procedures we have employed. The percent bp mismatch in the lower left part of Table 3 was based on the conversion we have recently empirically determined (CACCONE, DESALLE and POWELL 1988), namely a ΔTm of 1° corresponds to 1.7% bp mismatch. This estimate was then corrected for multiple substitutions by the formula of JUKES and CANTOR (1969).

Figure 1 is a phylogeny of the species based on scnDNA divergence. The unweighted pair-group method using arithmetic averages, UPGMA (SOKAL and MICHENER 1958), was used in this figure. Identical branching patterns are obtained using the FITCH and KITSCH programs of PHYLIP (FELSENSTEIN 1985) which are based on the least squares methods (CAVALLI-SFORZA and EDWARDS 1967; FITCH and MARGOLIASH 1967); the quantitative placement of the nodes change very slightly.

As the ΔTm's have error estimates between around 0.1° to 0.2°, all nodes in Figure 1 are obviously significant (*i.e.*, unambiguously define dichotomous nodes) with the exception of the nodes connecting SIM, SEC, and MAU. While all grouping methods (UPGMA, FITCH, and KITSCH) agree that SEC and MAU form a monophyletic group branching from SIM, the branch length connecting the two nodes varies from 0.2° to 0.6° depending upon the algorithm used to construct the tree. Therefore, we studied three additional strains of MAU in an attempt to clarify this split. Table 4 presents the results of all the independent measures through the nodes. While these may not be strictly independent measures as in some cases the same tracer is used in different comparisons, nevertheless they can be considered quasi-independent. A nonparametric Mann-Whitney test is also presented in Table 4; this indicates that the node connecting SEC and MAU is significantly smaller than that connecting these two species with SIM. If there is an effect of tracer preparation, then we can test the nodes only for those ΔTm's using the same tracer. This would be SEC-MAU and MAU-
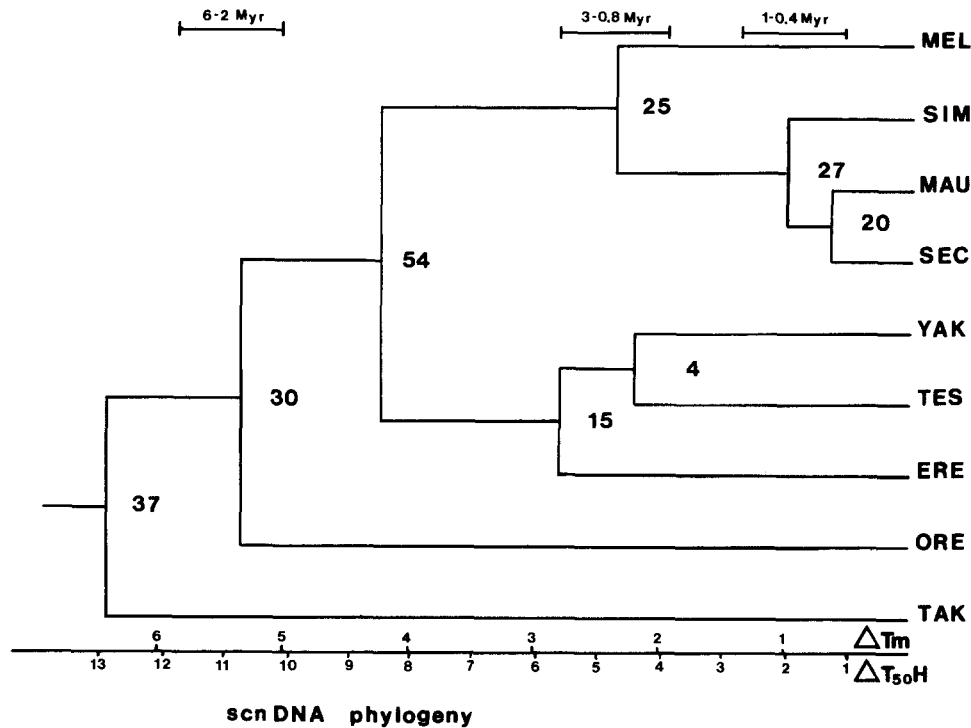
scn DNA phylogeny

FIGURE 1.—UPGMA dendogram based on scnDNA. The numbers at the nodes are the number of determinations through that node. The time estimates given on top are from LEMEUNIER et al. (1986). The difference in the two scales, ΔTm and ΔT$_{50}$H, is discussed in the text.

TABLE 4

Mann-Whitney test of significance of nodes connecting SIM, SEC and MAU[a]

Comparison	ΔTm	Rank	Comparison	ΔTm	Rank
SIM-SEC	0.89	3	SEC-MAU	0.57	12
SEC-SIM	1.14	1	G	0.74	6
SIM-MAU	0.73	7	H	0.60	10
G	0.59	11	I	0.67	8
H	0.80	4	MAU-SEC	0.63	9
I	0.79	5			
MAU-SIM	1.03	2			
	U = 30		P = 0.025		

[a] Species are designated as tracer-driver.

SEC for one node and SEC-SIM and MAU-SIM, that is, the same SEC and MAU tracers were used in both comparisons. Using the weighted averages and standard error calculations presented in CACCONE, AMATO and POWELL (1987), the two nodes are 0.58 $\pm$ 0.08 and 1.13 $\pm$ 0.06, clearly significantly different.

Table 5 presents the ΔTm matrix for mtDNA. Figure 2 is the UPGMA tree based on these data. This tree differs from that based on scnDNA in two ways. First the split among SIM-MAU-SEC appears different although the error on these nodes is such that the data cannot reject a trichotomy for these three species. We should note that the strains of SIM and MAU we used possessed mtDNA's which were

clearly different on the basis of restriction endonuclease patterns generated. Thus they did not contain the shared mtDNA type reported in SOLIGNAC, MONNEROT and MOUNOLOU (1986). The second difference is the placement of the outgroup, TAK. It appears ORE is more distant, although again these nodes are not significantly different from one another. A final point of interest is the clear distinction between YAK and TES, a ΔTm of 1.89 indicating about 3.3% bp mismatch. SOLIGNAC, MONNEROT and MOUNOLOU (1986) could not distinguish these species by restriction endonuclease digestions.

Another statistic which should reflect degree of genetic divergence is the percent of the DNA which reassociates; this is generally normalized to that for

TABLE 5

ΔTm matrix for mtDNA[a]

	MEL	SIM	MAU	SEC	YAK	TES	ERE	ORE	TAK
MEL	0	1.95* 0.02	2.26* 0.02	2.25 0.14	3.71 0.05	3.91 0.17	3.99 0.02	—[b]	—
SIM	3.39	0	0.93* 0.09	1.27 0.13	3.60 0.15	—	4.18 0.16	8.62 0.27	7.54 0.15
MAU	3.94	1.60	0	1.56 0.18	4.54 0.13	4.93 0.14	—	—	—
SEC	3.93	2.19	2.70	0	4.33 0.12	—	—	—	—
YAK	6.59	6.39	8.15	7.75	0	1.89 0.17	3.45 0.09	7.57 0.11	7.52 0.18
TES	6.96	—	8.89	—	3.28	0	—	—	—
ERE	7.11	7.47	—	—	6.11	—	0	—	—
ORE	—	16.31	—	—	14.12	—	—	0	—
TAK	—	14.06	—	—	14.02	—	—	—	0

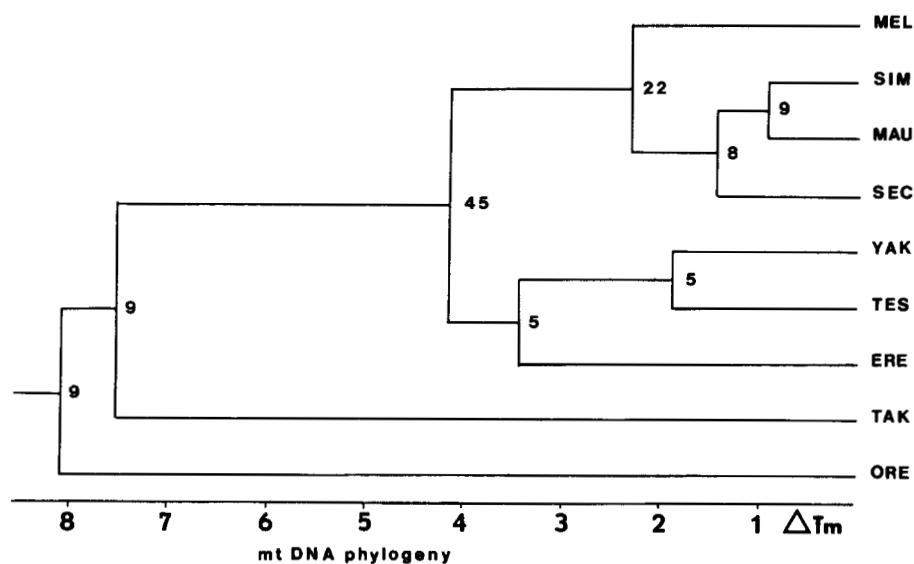[a] Arrangement of data is the same as in Table 3.
[b] Comparison was not done.

FIGURE 2.—UPGMA dendogram based on mtDNA. The numbers at the nodes are the number of determinations through that node.

homoduplex taken to be 100%. This normalized percent reassociation, NPR, is given in the tables in the APPENDIX. Unfortunately this statistic is not nearly as repeatable as the Tm. In replicate experiments discussed in CACCONE and POWELL (1987), the error associated with NPR is on the order of 10%. However, when one averages over intervals, the relationship between NPR and ΔTm is clear, Figure 3. For several groups of organisms the relationship is remarkably linear at least over some range. It is interesting to note that the data of SCHULZE and LEE (1986) using the hydroxlapatite (HAP) method on the *mulleri* group of *Drosophila* closely parallels our data using TEACL on the *melanogaster* subgroup. The somewhat

lower line for the TEACL method is expected due to the fact that the percent reassociation is determined by S1 digestion. Thus overhanging single-stranded ends and loops due to insertions/deletions will be digested and considered nonreassociated. With the HAP method, reassociation is determined by binding to HAP. Thus if two molecules form enough stable duplex to bind to HAP, all of the radioactivity in overhanging ends and loops will be considered reassociated. It is heartening that the two *Drosophila* lines are so parallel and the difference is in the direction expected on theoretical grounds.

Some authors express the divergence between two taxa taking into consideration that fraction of the
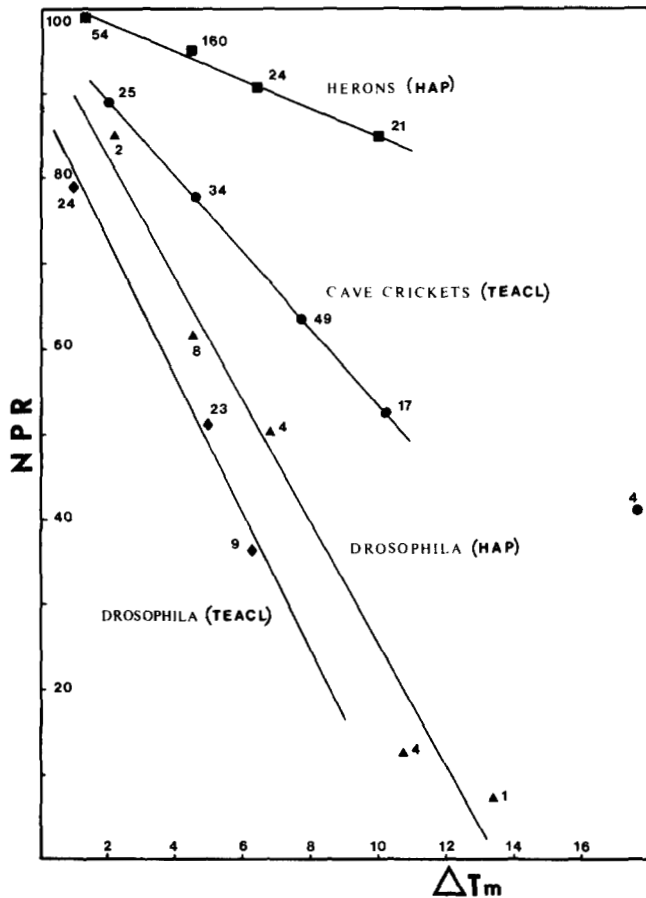
FIGURE 3.—The relationship between ΔTm and normalized percent reassociation, NPR. Data were grouped into ΔTm intervals, 0 to 3, 3 to 6, 6 to 9, etc. The mean NPR and mean ΔTm for each interval are plotted. The numbers next to the points are number of data in that interval. HAP indicates the hydroxylapatite method was used and TEACL indicates the TEACL method. The heron data are from SHELDON (1986); the cricket data from CACCONE and POWELL (1987); the Drosophila HAP data from SCHULZE and LEE (1986); and the Drosophila TEACL data from the present study.

genome which did not reassociate. This is the $T_{50}H$ of SIBLEY and AHLQUIST (1984), the T median of BRITTEN (1986), or T_mR of BENVENISTE (1985). Using the linear relationship between ΔTm and NPR, we can rescale the degree of divergence incorporating NPR; this is given in Figure 1 with the second scale $\Delta T_{50}H$. In the case of mtDNA, the NPR is very high for all comparisons. It averaged 91% with a standard deviation of 9%. There is no clear pattern of relationship between ΔTm and NPR for mtDNA probably because of the small range of NPR; thus we do not present $\Delta T_{50}H$ for mtDNA.

DISCUSSION

Intraspecific variation: In a previous publication (CACCONE, AMATO and POWELL 1987) we documented that two genomes of *D. mercatorum* differ from one another by a ΔTm of 1.3°. The strains used in that study were completely homozygous due to the man-
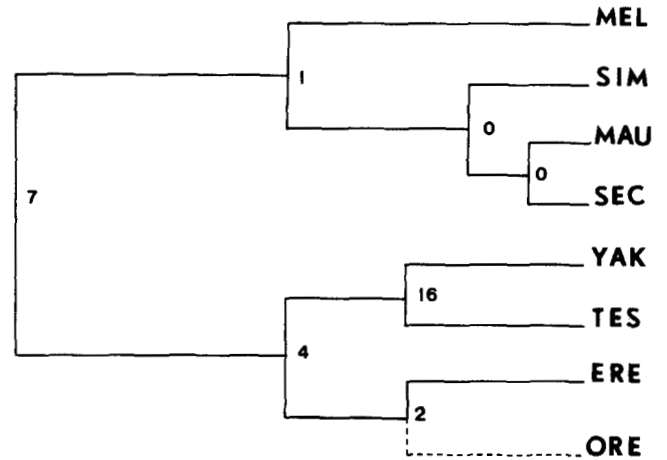

FIGURE 4.—The phylogeny proposed by LEMEUNIER et al. (1986) based primarily on polytene banding patterns. The numbers of fixed inversions are shown. ORE is connected by a dotted line to indicate the revised phylogeny proposed by this group places ORE as the most distantly related member of the subgroup (LACHAISE et al. 1988). The relationship of the three homosequential species was based on similarity of metaphase chrosmosomes (LEMEUNIER and ASHBURNER 1984).

ner in which parthenogenesis occurs: a haploid genome doubles and fuses giving rise to all female parthenogenetic lines. In the present study we could detect little significant divergence among sexual strains of MEL and SIM (Table 2). All the mean ΔTm's are positive indicating there is probably some divergence, though not as much as detected in *D. mercatorum*. This difference may be due to two, not mutually exclusive, factors. *D. mercatorum* may just be more variable than MEL and SIM. Alternatively, the heterozygosity inherent in sexual strains may obscure much of the differences between strains. The important point for the present study is that the degree of intraspecific variation detectable by DNA-DNA hybridization for the *melanogaster* subgroup is negligible. Thus, hybridization using a single strain of each species yields an accurate picture of species relationships.

Phylogenies: The phylogenies derived from scnDNA and mtDNA are congruent with two exceptions. Our mtDNA data could not resolve the relationship of SIM, SEC and MAU. The repeated tests for scnDNA did resolve this in favor of SEC and MAU being monophyletic (Table 4). Obviously, these species are very closely related and must be pushing the very limits of the resolving power of DNA-DNA hybridization data. COYNE and KREITMAN (1986) also had difficulty resolving this relationship based on nucleotide sequences of the alcohol dehydrogenase region. Their data tended to favor SIM and SEC as a monophyletic group with MAU being the more distant. However, the statistical support for this arrangement was not strong and they could not reject alternative relationships. Also COYNE and KREITMAN

had difficulty reconciling their proposed phylogeny with morphological and behavioral data. They had to invoke convergent evolution for several traits in the island species. The scnDNA phylogeny proposed here circumvents these difficulties as the island species are monophyletic. Thus the scnDNA phylogeny is congruent with morphological and behavioral data.

A second difference between the mtDNA and scnDNA phylogenies is the placement of TAK, the outgroup. ScnDNA clearly agrees with standard taxonomy that TAK is outside the *melanogaster* subgroup, while mtDNA shows TAK and ORE to be about equally distant. One possible reason for this is that these species are 14–16% divergent at the nucleotide level (Table 5) and thus might be approaching a saturation effect for substitutions with little or no selective constraints (*e.g.*, BROWN, 1983).

Figure 4 presents the phylogeny of the *melanogaster* subgroup based primarily on the banding pattern of the polytene chromosomes as presented in LEMEU-NIER *et al.* (1986). LACHAISE and colleagues (1988) have been reexamining the proposed phylogeny and have now concluded that ORE is the most distant member of the subgroup and should be placed outside the tree connecting the other species. Thus, the revised phylogeny based on chromosomes is completely congruent with that derived from DNA-DNA hybridization for scnDNA.

For three reasons we feel that scnDNA is superior to mtDNA for reconstructing phylogenetic relationships. First, scnDNA hybridization experiments are based on millions of nucleotides. MtDNA is a much smaller molecule and thus more subject to stochastic inconsistencies. Second, there is increasing evidence that mtDNA can cross species boundaries more easily than scnDNA thus obscuring phylogeny. For the *melanogaster* subgroup, just such a process has been invoked to explain patterns of mtDNA variation (SOLIGNAC, MONNEROT and MOUNOLOU 1986). Third, the mode of inheritance of mtDNA may make it more sensitive to population processes such as bottlenecks, and thus may be a less accurate molecular clock compared to scnDNA.

Therefore, we believe the phylogeny presented in Figure 1 reflects the true phylogeny as well as present data allow. It is consistent with morphological, behavioral, and chromosomal data. Where it disagrees with the phylogeny based on mtDNA (Figure 2), the differences are not significant. Finally, while the phylogeny in Figure 1 was constructed using the UPGMA algorithm, using the least-squares algorithms in PHYLIP, FITCH and KITSCH, identical branching patterns are obtained with only minor quantitative differences, *i.e.*, changes in branch lengths are minimal.

Evolutionary rates: One of the more remarkable findings with mammalian mtDNA is that it evolves five to ten times faster than nuclear DNA (summarized in BROWN 1985). In comparing Figures 1 and 2, it is clear this is not the case with *Drosophila*. Our previous study on just MEL and YAK indicated scnDNA and mtDNA evolve at about the same rate (POWELL *et al.* 1986); the results here confirm and extend this observation to the entire subgroup. If one just considers ΔTm based on that fraction of the scnDNA which hybridized under the stringency of the reassociation conditions (greater than about 75% homology required), the two types of DNA evolve at remarkably similar rates. However, if one considers divergence based on $\Delta T_{50}H$, scnDNA appears to evolve faster than mtDNA. Converting ΔTm to $\Delta T_{50}H$ for the mtDNA data would change the values negligibly as the NPR averaged greater than 90%. As discussed below, this difference in ΔTm and $\Delta T_{50}H$ probably reflects the fact that there are two discrete fractions of the nuclear genome with respect to rates of evolution. MtDNA appears to evolve at about the same rate as the more conserved fraction.

One question of great interest is whether measurements from DNA-DNA hybridization studies behave in a clock-like manner. FELSENSTEIN (in the PHYLIP manual) suggests one possible test for comparing whether trees based on the assumption of asynchrony give a significantly better fit than trees based on the assumption of branch lengths increasing synchronously, one of the assumptions of the molecular clock. However, this test cannot be validly directly applied to ΔTm's as there is covariance in the distance matrix (J. FELSENSTEIN, personal communication). We are developing other tests which circumvent this problem (J. R. POWELL, J. HARTIGAN and A. CACCONE, unpublished data). So far from our preliminary results we have not been able to reject the synchronous assumption for scnDNA.

If we accept the molecular clock for the present data, is there any way of calibrating the clock with respect to absolute time? It is extremely difficult to date times of divergence with insects for which there is a very poor fossil record. Nevertheless, based on biogeographical and geological considerations, LEU-MENIER *et al.* (1986) do suggest ranges of divergence times for some of the branches of the *melanogaster* subgroup. The split of the island species from SIM was proposed to have occurred 0.4 to 1 million years (myr) ago; MEL split from the SEC-MAU-SIM triad 0.8 to 3 myr ago; and ORE from the rest of the group 2 to 6 myr ago. While these are considerable ranges, if we even take the oldest suggested dates, the rate of DNA evolution is very high for these species. Approximately 1° ΔTm reflects one million years since last common ancestor. This is about five times the rate suggested for birds and primates (SIBLEY and AHLQUIST 1984). Converting to bp mismatch, this indicates 1.7% base substitutions per myr,

a relatively very high rate compared to estimates based on sequence data for mammals; even for introns, pseudogenes, and synonymous substitutions, the rate does not exceed 0.5% per myr (LI *et al.* 1985). If we consider the genome as a whole, as reflected in $\Delta T_{50}H$, the rate would be about twice as fast. If we take the more recent suggested divergence times, this would increase the rate another threefold. Thus, depending upon assumptions and interpretations, our data are consistent with a rate of change of 1.7% to 10% per myr. We do not place much faith in exact numbers, but whatever the interpretation, *Drosophila* evolve rapidly with respect to nucleotide substitutions (and probably insertions/deletions, see below). Rate estimates for Hawaiian *Drosophila* are similarly very rapid (HUNT, HALL and BRITTEN 1981).

The rate of change in scnDNA is not well correlated with the rate of accumulation of chromosomal rearrangements (compare Figures 1 and 4). For example MEL and SIM-SEC-MAU separated about the same time as did YAK and TES; the former have only one inversion separating them while the latter have 16. We have no explanation but only point out that the two processes appear to be uncoupled. Even within this rather limited subgroup of sibling species, lineages vary greatly with respect to rate of accumulation of chromosomal rearrangements.

Two component genomes: The rapid decrease in NPR in Figure 3 indicates the *Drosophila* nuclear genome may be composed of (at least) two components with respect to rate of DNA evolution. For example, comparing MEL and SIM, there is a ΔTm of a little over 2°, indicating that portion of the genome which hybridized had less than 4% bp mismatch. Yet about 25% of the genome did not hybridize under the conditions requiring about 75% or greater match. This indicates that there are parts of the genome which are accumulating changes very rapidly and parts which remain relatively conserved. This phenomenon has been observed by others for *Drosophila* (HUNT, HALL and BRITTEN 1981; ZWIEBEL *et al.* 1982; SHULTZE and LEE 1986) and cave crickets (CACCONE and POWELL 1987). It also occurs between genomes within a species of *Drosophila* (CACCONE, AMATO and POWELL 1987).

A control that this drop-off of NPR is not just an artifact of the technique comes from the mtDNA data. As mentioned in the RESULTS, this rapid decrease in NPR does not occur; NPR for mtDNA averaged over 90% and was not detectably correlated with ΔTm. This is expected as there is no indication from sequence data that mtDNA has two discrete components with respect to evolutionary change, nor is there evidence for large insertion/deletion changes (*e.g.*, WOLSTENHOLME and CLARY 1985). The exception to this is the A + T-rich control region where changes in length are common among *Drosophila* (*e.g.*,

SOLIGNAC, MONNEROT and MOUNOLOU 1986). This region could account for the fact that we did not observe 100% NPR for the mtDNA, although we attempted to minimize the effect of this region by labeling with CTP.

Is the highly variable fraction of scnDNA just a tail of a continuous distribution or is it a distinct component? This can be addressed by considering the shape of DNA melting curves. The cumulative curve, as generated by the TEACL method, is S-shaped. If one plots the data as the additional DNA made single-stranded at each temperature, it very nearly approaches a normal distribution. Thus the mean and mode are very nearly equal. Therefore, the melting of the slow fraction (that which hybridized) gives little indication of a skewed distribution and thus the fast evolving fraction must really be a distinct component. Further indication that the *Drosophila* genome is discretely heterogenous with respect to rates of divergence comes from sequence data of the glue gene cluster; the transition between a conserved region and a highly divergent region can occur over less than 50 bp (MARTIN and MEYEROWITZ 1986).

This rapid drop-off in proportion of the genome which will cross-hybridize may be due to two, not mutually exclusive, processes. First, there may be a portion of the genome which accumulates point mutations very rapidly and quickly crosses the 25% divergence cut-off. Second, there may be many insertions/deletions accumulating in the highly variable fraction. While this latter process may account for much of the intraspecific divergence observed by CACCONE, AMATO and POWELL (1987), it seems unlikely to account for all the reduced NPR in interspecific comparisons. Because the tracer is made single-copy, the effects of middle repetitive sequences such as transposons will be suppressed. It is difficult to believe that more than 60% of the single-copy genome could have changed via insertions/deletions between ORE and its closely related relatives in the *melanogaster* subgroup (Figure 4). Thus, it must be a combination of single base substitutions, insertions/deletions, and perhaps some as yet unidentified mutational process which contributes to the rapid rate of divergence of part of the *Drosophila* genome.

The relative proportions of the genome evolving rapidly and slowly varies from taxon to taxon (Figure 3). *Drosophila* appear to have the largest portion rapidly evolving; *i.e.*, for a given degree of divergence as indicated by ΔTm, there is a greater proportion of the genome which does not hybridize. Cave crickets appear intermediate. Birds, as exemplified by herons, do not exhibit the rapid drop-off of NPR. In fact, it is probably unnecessary to invoke two discrete components to explain the bird data. Whether this is due to genome structure and organization differences or
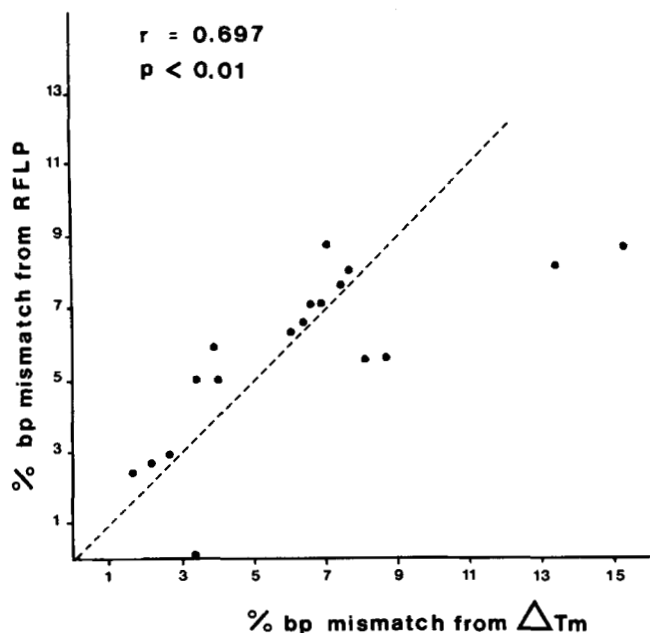
r = 0.697
p < 0.01

% bp mismatch from RFLP

% bp mismatch from $\triangle$Tm

FIGURE 5.—The relationship between mtDNA nucleotide substitution estimates based on restriction site variation (RFLP) and $\triangle$Tm. The dotted line has a slope of one, *i.e.*, the expected relationship if the two methods give the same result. The correlation coefficient, *r*, is significant. RFLP data are from SOLIGNAC, MONNEROT and MOUNOLOU (1986).

to differences in mutational mechanism is not clear. However, it does seem clear that not only do rates of DNA evolution vary greatly among taxa (BRITTEN 1986), the actual mechanisms of change may also be very different.

RFLP *vs* Tm: SOLIGNAC, MONNEROT and MOUNO-LOU (1986) studied these same species for mtDNA divergence as revealed by restriction fragment length polymorphism (RFLP). They used four enzymes recognizing four-base sequences and nine enzymes recognizing six-base sequences. An average of 40 cleavage sites per species was studied. Using the method of NEI and LI (1979) and NEI and TAJIMA (1981) they estimated the percent bp differences among species. It is of interest to compare their estimates based on RFLP to those obtained here based on $\triangle$Tm. Figure 5 plots the relationship. At low levels of divergence, the RFLP data tend to overestimate percent nucleotide substitutions while at the higher end, they underestimate it. Alternatively, one could argue $\triangle$Tm is inaccurate in the complimentary manner. However, our empirical results on DNA of known sequence indicate a very tight and accurate relationship between $\triangle$Tm and percent bp mismatch (CAC-CONE, DESALLE and POWELL 1988); the correlation coefficient is greater than 0.98. Thus we would argue that the inaccuracy lies in the estimates from RFLP data. The basis of this discrepancy between RFLP and $\triangle$Tm estimates of nucleotide divergence is not clear.

Given this less than perfect correlation, it is not surprising that the phylogenetic relationship derived by SOLIGNAC, MONNEROT and MOUNOLOU does not agree entirely with our mtDNA phylogeny. There are two major differences. First, they were unable to distinguish YAK and TES mtDNAs while we obtained a $\triangle$Tm of $1.89° \pm 0.17$ between these two species. By our calculations this means the species differ by a little over 3%. Assuming a random distribution of substitutions and a random distribution of restriction sites, the probability that a sample of 40 sites (about 200 bases) would miss a 3% difference is about 0.002. Thus it seems likely that the substitutions are not randomly distributed with respect to restriction sites. A second difference between the phylogenies concerns the placement of ERE; they place ERE with ORE as a monophyletic group while we place ERE with TES and YAK. As discussed above, all other data indicate ERE is in a phylad with TES and YAK and that ORE is the most distant.

LITERATURE CITED

BENVENISTE, R. E., 1985 The contribution of retroviruses to the study of mammalian evolution. pp 359–417. In: *Molecular Evolutionary Genetics*, Edited by R. J. MACINTYRE. Plenum Press, New York.

BRITTEN, R. J., 1986 Rates of DNA sequence evolution differ between taxonomic groups. Science **231:** 1393–1398.

BRITTEN, R. J., A. CETTA and E. H. DAVIDSON, 1978 The single-copy DNA sequence polymorphism of the sea urchin *Strongylocentrotus purpuratus*. Cell **14:** 1175–1186.

BROWN, W. M., 1983 Evolution of animal mitochondrial DNA, pp 62–88. In: *Evolution of Genes and Proteins*, Edited by M. NEI and R. K. KOEHN. Sinauer Press, Sunderland, Mass.

BROWN, W. M., 1985 The mitochondrial genome of animals, pp 95–130. In: *Molecular Evolutionary Genetics*, Edited by R. J. MACINTYRE. Plenum Press, New York.

CACCONE, A., and J. R. POWELL, 1987 Molecular evolutionary divergence among North American cave crickets. II. DNA-DNA hybridization. Evolution **41:** 1215–1238.

CACCONE, A., G. D. AMATO and J. R. POWELL, 1987 Intraspecific DNA divergence in *Drosophila:* a study on parthenogenetic *D. mercatorum*. Mol. Biol. Evol. **4:** 343–350.

CACCONE, A., R. DESALLE and J. R. POWELL, 1988 Calibration of the change in thermal stability of DNA duplexes and degree of base-pair mismatch. J. Mol. Evol. In press.

CAVALLI-SFORZA, L., and A. W. F. EDWARDS, 1967 Phylogenetic analysis: models and estimation procedures. Evolution **21:** 550–570.

COYNE, J., and M. KREITMAN, 1986 Evolutionary genetics of two sibling species *Drosophila simulans* and *D. sechellia*. Evolution **40:** 673–691.

FELSENSTEIN, J., 1985 Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:** 783–791.

FITCH, W. M., and E. MARGOLIASH, 1967 Construction of phylogenetic trees. Science **155**: 279–284.

HALL, T. J., J. W. GRULA, E. H. DAVIDSON and R. J. BRITTEN, 1980 Evolution of sea urchin non-repetitive DNA. J. Mol. Evol. **16**: 95–110.

HUNT, J. A., T. J. HALL and R. J. BRITTEN, 1981 Evolutionary distance in Hawaiian *Drosophila* measured by DNA reassociation. J. Mol. Evol. **17**: 361–367.

JUKES, T. H., and C. CANTOR, 1969 Evolution of protein molecules. pp. 21–123. In: *Mammalian Protein Metabolism*, Edited by H. N. MUNRO. Academic Press, New York.

LACHAISE, D., M.-L. CARIOU, J. R. DAVID, F. LEMEUNIER and M. ASHBURNER, 1988 Historical biogeorgraphy of the *D. melanogaster* species subgroup. Evol. Biol. **22**: In press.

LEMEUNIER, F., and M. ASHBURNER, 1984 Relationship within the *melanogaster* species subgroup of the genus *Drosophila* (*Sophophora*). IV. The chromosomes of tow new species. Chromosoma **89**: 343–351.

LEMEUNIER, F., J. R. DAVID, L. TSACAS and M. ASHBURNER, 1986 The *melanogaster* species group. pp 147–256. In: *The Genetics and Biology of Drosophila* (Vol. 3e), Edited by M. ASHBURNER, H. L. CARSON and J. N. THOMPSON. Academic Press, New York.

LI, W.-H., C.-C. LUO and C-I. WU, 1985 Evolution of DNA sequences. pp 1–94. In: *Molecular Evolutionary Genetics*, Edited by R. J. MACINTYRE. Plenum Press, New York.

MARTIN, C. H., and E. M. MEYEROWITZ, 1986 Characterization of the boundaries between adjacent rapidly and slowly evolving genomic regions in *Drosophila*. Proc. Natl. Acad. Sci. USA **83**: 8654–8658.

NEI, M., and W-H. LI, 1979 Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA **76**: 5269–5273.

NEI, M., and F. TAJIMA, 1981 DNA polymorphism detectable by restriction endonucleases. Genetics **97**: 145–163.

POWELL, J. R., A. CACCONE, G. D. AMATO and C. YOON, 1986 Rates of nucleotide substitution in *Drosophila* mitochondrial DNA and nuclear DNA are similar. Proc. Natl. Acad. Sci. USA **83**: 9090–9093.

SCHULZE, D. H., and C. S. LEE, 1986 DNA sequence comparison among closely related *Drosophila* species in the *mulleri* complex. Genetics **113**: 287–303.

SIBLEY, C. G., and J. E. AHLQUIST, 1984 The phylogeny of the hominoid primates as indicated by DNA-DNA hybridization. J. Mol. Evol. **20**: 2–15.

SOKAL, R. R., and C. D. MICHENER, 1958 A statistical method for evaluating systematic relationships. Univ. Kansas Sci. Bull. **38**: 1409–1438.

SOLIGNAC, M., M. MONNEROT and J.-C. MOUNOLOU, 1986 Mitochondrial DNA evolution in the *melanogaster* species subgroup of *Drosophila*. J. Mol. Evol. **23**: 31–40.

WOLSTENHOLME, D. R., and D. O. CLARY, 1985 Sequence evolution of *Drosophila* mitochondrial DNA. Genetics **109**: 725–744.

ZWIEBEL, L. J., V. H. COHN, D. R. WRIGHT and G. P. MOORE, 1982 Evolution of single-copy DNA and the ADH gene in seven Drosophilids. J. Mol. Evol. **19**: 62–71.

Communicating editor: C. C. LAURIE

APPENDIX

Tables 6–17 present more complete data. The left column abbreviations are as follows, n, the number of replicates; tm, the uncorrected median melting temperature; SE, the standard error of tm; Tr. length, the tracer length in bp; T corr., the correction for tracer length; Tm, the corrected median melting temperature; % Reassoc., the percent of reassociation; NPR, the normalized percent reassociation taking the homoduplex to be 100%.

TABLE 6

scnDNA

Tracer: Driver:	MEL MEL	MAU	SEC	SIM	YAK	ERE	ORE	TES	TAK	MEL-A	MEL-B	MEL-C
n	5	4	5	5	5	5	5	5	4	4	4	4
tm	56.58	53.58	53.29	53.44	50.83	51.66	50.44	51.55	48.61	56.11	56.18	56.12
SE	0.05	0.12	0.09	0.08	0.15	0.07	0.06	0.16	0.15	0.03	0.07	0.07
Tr. length	315	233	185	198	169	189	192	189	121	257	257	257
T corr.	1.58	2.14	2.70	2.52	2.95	2.64	2.61	2.64	3.91	1.94	1.94	1.94
Tm	58.16	55.72	55.99	55.96	53.78	54.30	53.05	54.19	52.52	58.05	58.12	58.06
% Reassoc.	73.8	48.7	47.2	39.2	43.2	47.1	47.6	39.2	16.7	81.8	80.0	78.6
NPR	100	66.0	64.0	53.2	58.4	63.4	64.6	53.2	22.6	111.0	108.4	106.5

TABLE 7

scnDNA

Tracer: Driver:	SIM SIM	MEL	MAU	SEC	ORE	ERE	TES	SIM- D	SIM- E	SIM- F	MAU- G	MAU- H	MAU- I
n	4	3	4	4	3	3	3	4	4	4	4	4	4
tm	55.00	52.21	54.27	53.54	49.04	50.53	49.66	54.59	54.56	54.47	53.15	52.87	52.85
SE	0.07	0.27	0.14	0.08	0.18	0.14	0.26	0.10	0.08	0.19	0.09	0.04	0.06
Tr. length	292	194	285	219	196	296	184	252	257	250	168	164	163
T corr.	1.76	2.57	1.75	2.28	2.55	1.69	2.71	1.98	1.95	1.99	2.97	3.04	3.07
Tm	56.71	54.78	56.02	55.82	51.59	52.22	52.37	56.56	56.51	56.46	56.12	55.91	55.92
% Reassoc.	74.9	51.7	65.3	52.3	44.6	41.2	47.5	69.2	76.0	79.3	60.0	55.84	56.7
NPR	100	69.1	87.2	69.8	59.5	55.0	63.4	92.4	101.5	105.9	80.1	74.5	75.8

TABLE 8

scnDNA

Tracer: Driver:	YAK YAK	ERE	ORE	MEL	SIM	MAU	SEC	TAK
n	5	4	5	4	4	4	4	3
tm	56.62	52.82	50.28	51.21	51.10	51.05	51.18	48.25
SE	0.07	0.16	0.26	0.05	0.17	0.10	0.09	0.17
Tr. length	223	161	164	133	140	149	138	126
T corr.	2.24	3.10	3.05	3.74	3.36	3.35	3.63	3.96
Tm	58.86	55.92	53.33	54.95	54.46	54.40	54.81	52.21
% Reassoc.	74.2	34.0	28.7	35.7	31.2	33.3	33.4	20.8
NPR	100	45.8	38.7	48.1	42.0	45.0	45.0	28.0

TABLE 9

scnDNA

Tracer: Driver:	MAU MAU	SEC	SIM	MEL	ERE	TES
n	4	4	4	4	3	4
tm	55.49	54.68	54.26	52.51	50.66	50.50
SE	0.08	0.11	0.14	0.07	0.21	0.27
Tr. length	203	189	188	175	186	169
T corr.	2.46	2.64	2.66	2.85	2.68	2.95
Tm	57.95	57.32	56.92	55.36	53.34	53.45
% Reassoc.	77.6	63.3	60.5	60.5	42.8	33.6
NPR	100	81.6	77.9	77.9	55.1	43.3

TABLE 10

scnDNA

Tracer: Driver:	SEC SEC	MAU	SIM	MEL	ERE	TES	MAU-G	MAU-H	MAU-I
n	4	4	3	4	3	3	4	4	4
tm	54.52	53.55	53.05	51.04	50.12	49.89	53.43	53.44	53.66
SE	0.06	0.06	0.02	0.08	0.02	0.18	0.12	0.07	0.11
Tr. length	164	145	148	126	150	141	147	142	155
T corr.	3.04	3.44	3.37	3.96	3.34	3.55	3.39	3.52	3.23
Tm	57.56	56.99	56.42	55.00	53.46	53.44	56.82	56.96	56.89
% Reassoc.	74.7	51.6	41.4	49.1	20.1	35.9	59.1	56.9	50.1
NPR	100	69.1	55.5	65.8	26.9	48.1	79.1	76.2	67.1

TABLE 11

scnDNA

Tracer: Driver:	ORE ORE	ERE	TES	MAU	SEC	TES TES	ERE	YAK
n	4	4	4	3	3	4	4	4
tm	55.41	49.03	48.93	48.38	48.44	55.40	52.22	52.66
SE	0.11	0.17	0.12	0.14	0.13	0.05	0.09	0.06
Tr. length	254	168	170	134	136	266	215	207
T corr.	1.97	2.97	2.95	3.72	3.67	1.88	2.32	2.42
Tm	57.38	52.00	51.88	52.10	52.11	57.28	54.54	55.08
% Reassoc.	80.4	46.3	33.8	31.5	44.0	82.0	68.7	68.7
NPR	100	57.6	42.1	39.2	54.7	100	83.8	83.8

TABLE 12

scnDNA

Tracer: Driver:	ERE ERE	YAK	MAU	ORE
n	3	3	3	3
tm	56.67	53.28	51.85	51.56
SE	0.15	0.09	0.08	0.18
Tr. length	198	156	203	226
T corr.	2.52	3.19	2.46	2.21
Tm	59.19	56.47	54.31	53.77
%	72.4	53.8	54.0	47.7
Reassoc.	100	74.3	74.6	66.0
NPR				

TABLE 13

scnDNA

Tracer: Driver:	TAK TAK	ERE	YAK	TES	ORE	MEL	SIM	SEC	MAU
n	4	4	3	4	4	3	4	4	4
tm	55.23	47.81	49.41	48.41	47.81	48.52	47.65	48.21	48.14
SE	0.08	0.12	0.11	0.06	0.19	0.11	0.10	0.08	0.05
Tr. length	180	131	209	143	126	171	143	159	152
T corr.	2.77	3.81	2.39	3.50	3.98	2.92	3.50	3.15	3.28
Tm	58.00	51.62	51.80	51.91	51.79	51.44	51.15	51.36	51.42
% Reassoc.	83.0	28.0	27.5	33.2	19.4	35.2	25.0	33.9	34.8
NPR	100	33.7	33.1	40.0	23.4	42.4	30.1	40.9	41.9

TABLE 14

mtDNA

Tracer: Driver:	MEL MEL	SEC	MAU	SIM	YAK	TES	ERE
n	5	4	5	5	5	3	5
tm	55.47	53.13	52.94	53.75	51.31	51.56	51.29
SE	0.11	0.09	0.06	0.06	0.11	0.13	0.07
Tr. length	268	255	233	306	275	268	243
T corr.	1.87	1.96	2.15	1.64	1.82	1.87	2.06
Tm	57.34	55.09	55.09	55.39	53.13	53.43	53.35

TABLE 15

mtDNA

Tracer: Driver:	SIM SIM	SEC	MAU	MEL	ERE	ORE	TAK
n	3	4	5	4	4	4	4
tm	57.38	55.56	56.40	54.99	50.55	48.40	48.58
SE	0.07	0.11	0.09	0.14	0.15	0.13	0.13
Tr. length	275	211	270	207	111	230	162
T corr.	1.82	2.37	1.85	2.42	4.47	2.18	3.08
Tm	59.20	57.93	58.25	57.41	55.02	50.58	51.66

TABLE 16

mtDNA

Tracer: Driver:	YAK YAK	TES	ERE	MEL	SEC	MAU	SIM	ORE	TAK
n	5	5	5	5	5	4	5	5	5
tm	56.61	54.52	53.53	52.53	52.02	51.71	53.05	49.49	49.50
SE	0.04	0.12	0.09	0.05	0.12	0.12	0.14	0.11	0.17
Tr. length	439	374	649	314	357	334	455	725	683
T corr.	1.14	1.34	0.77	1.59	1.40	1.50	1.10	0.69	0.73
Tm	57.75	55.86	54.30	54.12	53.42	53.21	54.15	50.18	50.23

TABLE 17

mtDNA

Tracer: Driver:	MAU MAU	SEC	SIM	MEL	TES
n	4	4	4	4	4
tm	57.33	54.93	55.89	53.85	49.82
SE	0.13	0.12	0.06	0.10	0.07
Tr. length	421	247	288	249	133
T. corr	1.19	2.03	1.73	2.01	3.77
Tm	58.52	56.96	57.62	55.86	53.59