Mitochondrial DNA clones and matriarchal phylogeny within and among geographic populations of the pocket gopher, *Geomys pinetis**

(population structure/genetic variation/restriction enzymes)

JOHN C. AVISE[†‡], CECILIA GIBLIN-DAVIDSON[†], JOSHUA LAERM[†], JOHN C. PATTON[†‡], AND ROBERT A. LANSMAN[§,‡]

Departments of [†]Zoology and [§]Biochemistry, and [‡]Program in Genetics, University of Georgia, Athens, Georgia 30602

ABSTRACT   Restriction endonuclease assay of mitochondria DNA (mtDNA) and standard starch-gel electrophoresis of proteins encoded by nuclear genes have been used to analyze phylogenetic relatedness among a large number of pocket gophers (*Geomys pinetis*) collected throughout the range of the species. The restriction analysis clearly distinguishes two populations within the species, an eastern and a western form, which differ by at least 3% in mtDNA sequence. Qualitative comparisons of the restriction phenotypes can also be used to identify mtDNA "clones" within each form. The mtDNA clones interconnect in a phylogenetic network which represents an estimate of matriarchal phylogeny for *G. pinetis*. Although the protein electrophoretic data also differentiate the eastern and western forms, the data are of limited usefulness in establishing relationships among more local subpopulations. The comparison between these two data sets suggests that restriction analysis of mtDNA is probably unequalled by other techniques currently available for determining phylogenetic relationships among conspecific organisms.

Type II restriction endonucleases cleave duplex DNA at specific recognition sites usually containing four, five, or six nucleotides (1, 2). The fragment patterns produced by the digestion of two homologous DNA molecules can differ because they are differentially modified (i.e., by methylation of bases within the recognition sites), nucleotide substitutions have abolished cleavage sites or created new ones, or major sequence rearrangements have altered the relative positions of cleavage sites within the molecules. Restriction endonucleases have been used to compare mtDNA sequences from a variety of mammalian sources. These studies have shown that the population of mtDNA molecules within an individual animal appears homogeneous in sequence (3, 4), mtDNA does not appear to be modified in ways that affect digestion (3, 5), mtDNA is maternally inherited (3, 6–8), and mtDNA evolves rapidly enough to produce easily detectable sequence heterogeneity within species (3, 4, 6, 9, 10). These observations provide the rationale for the use of restriction analysis of mtDNA to identify mtDNA "clones" in nature and to estimate their evolutionary geneology (phylogeny).

Critical assessment of the utility of any new information used for phylogenetic reconstruction is hampered by the fact that the true evolutionary history of the populations or species being studied is seldom if ever known. In the absence of an absolute calibration, the potential of a new approach can be evaluated only by comparing results with those obtained by independent methods. Gel electrophoresis of proteins is the simplest and strongest technique in current use for estimating the genetic relatedness of closely similar organisms. We have, therefore, performed both protein electrophoretic and mtDNA restriction

analyses with a large number of individuals of the pocket gopher, *Geomys pinetis*, live-trapped throughout the range of the species (11).

Unless local populations are completely isolated, we do not expect the two methods of analysis to yield exactly the same kind of information. First, the level of allozyme differentiation among conspecific populations is usually low and consists largely of differing frequencies of the same electromorphs (12). Recently introduced techniques for protein survey analysis, such as gel sieving and heat lability, may, however, prove to refine discrimination of local populations (13, 14). Second, a given electromorph may often be polyphyletic, in fact representing an assemblage of different allelic products sharing electrophoretic mobility (15–17). We have suggested elsewhere that convergence to a common mtDNA restriction phenotype from unrelated phenotypes is unlikely (3). Third, in contrast to mtDNA, nuclear alleles encoding allozymes are segregated and recombined during each generation of sexual reproduction. For these reasons, allozymes rarely provide unambiguous information about the phylogeny of conspecific populations, much less of individuals within these populations. We will show that analysis of mtDNA by restriction enzymes can partially or totally circumvent many of these difficulties.

MATERIALS AND METHODS

Gophers were live-trapped from the localities listed in Table 1. mtDNA was purified as described (3) from livers of individual animals and digested with six restriction endonucleases. *Hin*cII, *Hin*dIII, and *Bam*HI were purchased from Bethesda Research (Rockville, MD). *Bgl* II and *Bst*EII were prepared and provided by Richard Meagher. *Eco*RI was purified by published procedures (18). Samples of mtDNA were electrophoresed through 1.1% agarose gels (19) with *Hin*dIII-digested phage λ DNA as molecular weight standards. Fragments smaller than 500 base pairs (200 base pairs in the *Hin*cII) were not included in the data analysis because they would not be detectable in the more dilute DNA samples. Proteins were electrophoresed on horizontal starch gels according to standard procedures (20, 21).

RESULTS

mtDNA divergence

The mtDNA fragment phenotypes produced by digestion with three of the six restriction endonucleases are pictured in Fig. 1. Estimates of the number of mtDNA base substitutions per

Table 1.   mtDNA digestion phenotypes observed in geographic samples of *G. pinetis*

Type	Composite mtDNA digestion phenotype*	Counties	No. of individuals mtDNA	No. of individuals Proteins	Mean heterozygosity†
1	1N2N3N4N5N6N	Pierce, Camden, Charlton, GA; Nassau, FL	20		
2	1N2N3N4N5M6N	Camden, GA	1	44	0.001
3	1Q2N3N4N5N6N	Pierce, GA	1		
4	1N2N3N4N5P6N	Camden, GA	3		
5	1N2N3N4N5O6N	Screven, Jenkins, Burke, Richmond, GA	16	22	0.003ϵ
6	1N2N3N4N5N6P	Alachua, FL	2		
7	1N2N3M4N5-6P	Alachua, FL	1	12	0.030
8	1N2N3N4N5S6L	Alachua, FL	1		
9	1N2N3N4N5L6M	Levy, FL	1		
10	1N2N3N4N5N6M	Citrus, FL	1	12	0.058
11	1N2N3N4M5N6M	Marion, FL	2		
12	1N2N3N4N5Q6P	Grady, GA	2	2	
13	1N2O3N4N5Q6P	Gadsden, FL	3	3	
14	1N2N3N4I5-6K	Leon, FL	1	2	
15	1N2N3N4K5-6P	Thomas, GA; Wakulla, FL	2	8	0.051
16	1M2M3O4O5R6Q	Walton, FL	1		
17	1L2M3O4O5R6Q	Walton, FL	1	13	0.022
18	1M2M3N4O5R6Q	Walton, FL			
		Macon, Autauga, AL; Crenshaw, AL; Taylor, Talbot, GA	19	9	0.009
				5	0.024
19	1M2M3N4O5R6O	Taylor, GA	1	15	0.020
20	1L2M3N4O5R6R	Taylor, GA	1		
21	1J2M3N4O5R6Q	Russell, AL	2	12	0.033
22	1M2M3N4P5R6Q	Escambia, AL	3	8	0.020
23	1K2M3N4O5R6Q	Baldwin, AL	2	3	
		Totals	87	171	$\bar{H} = 0.025$

* Numbers refer to restriction enzymes: 1, *Eco*RI; 2, *Bam*HI; 3, *Bst*EII; 4, *Hin*dIII; 5, *Hin*cII; 6, *Bgl* II. Letters refer to digestion phenotypes, as exemplified by Fig. 1.
† Estimates of mean genic heterozygosity (for samples $n \geq 5$) at 25 protein loci.

nucleotide (p) differentiating individual organisms were calculated from the total counted fraction of shared mtDNA digestion fragments produced by all restriction enzymes using Nei and Li's approach (22). Estimates of p calculated by an approach suggested by Upholt (23) were virtually identical. To divide the samples into natural subdivisions according to estimated mtDNA sequence divergence, we subjected this matrix of p values to cluster analysis (24). Results yielded two distinct

subdivisions, largely homogeneous internally but significantly different from one another in percentages of shared mtDNA fragments. All gophers collected in northwest Georgia, Alabama, and the western panhandle of Florida belong to one subdivision, whereas samples collected from the remainder of Georgia and Florida belong to the other cluster. From now on we will refer to these two sets of populations as the "western" and "eastern" forms of *G. pinetis*, respectively. Estimates of
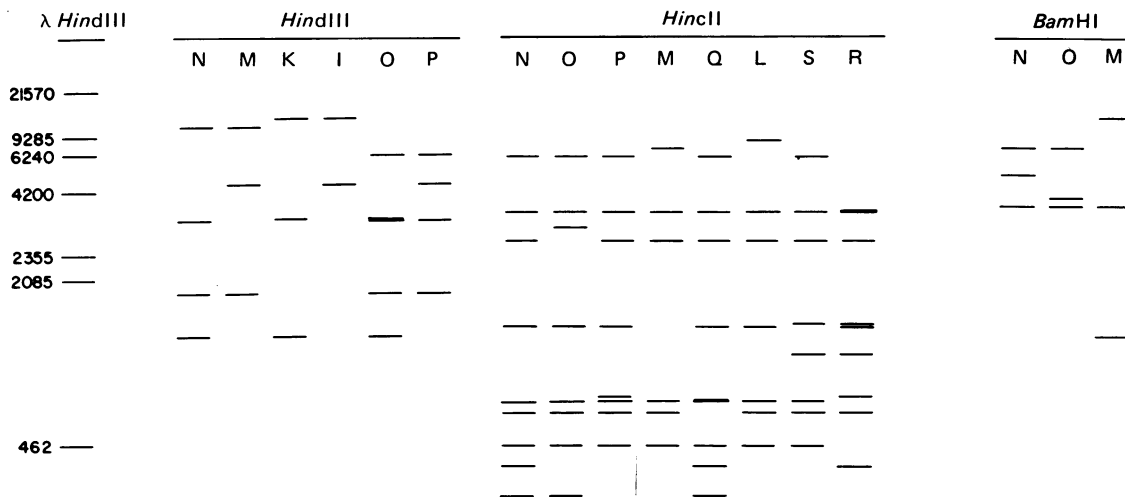

FIG. 1.   Diagram of digestion phenotypes of mtDNA observed in 86 specimens of *G. pinetis*. The molecular weight markers, for which sizes are given in base pairs, are the fragments generated by a *Hin*dIII digestion of phage λ DNA.

mtDNA base substitutions per nucleotide within and among these two forms of *G. pinetis* are summarized in Table 2. There is no overlap in p values in comparisons within compared to between the two forms.

A qualitative method of data analysis that does not submerge the digestion phenotypes in a quantitative summary statistic provides much additional information. An example of the geographic distribution of mtDNA phenotypes, produced by digestion with enzyme *Bam*HI, is shown in Fig. 2. All composite mtDNA digestion phenotypes produced by the six restriction enzymes are listed in Table 1. These composite phenotypes were connected into a most-parsimonious phylogenetic network. The basic procedure is similar to that used for estimation of phylogenies from another kind of qualitative data base—chromosomal inversions in *Drosophila* (25). For example, composite type 1 (abbreviated 1N2N3N4N5N6N) was found in 20 gophers collected from Pierce, Camden, and Charlton Cos., GA, and from Nassau Co., FL. This composite type differs from type 2 (observed in Camden Co., GA; Table 1) only in digestion phenotype produced by enzyme no. 5, *Hinc*II. Similarly, type 3 (observed in Pierce Co., GA) differs from type 1 only in digestion phenotype produced by *Eco*RI. Thus, types 2 and 3 connect separately to type 1 by a single fragment pattern change. In this fashion, and without any regard to geographic collection site, all 23 composite phenotypes listed in Table 1 were added to the phylogenetic network.

This unoriented network was subsequently superimposed on the geographic source of the collections, with results shown in Fig. 3. Each circle encompasses the geographic area within which a given composite phenotype was observed. Composite types are interconnected by branches of the network. In cases where two fragment patterns (e.g., *Bam*HI-O and *Bam*HI-M) cannot be interconverted by a single base substitution, we have added a solid line and a dashed line crossing the branches to indicate that the patterns differ in at least two restriction sites. Thus, the number of solid plus dashed lines indicates the minimum number of base substitutions required to account for the differences in the composite digestion phenotypes.

We have suggested earlier (3) that one advantage of a qualitative analysis of mtDNA digestion phenotypes lies in the low probability that complex phenotypes arise independently in the evolutionary process. In this study, we have observed only a single instance of probable parallelism: clones 17 and 20 both possess the digestion phenotype "L" produced by *Eco*RI, although they link in the phylogenetic network to clones 16 and 18, respectively, both of which exhibit phenotype 1M. 1M and 1L are relatable by a single base substitution; thus, parallelism in this case is not too surprising.

Most of the composite phenotypes in Fig. 3, on both a local
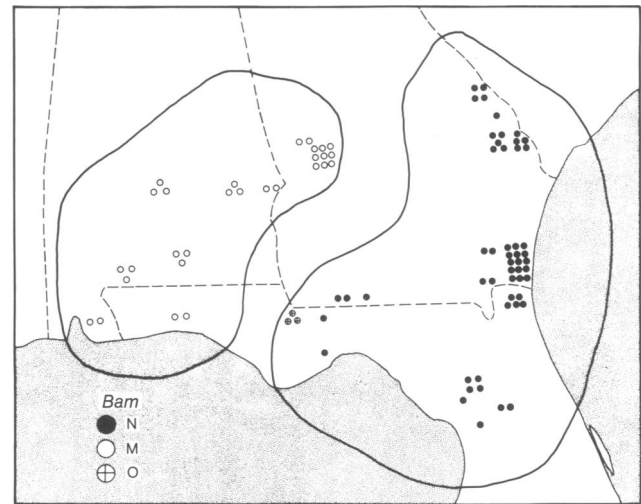

FIG. 2.   Geographic distribution of *Bam*HI digestion phenotypes. Each dot represents a single individual.

and regional basis, are clearly related by one or a few restriction site changes recognized by the restriction enzymes used. For example, type 18, which is found in a wide region within the range of the western form of *G. pinetis*, appears to have given rise by independent mutations to types 19, 20, and 16, which were observed in different localities within the range of 18. Type 17 may subsequently have evolved from type 16 (or conceivably vice versa; the tree is nondirectional) by an additional base change. If type 16 had not been included in our limited samples, the relationship of types 17 and 18 (two assayed base changes) would have appeared similar to the current connectedness of types 18 and 20.

The eastern and western forms of *G. pinetis* differ in digestion phenotypes produced by five of the six restriction enzymes used, and these phenotype differences represent a *minimum* of nine base changes. With the present data it is impossible to determine which particular populations of the eastern and western forms are best directly connected into the network, although a few populations (exhibiting composite types such as 7, 16, and 17) can be effectively eliminated from consideration by possession of different rare digestion phenotypes produced by *Bst*EII.

Protein divergence

There was relatively little allozymic variation, within or among populations, at proteins encoded by most of the 25 loci examined in this study. Counted heterozygosities (H, mean pro-.

Table 2.   Genetic distances between eastern and western forms of *G. pinetis*

	mtDNA		25 protein loci	
Comparison	No. pairwise comparisons of individuals	p,* mean (range)	No. pairwise comparisons of populations[†]	$\bar{D}$,[‡] mean (range)
Eastern	1378	0.005 (0.000–0.018)	10	0.009 (0.000–0.020)
Western	435	0.002 (0.000–0.010)	15	0.024 (0.001–0.049)
Eastern vs. western	1590	0.034 (0.025–0.047)	30	0.065 (0.040–0.098)
Totals	3403	0.018 (0.000–0.047)	55	0.044 (0.000–0.098)

* Base substitutions per nucleotide. To greatly simplify calculations, *Hinc*II was counted as a "six-base" enzyme; this will have only trivial effects on absolute values of p and should not alter relative values.
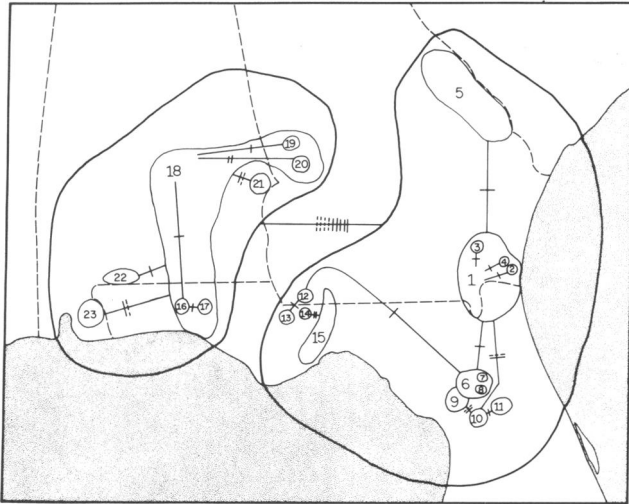[†] $n \geq 5$, see Table 1.
[‡] Codon substitutions per locus.

FIG. 3. Phylogenetic network of *G. pinetis* inferred from mtDNA fragment phenotypes produced by digestion with six different restriction endonucleases. Numbers refer to the composite mtDNA phenotypes listed in Table 1. Phenotypes 7, 14, and 15 are not connected to the network because information from *Hin*cII was not obtained. Phenotypes 8 and 23 were equally related to more than one other phenotype, and their placement in the network was decided by geographic contiguity.

portion of individuals heterozygous per locus in a local population) range from 0.000 to 0.058, with overall unweighted $\overline{H}$ equal to 0.025 (Table 1). Similar levels of electrophoretically assayed genic variation have previously been noted in various species and genera of pocket gophers (26–29).

Protein electromorphs encoded by two loci, 6-phosphogluconate dehydrogenase (*PGD*) and albumin (*ALB*), do show marked regional differentiation. Allele frequencies at these two loci are shown in Figs. 4 and 5. The "100" electromorph for albumin (ALB^{100}) is fixed in all populations of eastern *G. pinetis* defined by mtDNA and is also a common allele in a related congener *G. bursarius* (26). Hence, it is probably plesiomorphic (ancestral) to the *G. pinetis* group of gophers, distinct from ALB^{95}, which may be a derived allele unique to and nearly fixed in western *G. pinetis*. With the exception of peninsular
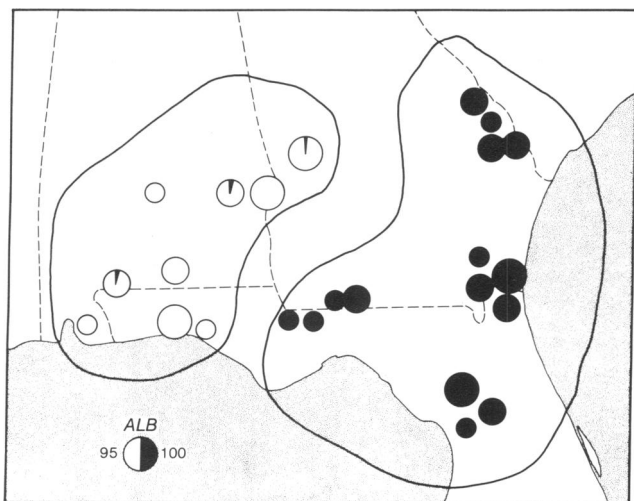

FIG. 4. Geographic distribution of electromorph frequencies of albumin in *G. pinetis*. Large circles represent samples of $n \geq 12$; medium circles, $5 \leq n \leq 11$; small circles, $n \leq 4$. Heavy lines encircle eastern and western forms of *G. pinetis* as defined by the mtDNA analysis.
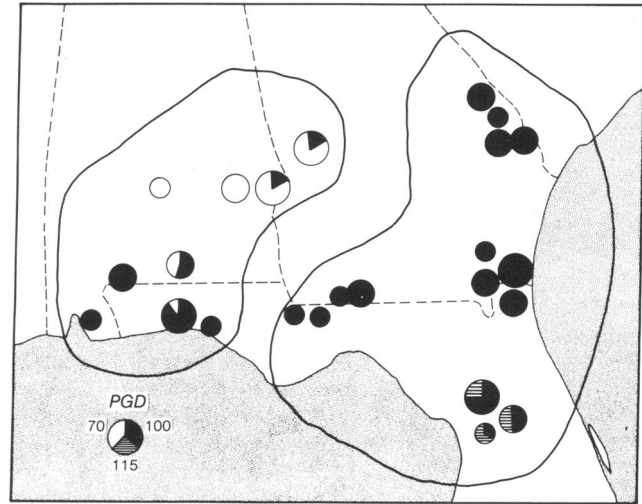

FIG. 5. Geographic distribution of electromorph frequencies of 6-phosphogluconate dehydrogenase, as in Fig. 4.

Florida populations, which exhibit both PGD^{115} and PGD^{100} in high frequency, all eastern *G. pinetis* populations (as well as several western populations) appear monomorphic for PGD^{100}. PGD^{70} was observed only in western populations, where it sometimes reaches high frequency. Because PGD^{100}, PGD^{70}, and PGD^{115} are also observed in the congener *G. bursarius* (26), they are likely ancestral to the entire *G. pinetis* assemblage. Hence, none may have arisen after the evolutionary separation of western from eastern *G. pinetis* stock.

Mean genetic distances (30) at 25 loci were calculated between all samples of $n > 5$ with results shown in Table 2. All distances are well within the range of values typical of conspecific populations in other vertebrates and invertebrates (12). Distances were subjected to a cluster analysis (24), and two major subdivisions corresponding perfectly to the eastern and western forms of *G. pinetis* previously defined by mtDNA were apparent. Apart from the distinctness of these two macrogeographic assemblages, these protein data were largely uninformative about possible phylogenetic relationships among local populations of *G. pinetis*.

DISCUSSION

The heart of the mtDNA data is summarized in Fig. 3. Each circled number represents one or a group of individuals sharing an identical mtDNA genotype as assayed by our restriction enzymes, distinct from all other such groups. If current belief is correct, that mitochondria are strictly maternally inherited, each group then represents a mtDNA clone. Organisms belonging to this clone must have evolved from a common female parent at some time in the past. Interconnectedness among clones, depicted by branches of the network in Fig. 3, provide strong estimates of matriarchal phylogeny.

The power and precision of this approach to natural population analysis are clear. Among 87 gophers examined, 23 clones were identified. Some of these clones (such as 5 and 18) are widespread, while others (such as 2 and 19) appear more local. The local clones are usually relatable by one or two assayed base substitutions to the widespread clone of that region. Clones in different geographic regions can also be readily related to one another. Because mtDNA is inherited through females, segregation and recombination during sexual reproduction do not confound attempts at reconstruction of mtDNA phylogeny. For these reasons also, an individual animal carries within its mitochondrial genome relatively unambiguous information about

the female lineage to which it belongs. In contrast to nuclear gene analyses, which are typically based on *population* allele frequencies, the natural and basic unit of analysis for mtDNA is the individual organism.

A striking contrast of the different kinds of information yielded by protein and mtDNA analyses is apparent in the data presented in Figs. 3 and 5 for samples collected from the Florida peninsula. This group of samples contains several identifiable mtDNA clones in two groups. Clones 9, 10, and 11 appear to be more closely related to the southeast Georgia assemblage (clone 1) than they are to clones 6 and 7, which appear in adjacent locales. One possible interpretation is that this area of Florida has been successfully colonized separately by gophers of rather distinct maternal lineages which now coexist and interbreed. Animals from both maternal lineages share a *PGD* electromorph (115) that has not been found in any other area. This allele probably arose *in situ* by mutation or represents a plesiomorph that has been spread by sexual reassortment so that it now appears in animals of distinct maternal phylogeny.

The choice of the restriction endonucleases used in this study was arbitrary. We expect that the use of a different group of enzymes would resolve a different set of mtDNA clones and would probably produce a slightly different representation of the relatedness of the samples we have used. Fig. 3 is, therefore, an estimation rather than an absolute evaluation of matriarchal phylogeny within *G. pinetis*. Clearly, more refined determinations of clonal diversity and interrelatedness could be obtained by including data from larger numbers of restriction endonucleases. The only limitation on the resolution obtained is the amount of mtDNA available from each sample. Even the rather limited data set presented here illustrates the wealth of phylogenetic information present in the distribution of mtDNA clones within a single species. It is clear to us that restriction analysis of mtDNA, either alone or in concert with conventional data on nuclear genes, will find a wide range of applications in population and evolutionary biology.

1. Boyer, H. W. (1971) *Annu. Rev. Microbiol.* **25**, 153–176.
2. Boyer, H. W. (1974) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **33**, 1125–1127.
3. Avise, J. C., Lansman, R. A. & Shade, R. O. (1979) *Genetics* **92**, 279–295.
4. Upholt, W. B. & Dawid, I. B. (1977) *Cell* **11**, 571–583.
5. Chang, A. C. Y., Lansman, R. A., Clayton, D. A. & Cohen, S. N. (1975) *Cell* **6**, 231–244.
6. Kroon, A. M., deVos, W. M. & Bakker, H. (1978) *Biochim. Biophys. Acta* **519**, 269–273.
7. Dawid, I. B. & Blacker, A. W. (1972) *Dev. Biol.* **29**, 152–161.
8. Conde, M. J., Pring, D. R. & Levings, C. S., III (1979) *J. Hered.* **70**, 2–4.
9. Brown, W. M., George, M., Jr. & Wilson, A. C. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 1967–1971.
10. Brown, W. M. & Wright, J. W. (1979) *Science* **203**, 1247–1249.
11. Hall, E. R. & Kelson, K. R. (1959) *The Mammals of North America* (Ronald, New York), Vol. 2.
12. Avise, J. C. (1974) *Syst. Zool.* **23**, 465–481.
13. Johnson, G. B. (1976) *Genetics* **83**, 149–167.
14. Singh, R. S., Lewontin, R. C. & Felton, A. A. (1976) *Genetics* **84**, 609–629.
15. King, J. L. & Ohta, T. (1975) *Genetics* **79**, 681–691.
16. Coyne, J. A. (1976) *Genetics* **84**, 593–607.
17. Singh, A. S., Lewontin, R. C. & Felton, A. A. (1976) *Genetics* **84**, 609–629.
18. Greene, P. J., Heyneker, H. L., Bolivar, F., Rodriguez, R. L., Betlach, M. C., Covarrbias, A. A., Backman, K., Russel, D. J., Tait, R. & Boyer, H. W. (1978) *Nucleic Acids Res.* **5**, 2373–2380.
19. Helling, R. B., Goodman, H. M. & Boyer, H. M. (1974) *J. Virol.* **14**, 1235–1244.
20. Ayala, F. J., Powell, J. R., Tracey, M. L., Mourao, C. A. & Perez-Salas, S. (1972) *Genetics* **70**, 113–139.
21. Selender, R. K., Smith, M. H., Yang, S. Y., Johnson, W. E. & Gentry, J. B. (1971) *Stud. Genet. VI Univ. Texas Publ.* **7103**, 49–90.
22. Nei, M. & Li, W.-H. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5269–5273.
23. Upholt, W. B. (1977) *Nucleic Acids Res.* **4**, 1257–1265.
24. Sneath, P. H. A. & Sokal, R. R. (1973) *Numerical Taxonomy* (Freeman, San Francisco).
25. Carson, H. L. & Kaneshiro, K. Y. (1976) *Annu. Rev. Ecol. Syst.* **7**, 311–345.
26. Penney, D. F. & Zimmerman, E. G. (1976) *Evolution* **30**, 473–483.
27. Nevo, E., Kim, Y. J., Shaw, C. R. & Thaeler, C. S. (1974) *Evolution* **28**, 1–23.
28. Selander, R. K. Kaufman, D. W., Baker, R. J. & Williams, S. L. (1974) *Evolution* **28**, 557–564.
29. Patton, J. L., Selander, R. K. & Smith, M. H. (1972) *Syst. Zool.* **21**, 263–270.
30. Nei, M. (1972) *Am. Nat.* **106**, 283–292.