In: Genetic Microdifferentiation in Human and Other Animal Populations.
Y.R. Ahuja & J. V. Neel (eds.) 1985 . Indian Anthropologist (Occasional Papers in Anthropology No. 1). p. 15-30.

Evolutionary Differentiation And The Sharing Of Alleles Between Populations

PAUL A FUERST[1]

SUMMARY

Theoretical results concerned with the persistance of alleles which are shared between two differentiating populations are presented. Two models are considered : the strict neutral mutation model of Kimura (1) and the varying mutation model of Nei, Chakraborty and Fuerst (10). It is shown that rare alleles (frequencies less than 0.05) and polymorphic alleles (0.10-0.90) are unlikely to be shared by two populations which show moderate levels of differentiation. Monomorphic (frequencies above 0.95) alleles may be retained in both populations for a considerable time period. The theoretical predictions are compared with data from two groups of natural populations. Discrepancies between theory and data can be accounted for by statistical sampling and by the occurrence of population bottlenecks.

INTRODUCTION

An important component of any science is the testing of hypotheses by the application of statistical criterion. In evolutionary genetics the null hypothesis available to us is the neutral mutation hypotheses and its resultant theoretical model (1). This model can provide expectations for a variety of population statistics, and these expectations can in turn be compared with estimates obtained from natural populations. Such an approach has been successfully applied in a series of studies which utilized data on electrophoretic variation of genetic loci in a wide array of different species (2-4). This report will extend the earlier studies by considering a new set of statistics, those related specifically to the sharing of alleles between a pair of populations. Some related theoretical studies were carried out by Nei and Li (5, 6). I will be principally concerned here with presenting an outline of the statistical results, and will present

[1]Department of Genetics, The Ohio State University, Columbus, Ohio 43210, U.S.A.

comparisons of these theoretical studies with only a limited set of observations from natural populations.

METHODS AND MATERIALS

The statistics which will be presented concern the proportion of those alleles which exist in two populations which will appear in samples taken from each of the populations. I shall identify a set of related statistics which are conditioned upon the frequency of the specified allele in a population. These two population statistics are related to statistics of alleles in single populations which have been studied previously (4). The frequency ranges being considered are (a) total alleles (including alleles from frequency 1/2N to 1.0), (b) rare alleles (from 1/2N to 0.01), (c) essentially monomorphic alleles (0.95 to 1.0); and (d) polymorphic alleles (0.10 to 0.90). Alleles are additionally scored as being either (1) in the same frequency class in both populations, (2) being in a specified frequency class in one population while appearing at a different frequency in the second population or, (3) not appearing in one of the population samples and therefore not being a shared allele.

Theoretical expectations were obtained using the stochastic integral method of Maruyama (7). To begin, a single population of size N=10,000 was simulated. Mutation rates are adjusted to yield results which will have a predetermined expected average heterozygosity. New mutant alleles are introduced by mutation each generation, and gene frequencies change under the influence of sampling caused by the finite size of the population. The population is permitted to reach an equilibrium state in the absence of selection (determined by the 4Nv value of the population being simulated, where N is population size and v is mutation rate). Following the attainment of genetic equilibrium, the population was split into two identical populations of size 10,000 and each population was permitted to evolve under the effects of mutation and sample for 300,000 generations (30 N generation). The identification of alleles in each of the two populations was tracked, and shared, alleles were recorded at various time points following the split. The total number of alleles in a particular frequency class in both populations was recorded. The proportion of shared alleles compared to this total number of alleles in a particular class was then determined. Relative divergence of alleles is plotted against the genetic identity measure of Nei (8). For the results presented in this paper, 400 replicate loci were generated at each average heterozygosity value which was studied.

The electrophoretic data which is used for comparison comes from two published studies. The first is from the study of geographic macrodifferentiation in the 23 populations of pocket gopher *Thomomys, bottae* (9),

while the second involved differentiation among 14 species of Caribbean lizards in the *Anolis bimaculatus* species group (10). In both cases a single population (the White Mountains population form of *Thomomys*, and the Nevis population from *Anolis bimaculatus*) was compared with all other populations to provide a manageable set of comparisons for analysis.

These two groups of natural populations were studied at 23 and 22 loci, respectively, with sample sizes ranging from 2N=20 genes for *A. acutus* up to 2N=280 for *T. bottae*, Hastings. The Hastings populations is the only one to have a sample size above 2N=100. The Nevis population of *A. bimaculatus* had a sample size of 50 genes, while the White Mountain population of *T. bottae* had 2N=30. Consequently, the comparison of populations for the statistic "shared rare alleles" cannot be made, nor will the comparison of total alleles be completely valid, since the simulation results contain alleles with frequency 5×10^{-5} and above while the data generally contain alleles only from about 0.02 and above.

An alternative model, still assuming no selection, was also analyzed. This model is the "varying mutation model" (11). The model assumes that neutral mutation rates are constant parameters at single loci, but that mutation rates over loci are distributed according to a gamma distribution. Assuming that the coefficient of variation of mutation rates is equal to 1.0, the model has been shown to lead to an improvement in the concordance between theory and data for several statistics of the neutral model (2, 3). Theoretical analyses were performed in the same manner for the varying mutation model as for the constant mutation model.

RESULTS

1. Theoretical Predictions

The rate of retention of rare alleles shared between two populations is given in Table 1 and 2. Table 1, which gives the percentage of rare

TABLE 1

PROPORTION OF RARE ALLELES WHICH ARE RARE IN EACH OF TWO POPULATIONS

Time (N-Generations)	Average Heterozygosity 0.05	0.10	0.20	0.50
0.05	0	.013	.134	.066
0.10	0	.011	.046	.035
0.50	0	0	.013	.002
1.0	0	0	0	.002
2.0	0	0	0	0

Calculated using the stochastic integral method of Maruyama (7), with population size, N, equal to 10,000.

TABLE 2

PROPORTION OF RARE ALLELE IN ONE POPULATION WHICH OCCUR AT ANY FREQUENCY IN THE SECOND POPULATION

Time (N-Generations)	Average Heterozygosity 0.05	0.10	0.20	0.50
0.50	.100	.164	.304	.209
0.10	.077	.086	.168	.119
0.50	.037	.029	.025	.020
1.0	0	.008	.006	.014
2.0	0	.034	0	.003
5.0	.024	.008	.011	.002
10.0	0	.016	.006	0
15.0	0	.007	0	0

alleles which will be rare in both populations, illustrates that this class of alleles will be extremely infrequent. Even the observation of a shared allele which is rare in only one of the populations (Table 2) is an unusual occurrence once a relatively short period of divergence has elapsed. There is a tendency for shared rare alleles to appear for a longer period of time when average heterozygosity is greater. This is probably a result of there simply being more rare alleles in such populations. Note that a very small number of rare alleles do occur in common between two populations at later periods, especially when considering alleles rare in only one population. These cases represent alleles which were polymorphic, or even monomorphic, in the two populations, at one time, but which are lost from the populations at a later time because of mutation to new alleles accompanied by random sampling.

Figure 1 presents the pattern of decay for shared polymorphic alleles. In accord with the results for rare alleles, the percentage of alleles which are polymorphic in both populations declines very rapidly. Thus, the proportion of alleles which are expected to be polymorphic in both populations will be low following even slight population differentiation. As a consequence, those rare loci at which the same polymorphic allele is retained in two species can be viewed as having the potential for retention of the allele by selection. There should then be strong reason to pursue experimental verification of such selection. Under neutral selection, even those situations in which the shared allele is polymorphic in only one of the two populations are expected to be infrequent if populations have been split for as little as 0.5 N generations. There is relatively little differenee between different average heterozygosity values when one considers alleles which are polymorphic in both populations. Greater differences can be observed when the allele is polymorphic in only one population; two populations with lower average heterozygosity

will share a greater proportion of their polymorphic alleles than more variable populations.

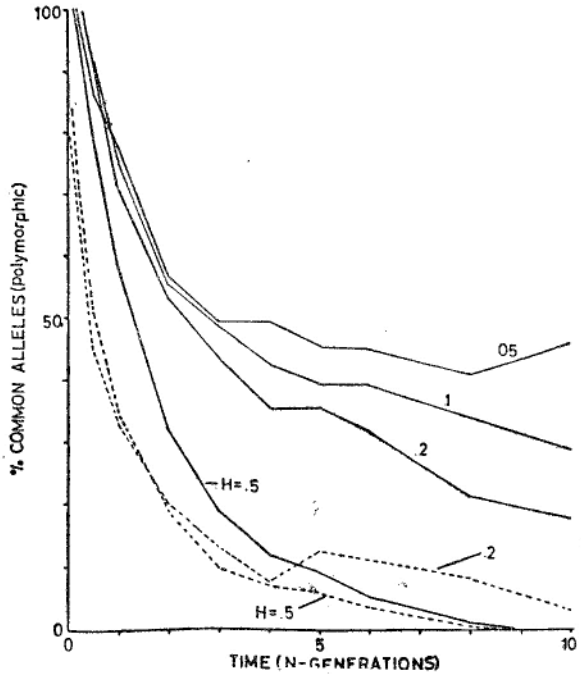

Fig 1. Decline in proportion of polymorphic alleles which are shared. Dotted line represent cases in which shared allele is polymorphic in both populations; solid lines represent cases in which shared allele is polymorphic in only one population.

Retention of shared monomorphic alleles in a population is shown in Figure 2. It can be seen that, in contrast to the retention of rare or polymorphic alleles, monomorphic alleles may be shared by two populations for an extended period of time, especially when average heterozygosity is low. Recall that the theory assumes that individual loci are replicates of each other with respect to mutation rate.

The monomorphic class of alleles makes up the greatest proportion of the allele frequency spectrum when heterozygosity is in the range found in most natural populations. This is illustrated in Table 3 which gives the expected number of alleles per locus in the different frequency ranges considered in this study for heterozygosities which bracket the two sets of natural data. It can be seen that monomorphic alleles make up the largest group. A substantial number of the alleles in a population will thus be retained as shared when at least one of the populations contains the allele in the monomorphic range. These results supplement the earlier findings of Li and Nei (5) on the retention of monomorphic alleles

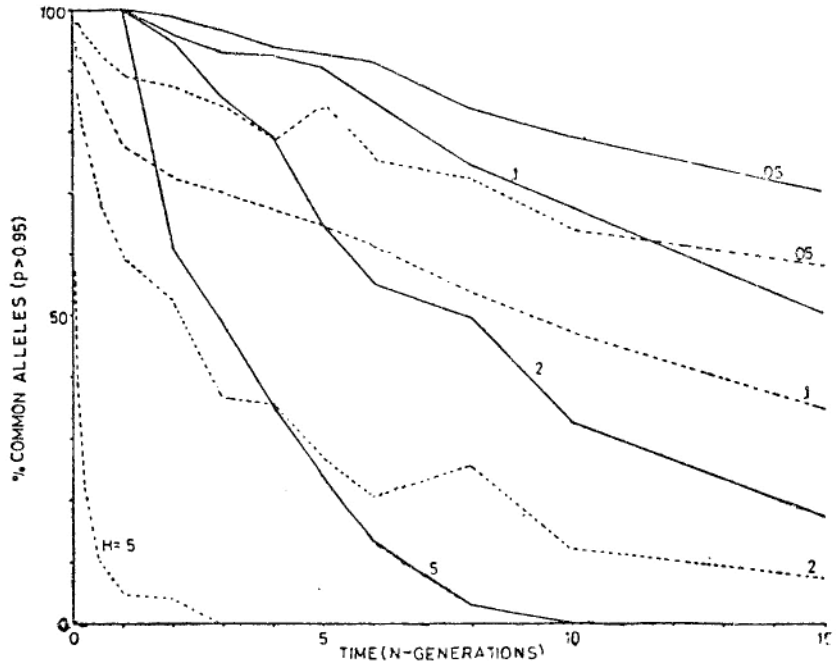

FIG. 2. Decline in percentage of monomorphic alleles which are shared. Dotted lines represent cases in which shared allele is monomorphic in both populations; solid lines represent cases in which shared allele is monomorphic in only one population.

in two populations. The class of loci in which the shared allele is monomorphic in both populations (shown by dotted lines in Figure 2) declines much more rapidly than the class in which the allele in common is

TABLE 3

EXPECTED NUMBER OF ALLELES PER LOCUS IN DIFFERENT FREQUENCY RANGES; EXPECTATIONS CALCULATED FOR ALLELES GREATER THAN 0.001 UNDER THE ASSUMPTION OF THE CONSTANT MUTATION MODEL

Frequency Class	Average Heterozygosity	Number of Alleles	
rare	(0.001-0.01)	0.05	0.10
polymorphic	(0.1-0.9)	0 12	0.26
monomorphic	(0.95)	0.87	0.75
total	(0.001)	1.38	1.77

monomorphic in only one population (shown by solid lines in Figure 2). Many cases in natural populations may thus exist in which an allele is shared by a pair of well differentiated populations; this shared possession, however, provides no evidence that the allele is being retained in

the populations for selective reasons. Nevertheless, there is an underlying selective component to the retention of shared monomorphic alleles. Although the model assumes that all alleles are identical in selective state (i.e. selective neutrality), one should keep in mind the original conceptualization of this model by Kimura (1) who assumed that neutral alleles make up only a small fraction of mutations occurring at the locus of interest. We assume that many other mutations occur during the time period considered by our model, but these other mutations are deleterious. Such deleterious mutations could never be incorporated to replace the monomorphic alleles in either population.

Figure 3 presents the pattern of decay for shared total alleles (fre-

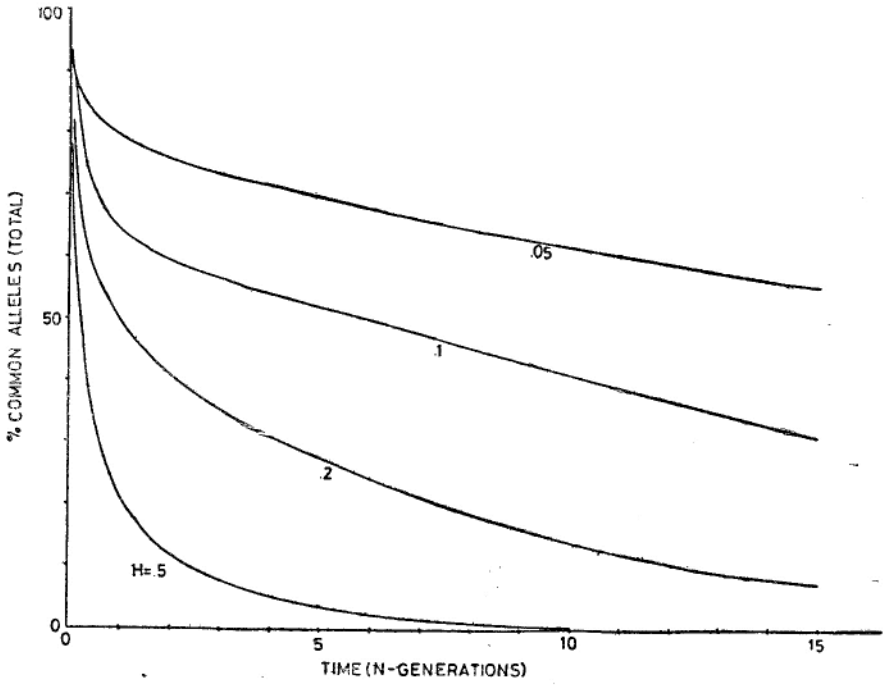

FIG. 3. Decline in proportion of total alleles which are shared.

quency from 0.001 to 1.0). Previous studies with a single population (12) have indicated that the simulation procedures which were used for this study will give very accurate results for all allele classes except those which are very rare (less than 0.001). There may thus be a slight underestimate of the class rare alleles, and this may also result in a minor effect on total shared alleles. This latter effect, however, should be very minor, because shared rare alleles are expected to be lost from the population at the earliest stages of differentiation.

Examination of Figure 3 indicates that there are two phases of population differentiation with respect to total shared alleles. In the first phase (less than 1N generations) the percentage of shared alleles drops rapidly for all heterozygosity values. This is a consequence of the rapid loss from at least one population of those alleles which were rare or polymorphic in the progenitor population. The second phase is a slow but steady decline of the percentage of shared alleles. During this phase, alleles which were monomorphic in the initial population are gradually being replaced by new alleles introduced through mutation. Some neutral alleles may be retained in a pair of populations despite very long divergence times from the common ancestral population. As population variability increases, the populations show rapid loss of shared alleles because of the greater mutational input of new alleles in such populations.

2. *Comparison of Theory with Data from Natural Populations*

In the foregoing section, the rate of decline in the percentage of shared alleles was expressed as a function of time, measured in N generations.

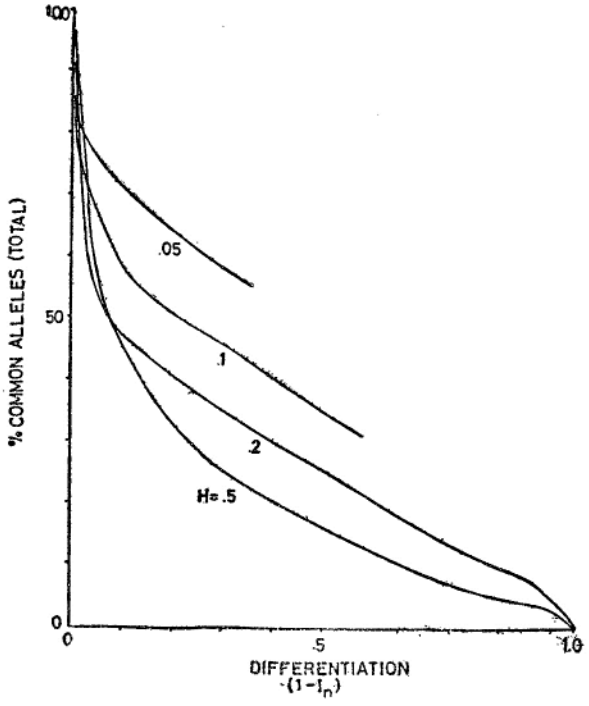

FIG. 4. Decline in proportion of total alleles which are shared as a function of population differentiation measured by $(1-I_n)$, where I_n is the genetic ientity of Nei-(8).

Any attempt to compare theory with data would be futile if expressed in such terms, however, since for no species pair do we accurately know the value of time measured in this way. What we do know is the genetic distance between a pair of species, and this measure can be used to estimate time under certain assumptions (13). We can easily express time for the theoretical results in terms of the genetic identity which is measured between a pair of populations (8). An example is given in Figure 4, where the decline in the percentage of total shared alleles is shown as a function of the genetic identity between two populations. The results can be compared to Figure 3 which expressed the same measure of shared alleles as a direct function of time. It can be seen that the same general patterns of decay are observable in the two figures.

We will now compare our theoretical results with data from two species groups, the pocket gophers classifed taxonomically as *Thomomys bottae*, and the lizards of the *Anolis bimaculatus* species group. Bacause of the natural history and geographic pattern of these populations, it is probable that bottlenecks of population size were important in the initiation of new isolated demes in each of these groups. As noted above, because of the sample sizes taken in the two surveys, no comparison of theory and data for the cases of shared rare alleles was possible.

The data have been plotted together with theoretical predictions computed for average heterozygosities 0.10 and 0.20. The average value of heterozygosity in the 15 *Anolis* populations was 0.056 while in the 23 *Thomomys* populations it was 0.093, so that we expect them to fall slightly beyond the expectations for $\bar{H} = 0.10$.

Figure 5 shows the patterns of decay in the percentage of polymorphic alleles which are polymorphic in both populations. The dotted lines represent values taken from Figure 1 for the constant mutation model, while the solid lines represent expectations based upon the varying mutation model. It can be seen that many of the data comparisons among populations show no shared polymorphic allelc. Those comparisons involving *Thomomys* populations which do have some alleles in common tend to follow the curves for the two models, while those for *Anolis* show a less rapid decay in the proportion of shared alleles which are present. It should be noted that the number of shared alleles which are expected to remain polymorphic in two populations is actually quite small. This is illustrated in Table 4, which presents the approximate number of shared polymorphic alleles which are expected to be observed in the populations which were studied. The table suggests that, despite the apparent scatter of points seen in Figure 5, the data and theory actually agree quite well. The principal reason for the scatter in Figure 8 seems to be the small number of alleles, especially in *Anolis*, which are polymorphic. With these small numbers, the sampling

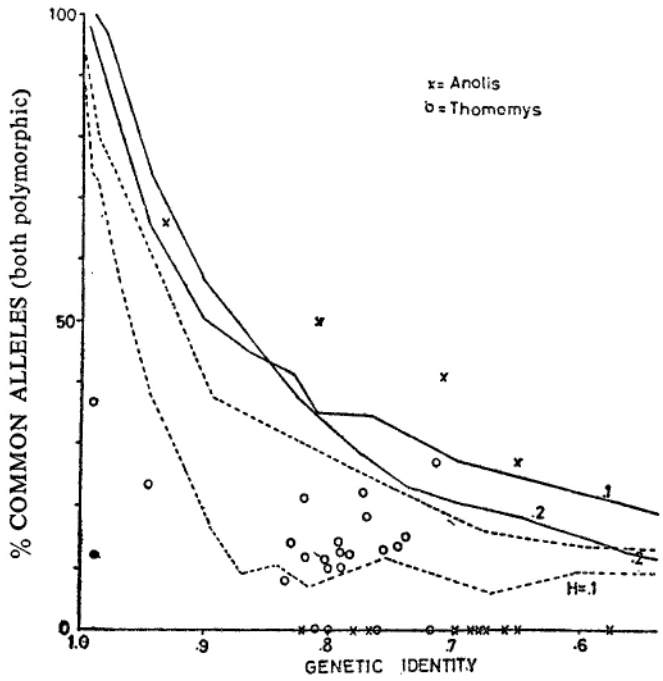

FIG. 5. Decline in percentage of polymorphic alleles which are polymorphic and are shared in both natural populations. Solid lines are expected rate of decay under the varying mutation model; dotted lines are expected rate of decay under the constant mutation model.

variance of the percentage of shared polymorphic alleles is expected to be large, and the distribution is expected to be skewed towards low values.

TABLE 4

OBSERVED AND EXPECTED NUMBER OF POLYMORPHIC ALLELES IN NATURAL POPULATIONS. EXPECTATIONS FOR *Thomomys* CALCULATED USING H=0.10; EXPECTATIONS FUR *Anolis* CALCULATED USING H=0.05, USING THE STOCHASTIC INTEGRAL METHOD

Average Number of Alleles in Each Population		Average Number of Polymorphic Alleles in Each Population		Number of Shared Polymorphic Alleles		
				(1)	(2)	(3)
Thomomys	expected (a)	39.0	10.9	1.4	1.3	—
	expected (b)	38.4	10.1	1.4	1.0	—
	observed	33.3	10.2	0.9±0.6	0.9±05	—
Anolis	expected (a)	28.7	4.5	0.4	0.2	1
	expected (b)	—	—	—	—	—
	observed	31.1	5.1	1.3±1.2	0.8±1.5	0

(a) constant mutation model
(b) varying mutation model
(1) genetic identity between 0.9—0.8
(2) genetic identity between 0.8—0.7
(3) genetic identity between 0.6—0.7

When we examine the situation for shared alleles which are polymorphic in only one of the populations the situation is similar. Firstly, the expected number of such shared alleles, given the heterozygosity, locus sample size and with genetic identity in the range 0.7-0.8 for the two species groups, should be about 4.1 for *Thomomys* and 2.3 for *Anolis*, under the constant mutation model. We expect only 3.2 allele under the varying mutation model for *Thomomys*. The observed data averaged 4.8 ± 1.8 and 2.5 ± 0.5, respectively. The total number of polymorphic alleles remains the same as that given in Table 4. Thus the problem of sampling error similar to that encountered for alleles polymorphic in both populations is not as important when the allele must be polymorphic in only a single population. Figure 6 shows the

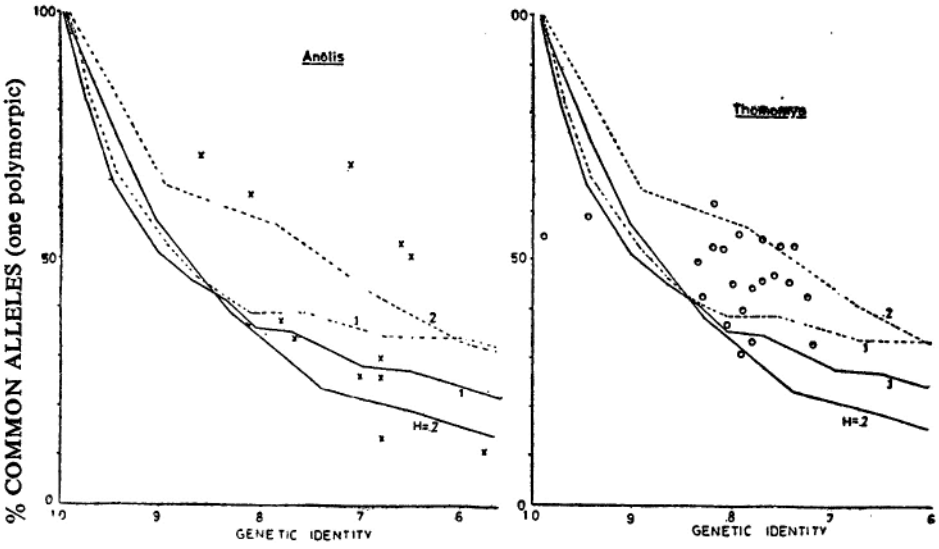

FIG. 6. Percentage of alleles which are shared but are polymorphic in only one population. Solid=Varying mutation model; Dotted=Constant mutation model.

decay of the percentage of alleles in which the shared allele is polymorphic in one population. Again, the solid line denotes the varying mutation model and the dotted line represents the constant mutation model. The *Thomomys* data seem to fit the constant mutation model better than the varying mutation model, and show much less scatter than the *Anolis* data, which could be construed to fit either model. Note, however, that both data sets have heterozygosities below 0.10, and should fit an expectation slightly above the 0.10 varying mutation curves. The mean of the *Anolis* data for populations with genetic identity 0.8-0.9 was 0.50, for I between 0.7-0.8 was 0.41 while for I=0.6-0.67 it was

0.34. For *Thomomys* populations with genetic identity between 0.8 and 0.9, the corresponding value was also 0.50, while for I between 0.7 and 0.8, the mean percentage of shared polymorphic alleles was 0.45.

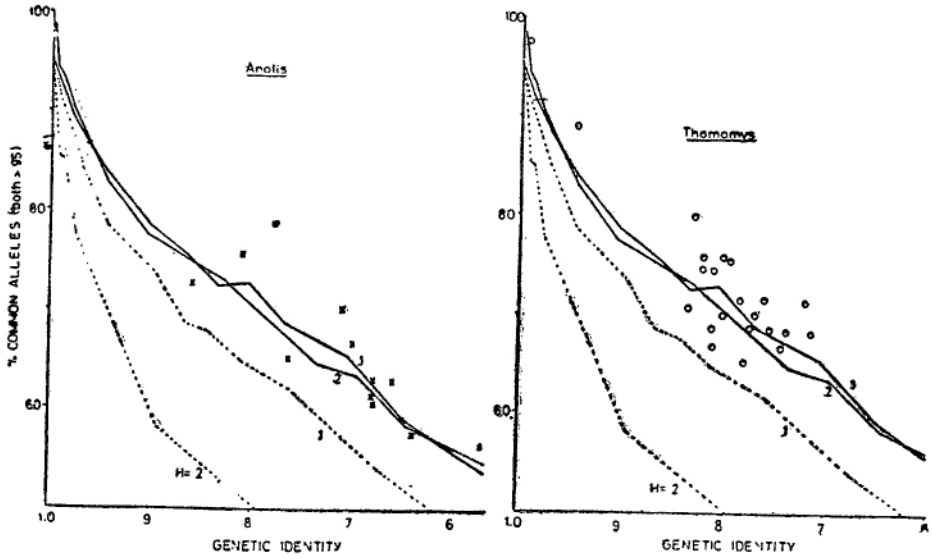

FIG. 7. Percentage of alleles which are shared and monomorphic in both populations. Solid=Varying mutation model; dotted-Constant mutation model.

The decay in the percentage of shared common monomorphic alleles is given in Figures 7 and 8. The varying mutation model retains alleles which have monomorphic frequencies in both populations longer than the constant mutation model. Both sets of data from natural populations show the elevated frequencies of common monomorphic alleles which are indicative of the varying mutation model. In contrast, when the common allele is at monomorphic frequency in only one population, the varying mutation model predicts a more rapid decline than the constant mutation model. Again; the data are more concordant with the predictions of the varying mutation model.

Figure 9 presents the final results, the decline of the percentage of total alleles shared by two populations. Both sets of data from natural populations show an inflated percentage of shared alleles compared to either theoretical model. Note that both models will yield an almost identical expectation when average heterozygosity of a population is about 0.10, which is expected to be closer to the observed values than the 0.20 curves. Keep in mind that the values for the theoretical curves were calculated for all alleles existing in the population. Since the

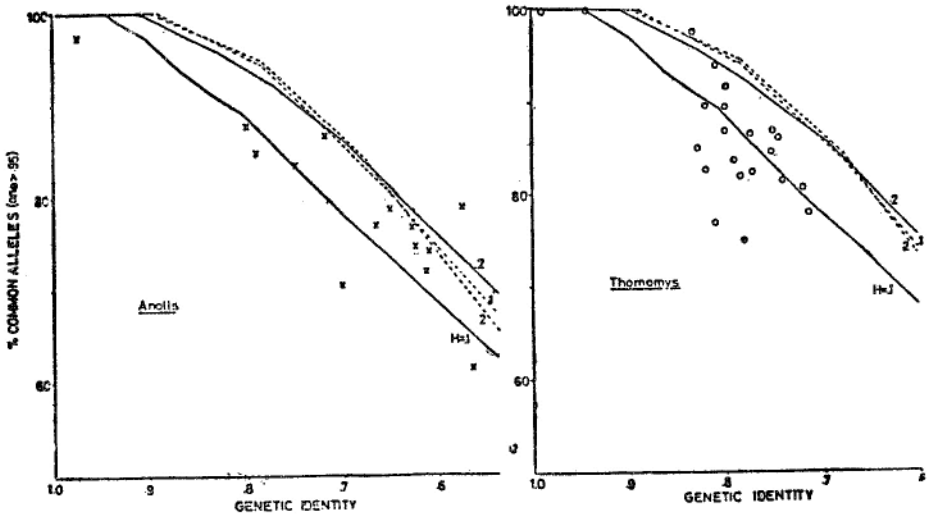

FIG. 8. Percentage of alleles which are shared and are monomorphic in only one population. Solid=Varying mutation model; Dotted=Constant mutation model.

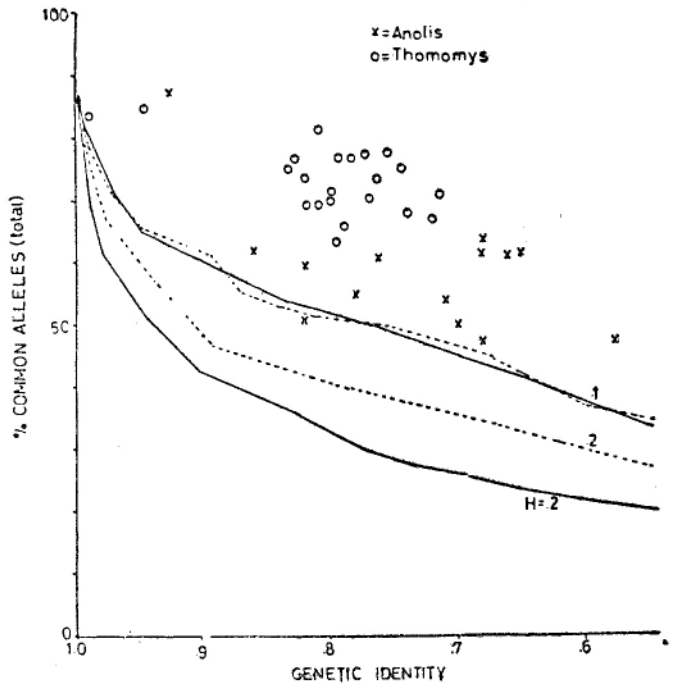

FIG. 9. Percentage of total alleles shared by two populations. Solid=Varying mutation model; Solid=Constant mutation model.

samples are much smaller than the population, many alelles, especially rare alleles, are excluded from the sample. We can estimate the number of alleles excluded from the sample by examining the allele frequency distribution. Under the constant mutation model, the average number of alleles less than 0.01 in a population with average heterozygosity of 0.10 will be 0.26 or 15% of the total alleles in the population. If we add this number of alleles to the number observed in our samples, and assume that there will be few or no shared rare alleles (see Tables 1 and 2 for justification) then the precentage of total alleles that are held in common will be about 60.7% for *Thomomys* populations with genetic identities of 0.7-0.8, 46.7% for *Anolis* populations with genetic identities between 0.7-0.8 and, 50.5% for *Anolis* with I=0.6-0.7. This correction suggests that much of the discrepancy in Figure 9 may be due to the statistical effect on rare alleles caused by the small number of genes sampled for each population.

DISCUSSION

The patterns of decline in the numbers of alleles held in common between two populations have been investigated theoretically using two models. One assumes equal mutation rates at all loci (constant mutation model), the second assumes that these rates are locus specific (varying mutation model). Comparison of these theoretical predictions with a very limited set of data suggests reasonable concordance specially when comparing the varying mutation model with alleles in some frequency ranges. There is especially good concordance for alleles which are essentially monomorphic, but poorer fit for other classes, such as total alleles. Some lack of fit can be attributed to the sample size of genes studied for natural populations. Other reasons for, lack of fit can be invoked, however. Two important factors can be considered, One must first ask whether balancing selection could be acting to distort the pattern of allele-sharing in natural populations. Examination of Table 4 suggests that this is not likely to be an important factor, since the balanced alleles should inflate the number of polymorphic shared alleles. The lack of any excess could be due to balanced alleles with a low (less than 0 2) or high (greater than 0.8) equilibrium frequency. For such alleles, balancing selection in association with finite population size can actually accelerate the allele towards loss or fixation (14).

A second potential cause of discrepancies could be the historical occurrence of population bottlenecks during the founding of demes which comprise the two sets of natural populations which were utilized for the present comparisons. Both *Thomomys*, which has a patchy distribution of populations in the western United States, and the *Anolis* populations,

which clearly arose through chance colonization of the islands in the eastern Caribbean, have population histories which suggest that bottlenecks could have played an important role in determining the pattern of genetic differentiation. Studies to be presented elsewhere (12), on the effects of bottlenecks in the allele frequency spectrum show that rare alleles will be lost from the population following a bottleneck, while a large proportion of orginally polymorphic alleles may be shifted towards either loss or fixation at a locus. Such a pattern, if the bottleneck occurs in a colonizing deme while the progenitor deme remains unchanged, could act to reduce total allele number and to accelerate loss of polymorphic alleles shared by both populations. We will show (11) that this also acts to slightly increase alleles polymorphic in one population and monomorphic in a second. This is the pattern of allele sharing which we see in the data (Figure 6). Further theoretical studies on this problem are being carried out. An extended comparison of our theoretical results with the large set of data from electrophoretic surveys, which we have used for other analyses (2-4), will also be reported elsewhere.

ACKNOWLEDGEMENTS

This study was supported by a grant from the National Science Foundation DEB-8110220. We acknowledge the help of the staff of the National Institute of Genetics of Japan, Dr. Takeo Maruyama, Head of the Laboratory of Physiological Genetics. and the Director of the Institute, Dr. E. Matsunaga for allowing the use of their facilities for these studies.

REFERENCES

1. KIMURA, M: *The Neutral Theory of Molecular Evolution.* Cambridge University Press, Cambridge, 1983
2. FUERST PA, R CHAKRABORTY, M NEI: Statistical studies on protein polymorphism in natural populations. I. Distribution of single locns heterozyosisy. *Genetics* 86:455-483, 1977
3. CHAKRABORTY R, PA FUERST, M NEI: Statistical studies of protein polymorphism in natural populations. II. Gene differentiation between populations. *Genetics* 88:367-390, 1978
4. CHAKRABORTY R, PA FUREST, M NEI: Statistical studies of protein polymorphism in natural populations. III. The distribution of allele frequencies and the number of alleles per locus. *Genetics* 94:1039-1063, 1980
5. NEI M, WH LI: Probability of identical monomorphism in related species. *Genet Res* 25:31-43, 1975
6. LI WH, M NEI: Persistance of common alleles in two related species. *Genetics* 86:901-914, 1977
7. MARUYAMA T: Stochastic integrals and their application to population genetics In: Kimura, M. (ed.) *Molecular Evolution, Protein Polymorphism and the Neutral Theory.* Springer Verlag, New York, p. 151-167, 1982

8. NEI M: Genetic distance between populations. *Amer Nat* 106:283-292, 1972
9. PATTON JL, SY YANG: Genetic variation in *Thomomys bottae* pocket gophers, macrogeographic patterns. *Evolution* 31:697-720, 1977
10. GORMAN GC, YJ KIM: *Anolis* lizards of the Eastern Caribbean: a case study in evolution. II. Genetic relationships and genetic variation of the *bimaculatus* group. *Systemat Zool* 25:62-67, 1976
11. NEI M, R CHAKRABORTY, PA FUERST: Infinite allele model with varying mutation rate. *Proc Nat Acad Sci USA* 73:4164-4168, 1976
12. FUERST PA, T MARUYAMA: Population bottlenecks and non-equilibrium models in population genetics. III. The use of stochastic integrals to study allele numbers in growing post-bottleneck populations. *Genetics* (submitted), 1985
13. NEI M: *Molecular Population Genetics and Evolution* North-Holland, Amsterdam, 1975
14. ROBERTSON A: Selection for Heterozygotes In Small Populations. *Genetics* 47:1291-1300, 1962