Molecular Population Genetics of mtDNA Size Variation in Crickets

David M. Rand¹ and Richard G. Harrison²

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 (*Grilllus 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 *Grilllus pennsylvanicus* approximately 60% of the former and 45% of the latter were heteroplasmic. From densitometry of autoradiograms 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 (heteroplasy) (Avise et al. 1979; Lansman et al. 1983; Ferris et al. 1983). In recent years, however, heteroplasm 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 Bonnehomme 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 heteroplasm 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.

Genetics 121: 551–569 (March, 1989)
however, from physiological comparisons. Homeo-
thermic 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; Mortiz, 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 mech-
anisms 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 distribu-
tions 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 vari-
ation 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 popula-
tion 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 sys-
tem 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 molecu-
lar basis of mtDNA size variation in Gryllus, (2)
decribe 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 se-
quence 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 indi-
viduals 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 statis-
tics of Birky, Maruyama and Fuerst (1983) and
Birky, Fuerst and Mayuyama (1989). The samples of
cripples 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 popu-
lations, 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 segre-
gation during germ cell divisions) will decrease diver-
sity. 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 (hetero-
plasmy); 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 vari-
ation to indicate the change in size of a mtDNA mole-
cule.)

**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 pop-
ulations of these crickets are discussed. "Pure" popu-
lations of *G. firmus* are found along the coast of
Connecticut while "pure" populations of *G. pennsyl-
vanicus* are found inland. In central Connecticut var-
ious hybrid populations can be found, the nature of
which depends on the ecological setting of the popu-
lation (Rand and Harrison 1989).

The two species show significant differences in sev-
eral 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
caracteristic 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 Apal, HinClI, HindIII and
XbaI (Harrison, Rand and Wheeler 1987). Using
these enzymes the composite restriction pattern of *G.
**TABLE 1**

<table>
<thead>
<tr>
<th>Locality name (region)</th>
<th>Town/ Locality</th>
<th>Frequency of A mtDNA</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. BRHCT (L)</td>
<td>Bristol</td>
<td>1.00</td>
<td>1984, 1985</td>
</tr>
<tr>
<td>2. CHTCT (U)</td>
<td>Chester</td>
<td>1.00</td>
<td>1984</td>
</tr>
<tr>
<td>3. DURCT (L)</td>
<td>Durham</td>
<td>1.00</td>
<td>1984</td>
</tr>
<tr>
<td>4. EHDCT (L)</td>
<td>East Haddam</td>
<td>1.00</td>
<td>1984</td>
</tr>
<tr>
<td>5. ERPCT (L)</td>
<td>New Haven/East Rock Park</td>
<td>1.00</td>
<td>1986</td>
</tr>
<tr>
<td>6. FLIME (U)</td>
<td>Flag Island, Harpswell, Maine</td>
<td>1.00</td>
<td>1984</td>
</tr>
<tr>
<td>7. GRICT (C)</td>
<td>Guilford/Grass Island</td>
<td>0.17</td>
<td>1984</td>
</tr>
<tr>
<td>8. GU2CT (C)</td>
<td>Guilford</td>
<td>0.10</td>
<td>1985</td>
</tr>
<tr>
<td>9. HDRC (L)</td>
<td>Haddam</td>
<td>1.00</td>
<td>1985</td>
</tr>
<tr>
<td>10. HRPME (U)</td>
<td>Harpswell, Maine</td>
<td>1.00</td>
<td>1984</td>
</tr>
<tr>
<td>11. HUMCT (I)</td>
<td>Cornwall/Housatonic Meadows</td>
<td>1.00</td>
<td>1985</td>
</tr>
<tr>
<td>12. MAVMA (U)</td>
<td>Martha's Vineyard</td>
<td>0.00</td>
<td>1986</td>
</tr>
<tr>
<td>13. MFPT (C)</td>
<td>Milford Point</td>
<td>0.00</td>
<td>1986</td>
</tr>
<tr>
<td>14. NCAGT (I)</td>
<td>North Caanan</td>
<td>0.69</td>
<td>1986</td>
</tr>
<tr>
<td>15. NGRCT (I)</td>
<td>North Granby</td>
<td>1.00</td>
<td>1987</td>
</tr>
<tr>
<td>16. PRBCT (S)</td>
<td>Plainville</td>
<td>0.56</td>
<td>1984, 1986</td>
</tr>
<tr>
<td>17. PR2CT (L)</td>
<td>Prospect</td>
<td>1.00</td>
<td>1984</td>
</tr>
<tr>
<td>18. SAPCT (C)</td>
<td>Old Saybrook/Saybrook Point</td>
<td>0.06</td>
<td>1984</td>
</tr>
<tr>
<td>19. SECT (S)</td>
<td>Seymour</td>
<td>0.75</td>
<td>1984-1986</td>
</tr>
<tr>
<td>20. SERCT (L)</td>
<td>Seymour</td>
<td>1.00</td>
<td>1984-1986</td>
</tr>
<tr>
<td>21. SFACCT (C)</td>
<td>Madison/Seafield Association</td>
<td>0.00</td>
<td>1984</td>
</tr>
<tr>
<td>22. SHACT (I)</td>
<td>Sharon</td>
<td>1.00</td>
<td>1985</td>
</tr>
<tr>
<td>23. STRMA (U)</td>
<td>Sturbridge, Massachusetts</td>
<td>1.00</td>
<td>1984</td>
</tr>
<tr>
<td>24. SXCT (S)</td>
<td>Essex</td>
<td>0.28</td>
<td>1984-1986</td>
</tr>
<tr>
<td>25. SXPT (L)</td>
<td>Essex</td>
<td>1.00</td>
<td>1984-1986</td>
</tr>
<tr>
<td>26. TBYCT (S)</td>
<td>Tylerville</td>
<td>0.75</td>
<td>1984</td>
</tr>
<tr>
<td>27. WIDCT (I)</td>
<td>Winchester</td>
<td>1.00</td>
<td>1986</td>
</tr>
<tr>
<td>28. WLFCT (L)</td>
<td>Wallingford</td>
<td>1.00</td>
<td>1984, 1986</td>
</tr>
<tr>
<td>29. WLRCT (S)</td>
<td>Wallingford</td>
<td>0.65</td>
<td>1984-1986</td>
</tr>
<tr>
<td>30. WOLCT (L)</td>
<td>Wolcott</td>
<td>1.00</td>
<td>1986</td>
</tr>
</tbody>
</table>

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 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.

**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 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-
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 EcoRI 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 EcoRI fragments cloned from females B and D (see Figure 2). The sequence of the size-variable region of both female B and D mtDNA were determined in one direction using the chain termination technique (SANGER, NICKLEN and COULSON 1977). An individual repeat unit (cut out by BglII; see Figure 2) was subcloned into the BamHI 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 *PvuII-EcoRI* 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 EcoRI (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 *EcoRI* 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 assigned a frequency of 0.05, a value “near a conservative lower limit of detectability…” on gels (AVISE and VRJENHOEK 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 homoplastic 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 \( i \)th allele (=size class) in the "population." Considering different levels of organization as "populations," Birky, Maruyama and Fuerst (1983) defined the following \( K \)s:

- \( K_a \): diversity within a cell
- \( K_i \): diversity within an individual
- \( K_c \): diversity within a deme
- \( K_r \): 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. \( K_i \) 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 \( K^* \) from our data, \( K_i \) was estimated using Equation 5 of Birky, Fuerst and Maruyama (1989): \( K_i = K_i / (N_s - 1)(N_s + 1) \). With \( N_s \) (the effective number of organelle genomes) estimated from 87 to 395 (Rand and Harrison 1986), the denominator 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 \( K_i \) of Birky, Fuerst and Maruyama (1989, Equation 2) will be referred to in this paper as \( K^* \) to indicate the region (= intraspecific) level of diversity:

\[
K^*_i = [\bar{K}^*_i + K_i(L - 1)]/L.
\]  

(1)

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

\[
K^*_s = [\bar{K}^*_s + K_s(S - 1)]/S
\]  

(2)

where \( K^*_s \) is the average \( K^*_i \) among \( S \) species, and \( K_s = 1 - \sum x_i^2 \) where \( x_i \) = frequency of the \( i \)th 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, NCAC7, NGRCT, WIDCT) and four G. firmus populations, three of which are from the "Coastal" region (GRCT, MFPCT, SFACT) and one from Martha's Vineyard Island (MAMVA; 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 NCAC7 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 GRCT 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**

\[
C_i = \bar{K}_i/K_d
\]  

(3a)

Among individuals within populations

\[
C_{IP} = (\bar{K}_i - \bar{K}_d)/K_d
\]  

(3b)

Among individuals within lineages

\[
C_{IL} = (K_d - \bar{K}_d)/K_d
\]  

(3c)

Among populations within lineages

\[
C_{PL} = (K_s - \bar{K}_d)/K_d
\]  

(3d)

where \( \bar{K}_i \) is the average \( K_i \) among the sampled populations of the species, \( \bar{K}_d \) is the average \( K_d \) 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. \( C_i \) 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 \( K^*_s \) (from Equation 1 of Birky, Fuerst and Maruyama 1989) and \( K^*_T \) (from Equation 1, this paper) are put in place of \( K_i \) 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
Among population samples from each species:

\[ C_{P} = \frac{(\bar{K}_e - \bar{K}_o) / K_e}{c_r} \]  

Among individual samples within populations:

\[ C_{P I} = \frac{(\bar{K}_e - \bar{K}_o) / K_e}{c_r} \]  

Among populations in the total sample:

\[ C_{P T} = \frac{(\bar{K}_e - \bar{K}_o) / K_e}{c_r} \]  

Among individual samples within lineages:

\[ C_{L} = \frac{(\bar{K}_e - \bar{K}_o) / K_e}{c_r} \]  

Among lineages in the total sample:

\[ C_{L T} = \frac{(\bar{K}_e - \bar{K}_o) / K_e}{c_r} \]  

where \( K_e, K_o, \) and \( K_a \) are the average \( K_e, K_o, \) and \( K_a \) among, respectively, all individuals, populations, and lineages in the total sample of eight \( G. \) pennsylvanicus and \( G. \) firmus populations. \( C_{L}, C_{P I}, C_{P T}, \) and \( C_{L T} \) should, and did, sum to 1.0 in all calculations (as with the three-level statistics, \( C_{L}, C_{P I}, \) and \( C_{P T} \) 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 \( K^* \) [from Birky, Fuernst, and Maruyama's (1989) Equation 1], \( K^* \) (from Equation 1, this paper) and \( K^* \) (from Equation 2, this paper) are put in place of \( K, K_o, \) and \( K_n \), 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:

\[ V = 4 \sum (n_i - A)^2 \]

where \( a_i \) is the transformed frequency in the \( i \)th sample (i.e., population), \( A = (\sum n_i) / N \) and \( N = \sum n_i \), \( V \) is determined by summing across \( r \) samples and is distributed as a \( \chi^2 \) 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'GGGGCATGCCCGCCG 3' which demonstrates dyad symmetry. Three base pairs beyond this symmetric sequence is a BgII site. We will define a repeat as the 220-bp sequence running from the first G in the symmetric sequence to the T immediately preceding 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 Wostenholme's (1987) Figure 2. This highly conserved stretch lies ~330 bp from the beginning of
**mtDNA Length Polymorphism**

**Figure 2.**—Restriction map of the repeat region of female B. The sequence in Figure 3 runs from the EcoRI site to the PvuII 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 IrRNA 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 = BglII, E = EcoR1, P = PvuII and X = XbaI. 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 GAATTCATATA AAGATAATT TTCTTTTTTT GAAAGAAAA AAAAAAAAAG GAAAAATTAG
61 TAGTAATATT AATTTATATC AGTTATATGA AGATATATGA AATATTATA TATACGAAT
121 AATTTGCGG GTGTTGCAG CTTAAGATT TGGTTGGGGGC ATGCCCTCCGA AGATCACTTTT
181 AACCGTAAACA CAGTTAGGA AATTTAATTT AAGATAAGA TTTGATTTCGG TGAATATAT
241 TGTTGATTGA TTATTTGATA TAATATTGGA TAGGTATCC ATGGGGCGTG CGATGAAAAAG
301 TAAAGTTTG GTTTAGGGGA GTGATAGGA GTATGACGAT ATTTTGGCGTG GTTGGCGAG
361 CTTAAGATT TGTTGGGGGC ATGCCCTCCGA AGATCACTTTT ACGCTAACAA CAGTTAGGA
421 ATTTAATTGG AAGGATAAAA TTTGATTTCGG TGAATATATGA AATTTGATA AATTTGATA
481 TAATATTAG TATATTGATCC ATGGGATCCTG ATGCGGAAGAA TAAATTGTTG TTGATAGGA
541 GTGATGGAA GTGATGATCC ATTTGCGGGG TGGTTGGCGTG CTTAAGATT TGTTGGGGGC
601 ATGCCCTCCGA AGATTCTTTT AACCCTAACA CAGTTAGGA AATTTTATTT AAGGATAAGA
661 TTGATGGATAT TAGGTATTGA AAATTTGATA TAATATTGAT TAGGATATCC
721 ATGGGATCGTG ATGGGATTTA TAAAGTTTG TTATTTGCG TGGAGAAAT
781 TATACATATA TTATATTCTA AAAAGATGTA TAATGTAGTA ATTATATTAT TAAATTAAG
841 TGGATTGGG AATACATATT TGGATGATCC ATGGGATCCTG ATGCGGAAGAA TAAATTGTTG

**Figure 3.**—Nucleotide sequence of the repeat region in female B. The sequence is numbered starting with the first base in the 5' EcoRI cloning site. The symmetric sequences bounding the repeats are boxed and the BglII 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 the srRNA gene in cricket mtDNA is of similar length, the beginning of this gene would fall near position CLARY and WOLSTENHOLME'S (1987) (Figure 2).
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 (rRNA) 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* mitochondrial gene organization is the same in crickets and flies, 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 *BgII*. The sequence data show that an S mtDNA has three full repeats. This indicates that a partial digest with *BgII* would produce three-run “ladders” built off both of the adjacent *BgII* fragments as well as a ladder of repeats themselves (see Figure 2). The prediction for an L mtDNA is a five-run 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. pennsylvaniae* 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, heteroplasmic 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 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 heteroplasmic 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: 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: 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: in moving from pure to hybrid populations, the frequency of M/S decreases significantly in the A lineage (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

<table>
<thead>
<tr>
<th>Genotype (n)</th>
<th>&quot;Pure&quot; Total (319)</th>
<th>&quot;Pure&quot; (G. pennsylvaniae)</th>
<th>&quot;Pure&quot; (G. firmus)</th>
<th>&quot;Hybrid&quot; Total (52)</th>
<th>&quot;Hybrid&quot; (G. loam)</th>
<th>&quot;Hybrid&quot; (G. sand)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>12.6</td>
<td>15.1</td>
<td>23.1</td>
<td>3.3</td>
<td>4.1</td>
<td>16.7</td>
</tr>
<tr>
<td>S/VS</td>
<td>0.3</td>
<td>0.0</td>
<td>1.9</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>S/VS/T</td>
<td>0.3</td>
<td>0.0</td>
<td>1.9</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>M</td>
<td>40.4</td>
<td>39.6</td>
<td>15.4</td>
<td>58.7</td>
<td>59.2</td>
<td>25.0</td>
</tr>
<tr>
<td>M/S</td>
<td>25.4</td>
<td>28.3</td>
<td>38.5</td>
<td>14.1</td>
<td>16.3</td>
<td>43.8</td>
</tr>
<tr>
<td>M/S/VS</td>
<td>1.9</td>
<td>0.0</td>
<td>1.9</td>
<td>3.3</td>
<td>2.0</td>
<td>0.0</td>
</tr>
<tr>
<td>M/VS</td>
<td>0.6</td>
<td>1.9</td>
<td>1.9</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>M</td>
<td>0.6</td>
<td>0.0</td>
<td>0.0</td>
<td>1.1</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>L/M</td>
<td>8.5</td>
<td>5.7</td>
<td>5.8</td>
<td>8.7</td>
<td>14.3</td>
<td>6.2</td>
</tr>
<tr>
<td>L/M/S</td>
<td>4.8</td>
<td>5.7</td>
<td>3.9</td>
<td>4.4</td>
<td>4.1</td>
<td>8.3</td>
</tr>
<tr>
<td>L/S</td>
<td>1.3</td>
<td>0.0</td>
<td>1.9</td>
<td>2.1</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>V/L</td>
<td>0.3</td>
<td>0.0</td>
<td>0.0</td>
<td>1.1</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>V/L/L/M</td>
<td>0.6</td>
<td>1.9</td>
<td>0.0</td>
<td>1.1</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>V/L/M</td>
<td>0.9</td>
<td>1.9</td>
<td>1.9</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>V/L/M/S</td>
<td>0.3</td>
<td>0.0</td>
<td>1.9</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>X/L/M</td>
<td>0.3</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>X/V/L/M</td>
<td>0.3</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>f (heteroplasmy)</td>
<td>46.1</td>
<td>45.3</td>
<td>61.5</td>
<td>36.9</td>
<td>36.7</td>
<td>58.3</td>
</tr>
</tbody>
</table>

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 x 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

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

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...
frequencies in the Sand A and Sand B lineages from a Wilcoxon signed-rank test is a significant correction. 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 $K_s$ and $K_r$ (in Table 4 $K_r$ replaces $K_s$ 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: $G_{ip} = (K_r - \bar{K}_d)/K_r$ (in Table 4 the appropriate notation is $G_{ip} = (k_d - \bar{K}_d)/k_d$ as populations are lumped into larger categories). These values show that in all populations (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 heteroplasmic presented in Table 2. Higher levels of heteroplasmic 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 for “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 ($C_3$). Consistent with the observation that...
heteroplasmy is more frequent in the B lineage (Table 2). The $C_r$ 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 ($C_{PI}$ 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

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Sampling among or within</th>
<th>Sampling among</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C_{IP}</td>
<td>38.4 ± 6.0</td>
<td>37.2 ± 6.1</td>
</tr>
<tr>
<td>C_{IPN}</td>
<td>59.7 ± 9.4</td>
<td>64.0 ± 9.9</td>
</tr>
<tr>
<td>C_{IPB}</td>
<td>55.0 ± 4.4</td>
<td>58.7 ± 4.5</td>
</tr>
<tr>
<td>C_{IPN}</td>
<td>66.3 ± 10.3</td>
<td>67.4 ± 10.1</td>
</tr>
<tr>
<td>C_{IPMB}</td>
<td>61.7 ± 6.0</td>
<td>62.5 ± 6.1</td>
</tr>
<tr>
<td>G_{IPN}</td>
<td>6.6 ± 1.1</td>
<td>3.4 ± 1.0</td>
</tr>
<tr>
<td>G_{IPMB}</td>
<td>6.6 ± 1.7</td>
<td>4.1 ± 2.2</td>
</tr>
<tr>
<td>4-Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C_I</td>
<td>33.7 ± 5.5</td>
<td>31.2 ± 5.6</td>
</tr>
<tr>
<td>C_P</td>
<td>52.1 ± 3.8</td>
<td>53.5 ± 4.5</td>
</tr>
<tr>
<td>C_L</td>
<td>65.0 ± 5.5</td>
<td>56.8 ± 5.5</td>
</tr>
<tr>
<td>C_T</td>
<td>66.5 ± 5.5</td>
<td>68.8 ± 5.5</td>
</tr>
<tr>
<td>G_I</td>
<td>6.1 ± 1.1</td>
<td>3.4 ± 1.3</td>
</tr>
<tr>
<td>G_T</td>
<td>14.1 ± 3.4</td>
<td>15.5 ± 4.9</td>
</tr>
<tr>
<td>G_L</td>
<td>8.1 ± 2.4</td>
<td>12.1 ± 4.0</td>
</tr>
</tbody>
</table>

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 (l = individual, P = population, L = lineage, T = total sample). Sampling 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).

There is noticeably more variation among populations within the total sample (C_{IP}) than among populations within lineages (C_{PL}). Standard deviations from the jackknife runs suggest that this is a significant difference. As indicated by C_{IT}, this among-population variation in the total sample of eight populations (C_{IP}) 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_{IT} 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 lineage} = 3.755, V_{B 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 heteroplasm, 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_r given that K_s is small and that K_r \gg u (mutation rate):

\[ K_r \sim 2N_m u/(2N_m u + 1) \]

where N_m 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_r and N_m:

\[ u \sim 1/(1/K_r - 1)(2N_m). \]

Birky, Maruyama and Fuerst (1983) have shown that with strict maternal transmission, N_m reduces to N_r, 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_r 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_r is approximately 0.2−0.3. Thus, from Equation 6, u is estimated to be 1.25 \times 10^{-4} to 2.14 \times 10^{-4}. The relationship between N, K, and u for a range of values is illustrated in Figure 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, $N_{me}$. 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 $1 - K$) within an equilibration population assuming maternal inheritance is:

$$Q \sim 1/[1 + (2N_g + 2N_{me})v].$$

(7)

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

$$v \sim (1/Q - 1)/(2N_g + 2N_{me}).$$

(8)

Since $Q$ is an identity measure, $1 - K = Q$ which ranges from 0.7 to 0.8 ($K_c = 0.3 - 0.2$). From Equation 8, with $N_e = 1000$ and $N_{me} = 87$ to 395, $v$ is estimated to be $1.19 \times 10^{-5}$ to $2.12 \times 10^{-5}$. 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; EFSRADIADIS 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 EFSRADIADIS 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 PARDEU 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.

However, 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 GGGGCGCATGCCCC 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 GGGGCCATGCCCC 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 MORTIZ, 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 canonical 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 KANE CRAIG 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^{-10}), 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^{-5}) 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}, N_c = 0.0019, 0.1667 and 0.9524, respectively (N_c = 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 variation due to differences among populations (G_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 G_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 VRJENHOEK 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 \(G.\) \(f\)irmus along the coast are generally denser and appear to be larger than populations of \(G.\) \(p\)ennsylvanicus in fields in Northwestern Connecticut (D. RAND, personal observation). If \(G.\) \(f\)irmus 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 cytoplasmics. 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 \(G.\) \(p\)ennsylvanicus and \(G.\) \(f\)irmus 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 \(G.\) \(f\)irmus) populations of the B lineage are different. If the enzymes responsible for the replication of mtDNA in \(G.\) \(p\)ennsylvanicus 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 Averianov-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. MOOSER 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**


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.


Communicating editor: J. AVISE