Molecular Population Genetics of mtDNA Size Variation in Crickets

David M. Rand[1] and Richard G. Harrison[2]

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

Manuscript received July 6, 1988
Accepted for publication November 21, 1988

ABSTRACT

Nucleotide sequence analysis of a region of cricket (*Gryllus firmus*) mtDNA showing discrete length variation revealed tandemly repeated sequences 220 base pairs (bp) in length. The repeats consist of 206 bp sequences bounded by the dyad symmetric sequence 5′GGGGGCATGCCCCC3′. The sequence data showed that mtDNA size variation in this species is due to variation in the number of copies of tandem repeats. Southern blot analysis was used to document the frequency of crickets heteroplasmic for two or more different-sized mtDNAs. In New England populations of *G. firmus* and a close relative *Gryllus pennsylvanicus* approximately 60% of the former and 45% of the latter were heteroplasmic. From densitometry of autoradiographs the frequencies of mtDNA size classes were determined for the population samples and are shown to be very different in the two species. However, in populations where hybridization between the two species has occurred, the frequencies of size classes and cytoplasmic genotypes in each species' distinct mtDNA lineage were shifted in a manner suggesting nuclear-cytoplasmic interactions. The data were applied to reported diversity indices and hierarchical statistics. The hierarchical statistics indicated that the greatest proportion of variation for mtDNA size was due to variation among individuals in their cytoplasmic genotypes (heteroplasmic or homoplasmic state). The diversity indices were used to estimate a per-generation mutation rate for size variants of 10^{-4}. The data are discussed in light of the relationship between genetic drift and mutation in maintaining variation for mtDNA size.

EMERGING from the many recent studies of animal mitochondrial DNA (mtDNA) are characteristics of this molecule which were not apparent from the initial studies in vertebrates (*e.g.*, UPHOLT and DAWID 1977; BROWN, GEORGE and WILSON 1979; AVISE *et al.* 1979; LANSMAN *et al.* 1983). The organization of the genome is very different in nematodes, insects and vertebrates (WOLSTENHOLME *et al.* 1987; CLARY and WOLSTENHOLME 1985; ANDERSON *et al.* 1981) and the rates of evolution (relative to the nuclear genome) can vary considerably among lineages (VAWTER and BROWN 1986; POWELL *et al.* 1986; CACCONE, AMATO and POWELL 1988). Moreover, while early restriction endonuclease surveys revealed extensive variation in mtDNA sequences among individuals, little or no variation was identified within the cells of individuals (heteroplasmy) (AVISE *et al.* 1979; LANSMAN *et al.* 1983; FERRIS *et al.* 1983). In recent years, however, heteroplasmy has been described (or inferred) in a variety of animals (SOLIGNAC, MONNEROT and MOUNOLOU 1983; BROWN and DESROSIERS 1983; MONNEROT, MOUNOLOU and SOLIGNAC 1984; HAUSWIRTH *et al.* 1984; HARRISON, RAND and WHEELER 1985; DENSMORE, WRIGHT and BROWN 1985; BERMINGHAM, LAMB and AVISE 1986; MORITZ and BROWN 1987; HALE and SINGH 1986; WALLIS 1987; BOURSOT, YONEKAWA and BONHOMME 1987; SNYDER *et al.* 1987). In the vast majority of these cases the mtDNA molecules comprising the mixed cytoplasmic population have differed in *size* rather than in restriction enzyme recognition sites. The proliferation of reports describing naturally occurring mtDNA size variation and heteroplasmy in lower animals now indicates that (with the exception of mammals) these types of genetic variation are not uncommon.

Three general types of size variation can be identified (reviewed in MORITZ, DOWLING and BROWN 1987): (1) variation in the number of nucleotides in a "homopolymer run" of the same nucleotide (BROWN and DESROSIERS 1983; HAUSWIRTH *et al.* 1984); (2) variation in the copy number of tandemly repeated sequences (DENSMORE, WRIGHT and BROWN 1985; SOLIGNAC, MONNEROT and MOUNOLOU 1986; SNYDER *et al.* 1987; this paper); and (3) tandem duplication or deletion of large (1–8 kb) regions of the genome (MORITZ and BROWN 1986, 1987; WALLIS 1987; BOURSOT, YONEKAWA and BONNEHOMME 1987). Yet the nature of mtDNA size variation shows no clear phylogenetic patterns (BROWN 1983, 1985). There is as much variation within nematodes or amphibians as in most other animals combined (MORITZ, DOWLING and BROWN 1987). A loose pattern does emerge,

[1] Present address: Museum of Comparative Zoology, 26 Oxford Street, Harvard University, Cambridge, Massachusetts 02138.
[2] Present address: Section of Ecology and Systematics, Corson Hall, Cornell University, Ithaca, New York 14853.

however, from physiological comparisons. Homeothermic animals generally have the smallest and least variable mitochondrial genomes while poikilotherms have slightly larger mtDNAs which are considerably more variable in size (WALLACE 1982; SEDEROFF 1984; LEWIN 1985; WALLIS 1987; MORITZ, DOWLING and BROWN 1987). Whether this reflects stronger selection for smaller mtDNAs in animals with higher metabolic rates needs to be tested rigorously.

While much remains to be learned about the mechanisms generating mtDNA length mutations, the existence of mtDNA size variation and heteroplasmy provide relatively simple systems in which to study genetic drift, mutation and selection. The work of BIRKY (1978, 1983), THRAILKILL *et al.* (1980), SOLIGNAC *et al.* (1984) and RAND and HARRISON (1986), have shown that drift is a fundamental aspect of the transmission genetics of mtDNA. Moreover, there is evidence from transmission studies (SOLIGNAC *et al.* 1984; RAND and HARRISON 1986; but see SOLIGNAC *et al.* 1987), population cage experiments (MACRAE and ANDERSON 1988) and from frequency distributions in natural populations (HALE and SINGH 1986) that smaller mtDNAs have a selective advantage over larger molecules. Although it may be premature to claim that selection for smaller mtDNAs is a general phenomenon, both drift and selection tend to reduce genetic variation. The presence of mtDNA size variation and heteroplasmy in many species indicates that the mutation rate for size variation is sufficiently high to overcome the effects of drift and selection. With the increased use of mtDNA as a marker for population and evolutionary studies, it is essential that we understand the genetics of this molecule to interpret properly its patterns of variation.

Crickets in the genus *Gryllus* provide a simple system in which to describe the population genetics of mtDNA size polymorphism as the structural basis of the variation is restricted to a specific region of the mitochondrial genome (HARRISON, RAND and WHEELER 1985; *cf.* CANN and WILSON 1983). The goals of this paper are to (1) characterize the molecular basis of mtDNA size variation in Gryllus, (2) describe the patterns of mtDNA size variation in New England populations of *Gryllus firmus* and *Gryllus pennsylvanicus*, and (3) use this system to make more general statements about the integration of genetic drift, mutation and selection in the maintenance of mtDNA size variation.

The first of these aims is addressed through sequence analysis of the size-variable region of two different-sized mtDNA molecules from two isofemale lines of *G. firmus*. The patterns of population variation are revealed through Southern blot analysis using a cloned probe to determine the size(s) of the mtDNA molecules within individuals. From these data the frequencies of homoplasmic and heteroplasmic individuals are estimated from a number of pure and hybrid populations of the two cricket species. Through densitometry of bands on autoradiographs, estimates of the frequencies of mtDNA size classes were determined for the samples of individuals.

The final objective is addressed through an analysis of diversity for mtDNA size using hierarchical statistics of BIRKY, MARUYAMA and FUERST (1983) and BIRKY, FUERST and MAYUYAMA (1989). The samples of crickets were selected such that a hierarchy of organizational levels was clearly represented in the data: individuals, populations, species and the total sample. Using an approach similar to that employed by LEWONTIN (1972) in his analysis of human populations, the diversity for mtDNA size can be apportioned within or among individuals, populations, and species. Length mutations will increase the diversity for mtDNA size; genetic drift (*e.g.*, vegetative segregation during germ cell divisions) will decrease diversity. The manner in which diversity is apportioned within or among levels of the hierarchy depends on the relative strengths of these two opposing forces. If the mutation rate is very high relative to drift, most of the diversity will be within individuals (heteroplasmy); if drift is strong relative to mutation, most of the diversity will be among populations. (The term mutation is used in the context of mtDNA size variation to indicate the change in size of a mtDNA molecule.)

Biology of *Gryllus*. The ranges of *G. pennsylvanicus* and *G. firmus* overlap in a zone of hybridization which extends from Connecticut to Virginia (HARRISON and ARNOLD 1982). In this paper only New England populations of these crickets are discussed. "Pure" populations of *G. firmus* are found along the coast of Connecticut while "pure" populations of *G. pennsylvanicus* are found inland. In central Connecticut various hybrid populations can be found, the nature of which depends on the ecological setting of the population (RAND and HARRISON 1989).

The two species show significant differences in several morphologic characters and in the frequencies of alleles at loci coding for soluble enzymes (esterase, peptidase-1, peptidase-3, phosphoglucose isomerase). An individual cricket, however, cannot generally be assigned to either species on the basis of a single electrophoretic or morphologic character (HARRISON 1979; HARRISON and ARNOLD 1982). The best single character with which to determine the species identity of an individual's maternal lineage is mtDNA. The pure forms of *G. pennsylvanicus* and *G. firmus* possess mtDNAs which can be distinguished by restriction analysis with the enzymes *Apa*I, *Hinc*II, *Hind*III and *Xba*I (HARRISON, RAND and WHEELER 1987). Using these enzymes the composite restriction pattern of *G.*

TABLE 1

Collecting localities sampled in the analysis of mtDNA size variation

Locality name (region)	Town/ Locality	Frequency of A mtDNA	Year
1. BRHCT (L)	Bristol	1.00	1984, 1986
2. CHTCT (U)	Chester	1.00	1984
3. DURCT (L)	Durham	1.00	1984
4. EHDCT (L)	East Haddam	1.00	1984
5. ERPCT (L)	New Haven/East Rock Park	1.00	1986
6. FLIME (U)	Flag Island, Harpswell, Maine	1.00	1984
7. GRICT (C)	Guilford/Grass Island	0.17	1984
8. GU2CT (C)	Guilford	0.10	1985
9. HDRCT (L)	Haddam	1.00	1985
10. HRPME (U)	Harpswell, Maine	1.00	1984
11. HUMCT (I)	Cornwall/Housatonic Meadows	1.00	1985
12. MAVMA (U)	Martha's Vinyard	0.00	1986
13. MFPCT (C)	Milford Point	0.00	1986
14. NCACT (I)	North Caanan	0.69	1986
15. NGRCT (I)	North Granby	1.00	1987
16. PRBCT (S)	Plainville	0.56	1984, 1986
17. PR2CT (L)	Prospect	1.00	1984
18. SAPCT (C)	Old Saybrook/Saybrook Point	0.00	1984
19. SEDCT (S)	Seymour	0.73	1984–1986
20. SERCT (L)	Seymour	1.00	1984–1986
21. SFACT (C)	Madison/Seafield Association	0.00	1984
22. SHACT (I)	Sharon	1.00	1985
23. STRMA (U)	Sturbridge, Massachusetts	1.00	1984
24. SXDCT (S)	Essex	0.28	1984–1986
25. SXPCT (L)	Essex	1.00	1984–1986
26. TYBCT (S)	Tylerville	0.75	1984
27. WIDCT (I)	Winchester	1.00	1986
28. WLFCT (L)	Wallingford	1.00	1984, 1986
29. WLRCT (S)	Wallingford	0.63	1984–1986
30. WOLCT (L)	Wolcott	1.00	1986

The Region refers to the four sampling regions to which localities were assigned for the analysis of the hybrid zone as described in MATERIALS AND METHODS [see also RAND and HARRISON (1989); I = Inland; L = Loam; S = Sand; C = Coastal; U = Unassigned]. Frequency of the A mtDNA serves as an indication of species composition of the population. The A mtDNA type is fixed or in high frequency in *G. pennsylvanicus* populations. Loam localities (all fixed for the A mtDNA type) are geographically within the hybrid zone.

pennsylvanicus is referred to as AAAA (hereafter the A mtDNA lineage). In *G. firmus* the corresponding composite restriction pattern is BBBB (hereafter the B mtDNA lineage). Rare mtDNA molecules have been identified which are a result of loss of restriction sites (BAAA, ABAA, ADAA, BCBB) (HARRISON, RAND and WHEELER 1987). As defined by the *Hin*dIII and *Xba*I digests, these mtDNA genotypes are clearly derived from either of the two major A or B lineages.

MATERIALS AND METHODS

Collecting: The crickets used in this study were collected from 30 different localities during the late summer and early fall of 1984–1986. Twenty-six of these population samples are from Connecticut, two from Maine and two from Massachusetts (Table 1). A five-letter code is used to name the collecting sites. The first three letters indicate the name of the town or locality and the last two letters indicate the state. The Connecticut collections were also used to document patterns of genetic variation across a hybrid zone between *G. pennsylvanicus* and *G. firmus*, and are grouped into regions based on the geographic location and ecology of the collecting locality (RAND and HARRISON 1989). The "Inland" sites are samples from northwestern Connecticut and in most cases represent "pure" *G. pennsylvanicus*. The "Coastal" sites are samples from along the coast of Connecticut and represent "pure" *G. firmus*. Populations from within the hybrid zone, where species identity is uncertain, were classified as in either the "Sand" sites or the "Loam" sites depending on the soil characteristics of the locality at which the crickets were collected.

While each of the Sand sites and Loam sites show evidence of hybridization and can be classified as hybrid populations, it is necessary to distinguish them for the following reason. At each of the Sand sites both the A (*G. pennsylvanicus*) and B (*G. firmus*) mtDNAs are present, whereas not a single B mtDNA has been found at any of the Loam sites (RAND and HARRISON 1989). It has been shown that this is due to the asymmetric outcome of reciprocal hybrid crosses between males and females of the two species (HARRISON 1983; RAND and HARRISON 1989). The Loam and Sand sites are analyzed separately in the current study as the history of hybridization and the genetic makeup of the two types of populations are clearly different. Some population samples were not assigned to the Inland, Coastal, Loam or Sand categories, and therefore the combined sample of all crickets is larger than that for the samples assigned to these four categories. The Maine populations are *G. pennsylvanicus* and the one popu-

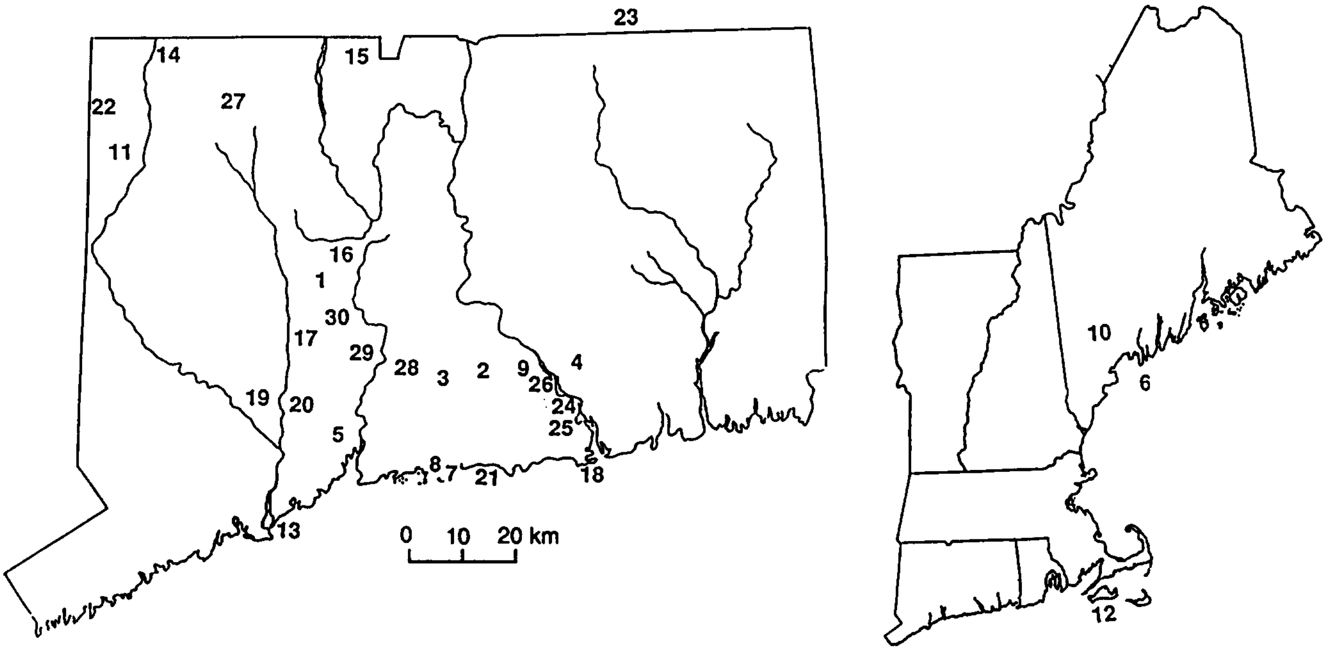

FIGURE 1.—Map of collecting localities listed in Table 1.

lation from Martha's Vineyard, Massachusetts, is *G. firmus* (Figure 1).

Cloning and sequencing: Pure mtDNA was isolated from isofemale lines of *G. firmus* as described previously (HARRISON, RAND and WHEELER 1985). It was determined by restriction analysis that two isofemale lines (hereafter female B and female D) contained mtDNA which differed in size by approximately 220 bp. When compared to restriction profiles of DNA from individuals whose mtDNA size had been determined previously (HARRISON, RAND and WHEELER 1985), females B and D possessed, respectively, mtDNA of the size classes "VS" and "S" (for "very small" and "small"; see also RAND and HARRISON 1986). The pure samples of female B and D mtDNA were digested with *Eco*RI and the ~3 kilobase (kb) size-variable bands were cloned (in separate reactions) into the sequencing vector pEMBL in *Escherichia coli* strain JM101. Restriction analysis of small scale preparations of putative recombinant plasmids containing female B or D mtDNA revealed insert bands which, as expected, differed in size by 220 bp. Southern blot analysis of the restricted plasmids using pure *G. firmus* mtDNA as a probe confirmed that the insert bands were mtDNA.

Nested deletions were carried out on CsCl purified large scale plasmid preparations following the techniques of HENIKOFF (1982). This generated a series of overlapping subclones which spanned the size-variable region of the *Eco*RI fragments cloned from females B and D (see Figure 2). The sequence of the size-variable region of both female B and D were determined in one direction using the chain termination technique (SANGER, NICKLEN and COULSON 1977). An individual repeat unit (cut out by *Bgl*II; see Figure 2) was subcloned into the *Bam*HI site of pEMBL and its sequence determined in both directions. The sequence of the individual repeat agreed with both sequences from females B and D.

Additional sequence data were obtained from an m13mp18 clone of the ~2-kb *Pvu*II-*Eco*RI fragment from female D mtDNA.

mtDNA analysis. Southern blot methods were used for all population samples following the general protocols of MANIATIS, FRITSCH and SAMBROOK (1982) and as described previously (HARRISON, RAND and WHEELER 1987). Total cricket DNA was digested with *Eco*RI (New England Biolabs), electrophoresed in horizontal 0.7% agarose gels and blotted to nitrocellulose for hybridization. The DNA used as a hybridization probe was the entire recombinant plasmid containing the 3-kb *Eco*RI fragment of *G. firmus* female D used for sequence analysis of the size variable region of cricket mtDNA. The size(s) of the mtDNA within individual crickets were determined in reference to cricket mtDNA of known size which was run as a standard on all gels. The size of this standard was determined by electrophoresing it next to the 3-kb fragment used for sequence analysis.

Densitometry: The frequencies of mtDNA size classes present within heteroplasmic individuals were estimated by densitometry of autoradiographs (see Figures 2 and 3 in RAND and HARRISON 1986). The frequency estimate of a given size class within an individual was based on the height of the densitometric peak for that size class relative to the sum of the peak heights for all size classes visible in that lane of the autoradiograph. All frequency estimates were rounded to the nearest 0.05 value. In a number of instances rare size variants were clearly visible on the autoradiographs, but produced barely perceptible peaks on the autoradiograph. These rare size classes were arbitrarily assigned a frequency of 0.05, a value "near a conservative lower limit of detectability . . ." on gels (AVISE and VRIJENHOEK 1987, p. 520). A few lanes of autoradiographs were overexposed and produced square peaks on the densitometer. Frequency estimates for such individuals were made by visual comparison to a series of autoradiographs of other individuals for which reliable densitometer tracings were obtained. Individuals which were visibly homoplasmic were not densitometer traced.

Size class frequencies for groups of individuals (populations, species, etc.) were calculated as the mean for each mtDNA size class across all individuals in the group. For example, a heteroplasmic individual with 60% small mtDNA and 40% large mtDNA would contribute the value 0.6 to the small size class frequency tabulation and 0.4 to the large size class, whereas a homoplasmic small mtDNA individual would contribute the value 1.0 to the small size class frequency tabulation. After all individuals in the sample were tabulated, the sums of the values for each size class were divided by the number of individuals in the sample (this is equivalent to a mean for each size class across all individuals in the sample).

Diversity indices and hierarchical statistics: BIRKY, MARUYAMA and FUERST (1983) proposed K indices to characterize the variation within and among samples of organelle genomes. These K values can be calculated as a standard measure of heterozygosity or "gene diversity": $K = 1 - \sum x_i^2$, where x_i is the frequency of the ith allele (=size class) in the "population." Considering different levels of organization as "populations," BIRKY, MARUYAMA and FUERST (1983) defined the following Ks:

K_a = diversity within a cell
K_b = diversity within an individual
K_c = diversity within a deme
K_d = diversity within a region (or array of demes).

In this study no estimate of mtDNA size class frequencies were obtained from single cells, hence K_a was not estimated directly. K_b was estimated from the mtDNA size class frequencies within each individual. The K values for higher levels of organization were calculated from the mean frequencies of size classes in the collective sample of individuals representing that level of organization.

More recently BIRKY, FUERST and MARUYAMA (1989) have defined K^* values which permit genes to be sampled from the same "population" as well as from two different "populations" (Equations 1 and 2 of BIRKY, FUERST and MARUYAMA 1989). To obtain K_c^* from our data, K_z was estimated using Equation 5 of BIRKY, FUERST and MARUYAMA (1989): $K_z \sim K_b/[(N_{eo} - 1)/(N_{eo} + 1)]$. With N_{eo} (the effective number of organelle genomes) estimated from 87 to 395 (RAND and HARRISON 1986), the denomenator in this equation ranges from 0.977 to 0.995; the value 0.985 was used.

It is apparent from Equations 1 and 2 of BIRKY, FUERST and MARUYAMA (1989) that K^* indices can be extended through any number of hierarchical levels. In the current study we consider four levels: individuals, populations, species, total. Therefore K_t of BIRKY, FUERST and MARUYAMA (1989, Equation 2) will be referred to in this paper as K_d^* to indicate the region (= intraspecific) level of diversity:

$$K_d^* = [\bar{K}_c^* + K_d(L - 1)]/L. \qquad (1)$$

An additional K^* index is proposed (K_t^*) allowing genes to be sampled from anywhere within a species, as well as from anywhere within any number of other species:

$$K_t^* = [\bar{K}_d^* + K_t(S - 1)]/S \qquad (2)$$

where $\bar{K}_d^*$ is the average K_d^* among S species, and $K_t = 1 - \sum x_i^2$ where x_i = frequency of the ith size class averaged across all individuals in the sample of S species.

For analysis of the hierarchical structure of mtDNA size variation a subset of the entire data set was used: four *G. pennsylvanicus* populations from the "Inland" region (HUMCT, NCACT, NGRCT, WIDCT) and four *G. firmus* populations, three of which are from the "Coastal" region (GRICT, MFPCT, SFACT) and one from Martha's Vineyard Island (MAVMA; see Tables 1 and 5). The data were selected in this manner to test whether there are differences between "pure" samples of the A (*G. pennsylvanicus*) and B (*G. firmus*) mtDNA lineages in the nature of mtDNA size variation. While the samples HUMCT, NGRCT and WIDCT were fixed for the A mtDNA type, five B mtDNA individuals were present in the NCACT sample. These five individuals were removed in the hierarchical analysis so that *G. pennsylvanicus* would be represented only by individuals with A mtDNA. Similarly, the samples MAVMA, MFPCT and SFACT were fixed for the B mtDNA type, but two individuals in the GRICT sample possessed A mtDNAs. These two individuals were removed in the analysis so that *G. firmus* would be represented only by individuals with B mtDNA.

Thus, using this subset of populations there are three hierarchical levels within each species for which mtDNA size class frequencies were obtained [indicated by the subscripts I = individuals, P = populations, L = mtDNA lineages (=species)]. To quantify population subdivision, mtDNA diversity was apportioned to within-individual, among-individual and among-population components (indicated by C with the appropriate subscript) in a manner analogous to that used by LEWONTIN (1972) (see also NEI 1973):

Within individuals

$$C_I = \bar{K}_b/K_d \qquad (3a)$$

Among individuals within populations

$$C_{IP} = (\bar{K}_c - \bar{K}_b)/K_d \qquad (3b)$$

Among individuals within lineages

$$C_{IL} = (K_d - \bar{K}_b)/K_d \qquad (3c)$$

Among populations within lineages

$$C_{PL} = (K_d - \bar{K}_c)/K_d \qquad (3d)$$

where $\bar{K}_c$ = the average K_c among the sampled populations of the species, $\bar{K}_b$ = the average K_b among all sampled individuals in the species and K_d is the probability that two mtDNA molecules drawn from two different populations in the species are of different size. The sum of C_I, C_{IP} and C_{PL} should, and did, equal to 1.0 in all calculations (C_{IL} is not included in this sum as it "skips" a level in the hierarchy and would therefore be redundant in accounting for the total diversity).

These C statistics are calculated separately for the two species from the four population samples selected to represent each species. In Table 6, Equations 3a–d will be referred to as C statistics based on the operation of "sampling among" since the diversity measures on which they are based are defined by the operation of drawing two "alleles" from *different* populations. A second set of intraspecific C statistics can be defined from the K^* diversity measures. These C^* statistics are calculated as in Equations 3a–d with the exception that K_c^* [from Equation 1 of BIRKY, FUERST and MARUYAMA (1989)] and K_d^* (from Equation 1, this paper) are put in place of K_c and K_d, respectively. In Table 6 the C^* statistics are referred to under "sampling among or within" as the diversity measures allow for the two sampled alleles to be derived from two different populations *or* from within the same population.

When the eight representative populations from the two species are considered together, a fourth level is added (indicated by the subscript T = total). These levels would correspond, respectively, to D = demes, R = regions, S = subdivisions and T = total in the traditional sense of hierarchical population analysis (WRIGHT 1978). With all eight populations from both species combined, there are seven C statistics which can be calculated across the four hierarchical levels:

Within individuals

$$C_I = \bar{K}_b/K_t \quad (4a)$$

Among individuals within populations

$$C_{IP} = (\bar{K}_c - \bar{K}_b)/K_t \quad (4b)$$

Among individuals within lineages

$$C_{IL} = (\bar{K}_d - \bar{K}_b)/K_t \quad (4c)$$

Among individuals within the total sample

$$C_{IT} = (K_t - \bar{K}_b)/K_t \quad (4d)$$

Among populations within lineages

$$C_{PL} = (\bar{K}_d - \bar{K}_c)/K_t \quad (4e)$$

Among populations within the total sample

$$C_{PT} = (K_t - \bar{K}_c)/K_t \quad (4f)$$

Among lineages within the total sample

$$C_{LT} = (K_t - \bar{K}_d)/K_t \quad (4g)$$

where $\bar{K}_b$, $\bar{K}_c$ and $\bar{K}_d$ are the average K_b, K_c and K_d among, respectively, all individuals, populations and lineages in the total sample of eight *G. pennsylvanicus* and *G. firmus* populations. C_I, C_{IP}, C_{PL} and C_{LT} should, and did, sum to 1.0 in all calculations (as with the three-level statistics, C_{IL}, C_{IT} and C_{PT} are not included in this calculation since they skip levels of the hierarchy and would be redundant in accounting for the total diversity).

A second set of four-level (interspecific) C statistics can be defined from the K^* diversity measures. These C^* statistics are calculated as in equations (4a–g) with the exception that K_c^* [from BIRKY, FUERST and MARUYAMA's (1989) Equation 1), K_d^* (from Equation 1, this paper) and K_t^* (from Equation 2, this paper) are put in place of K_c and K_d and K_t, respectively. These four-level C^* statistics are referred to as "sampling among or within" in Table 6 for the reasons discussed above.

When other population samples were selected to represent the *A* and *B* mtDNA lineages only slight quantitative differences in the K and C statistics were observed. The discussion of the results presented below would not be affected by the presentation of analyses derived from a different subset of the complete data set.

To approximate the error associated with the hierarchical statistics, a jackknife approach was used where an entire population sample was removed from *each* species. Means and standard deviations for the values in Equations 3 and 4 were obtained from four different jackknife runs each of which consisted of three populations from each species. For the three-level, intraspecific statistics (Equations 3a–d), three populations were used in each species' jackknife run. For the four-level, interspecific statistics (Equations 4a–g), six populations were used in each jackknife run (a combined sample of three populations from each species).

For comparison a statistic of heterogeneity is calculated following the methods used by DESALLE *et al.* (1987) (see also WEIR and COCKERHAM 1984). Frequencies of mtDNA size classes in populations were arcsin-square root transformed and tested for significant among-sample heterogeneity with the following statistic:

$$V = 4 \sum n_i(a_i - A)^2$$

where a_i = the transformed frequency in the ith sample (*i.e.*, population), $A = (\sum n_i a_i)/N$ and $N = \sum n_i$. V is determined by summing across r samples and is distributed as a χ^2 with $r - 1$ d.f. When significant among-sample variation is indicated, the variance can be further partitioned: the total variance $= rV/(4N(r-1))$, the between population variance (Wahlund variance) $= r(V - (r-1))/(4N(r-1))$ and the proportion of the total variance which is due to between population variation $= (V - (r-1))/V$.

RESULTS

Sequence analysis: The two isofemale lines of *G. firmus* used in this study possessed mtDNA of different sizes: female B mtDNA = 16.04 kb, female D mtDNA = 16.26 kb. The nucleotide sequence of the region containing discrete size variation was determined for both female B and D (Figures 2 and 3). The complete sequence of the repeat region of female B mtDNA (Figure 3) reveals two tandem 220-bp repeats, while the female D sequence has three repeats. The repeats contain the 14-bp sequence 5′GGGGGCATGCCCCC 3′ which demonstrates dyad symmetry. Three base pairs beyond this symmetric sequence is a *Bgl*II site. We will define a repeat as the 220-bp sequence running from the first G in the symmetric sequence to the T immediately preceeding the next symmetric sequence (Figure 3). Defined as such, the repeats are flanked by segments of themselves: 5′ to the first repeat is a 41-bp sequence (bp 114–154, Figure 3) corresponding to the *last* 41 bp in an individual repeat; 3′ to the last repeat is a 153-bp sequence (bp 595–747, Figure 3) identical to the *first* 153 bp of an individual repeat (hereafter the 5′ and 3′ "flanking segments"). Thus, 26 bp of a complete repeat is "missing" from the flanking segments.

In female B the two 220-bp repeats are identical with two exceptions: (1) the corresponding nucleotides at positions 293 and 513 are A and G, respectively, and (2) the corresponding nucleotides at positions 308 and 528 are G and C, respectively (Figure 3). Interestingly, the variable sites 308 and 528 lie in the position of the first base of the 26-bp sequence which is "missing" from the flanking segments. Female D possesses a third copy of the repeat and all three copies are identical. In addition to the extra repeat, the B and D sequences differ at three positions: (1) position 128 of the 5′ flanking segment (C in female B, T in female D); (2) position 293 in the first repeat (A in female B, G in female D); and (3) position 528 in the second repeat (C in female B, G in female D).

The repeats and their flanking segments have a base composition of 64% A+T. These sequences are situated in DNA that is 88% A+T on one side (positions 1–113) and 80% A+T on the other side (positions 748–892; Figure 3). Sequence data from the pEMBL deletions and the m13 clone extend into the small ribosomal RNA (srRNA) gene as evidenced by 100% similarity to positions 1757 to 1778 of CLARY and WOLSTENHOLME's (1987) Figure 2. This highly conserved stretch lies ~330 bp from the beginning of

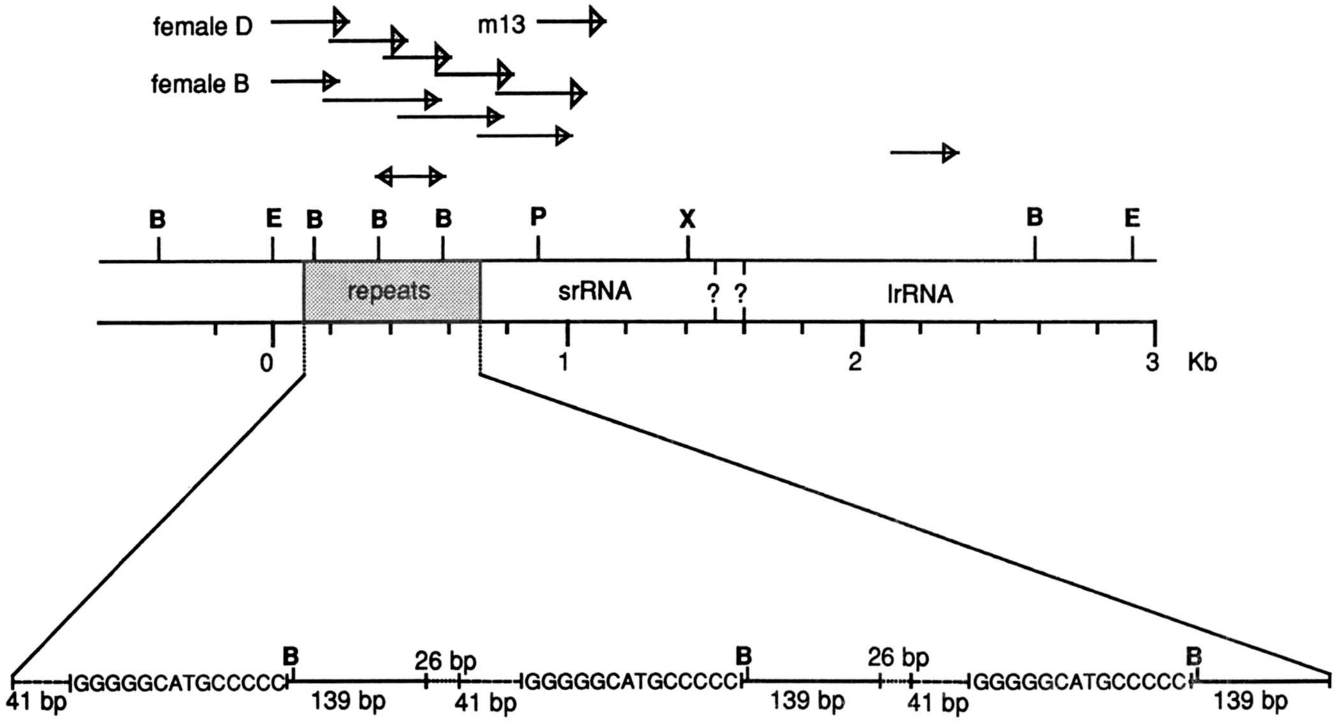

FIGURE 2.—Restriction map of the repeat region of female B. The sequence in Figure 3 runs from the *Eco*RI site to the *Pvu*II site. Repeat DNA is indicated by the shaded area which is expanded below. The arrows above the map indicate sequence from deletion subclones of female B and female D (the additional repeat in female D is not shown in the restriction map). The srRNA and lrRNA genes were identified as described in MATERIALS AND METHODS. The "?" indicates that the precise 5′ and 3′ boundaries of these genes have not been located. B = *Bgl*II, E = *Eco*RI, P = *Pvu*II and X = *Xba*I. Positions of restriction sites not identified by sequencing are from HARRISON, RAND and WHEELER (1987) and unpublished data. Scale below the map is in kilobases.

1 GAATTCATAA AAGATAATTT TTCTTTTTTT GTAAGAAAAA AAAAAAAAAG GAAAAATTAG

61 TAGTAATATT AATTTATATC AGTTTATTGA AAGTAATGTA AAATATTATA TAAATCGAAT

121 AATTTGGCTG GTTGTTCGAG CTTAAAGATT TGTTGGGGGC ATGCCCCCCA AGATCTTTTG

181 AACCGTACAA CAGTTAGGAA ATTTTAATTG AAGGATAAGA TTTGATTTCC TGGAGTATAT

241 TGTTGATTGA AATTTGAATA TAATATTGAT TGATTGATCC ATGGGTCGTG ATATGAAAAG

301 TAAAGTTGTG TTGTAAGGGA GGTAATTGAA GTGATCGAAT AATTTGGCTG GTTGTTCGAG

361 CTTAAAGATT TGTTGGGGGC ATGCCCCCCA AGATCTTTTG AACCGTACAA CAGTTAGGAA

421 ATTTTAATTG AAGGATAAGA TTTGATTTCC TGGAGTATAT TGTTGATTGA AATTTGAATA

481 TAATATTGAT TGATTGATCC ATGGGTCGTG ATGTGAAAAG TAAAGTTCTG TTGTAAGGGA

541 GGTAATTGAA GTGATCGAAT AATTTGGCTG GTTGTTCGAG CTTAAAGATT TGTTGGGGGC

601 ATGCCCCCCA AGATCTTTTG AACCGTACAA CAGTTAGGAA ATTTTAATTG AAGGATAAGA

661 TTTGATTTCC TGGAGTATAT TGTTGATTGA AATTTGAATA TAATATTGAT TGATTGATCC

721 ATGGGTCGTG ATGTGAAAAG TAAAGTTTTA TTCTTGCTTT TTTATTTGCG TGGAGAATGA

781 TTTACATATA TTTATTATCT AAAAGATGTA TAATGTAGTA ATTAATATTA TACAATTAAG

841 TTGATTTGGA TATAGTATTT CTTATAGTAT TGGTTAAATG CGTGCCCAGC TG

FIGURE 3.—Nucleotide sequence of the repeat region in female B. The sequence is numbered starting with the first base in the 5′ *Eco*RI cloning site. The symmetric sequences bounding the repeats are boxed and the *Bgl*II sites 3′ to the symmetric sequences are underlined. The boundaries between "A+T-rich" DNA and the 5′ 41-bp and 3′ 139-bp flanking segments of Bgl repeat DNA are marked by vertical lines. Single base differences between the two repeats in female B are boxed. The sequence of the above region was determined for female D as well which possesses a third copy of the 220-bp *Bgl* repeat (and a T in place of C at position 128).

the srRNA gene in Drosophila (there are additional short stretches of lower sequence similarity between the cricket sequence and positions 1430–1730 of CLARY and WOLSTENHOLME's (1987) (Figure 2). If the srRNA gene of cricket mtDNA is of similar length, the beginning of this gene would fall near position

720 of the cricket sequence (Figure 3, this paper). Approximately 1 kb beyond the srRNA sequences, cricket mtDNA shows 75% similarity to the large ribosomal RNA (lrRNA) gene of *Drosophila yakuba* and the mosquito *Aedes albopictus* (Figure 2). If the mitochondrial gene organization is the same in crickets and flies, the repeat region lies in a position corresponding to the A+T rich region of Drosophila which is known to contain the origin of replication.

Partial digest analysis of mtDNA size classes: In the sample of 319 crickets reported here, seven different-sized mitochondrial genomes were detected. In increasing size these mtDNAs are referred to as T, VS, S, M, L, VL and X (indicating the so-called "tiny," "very small," "small," "medium," "large," "very large" and "extra large" size classes). The sequence data derive from mtDNAs of the two next-smallest size classes (female B = VS, female D = S). To determine whether larger mtDNAs possess additional copies of the repeats, a partial restriction digest experiment was conducted. Total DNA from crickets whose mtDNA had been determined previously as S and L (16.26 and 16.70 kb) were digested incompletely with *Bgl*II. The sequence data show that an S mtDNA has three full repeats. This indicates that a partial digest with *Bgl*II would produce three-rung "ladders" built off both of the adjacent *Bgl*II fragments as well as a ladder of repeats themselves (see Figure 2). The prediction for an L mtDNA is a five-rung ladder of partially digested repeat fragments. The autoradiograph obtained from this partial digest experiment is consistent with the predictions (Figure 4). Thus copy number of repeats varies from one to seven in this sample of crickets.

From restriction analyses, mtDNA size variation in *G. pennsylvanicus* is indistinguishable from that in *G. firmus* (Harrison, Rand and Wheeler 1985). Sequence analysis of the repeat region in mtDNA of other species in the genus *Gryllus* is currently in progress. For the analyses below it will be assumed that mtDNA molecules of different size have different numbers of a tandemly repeated sequence. These molecules will be referred to as "size classes" and will be treated as "alleles."

Frequencies of heteroplasmic and homoplasmic genotypes: Of the 319 crickets sampled, 46.1% of the individuals were heteroplasmic (Table 2). When comparing pure populations, heteroplasmy is less frequent (but not significantly so) in the A mtDNA lineage than in the B lineage. While there appear to be a number of other differences between pure populations of the two species (*i.e.*, A and B mtDNA lineages) in the frequencies of heteroplasmic and homoplasmic genotypes (Table 2), the only significant difference in the current sample is in the frequencies of homoplasmic

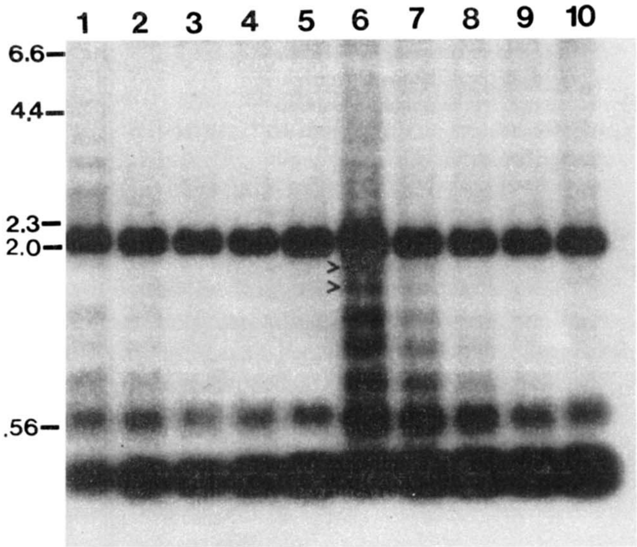

FIGURE 4.—*Bgl*II partial digest experiment. Two samples of total DNA known to contain S and L mtDNA were digested separately with *Bgl*II. At 10-min intervals aliquots were removed from the digest reaction. The aliquots were electrophoresed, blotted to nitrocellulose and probed within the 220-bp *Bgl*II fragment of *G. firmus* mtDNA. Because the *Bgl*II fragments 5′ and 3′ to the repeats contain segments of repeat DNA, the autoradiograph of this filter should reveal three different "ladders": one built off the 2.0-kb *Bgl*II fragment 3′ to the repeats, one built off the 560-bp *Bgl*II fragment 5′ to the repeats and a ladder of *Bgl* repeats themselves. In each case the "rungs" of the ladder should be at 220 bp intervals. Lanes 1–5: 10-, 20-, 30-, 40- and 50-min aliquots from the *Bgl*II digest of total cricket DNA containing S mtDNA (3 repeats). Lanes 6–10: the same time point aliquots removed from the *Bgl*II digest of total cricket DNA containing L mtDNA (which should have 5 repeats). As predicted there are two extra bands in lane 6 (arrowheads).

M genotypes [$G = 7.934$, d.f. = 1, $P < 0.01$, G test, Sokal and Rohlf (1981) p. 737].

The data from hybrid populations indicate that hybridization does affect mtDNA genotype frequencies, but does so to a greater extent in the A lineage. There is a slight (but nonsignificant) decrease in the frequency of heteroplasmic individuals in hybrid *vs.* pure populations. This decrease is greater in the A lineage than in the B lineage such that, in hybrid populations, the difference between the A and B lineages in the frequency of heteroplasmy is significant ($G = 4.572$, d.f. = 1, $P < 0.05$). This is most evident in the frequency of the M/S genotype: in moving from pure to hybrid populations, the frequency of M/S decreases significantly in the A lineage ($G = 4.200$, d.f. = 1, $P < 0.05$) but increases nonsignificantly in the B lineage. These shifts increase the differences between the A and B lineages in the frequency of the M/S genotype [A (Loam) *vs.* B (Sand): $G = 8.933$, d.f. = 1, $P < 0.005$]. A comparable pattern is observed in the frequency of the homoplasmic M genotype. While homoplasmic M individuals are more frequent in hybrid than in pure populations, the increase is significant in the A lineage [A (Pure) *vs.* A (Loam): $G =$

TABLE 2

mtDNA genotype frequencies in the total sample of crickets and in the A and B lineages from pure and hybrid populations

		"Pure"		"Hybrid"		
Genotype (*n*)	Total (319)	A (*G pennsylvanicus*) (53)	B (*G. firmus*) (52)	A (loam) (92)	A (sand) (49)	B (sand) (48)
S	12.6	15.1	23.1	3.3	4.1	16.7
S/VS	0.3	0.0	1.9	0.0	0.0	0.0
S/VS/T	0.3	0.0	1.9	0.0	0.0	0.0
M	40.4	39.6	15.4	58.7	59.2	25.0
M/S	25.4	28.3	38.5	14.1	16.3	43.8
M/S/VS	1.9	0.0	1.9	3.3	2.0	0.0
M/VS	0.6	1.9	1.9	0.0	0.0	0.0
L	0.6	0.0	0.0	1.1	0.0	0.0
L/M	8.5	5.7	5.8	8.7	14.3	6.2
L/M/S	4.8	5.7	3.9	4.4	4.1	8.3
L/S	1.3	0.0	1.9	2.1	0.0	0.0
VL	0.3	0.0	0.0	1.1	0.0	0.0
VL/L	0.6	0.0	0.0	2.1	0.0	0.0
VL/L/M	0.6	1.9	0.0	1.1	0.0	0.0
VL/M	0.9	1.9	1.9	0.0	0.0	0.0
VL/M/S	0.3	0.0	1.9	0.0	0.0	0.0
X/L/M	0.3	0.0	0.0	0.0	0.0	0.0
X/VL/M	0.3	0.0	0.0	0.0	0.0	0.0
f (heteroplasmy)	46.1	45.3	61.5	36.9	36.7	58.3

Sample sizes of the A and B lineages do not add up to the total sample as there were some individuals not scored for the A or B composite genotype. Single letters represent homoplasmic genotypes, letters separated by a "/" indicate heteroplasmic genotypes. See text for description of size classes.

4.924, $P < 0.05$] but not significant in the B lineage. The difference in the frequencies of the M genotype between the A and B lineages in hybrid populations is very significant [A (Loam) *vs.* B (Sand): $G = 14.897$, d.f. = 1, $P < 0.001$].

The distributions of frequencies of the M size class are presented in Table 3. The distribution of the total sample is generally U-shaped but skewed toward frequencies of 1.0 (homoplasmic for *M*). The distributions of the pure A and pure B samples are skewed towards opposite ends while the distributions of hybrid A and hybrid B are skewed in the same direction. There is a slight indication of an excess of heteroplasmic individuals in the middle frequencies (0.4–0.6), this being most evident in the hybrid B sample.

Frequencies of size classes: The data presented in Tables 4 and 5, and shown graphically in Figure 5, are mean frequencies of the seven mtDNA size classes among individuals in various subdivisions of the complete data set. The frequencies of size classes in the A and B lineages from pure populations are significantly different [Figure 5A; 2 × 4 contingency test with the rare size classes (T, VS, VL X) lumped: $G = 14.894$, d.f. = 3, $P < 0.005$; SOKAL and ROHLF (1981) p. 745]. From jackknife analysis of individuals within population samples standard deviations of the frequency estimates range from less than 1% to 7.6% and generally are in the range of a few percent (data not shown).

TABLE 3

Distributions of the M size class frequencies within individuals

	Frequency counts				
Frequency of M	Total sample	Pure A	Pure B	Hybrid A	Hybrid B
0.00	49	8	15	10	6
0.05–0.15	11	0	4	4	1
0.20–0.25	4	0	3	1	0
0.30–0.35	12	1	1	4	5
0.40–0.45	13	2	4	3	4
0.50–0.55	16	1	4	3	7
0.60–0.65	16	0	5	7	4
0.70–0.75	9	1	1	2	3
0.80–0.85	8	3	0	4	1
0.90–0.95	51	16	7	20	4
1.00	130	21	8	83	13

Frequency counts are grouped in the same manner as in Table 2 with the exception that hybrid A categories (Loam A and Sand A) are pooled. Estimates of frequencies from densitometric scans were rounded to the nearest 0.05. Note that the frequency classes 0.0 and 1.0 include one value, the class 0.05–0.15 includes three values, and all other classes include two values.

The frequencies of size classes in hybrid populations are presented in Figure 5B. As described in the MATERIALS AND METHODS, three hybrid categories are defined on the basis of the population's locality and mtDNA composition: Loam populations which possess only A mtDNA, and Sand populations which are polymorphic for mtDNA type and thus can be divided

TABLE 4

Frequencies of mtDNA size classes and diversity indices in pooled samples

Level (n)	f(VS)	f(S)	f(M)	f(L)	f(VL)	f(X)	K_c	$\bar{K}_b$	G_{IP}
Total (319)	0.002	0.213	0.686	0.067	0.012	0.000	0.4706	0.1486	0.6842
AmtDNA (209)	0.002	0.124	0.775	0.083	0.015	0.001	0.3774	0.1152	0.6948
BmtDNA (106)	0.001	0.451	0.504	0.037	0.007	0.000	0.5423	0.2155	0.6026
Inland (50)	0.001	0.209	0.763	0.024	0.003	0.000	0.3735	0.1088	0.7087
Loam (92)	0.002	0.074	0.813	0.083	0.028	0.000	0.3258	0.0987	0.6971
Sand (100)	0.002	0.208	0.720	0.069	0.000	0.000	0.4326	0.1869	0.5679
Coastal (40)	0.006	0.480	0.459	0.035	0.019	0.000	0.5575	0.2367	0.5745
Unass'd (25)	0.000	0.336	0.468	0.170	0.016	0.010	0.6388	0.1390	0.7824
A Inland (44)	0.001	0.158	0.818	0.019	0.004	0.000	0.3052	0.0997	0.6733
B Inland (5)	0.000	0.700	0.280	0.020	0.000	0.000	0.4312	0.1360	0.6846
A Sand (49)	0.004	0.083	0.821	0.092	0.000	0.000	0.3099	0.1390	0.5515
B Sand (48)	0.000	0.349	0.601	0.050	0.000	0.000	0.5144	0.2455	0.5227
A Coastal (2)	0.025	0.900	0.075	0.000	0.000	0.000	0.1837	0.1675	0.0882
B Coastal (38)	0.005	0.458	0.479	0.037	0.020	0.000	0.5591	0.2404	0.5700
A Unas'd (22)	0.000	0.291	0.486	0.193	0.018	0.012	0.6410	0.1579	0.7537
B Unass'd (3)	0.000	0.667	0.333	0.000	0.000	0.000	0.4444	0.0000	1.0000

Level indicates the level of grouping for the pooled samples (*e.g.*, AmtDNA is the A mtDNA lineage; Inland, Loam, Sand, and Coastal are the four regions described in MATERIALS AND METHODS; Unass'd are crickets not assigned to these four regions; A Inland, B Inland, etc., indicate the sample of crickets separated by the A or B mtDNA type). n = sample size, f(VS) = frequency of the VS size class, K_c = diversity measure for the combined sample of size classes from the level of grouping, $\bar{K}_b$ = mean diversity measure for all individuals in the level of grouping, $G_{IP} = (K_c - \bar{K}_b)/K_c$ where K_c and $\bar{K}_b$ are calculated from the individuals within a *single* population. G_{IP} is meant to be distinguished from C_{IP} in Equations 3 and 4 as the latter is the among-individual within-population component of the total diversity in the sample. Data for the T size class are not listed as it was found only in the coastal region (f(T) = 0.001). Frequencies are rounded to the third decimal place but diversity measures are calculated from unrounded numbers.

further into Sand A and Sand B categories. The frequencies of size classes in the Loam A and Sand A samples do not differ significantly, nor do either of them differ from the frequencies in the Pure A sample. The frequencies in the Sand A and Sand B samples are significantly different ($G = 11.132$, d.f. = 3, $P < 0.025$), however this difference is less than that between pure samples of the A and B lineage (Figure 5A).

These data indicate that hybridization only affects the frequencies of size classes in the B lineage. This effect is illustrated in Figure 5C which presents the frequencies of size classes in hybrid and pure samples of the B lineage (Sand B and Pure B, respectively). While a 2 × 4 contingency test does not reveal a significant difference, the pure populations consistently have a higher frequency of the S size class which, from a Wilcoxon signed-rank test, is a significant pattern [$T_s = 0.0$; $P < 0.03$, SOKAL and ROHLF (1981) p. 448]. Five populations from the Sand region [SEDCT, WLRCT, PRBCT, TYBCT, SXDCT] and five pure populations [MFPCT, GRICT, SFACT, SAPCT, MAVMA] were paired in a west-to-east direction. None of the 120 possible pairwise combinations of five Sand and five pure populations results in a P value greater than 0.05.

There were no significant differences among years or between sexes in the frequencies of size classes (data not shown).

Diversity indices: The data presented in the right-hand columns of Tables 4 and 5 are estimates of K_b and K_c (in Table 4 K_d replaces K_c as the samples are pooled from several populations). These values indicate that there is considerably more diversity within populations (or larger groupings in the case of Table 4) than within individuals. This is expressed in another way by the G statistic listed in the last column of Tables 4 and 5: $G_{IP} = (K_c - \bar{K}_b)/K_c$ (in Table 4 the appropriate notation is $G_{IP} = (K_d - \bar{K}_b)/K_d$ as populations are lumped into larger categories). These values show that in all populations (or larger categories) greater than 50% of the diversity present in a population is due to variation among individuals in their cytoplasmic genotypes (*i.e.*, homoplasmic or heteroplasmic state). The patterns of G_{IP} values are generally consistent with the data on the frequency of heteroplasmy presented in Table 2. Higher levels of heteroplasmy should result in lower G_{IP} values as a greater proportion of the size class diversity is present within individuals. The G_{IP} values for the A and B lineages in Table 4 reveal this effect (compare also the G_{IP} values of "A Inland" with "B Coast").

Hierarchical structure of mtDNA size variation: Table 6 lists the hierarchical statistics calculated among the four populations within each lineage ("3-level statistic") and across the four hierarchical levels in the combined sample of eight populations ("4-level statistic"). The 3-level statistics show that about 35% of the total diversity for mtDNA size lies within the individuals (C_I). Consistent with the observation that

TABLE 5

Frequencies of mtDNA size classes and diversity indices in New England Populations

Population (n)	f(VS)	f(S)	f(M)	f(L)	f(VL)	f(X)	K_c	$\bar{K}_b$	G_{IP}
BRHCT (5)	0.000	0.000	1.000	0.000	0.000	0.000	0.0000	0.0000	
CHTCT (3)	0.000	0.000	0.134	0.683	0.100	0.083	0.4983	0.4350	0.1270
DURCT (7)	0.000	0.157	0.829	0.014	0.000	0.000	0.2885	0.0529	0.8166
EHDCT (4)	0.000	0.025	0.875	0.100	0.000	0.000	0.2237	0.1450	0.3518
ERPCT (3)	0.000	0.000	0.666	0.334	0.000	0.000	0.4444	0.0000	1.0000
FLIME (5)	0.000	0.200	0.800	0.000	0.000	0.000	0.3200	0.0000	1.0000
GRICT (10)	0.015	0.495	0.450	0.040	0.000	0.000	0.5506	0.1365	0.7521
GU2CT (4)	0.000	0.337	0.663	0.000	0.000	0.000	0.4471	0.1288	0.7119
HDRCT (12)	0.000	0.000	0.908	0.067	0.025	0.000	0.1698	0.0508	0.7008
HRPME (2)	0.000	1.000	0.000	0.000	0.000	0.000	0.0000	0.0000	
HUMCT (11)	0.000	0.095	0.900	0.005	0.000	0.000	0.1808	0.0173	0.9043
MAVMA (12)	0.000	0.671	0.329	0.000	0.000	0.000	0.4416	0.1037	0.7652
MFPCT (12)	0.004	0.500	0.471	0.021	0.004	0.000	0.5278	0.3075	0.4174
NCACT (14)	0.004	0.421	0.561	0.011	0.003	0.000	0.5078	0.1125	0.7785
NGRCT (11)	0.000	0.204	0.764	0.032	0.000	0.000	0.3740	0.0909	0.7570
PRBCT (23)	0.000	0.174	0.733	0.093	0.000	0.000	0.4243	0.1219	0.7127
PR2CT (3)	0.000	0.000	0.966	0.000	0.034	0.000	0.0644	0.0600	0.0683
SAPCT (6)	0.000	0.658	0.225	0.000	0.117	0.000	0.5023	0.2708	0.4609
SEDCT (20)	0.000	0.065	0.905	0.030	0.000	0.000	0.1758	0.0680	0.6132
SERCT (29)	0.007	0.134	0.843	0.016	0.000	0.000	0.2708	0.1383	0.4893
SFACT (12)	0.004	0.358	0.571	0.063	0.000	0.000	0.5418	0.2325	0.5709
SHACT (3)	0.000	0.550	0.250	0.200	0.000	0.000	0.5950	0.3317	0.4425
STRMA (1)	0.000	0.400	0.000	0.600	0.000	0.000	0.4800	0.0000	1.0000
SXDCT (36)	0.000	0.342	0.571	0.087	0.000	0.000	0.5497	0.2876	0.4768
SXPCT (28)	0.000	0.023	0.684	0.211	0.082	0.000	0.4805	0.1254	0.7390
TYBCT (4)	0.000	0.000	0.750	0.250	0.000	0.000	0.3750	0.2400	0.3600
WIDCT (14)	0.000	0.089	0.857	0.047	0.007	0.000	0.2551	0.1911	0.2509
WLFCT (6)	0.000	0.167	0.833	0.000	0.000	0.000	0.2777	0.0000	1.0000
WLRCT (17)	0.012	0.191	0.797	0.000	0.000	0.000	0.3280	0.1891	0.4230
WOLCT (2)	0.000	1.000	0.000	0.000	0.000	0.000	0.0000	0.0000	

See Table 3 for details.

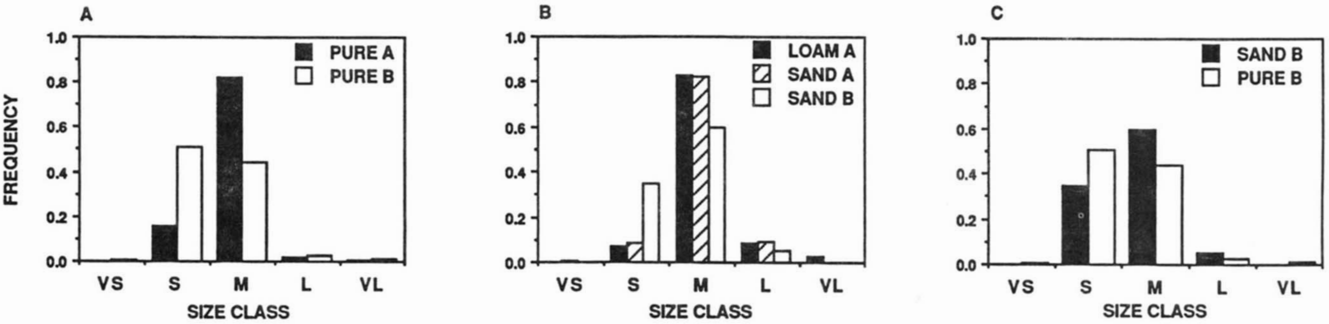

FIGURE 5.—Frequencies of mtDNA size classes in pure and hybrid samples of *G. pennsylvanicus* and *G. firmus*. A, Pure populations of the A and B mtDNA lineages. B, Hybrid populations divided by type of locality (Loam *vs.* Sand) and mtDNA type. C, Comparison of frequencies in hybrid and pure populations of the B lineage. Two very rare size classes (T, the smallest and X, the largest) are not shown as their frequencies are too low to be perceptable in the figure.

heteroplasmy is more frequent in the B lineage (Table 2), the C_I values are slightly higher in the B lineage than in the A lineage, although the standard deviations from the jackknife runs suggest no significant difference. About 60% of the total diversity can be attributed to variation among individuals within populations (C_{IP}) and slightly more among individuals within lineages (C_{IL}). Again, these values are lower in the B mtDNA lineage than in the A lineage, consistent with the data on heteroplasmy (Table 2). In both lineages, however, population differentiation accounts for a very small proportion of the genetic diversity of mtDNA size (C_{PL} is small).

While similar results are obtained from the 4-level analysis (Table 6), the differences between the two species indicated in Table 2 and Figure 5 are expressed in another way. About 33% of the total diversity lies within individuals (C_I). Most of the diversity (52–66%) can be attributed to variation among individuals within populations, lineages or the total sample (C_{IP}, C_{IL}, C_{IT}). A very small proportion of the diversity within lineages is due to variation among populations

TABLE 6

Hierarchical diversity statistics of mtDNA size variation

Statistic	Sampling among or within	Sampling among
3-Level		
$C_{I(A)}$	33.7 ± 10.3	32.6 ± 10.1
$C_{I(B)}$	38.4 ± 6.0	37.2 ± 6.1
$C_{IP(A)}$	59.7 ± 9.4	64.0 ± 9.9
$C_{IP(B)}$	55.0 ± 4.4	58.7 ± 4.5
$C_{IL(A)}$	66.3 ± 10.3	67.4 ± 10.1
$C_{IL(B)}$	61.7 ± 6.0	62.8 ± 6.1
$C_{PL(A)}$	6.6 ± 1.1	3.4 ± 1.0
$C_{PL(B)}$	6.6 ± 1.7	4.1 ± 2.2
4-Level		
C_I	33.7 ± 5.5	31.2 ± 5.6
C_{IP}	52.1 ± 3.8	53.3 ± 4.3
C_{IL}	63.0 ± 5.5	56.8 ± 5.5
C_{IT}	66.3 ± 5.5	68.8 ± 5.5
C_{PL}	6.1 ± 1.1	3.4 ± 1.3
C_{PT}	14.1 ± 3.4	15.5 ± 4.9
C_{LT}	8.1 ± 2.4	12.1 ± 4.0

The three-level statistics are calculated separately for four pure populations of the A lineage and four pure populations of the B lineage [note the subscripts (A) or (B)]. The four-level statistics are calculated from the total sample of all eight populations used in the three-level calculations, hence there are four hierarchical levels (I = individual, P = population, L = lineage, T = total sample). Sampling "among" *vs.* "among or within" indicates whether the diversity measures are based on the operation of drawing two different copies of a gene from two different populations ("among"), or from either a different population or from the original population ("among or within"). Values listed are means ± one standard deviation of jackknife runs and indicate the percent of the total diversity (see MATERIALS AND METHODS for details).

(C_{PL}). There is noticeably more variation among populations within the total sample (C_{PT}) than among populations within lineages (C_{PL}). Standard deviations from the jackknife runs suggest that this is a significant difference. As indicated by C_{LT}, this among-population variation in the total sample of eight populations (C_{PT}) is associated with differences between the two lineages.

The two different methods of determining the diversity measures ("sampling among" versus "sampling among or within") have a only slight effect on the hierarchical statistics. As expected, when the "sampling among or within" approach is used, the C_{IP}, C_{IL} and C_{IT} values are lower and the C_{PL} values are higher. The greatest difference is seen in the C_{LT} values (the between-species component of diversity). These differences are a result of the additional within-group diversity which can contribute to the diversity measures in the operation of drawing two "alleles."

The same patterns are revealed by the V statistic based on the arcsin-square root transformed frequencies of mtDNA size classes. There is no significant variation among populations within each of the two mtDNA lineages (for the S size class, $V_{A\ \mathrm{lineage}} = 3.755$, $V_{B\ \mathrm{lineage}} = 2.820$; $P > 0.1$). When the populations are combined in the analysis, there is significant among-population variation (for the S size class V = 33.88, $P < 0.001$; total variance = 0.0663, Wahlund variance = 0.0487, between population proportion of total = 0.7343). It should, and does, follow from these tests that there is significant variation between the A and B lineages in transformed frequencies of size classes (for the S size class, V = 28.29, $P < 0.001$; total variance = 0.0996, Wahlund variance = 0.0961, between lineage proportion of total = 0.9647). When the analysis is done on the M size class, the same general patterns are revealed (data not shown).

Thus, the data on heteroplasmy, frequencies of size classes, diversity indices and the G statistics show that there are subtle but consistent differences between the two species in the nature of mtDNA size variation.

Estimates of mutation rates: BIRKY, MARUYAMA and FUERST (1983) provide an equation for the equilibrium value of K_c given that K_a is small and that $K_c \gg u$ (mutation rate):

$$K_c \sim 2N_{eo}u/(2N_{eo}u + 1) \tag{5}$$

where N_{eo} is the effective number of organelle genes under conditions where gene diversity is decaying at a steady rate. This can be rearranged to express u in terms of K_c and N_{eo}:

$$u \sim 1/(1/K_c - 1)(2N_{eo}). \tag{6}$$

BIRKY, MARUYAMA and FUERST (1983) have shown that with strict maternal transmission, N_{eo} reduces to N_f, the effective number of females in the population. Mark-recapture data from a coastal population of *G. firmus* indicate that this population consisted of about 1500 individuals of both sexes (D. RAND, unpublished data). A rough estimate of N_f would be on the order of 10^3; the data from Tables 4 and 5 indicate that in samples greater than about 10 individuals, K_c is approximately 0.2–0.3. Thus, from Equation 6, u is estimated to be 1.25×10^{-4} to 2.14×10^{-4}. The relationship between N, K_c and u for a range of values is illustrated in Figure 6.

The major sources of error in this estimate of $u = 10^4$ are (1) population size, which was estimated from a single field experiment and (2) K_a, which as shown by BIRKY, FUERST and MARUYAMA (1989) must be low for an accurate estimate. K_a can be estimated using Equations 4 and 5 of BIRKY, FUERST and MARUYAMA (1989) and available data. RAND and HARRISON (1986) estimated N_{eo} as ranging from 87 to 395 in *G. firmus*. Using $g = 10$ [the number of germ cell generations per animal generation; referred to as c in BIRKY, FUERST and MARUYAMA (1989)], $K_a = (K_b)D$ [modified from Equations 4 and 5, BIRKY, FUERST and MARUYAMA (1989)] where D varies with N_{eo} from about 0.91 to 0.98. With K_b = 0.2–0.3, $K_a \sim$ 0.18–0.29. In this range of K_a values the error can be as high as 25% (see Figure 3 of BIRKY, FUERST and

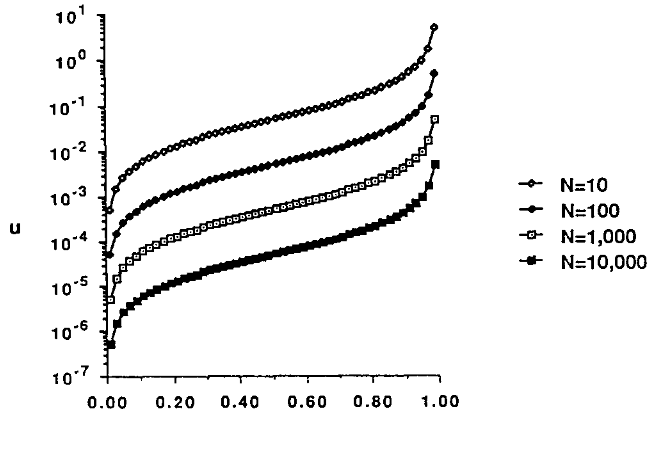

FIGURE 6.—Relationships between gene diversity (K) and mutation rate (u) for four different effective population sizes (N) based on Equation 6.

MARUYAMA 1989). Additional sources of error are (1) the sample estimates of K_b and K_c (from the jackknife analysis K_b and K_c can vary by ±1 to 8%); (2) the densitometric error [estimated at about 1% (RAND and HARRISON 1986)]; and (3) the assumption of an infinite alleles model of mutation. For the current data, a finite alleles model is more appropriate. However, since the L, M and S size classes sum to greater than 0.95 in most cases, estimates of gene diversity with three major alleles and an "infinite" number of alleles will not differ greatly from the estimates presented above.

TAKAHATA and MARUYAMA (1981) used a slightly different approach in which the effective population size of individuals, N_e was distinguished from the effective population size of organelle genomes within cell lineages, N_{eo}. Moreover, they incorporate a term for the number of cell generations per animal generation, g, and use a mutation rate per *cell* generation (v) rather than per animal generation. Their equation for the sum of squares of the frequencies of different mtDNA types (*i.e.*, gene identity or $1 - K_c$) within an equilibration population assuming maternal inheritance is:

$$Q \sim 1/[1 + (2N_e g + 2N_{eo})v]. \quad (7)$$

This can be rearranged as before to express the mutation rate (v) in terms of N_e, n, g and Q:

$$v \sim (1/Q - 1)/(2N_e g + 2N_{eo}). \quad (8)$$

Since Q is an identity measure, $1 - K_c = Q$ which ranges from 0.7 to 0.8 ($K_c = 0.3 - 0.2$). From Equation 8, with $N_e = 1000$ and $N_{eo} = 87$ to 395, v is estimated to be 1.19×10^{-5} to 2.12×10^{-5}. This is the mutation rate per *cell* generation; with $g = 10$, the mutation rate per animal generation is in close agreement with the estimates derived from Equation 6.

DISCUSSION

Molecular basis of mtDNA size variation: The sequence data clearly show that size variation in the mtDNA of *G. firmus* is due to differences among molecules in the number of 220-bp repeats. A likely mechanism which could generate length mutations is the slippage and mismatching of single strands during replication (STREISINGER *et al.* 1966; EFSTRADIADIS *et al.* 1980). Since portions of the mitochondrial genome are exposed as single strands for considerable periods of time during replication (CLAYTON 1982), slip-mismatch across entire repeats may occur. The G+C-rich dyadic sequence in cricket mtDNA could act as "landmarks" to stabilize slipped strands (see Figure 11 in EFSTRADIADIS *et al.* 1980). Alternatively, the potential cruciform secondary structure of the G+C-rich dyadic sequence may play a role in length mutations. In Cnemidophorus lizards it appears that the ends of large duplicated regions of mtDNA lie near transfer RNA (tRNA) genes (MORITZ and BROWN 1987). There may be enough primary or secondary structural similarity between tRNAs that they could serve as recognition sites for strand matching or breakage and ligation during the duplication process (MORITZ and BROWN 1987; CANTATORE *et al.* 1987). These mechanisms may apply in general to species where repeated sequences appear to be the source of variation in mtDNA size (FAURON and WOLSTENHOLME 1976; POTTER *et al.* 1980; MERTENS and PARDUE 1981; DENSMORE, WRIGHT and BROWN 1985; SOLIGNAC, MONNEROT and MOUNOLOU 1986; SNYDER *et al.* 1987).

The repeated sequences could allow for recombination. Intramolecular recombination could loop out repeat(s) making the resultant mtDNA molecules smaller. Intermolecular recombination could produce a unicircular dimer which, if resolved into monomers, could release molecules larger or smaller than either of the parent molecules (Figure 7). Although mtDNA has been shown to exist as a unicircular dimer in cell culture (CLAYTON and VINOGRAD 1967; CLAYTON, DAVIS and VINOGRAD 1970), previous analyses have revealed no clear evidence for a 32-kb species of mtDNA (HARRISON, RAND and WHEELER 1985, 1987; RAND and HARRISON 1986). Moreover, several reviewers have suggested that recombination is unlikely in animal mtDNA (CLAYTON 1982; BROWN 1985; MORITZ, DOWLING and BROWN 1987). The lack of evidence for recombination, however, may be due to the absence of informative markers with which to identify the products of recombination (*e.g.*, restriction site or nucleotide sequence differences between repeated regions of mtDNA).

One further possible (although remote) mechanism is transposition. The general structure of the *Bgl* repeat is suggestive of a transposable element. How-

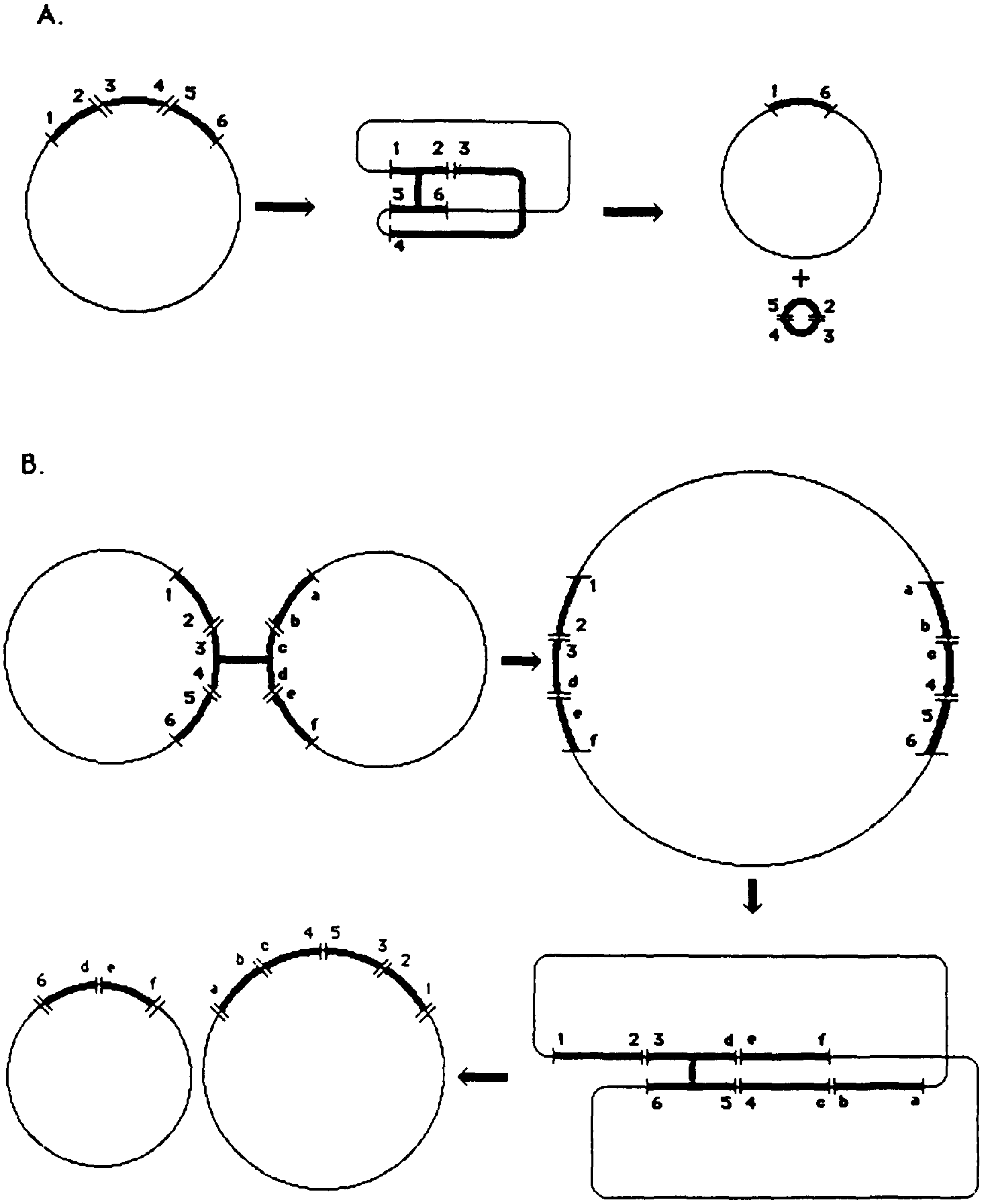

FIGURE 7.—Possible mechanisms of recombination generating length variants in cricket mtDNA. Bold lines represent repeated DNA. Numbers or letters serve as landmarks with which to identify ends of different repeats. A bold line running perpendicular to repeats indicates the site of recombination. A, Intramolecular recombination; B, intermolecular recombination. These are meant to serve as examples; other intermediates and products could be drawn.

ever, there is no meaningful open reading frame in the *Bgl* repeat region and, moreover, there is no evidence for transposable elements in animal mtDNA (BROWN 1985; MORITZ, DOWLING and BROWN 1987).

Features of animal mtDNA regulatory sequences: Although the sequence data alone cannot be used to assign a specific function to the repeats, characteristics of the sequence are suggestive of mtDNA control regions. The dyad symmetric sequence GGGGGCATGCCCCC has the potential to form a 7-bp cruciform structure. Dyad symmetry has been found to exist surrounding the origin of light strand replication in the mtDNAs of human, mouse, and Xenopus (CLAYTON 1982; WONG *et al.* 1983). Cruciform structures have also been shown to play a role in the initiation of DNA replication (ZANNIS-HADJO-

POULOS *et al.* 1988). The GGGGGCATGCCCCC sequence could stabilize a 234-bp hairpin structure. CLARY and WOLSTENHOLME (1987) have identified a conserved potential hairpin structure in A+T-rich regions of *Drosophila virilis* and *Drosophila yakuba* where mtDNA replication is initiated. It also has been shown that the replication origin of many organelle DNAs is located close to a region of variable size (reviewed in MORITZ, DOWLING and BROWN 1987).

The region may also be important in the initiation of transcription. A "TATA box":TATAA lies immediately adjacent to the 5′ flanking segment (bases 109–113, Figure 3) and within the repeat region itself (bases 259–263, 479–483, 699–703, Figure 3). This corresponds very well with the cannonical TATA box believed to be a transcription initiation signal associated with procaryotic and eucaryotic nuclear genes (LEWIN 1985). Although TATA-like sequences are found near transcription initiation sites in the yeast mitochondrial genome (OSINGA and TABAK 1982) and near the light strand transcription start site in mouse mtDNA (CHANG and CLAYTON 1986), only poor correspondence to such sequences can be identified near known transcription initiation sites in the mitochondrial genome of humans (CHANG and CLAYTON 1984, 1986). A peculiar TTGA sequence is repeated once, twice and three times within each repeat of cricket mtDNA (see positions 208–277, Figure 3). This tetramer is also found at a conserved position in the 5′ regulatory regions of chorion genes in *Drosophila melanogaster* and *B. mori* (KAFATOS *et al.* 1987).

A balance of genetic drift and mutation: The analysis of mtDNA size variation in natural cricket populations has revealed that about 35% of the total diversity for mtDNA size lies within individuals and that greater than 50% of the total diversity is due to variation among individuals within local populations. Moreover, there is very little between population heterogeneity for mtDNA size variants. This is in contrast to the nature of variation for restriction enzyme recognition sites. Although heteroplasmy for restriction sites has been observed directly (HALE and SINGH 1986) or inferred (HAUSWIRTH and LAIPIS 1982), it is a rare phenomenon (however, it is more difficult to detect than size heteroplasmy) (BERMINGHAM, LAMB and AVISE 1986). And while restriction enzyme polymorphisms have been reported within local populations of the same species, most of the variation for these polymorphisms exists between populations (AVISE *et al.* 1979; LANSMAN *et al.* 1983; FERRIS *et al.* 1983; DESALLE, GIDDINGS and KANESHIRO 1986; DESALLE, GIDDINGS and TEMPLETON 1986; ASHLEY and WILLIS 1987; NELSON, BAKER and HONEYCUTT 1987; MACNEIL and STROBECK 1987).

These differences provide an illustration of the role of mutation and genetic drift in determining the structure of mtDNA variation. BIRKY, MARUYAMA and FUERST (1983) show that levels of heteroplasmy are determined by the mean time of occurrence of mutations relative to the mean time required to eliminate diversity through vegetative segregation. If drift affects restriction site variants in much the same way it affects size variants, the very high frequency of size heteroplasmy can be explained by the higher mutation rate for size variation relative to that for single base-pair changes which might alter a restriction fragment pattern. However, the mutation rate is not so high that the greatest proportion of diversity lies within individuals. Genetic drift during the vegetative segregation in germ cell lineages produces crickets with different mtDNA genotypes (*i.e.*, different heteroplasmic or homoplasmic states). Yet the balance between drift and mutation is not one that allows for significant differentiation among populations.

An important difference between the size variation described here and restriction site polymorphism is that in the former there are a finite number of size classes whereas the latter is more closely approximated by an infinite "alleles" model (WHITTAM *et al.* 1986). Mutations for size variation will shuffle molecules between size classes (CLARK 1988) while mutations affecting restriction sites are likely to generate new alleles. In either scenario, however, with a low mutation rate (10^{-9}–10^{-6}), random segregation and lineage extinction (AVISE, NEIGEL and ARNOLD 1984) would tend to result in the fixation of different mtDNA types in different populations. With a higher mutation rate (10^{-4}) variation within populations would account for a larger proportion of the total mtDNA variation. With extremely high mutation rates (10^{-2}) variation within individuals would begin to account for much of the variation. These effects can be illustrated for intrapopulation variation using Equation 6. With mutation rates 10^{-6}, 10^{-4} and 10^{-2}, $K_c = 0.0019$, 0.1667 and 0.9524, respectively ($N_e = 1000$). mtDNA size variation in crickets is best characterized by the intermediate case above: the mutation rate is sufficiently high to maintain a high level of heteroplasmy with little population differentiation, but the effects of drift within cell lineages are evident as individuals tend to have different frequencies of size classes. That very little of the total variation within the mtDNA lineages is due to variation among populations (C_{PL}, Table 6) may be in part an effect of the finite number of size classes. With an equivalent mutation rate under an infinite alleles model, populations might tend to have different arrays of alleles and the C_{PL} values would be higher.

An alternative interpretation of the hierarchical statistics could invoke biparental inheritance of mtDNA and high migration rates between populations. High levels of heteroplasmy (C_I) could be due to paternal leakage, while low levels of interpopulation differentiation (C_{PL}, C_{PT}) could be the result of the homogenizing effects of gene flow. There is, however, convincing evidence for maternal transmission of animal mtDNA (LANSMAN, AVISE and HUETTEL 1983; AVISE and VRIJENHOEK 1987). Moreover, direct measurements of dispersal in Gryllus, and the presence of formidable barriers to gene flow in southern New England, indicate that migration is very limited in these species (RAND and HARRISON 1989).

Species differences in mtDNA genotype and size class frequencies: The frequencies of the *M* homoplasmic genotype (Table 2) and the frequencies of mtDNA size classes (Figure 5) are very different in the two cricket species (A and B lineages). An informed discussion of the dynamics of these differences requires some knowledge of whether or not the distributions represent equilibrium conditions. If some event in the history of the two lineages perturbed their distributions, the current differences may simply be temporary as the two distributions return to the same equilibrium. Although the samples from each of the three years showed no significant differences, the approaches to an intermediate equilibrium would have to be very rapid to be detected over this short a period of time. If the current-day patterns do represent equilibrium conditions then the nature of genetic drift, mutation, selection or the integration of these forces must affect mtDNA size variants differently in the two species.

It is unlikely that drift at the cellular level is significantly different in the two species; the number of mitochondria per cell and the sampling regime during development must be very similar. In support of this statement is the close agreement of the transmission data from flies and crickets (SOLIGNAC *et al.* 1984; RAND and HARRISON 1986). At the population level, however, the effective population sizes may be very different which would affect the levels of gene diversity. Populations of *G. firmus* along the coast are generally denser and appear to be larger than populations of *G. pennsylvanicus* in fields in Northwestern Connecticut (D. RAND, personal observation). If *G. firmus* does have a larger effective population size, this could account for the higher incidence of heteroplasmy in its mtDNA lineage. Genetic drift is an unlikely explanation for the differences in the frequencies of size classes in the two species. If one were to argue for drift in this context the differences among populations would represent random fluctuations in size class frequencies. Yet the frequencies of size classes for the populations within each lineage are more similar to one another than they are to the frequencies from populations of the other lineage (Table 5). The probability of this being a result of drift is vanishingly small.

Alternatively, the mutation rate for size variation may be higher in the B than in the A lineage. While this could explain the higher incidence of heteroplasmy in the B lineage, one would have to invoke different mutational processes in each of the lineages to generate the observed size class frequencies. It may be that the mutation rate from M to S is higher in the B lineage whereas the mutation rates between the various other size classes are about the same in the two lineages. However unlikely, this would account for the very different frequencies of the M and S size classes but similar frequencies of the rare classes in the two lineages.

Differences in the nature of the selection regimes on mtDNAs in the two lineages could also explain the discordant frequency distributions. The selection could be due to fitness differences among individuals possessing different sized mtDNAs or the result of replicative differences among molecules within cytoplasms. Replication-based selection differences may well be a combined effect of the ability of a molecule with a given number of tandem repeats to engage replication enzymes relative to its kinetic disadvantage in a "race for replication" (RAND and HARRISON 1986; MORITZ and BROWN 1987; S. R. PALUMBI and A. C. WILSON, unpublished data). Irrespective of any distinction between individual *vs.* cytoplasmic selection, the frequencies of size classes under selection would depend on the nature of the mutational processes. If one makes the simplifying assumption that the mutation rates between adjacent size classes are equal and that mutations from the smallest size class to a "smaller" molecule and from the largest size class to a "larger" molecule result in loss of the new genome, then with selectively equivalent (or neutral) size classes the frequency distribution would approximate a normal distribution. Under these assumptions the selection coefficient for the S size class would have to be very different in the A and B lineages (Figure 5).

In the absence of experimental manipulations it is difficult to determine the relative contributions of drift, mutation and selection to the shapes of the frequency distributions. However, knowledge of the shapes of frequency distributions is an essential prerequisite to the design of functional assays which could shed light on the balance of forces maintaining the distribution (*e.g.*, see KEITH 1983; KEITH *et al.* 1985).

Nuclear-cytoplasmic interactions? Analysis of samples from pure and hybrid populations of *G. pennsylvanicus* and *G. firmus* provides an opportunity to investigate the effects of the mixing of nuclear genes on the frequencies of mtDNA size classes. The data presented in Figure 5 show that the frequencies in hybrid and coastal (pure *G. firmus*) populations of the B lineage are different. If the enzymes responsible for the replication of mtDNA in *G. pennsylvanicus* are most efficient at recognizing a molecule with four repeats (M size class) this size class would be the most frequent. If the enzymes of *G. firmus* are equally efficient at recognizing three- and four-repeat molecules (S and M size classes, respectively) then the two types of molecules would be in approximately equal frequency (assuming uniform mutation rates between size classes). It may be that hybrid populations have intermediate frequencies because their mtDNA replication machinery is, in fact, hybrid.

HARRISON (1986), HARRISON, RAND and WHEELER (1987) and RAND and HARRISON (1989) have reported evidence for introgression of nuclear alleles of *G. pennsylvanicus* into *G. firmus*-like hybrid zone populations. Interestingly, however, there is less evidence for introgression of *G. firmus* alleles into *G. pennsylvanicus*-like populations in the hybrid zone. As would be predicted from a hybrid-replication-machinery hypothesis, the frequencies of mtDNA size classes in these hybrid zone Loam populations are not significantly different from those of pure *G. pennsylvanicus* populations from the Inland region (northwestern Connecticut). However, these observations fail to explain why the distribution of size variants in the A lineage in Sand populations does not show the effects of introgression (Figure 5B).

Some recent reports have presented conflicting evidence on the effects of the nuclear genetic complement on the frequency of mtDNA size variants and heteroplasmy. In newts a weak effect of hybridization on the frequencies of size variants and heteroplasmy has been suggested (WALLIS 1987). In Cnemidophorus lizards there appears to be no effect of hybridization on the frequencies of mtDNA insertions and heteroplasmy, but these frequencies do differ between triploid and diploid individuals (DENSMORE, WRIGHT and BROWN 1985; MORITZ and BROWN 1987). The current data on crickets indicate that heteroplasmy is slightly *less* frequent in hybrid than in nonhybrid individuals. However, this pattern does not apply to all heteroplasmic genotypes in the two species: the frequency of M/S decreases in hybrid samples of the A mtDNA lineage but increases in hybrid samples of the B lineage. Thus hybridization appears to influence the frequency as well as the nature of heteroplasmy, *i.e.*, the relative frequency of size classes.

Although these data are observations of static patterns, the consistency of the patterns among populations and within lineages suggest that variation for mtDNA size is maintained by different nuclear-cytoplasmic interactions in the two cricket species. Clearly, the details of the dynamics of drift, mutation and selection which govern these interactions must be addressed experimentally at the cellular level. However, the focus of such experiments is sharpened by the knowledge of the populational patterns for which a body of theory already exists (GREGORIUS and ROSS 1984; CLARK 1984; ASMUSSEN, ARNOLD and AVISE 1987).

C. P. AQUADRO, J. C. AVISE, C. W. BIRKY, A. G. CLARK, L. R. HALE, C. MORITZ, C. M. SIMON and an anonymous reviewer provided very helpful comments on earlier drafts. We gratefully acknowledge K. WHARTON, T. XU and Dr. BOOT of the Artavanis-Tsakonis laboratory at Yale for expert technical advice on cloning and sequencing. Additional thanks go to S. DELLAPORTA for allowing this work to be completed in his laboratory, to M. MOOSEKER for the use of his densitometer and to R. C. LEWONTIN for computational facilities. Supported by National Institutes of Health predoctoral training grant GM 07499-08-10 to D.M.R. and National Science Foundation grant BSR 84-07474 to R.G.H.

LITERATURE CITED

ANDERSON, S., A. T. BANKIER, B. G. BARRELL, M. H. L. DE BRUIJN, A. R. COULSON, J. DROUIN, I. C. EPERON, D. P. NIERLICH, B. A. ROE, F. SANGER, P. H. SCHREIER, A. J. H. SMITH, R. STADEN and I. G. YOUNG, 1981 Sequence and organization of the human mitochondrial genome. Nature **290:** 457–465.

ASHLEY, M., and C. WILLS, 1987 Analysis of mitochondrial DNA polymorphisms among channel island deer mice. Evolution **41:** 854–863.

ASMUSSEN, M. A., J. ARNOLD and J. C. AVISE, 1987 Definition and properties of disequilibrium statistics for associations between nuclear and cytoplasmic genotypes. Genetics **115:** 755–768.

AVISE, J. C., and R. C. VRIJENHOEK, 1987 Mode of inheritance and variation of mitochondrial DNA in hybridogenetic fishes of the genus *Poeciliopsis.* Mol. Biol. Evol. **4:** 514–525.

AVISE, J. C., J. E. NEIGEL and J. ARNOLD, 1984 Demographic influences of mitochondrial DNA lineage survivorship in animal populations. J. Mol. Evol. **20:** 99–105.

AVISE, J. C., C. GILBIN-DAVIDSON, J. LAERM, J. C. PATTON and R. A. LANSMAN, 1979 Mitochondrial DNA clones and matriarchal phylogeny within and among geographic populations of the pocket gopher, *Geomys pinetis.* Proc. Natl. Acad. Sci. USA **76:** 6694–6698.

BERMINGHAM, E., T. E. LAMB and J. C. AVISE, 1986 Size polymorphisms and heteroplasmy in the mitochondrial DNA of lower vertebrates. J. Hered. **77:** 249–252.

BIRKY, C. W., 1978 Transmission genetics of mitochondria and chloroplasts. Annu. Rev. Genet. **12:** 471–512.

BIRKY, C. W., 1983 Relaxed cellular controls and organelle heredity. Science **222:** 468–475.

BIRKY, C. W., P. FUERST and T. MARUYAMA, 1989 Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. Genetics **121:** 613–627.

BIRKY, C. W., T. MARUYAMA and P. FUERST, 1983 An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. Genetics **103:** 513–527.

BOURSOT, P., H. YONEKAWA and F. BONHOMME, 1987 Heteroplasmy in mice with a deletion of a large coding region of mitochondrial DNA. Mol. Biol. Evol. **4:** 46–55.

BROWN, G. G., and L. J. DESROSIERS, 1983 Rat mitochondrial DNA polymorphism: sequence analysis of a hypervariable site for insertions and deletions. Nucleic Acids Res. **11:** 6699–6708.

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 Associates, Sunderland, Mass.

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

BROWN, W. M., M. GEORGE and A. C. WILSON, 1979 Rapid evolution of animal mitochondrial DNA. Proc. Natl. Acad. Sci. USA **76:** 1967–1971.

CACCONE, A., G. D. AMATO and J. R. POWELL, 1988 Rates and patterns of scnDNA and mtDNA divergence within the *Drosophila melanogaster* subgroup. Genetics **118:** 671–683.

CANN, R. L., and A. C. WILSON, 1983 Length mutations in human mitochondrial DNA. Genetics **104:** 699–711.

CANTATORE, P., M. N. GADALETA, M. ROBERTI, C. SACCONE and A. C. WILSON, 1987 Duplication and remoulding of tRNA genes during the evolutionary rearrangement of mitochondrial genomes. Nature **329:** 853–855.

CHANG, D. D., and D. A. CLAYTON, 1984 Precise identification of individual promoters for transcription of each strand of human mitochondrial DNA. Cell **36:** 635–643.

CHANG, D. D., and D. A. CLAYTON, 1986 Identification of primary transcriptional start sites of mouse mitochondrial DNA: accurate in vitro initiation of both heavy- and light-strand transcripts. Mol. Cell Biol. **6:** 1446–1453.

CLARK, A. G., 1984 Natural selection with nuclear and cytoplasmic transmission. I. A deterministic model. Genetics **107:** 679–701.

CLARK, A. G., 1988 Deterministic theory of heteroplasmy. Evolution **42:** 621–626.

CLARY, D. O., and D. R. WOLSTENHOLME, 1985 The mitochondrial DNA molecule of *Drosophila yakuba*. Nucleotide sequence, gene organization and genetic code. J. Mol. Evol. **22:** 252–271.

CLARY, D. O., and D. R. WOLSTENHOLME, 1987 *Drosophila* mitochondrial DNA: conserved sequences in the A+T-rich region and supporting evidence for a secondary structure model of the small ribosomal RNA. J. Mol. Evol. **25:** 116–125.

CLAYTON, D. A., 1982 Replication of animal mitochondrial DNA. Cell **28:** 693–705.

CLAYTON, D. A., R. W. DAVIS and J. VINOGRAD, 1970 Homology and structural relationships between the dimeric and monomeric circular forms of mitochondrial DNA from human leukemic leukocytes. J. Mol. Biol. **47:** 137–153.

CLAYTON, D. A., and J. VINOGRAD, 1967 Circular dimer and catenate forms of mitochondrial DNA from human leukaemic leukocytes. Nature **216:** 652–657.

DENSMORE, L. D., J. W. WRIGHT and W. M. BROWN, 1985 Length variation and heteroplasmy are frequent in mitochondrial DNA from parthenogenetic and bisexual lizards (genus *Cnemidophorus*). Genetics **110:** 689–707.

DESALLE, R., L. V. GIDDINGS and K. Y. KANESHIRO, 1986 Mitochondrial DNA variability in natural populations of Hawaiian *Drosophila*. II. Genetic and phylogenetic relationships of natural populations of *D. sylvestris* and *D. heteroneura*. Heredity **56:** 87–96.

DESALLE, R., L. V. GIDDINGS and A. R. TEMPLETON, 1986 Mitochondrial DNA variability in natural populations of Hawaiian *Drosophila*. I. Methods and levels of variability in *D. sylvestris* and *D. heteroneura* populations. Heredity **56:** 75–85.

DESALLE, R., A. TEMPLETON, I. MORI, S. PLETSCHER and J. S. JOHNSON, 1987 Temporal and spatial heterogeneity of mtDNA polymorphisms in natural populations of *Drosophila mercatorum*. Genetics **116:** 215–223.

EFSTRADIADIS, A., J. W. POSAKONY, T. MANIATIS, R. M. LAWN, C. O'CONNELL, R. A. SPRITZ, J. K. DE RIEL, B. G. FORGET, S. W. WEISMANN, J. L. SIGHTOM, A. E. BLECHEL, O. SMITHIES, F. E. BARALLE, C. C. SHOULDERS and N. J. PROUDFOOT, 1980 The structure and evolution of the human β-globin gene family. Cell **21:** 653–668.

FAURON, C. M. R., and D. R. WOLSTENHOLME, 1976 Structural heterogeneity of mitochondrial DNA molecules within the genus *Drosophila*. Proc. Natl. Acad. Sci. USA **73:** 3623–3627.

FERRIS, S. D., R. D. SAGE, E. M. PRAGER, U. RITTE and A. C. WILSON, 1983 Mitochondrial DNA evolution in mice. Genetics **105:** 681–721.

GREGORIUS, H.-R., and M. D. ROSS, 1984 Selection with gene-cytoplasm interactions. I. Maintenance of cytoplasm polymorphisms. Genetics **107:** 165–178.

HALE, L. R., and R. S. SINGH, 1986 Extensive variation and heteroplasmy in size of mitochondrial DNA among geographic populations of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **83:** 8813–8817.

HARRISON, R. G., 1979 Speciation in North American field crickets: evidence from electrophoretic comparisons. Evolution **33:** 1009–1023.

HARRISON, R. G., 1983 Barriers to gene exchange between closely related cricket species. I. Laboratory hybridization studies. Evolution **37:** 245–251.

HARRISON, R. G., 1986 Pattern and process in a mosaic hybrid zone. Heredity **56:** 337–349.

HARRISON, R. G., and J. ARNOLD, 1982 A narrow hybrid zone between closely related cricket species. Evolution **36:** 535–552.

HARRISON, R. G., D. M. RAND and W. C. WHEELER, 1985 Mitochondrial DNA size variation within individual crickets. Science **228:** 1446–1448.

HARRISON, R. G., D. M. RAND and W. C. WHEELER, 1987 Mitochondrial DNA variation in Field Crickets across a narrow hybrid zone. Mol. Biol. Evol. **4:** 144–158.

HAUSWIRTH, W. W., and P. J. LAIPIS, 1982 Mitochondrial DNA polymorphism in a maternal lineage of Holstein cows. Proc. Natl. Acad. Sci. USA **79:** 4686–4690.

HAUSWIRTH, W. W., M. J. VAN DEWALLE, P. J. LAIPIS and P. D. OLIVO, 1984 Heterogeneous DNA D-loop sequences in bovine tissue. Cell **37:** 1001–1007.

HENIKOFF, S., 1982 Unidirectional deletion of recombinant DNA clones with exonuclease III. Gene **28:** 351–352.

KAFATOS, F. C., S. A. MITSIALIS, H. T. NGUYEN, N. SPOEREL, S. G. TSITILOU and G. D. MAZUR, 1987 Evolution of structural genes and regulatory elements for the insect chorion, pp. 161–178 in *Development as an Evolutionary Process*, edited by R. RAFF and E. C. RAFF. Alan R. Liss, New York.

KIETH, T. P., 1983 Frequency distributions of esterase-5 alleles in two populations of *Drosophila pseudoobscura*. Genetics **105:** 135–155.

KIETH, T. P., L. D. BROOKS, R. C. LEWONTIN, J. C. MARTINEZ-CRUZADO and D. L. RIGBY, 1985 Nearly identical allelic distributions of xanthine dehydrogenase in two populations of *Drosophila pseudoobscura*. Mol. Biol. Evol. **2:** 206–216.

LANSMAN, R. A., J. C. AVISE and M. D. HUETTEL, 1983 Critical experimental test of the possibility of "paternal leakage" of mitochondrial DNA. Proc. Natl. Acad. Sci. USA **80:** 1969–1971.

LANSMAN, R. A., J. C. AVISE, C. F. AQUADRO, J. F. SHAPIRA and S. W. DANIEL, 1983 Extensive genetic variation in mitochondrial DNA's among geographic populations of the deer mouse, *Peromyscus maniculatus*. Evolution **37:** 1–16.

LEWIN, R., 1985 *Genes II*. John Wiley & Sons, New York.

LEWONTIN, R. C., 1972 The apportionment of human diversity. Evol. Biol. **6:** 381–398.

MACNEIL, D., and C. STROBECK, 1987 Evolutionary relationships among colonies of columbian ground squirrels as shown by mitochondrial DNA. Evolution **41:** 873–881.

MACRAE, A. F., and W. W. ANDERSON, 1988 Evidence for non-neutrality of mitochondrial DNA haplotypes in *Drosophila pseudoobscura*. Genetics **120:** 485–494.

MANIATIS, T., E. F. FRITSCH and J. SAMBROOK, 1982 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

MERTENS, S. H., and M. L. PARDUE, 1981 Mitochondrial DNA in *Drosophila*: an analysis of genome organization and transcription in *Drosophila melanogaster* and *Drosophila virilis*. J. Mol. Biol. **153:** 1–21.

MONNEROT, M., J-C. MOUNOLOU and M. SOLIGNAC, 1984 Intra-individual length heterogeneity of *Rana esculenta* mitochondrial DNA. Biol. Cell **52:** 213–218.

MORITZ, C., and W. M. BROWN, 1986 Tandem duplication of D-loop and ribosomal RNA sequences in lizard mitochondrial DNA. Science **233:** 1425–1427.

MORITZ, C., and W. M. BROWN, 1987 Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content. Proc. Natl. Acad. Sci. USA **84:** 7183–7187.

MORITZ, C., T. E. DOWLING and W. M. BROWN, 1987 Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annu. Rev. Ecol. Syst. **18:** 269–92.

NEI, M., 1973 Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA **70:** 3321–3323.

NELSON, K., R. J. BAKER and R. L. HONEYCUTT, 1987 Mitochondrial DNA and protein differentiation between hybridizing cytotypes of the white-footed mouse, *Peromyscus leucopus*. Evolution **41:** 864–872.

OSINGA, K. A., and H. F. TABAK, 1982 Initiation of transcription of genes for mitochondrial ribosomal RNA in yeast: comparison of the nucleotide sequences around the 5′-ends of both genes reveals a homologous stretch of 17 nucleotides. Nucleic Acids Res. **10:** 3617–3626.

POTTER, D. A., J. M. FOSTEL, M. BERNINGER, M. L. PARDUE and T. CECH, 1980 DNA-protein interactions in the *Drosophila melanogaster* mitochondrial genome as deduced from trimethylsporalen crosslinking patterns. Proc. Natl. Acad. Sci. USA **77:** 4118–4122.

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.

RAND, D. M., and R. G. HARRISON, 1986 Mitochondrial DNA transmission genetics in crickets. Genetics **114:** 955–970.

RAND, D. M., and R. G. HARRISON, 1989 Ecological genetics of a mosaic hybrid zone: Mitochondrial, nuclear and reproductive differentiation by soil type in crickets. Evolution (in press).

SANGER, F., S. NICKLEN and A. R. COULSON, 1977 DNA sequencing with chain terminating inhibiters. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SEDEROFF, R. R., 1984 Structural variation in mitochondrial DNA. Adv. Genet. **22:** 1–108.

SNYDER, M., A. R. FRASER, J. LAROCHE, K. E. GARTNER-KEPKAY and E. ZOUROS, 1987 Atypical mitochondrial DNA from the deep-sea scallop *Placopecten magellanicus*. Proc. Natl. Acad. Sci. USA **84:** 7595–7599.

SOKAL, R. R., and F. J. Rohlf, 1981 *Biometry*, Ed. 2. W. H. Freeman, San Francisco.

SOLIGNAC, M., M. MONNEROT and J.-C. MOUNOLOU, 1983 Mitochondrial DNA heteroplasmy in *Drosophila mauritiana*. Proc. Natl. Acad. Sci. USA **80:** 6942–6946.

SOLIGNAC, M., M. MONNEROT and J.-C. MOUNOLOU, 1986 Concerted evolution of sequence repeats in *Drosophila* mitochondrial DNA. J. Mol. Evol. **24:** 53–60.

SOLIGNAC, M., J. GENERMONT, M. MONNEROT and J.-C. MOUNOLOU, 1984 Genetics of mitochondria in Drosophila: inheritance in heteroplasmic strains of *D. mauritiana*. Mol. Gen. Genet. **197:** 183–188.

SOLIGNAC, M., J. GENERMONT, M. MONNEROT and *J.-C. Mounolou*, 1987 Drosophila mitochondrial genetics: Evolution of heteroplasmy through germ line cell divisions. Genetics **117:** 687–696.

STREISINGER, G., Y. OKADA, J. EMRICH, J. NEWTON, A. TSUGITA, E. TERZAGHI and M. INOUYE, 1966 Frameshift mutations and the genetic code. Cold Spring Harbor Symp. Quant. Biol. **31:** 77–84.

TAKAHATA, N., and T. MARUYAMA, 1981 A mathematical model of extranuclear genes and the genetic variability maintained in a finite population. Genet. Res. **37:** 291–302.

THRAILKILL, K. M., C. W. BIRKY, G. LUCKERMANN and K. WOLF, 1980 Intracellular population genetics: evidence for random drift of mitochondrial allele frequencies in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. Genetics **96:** 237–262.

UPHOLT, W. I., and I. B. DAWID, 1977 Mapping of mitochondrial DNA of sheep and goats: rapid evolution in the D-loop region. Cell **11:** 571–583.

VAWTER, L., and W. M. BROWN, 1986 Nuclear and mitochondrial DNA comparisons reveal extreme rate variation in the molecular clock. Science **234:** 194–196.

WALLACE, D. C., 1982 Structure and evolution of organelle genomes. Microbiol. Rev. **46:** 208–240.

WALLIS, G. P., 1987 Mitochondrial DNA insertion polymorphism and germline heteroplasmy in the *Trturus cristatus* species complex. Heredity **58:** 229–238.

WEIR, B. S., and C. COCKERHAM, 1984 Estimating F-statistics for the analysis of populations structure. Evolution **38:** 1358–1370.

WHITTAM, T. S., A. G. CLARK, M. STONEKING, R. L. CANN and A. C. WILSON, 1986 Allelic variation in human mitochondrial genes based on patterns of restriction site polymorphisms. Proc. Natl. Acad. Sci. USA **83:** 9611–9615.

WOLSTENHOLME, D. R., J. L. MACFARLANE, R. OKIMOTO, D. O. CLARY and J. A. WAHLEINTHNER, 1987 Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nemotode worms. Proc. Natl. Acad. Sci. USA **84:** 1324–1328.

WONG, J. F. H., E. P. MA, R. K. WILSON and R. K. ROE, 1983 DNA sequence of the *Xenopus laevis* mitochondrial heavy and light strand replication origins and flanking tRNA genes. Nucleic Acids Res. **11:** 4977–4995.

WRIGHT, S., 1978 *Evolution and the Genetics of Populations, Vol. 4: Variability Within and Among Natural Populations.* Chicago University Press, Chicago.

ZANNIS-HADJOPOULOS, M., L. FRAPPIER, M. KHOURY and G. B. PRICE, 1988 Effect of anticruciform DNA monoclonal antibodies on DNA replication. EMBO J. **7:** 1837–1844.

Communicating editor: J. AVISE