A Statistical Test for Detecting Geographic Subdivision[1]

Richard R. Hudson, Dennis D. Boos,† and Norman L. Kaplan‡*

*Department of Ecology and Evolutionary Biology, University of California at Irvine; †Department of Statistics, North Carolina State University; and ‡Division of Biometry and Risk Assessment, National Institute of Environmental Health Sciences.

A statistical test for detecting genetic differentiation of subpopulations is described that uses molecular variation in samples of DNA sequences from two or more localities. The statistical significance of the test is determined with Monte Carlo simulations. The power of the test to detect genetic differentiation in a selectively neutral Wright-Fisher island model depends on both sample size and the rates of migration, mutation, and recombination. It is found that the power of the test is substantial with samples of size 50, when $4Nm < 10$, where N is the subpopulation size and m is the fraction of migrants in each subpopulation each generation. More powerful tests are obtained with genes with recombination than with genes without recombination.

Introduction

The extent to which natural populations are subdivided into genetically differentiated subpopulations is of interest to a wide variety of biologists, including evolutionary biologists, conservation biologists, and plant and animal breeders. Traditional indexes, such as Wright's (1951) fixation indexes, indicating the amount of genetic differentiation between local populations are based on allele frequencies and have frequently been estimated by using data on enzyme polymorphisms (Nei 1987, pp. 159–166, 187–192; Weir 1990, Chap. 5). Modern techniques of molecular biology make it possible to gather detailed information about DNA-level genetic variation in populations. Recently, Nei (1982) and Lynch and Crease (1990) have introduced adaptations of Wright's fixation indexes, adaptations that are particularly suited for data on DNA sequence and restriction-map variation. Various ways of estimating gene flow with molecular data have also been studied (Slatkin and Barton 1989).

A natural question to ask is whether estimates of genetic differentiation are compatible with the null hypothesis that the subpopulations are not genetically different. Clearly, it would be desirable to have a statistical test for this purpose. Most estimates of genetic differentiation are not useful test statistics because their distribution under the null hypothesis is not known, and so it is not clear how to calculate significance levels. We describe here a simple nonparametric method for testing the null hypothesis of no genetic differentiation between subpopulations at different localities. The method is permutation based, using Monte Carlo simulations to estimate significance levels, and is particularly suited for data on nucleotide variation in samples from two or more localities. In a recent paper, Lynch and Crease (1990) suggested this approach

1. Key words: geographic subdivision, migration, permutation test, nucleotide polymorphism.

Address for correspondence and reprints: Richard R. Hudson, Department of Ecology and Evolutionary Biology, University of California, Irvine, Irvine, California 92717.

Mol. Biol. Evol. 9(1):138–151. 1992.

0737-4038/92/0901-0010$02.00

but did not investigate its properties. Stoneking et al. (1990) have also recently employed this approach in analyzing geographic variation in human mitochondrial DNA.

The method is based on the following simple idea: Suppose that DNA is collected from individuals from two or more localities. The nucleotide sequences (or restriction maps) of a genetic region are determined for each of the sampled individuals. These nucleotide sequences, referred to as "sampled DNA sequences," are the data on which a statistical test of geographic subdivision is based. A test statistic, which estimates a measure of the genetic differentiation of the subpopulations and which will be specified later, is calculated from the sampled DNA sequences. The problem is to determine whether the observed value of the test statistic is statistically significant. To do this we use a permutation-based procedure. If there is no genetic differentiation of subpopulations, then which locality a DNA sequence comes from is irrelevant. Thus randomly assigning the sampled DNA sequences to the different localities will produce a "new" sample that is distributionally equivalent to the original sample. By repeatedly randomly assigning the sequences in this way, the statistical significance of the observed value of the test statistic can be estimated.

The traditional method of testing for genetic differentiation of populations is a χ^2 test of allele frequencies in samples from different localities (Nei 1987, p. 227). This test can be directly adapted to use with nucleotide variation by treating each distinct haplotype as an allele. (A haplotype is a particular sequence or restriction map.) If the number of haplotypes is so large that the expected numbers of each haplotype from each locality are very small, then this test is not very powerful. In the extreme case, where each haplotype is unique in the sample, there is no information about genetic differentiation in the sample frequencies of haplotypes. This situation can occur when a large region of DNA is sequenced or when it is mapped with many different restriction enzymes. To examine this issue, we will compare the power of the permutation-based test with that of the χ^2 test, for a variety of sample sizes, mutation rates, and recombination rates.

Two distinct classes of test statistics will be considered. One class, referred to as "haplotype statistics," is based on haplotype frequencies in the sample, as in the χ^2 test. For these statistics, it does not matter whether two haplotypes differ by one nucleotide or by hundreds. The other class of test statistics, the sequence statistics, uses the information on the number of differences between sequences. The statistics γ_{ST} (Nei 1982) and N_{ST} (Lynch and Crease 1990) are sequence statistics. We will investigate several sequence statistics to compare their performance with that of haplotype test statistics.

Statistics

The traditional method of testing for genetic differentiation of samples from different localities is a χ^2 test based on allele frequencies (Nei 1987, p. 227). If each distinct haplotype (sequence or restriction map) is treated as an allele, then this test can be directly applied. The test statistic is

$$\chi^2 = \sum_{i=1}^{L} \sum_{j=1}^{K} \frac{(n_{ij} - n_i \hat{p}_j)^2}{n_i \hat{p}_j}, \tag{1}$$

where L is the number of localities, K is the number of haplotypes in the total sample,

n_i is the sample size from locality i, n_{ij} is the observed number of copies of haplotype j from locality i, and $\hat{p}_j$ is the frequency of haplotype j in the total sample. Indeed,

$$\hat{p}_j = \frac{\sum_{i=1}^{L} n_{ij}}{\sum_{i=1}^{L} n_i}.$$

This statistic is the standard χ^2 statistic used to test for homogeneity in an $L \times K$ contingency table. Under the null hypothesis, χ^2 should be approximately χ^2 distributed with $(L-1)(K-1)$ degrees of freedom, and thus the statistical significance of an observed value of χ^2 can be assessed by comparison with the critical values of the χ^2 distribution. In practice, rare haplotypes are lumped together so that the expected numbers of each haplotype in each locality are not too small. Results will be presented for this method where rare haplotypes are lumped so that the expected number of each haplotype in each locality is at least one.

In addition to χ^2, we will consider another haplotype statistic, H_{ST}, defined as

$$H_{ST} = 1-(H_S/H_T), \tag{2}$$

where H_S is a weighted average of estimated haplotype diversities in the subpopulations, and H_T is an estimate of the haplotype diversity of the total population; that is,

$$H_S = \sum_{i=1}^{L} w_i H_i \tag{3a}$$

and

$$H_T = \frac{n}{n-1}\left(1 - \sum_{j=1}^{K} \hat{p}_j^2\right), \tag{3b}$$

where

$$H_i = \frac{n_i}{n_i - 1}\left(1 - \sum_{j=1}^{K} \hat{p}_{ij}^2\right), \tag{4}$$

$\hat{p}_{ij}$ is the frequency of the jth haplotype in the sample from locality i, w_i is a weighting factor for population i to be specified later, and n is the total sample size, ($n = \sum_{i=1}^{L} n_i$). H_{ST} can be considered an estimator of Nei's G_{ST}. Nei has suggested an alternative estimator of G_{ST}, which could be used as a test statistic:

$$\hat{G}_{ST} = 1-(H_S/\tilde{H}_T), \tag{5}$$

where

$$\tilde{H}_T = 1 - \sum_{j=1}^{K} \hat{p}_j^2 + H_S/\tilde{n}L, \tag{6}$$

and where $\tilde{n}$ is the harmonic mean of the sample sizes from the localities (Nei 1987, pp. 190–192). $\tilde{H}_T$ is an unbiased estimate of total gene diversity (Nei and Chesser 1983). In testing for genetic differentiation, we have found that $\hat{G}_{ST}$ has power identical to that of H_{ST}, in every case that we have examined.

The other statistics that we will consider are sequence statistics and require some additional notation. To simplify the discussion, we consider only two localities, but the arguments can easily be generalized. Let $d_{ij,lk}$ denote the number of differences (restriction sites or nucleotide sites) between the jth sequence from locality i and the kth sequence from locality l. The average number of differences between sequences from within locality i is denoted K_i, $i = 1, 2$, and the average number of differences between two sequences regardless of their locality is denoted K_T. Then,

$$K_i = \frac{\sum_{j=1}^{n_i-1} \sum_{k=j+1}^{n_i} d_{ij,ik}}{\binom{n_i}{2}}, \quad i = 1, 2 \tag{7}$$

and

$$K_T = \frac{\sum_{i=1}^{2} \sum_{j=1}^{n_i-1} \sum_{k=j+1}^{n_i} d_{ij,ik} + \sum_{j=1}^{n_1} \sum_{k=1}^{n_2} d_{1j,2k}}{\binom{n_1+n_2}{2}}. \tag{8}$$

One sequence statistic that we will consider is

$$K_{ST} = 1-(K_S/K_T), \tag{9}$$

where K_S is a weighted average of the K_1 and K_2, and where

$$K_S = wK_1+(1-w)K_2, \tag{10}$$

where w is in the interval (0, 1). Simulations, carried out as described in the Simulation Methods section below, suggest that the most powerful test based on K_{ST} is obtained with $w = n_1/(n_1+n_2)$. Unless otherwise specified, K_{ST} will be calculated with this weighting. The statistic K_{ST} is identical to the sequence statistic γ_{ST} of Nei (1982), except for the choice of w, which for γ_{ST} is proportional to the actual subpopulation size—or, typically, in practice—1 divided by the number of subpopulations. The statistic N_{ST} of Lynch and Crease (1990) is also very similar to K_{ST}, differing in the weighting factor w and in replacing K_T by the average number of differences between sequences from *different* localities. (In contrast, K_T is the average number of differences between sequences, regardless of the locality from which they were drawn.)

The statistics γ_{ST} and N_{ST} are intended to provide a measure of the extent of genetic differentiation in a subdivided population. In contrast, our goal is to test whether two or more localities are genetically different. For this reason we will seek weightings which maximize the power that the tests have to detect genetic differentiation, while acknowledging that these weightings may not be appropriate for other purposes.

The statistics H_{ST} and H_S can be calculated with the same formulas as are K_{ST} and K_S if, for each i, j, k, and l, the value of $d_{ij,lk}$ is replaced by $\delta_{ij,lk}$, where $\delta_{ij,lk}$ is 0 if $d_{ij,lk}$ is 0 and is 1 otherwise. To study a statistic that takes account of the number of nucleotide differences between different haplotypes but does not give as much weighting to large numbers of differences, we considered the statistics K^*_{ST} and K^*_S. These are calculated as are K_{ST} and K_S, respectively, except that $d_{ij,lk}$ is replaced by $\log(1+d_{ij,lk})$. For example,

$$K_i^* = \frac{\sum_{j=1}^{n_i-1} \sum_{k=j+1}^{n_i} \log(1+d_{ij,ik})}{\binom{n_i}{2}}, \quad i = 1, 2. \tag{11}$$

The logarithm function is used to downweight the cases where $d_{ij,lk}$ is large. We have not corrected the $d_{ij,lk}$ for multiple mutations at the same site.

Two other statistics, Z and Z^*, referred to as "rank statistics," will also be examined. To calculate the rank statistic, one first rank orders all the $d_{ij,lk}$ values in order of increasing size. The Z statistic is a weighted sum of Z_1 and Z_2, where Z_i is the average of the ranks of all the $d_{ij,lk}$ values for pairs of sequences from within locality i. The null hypothesis is rejected if the Z statistic is too small. Z^* is the weighted sum of Z^*_1 and Z^*_2, where these are the average of the logarithm of 1 plus the rank of the $d_{ij,lk}$ values for pairs of sequences from within locality 1 and locality 2, respectively.

The permutation-based method for assessing the statistical significance of an observed value of a test statistic will be described in terms of two localities, but the generalization to more than two localities is straightforward.

Suppose that we have both a sample size n_1 from locality 1 and a sample of size n_2 from locality 2. The null hypothesis that we want to test is that the two subpopulations are not genetically differentiated. Under the null hypothesis, the two samples are distributionally equivalent to two samples obtained by randomly partitioning a sample of size n_1+n_2 from a single population into two subsets, of sizes n_1 and n_2, respectively. Thus, if we randomly partition the pooled sample many times, calculating the statistic each time, then we can estimate the P value of the observed value of the statistic. (The P value is the probability of obtaining either the observed value of the statistic or a more extreme value.) For example, consider the test statistic K_{ST}. We denote the observed value of the statistic for our sample by $\langle K_{ST} \rangle$. Under the null hypothesis, we expect K_{ST} to be near 0, and so we will reject the null hypothesis if $\langle K_{ST} \rangle$ is too big. The P value of $\langle K_{ST} \rangle$ is estimated by the proportion of random partitions that result in $K_{ST} \geq \langle K_{ST} \rangle$. Large values of $\langle K_{ST} \rangle$ have small P values. In all cases, we used 1,000 random partitions to estimate the P value. We are, in effect, estimating the P value from the outcome of 1,000 Bernoulli trials; thus the variance of our estimate of the P value is $p(1-p)/1{,}000$, where p is the true P value. If the estimated P value is small, say, <0.05, then we reject the null hypothesis that the two subpopulations are not genetically differentiated.

The quantity K_T is a constant over partitions, since the average number of differences between sequences is the same no matter how the sequences are assigned to locality. It follows from the definition of K_{ST} that K_S is equivalent to K_{ST} as a test statistic, except that a small observed value of K_S, rather than a large value, would

lead to rejection of the null hypothesis. To see this, note that, if, for a particular partition, $K_{ST} > \langle K_{ST} \rangle$, then it is necessarily the case that the value of K_S for that partition is less than the observed value of K_S, $\langle K_S \rangle$. So the P value of $\langle K_S \rangle$ is identical to the P value of $\langle K_{ST} \rangle$. By exactly the same argument, the P value of $\langle H_S \rangle$ is identical to the P value of $\langle H_{ST} \rangle$.

The extension to more than two localities is immediate. The P value is estimated in the same way as before, except that now $\sum_{i=1}^{L} ni$ DNA sequences are partitioned into L subsets of sizes $n_1, n_2, \ldots, n_L$.

The P values of the other statistics—K_S^*, Z, Z^*, and H_S—are calculated in the same way. In addition, the P value of χ^2 can also be determined by this permutation-based method rather than by using the critical values from the χ^2 distribution.

When more than two localities are considered and the null hypothesis is rejected, it is of interest to ask whether the localities of any subset are not genetically differentiated. To answer this question properly involves simultaneous inference, and so we suggest the simpler approach of just looking at the pairwise comparisons.

Simulation Methods

We consider a single alternative model—namely, a selectively neutral Wright-Fisher island model—with two diploid subpopulations, each of size N. Properties of samples under this model depend on $4Nm$, $4Nr$, and $4Nu$, where m is the fraction of each subpopulation which is made up of migrants in each generation, r is the expected number of recombination events (crossovers) per generation per gamete in the region sequenced, and u is the neutral mutation rate per generation in the region sequenced. All mutations are assumed to be selectively neutral and are assumed to occur at sites not already segregating in the population. This is a version of the infinite-sites neutral model proposed by Kimura (1969). This model has been extensively studied for completely linked regions. [e.g., see Watterson 1975; Hartl and Clark 1989, chap. 6).] Samples can be quickly generated under this model without recombination by using the coalescent approach (Strobeck 1987). Minor modifications of Hudson's (1983) method using Strobeck's approach allow uniform recombination to be incorporated. All the results concerning the power of the tests are based on simulations carried out in this way, using the coalescent approach. For each parameter combination, 1,000 samples were generated. Each sample consists of haplotypes drawn from a population at statistical equilibrium. For each sample, all the test statistics were calculated, and the null hypothesis was tested by generating 1,000 partitions of the sample. The probability of rejection of the null hypothesis by a particular test statistic—i.e., the power of the test—was estimated as the proportion of the 1,000 samples for which the null hypothesis was rejected by that statistic. Recall that the null hypothesis is rejected if the estimated P value of the observed value of the test statistic is <0.05. The P value is estimated using the permutation-based method described in the preceding section. In the following paragraphs, we do not attempt to explore the entire parameter space but will examine a small number of cases which illustrate some properties of the test. The results are presented in tables 1–3.

For each of the statistics K_S, K_S^*, Z, Z^*, and H_S, it is necessary to assign a value to w, a weighting parameter. To find approximate values of w for which the power of the test is maximized, all values of w that are between 0.05 and 0.95, at intervals of 0.05, were tried for each statistic and for each of the n_1 values shown in table 1. The results (not shown) indicated that the power of the tests is quite sensitive to the factor

Table 1
Power of Tests to Detect Subdivided Population

				POWER OF[a]						
n_1	n_2	$\bar{H}_T$	$\bar{G}_{ST}$	K_S^*	Z^*	H_S	χ^2	χ^2 (table)[b]	K_S	Z
35 ...	5	0.88	0.054	0.57	0.61	0.62	0.42	0.37	0.44	0.47
30 ...	10	0.90	0.048	0.79	0.83	0.84	0.70	0.66	0.65	0.67
25 ...	15	0.91	0.051	0.86	0.90	0.89	0.86	0.83	0.70	0.76
20 ...	20	0.91	0.052	0.87	0.91	0.90	0.92	0.87	0.75	0.80

NOTE.—For each row, 1,000 samples were generated under a Wright-Fisher island model with two subpopulations. For all cases, $4Nu = 5.0$ and $4Nm = 2.0$, and complete linkage was assumed. For each sample, 1,000 random partitions were produced, and the P value of the observed value of the statistic was calculated. $\bar{H}_T$ is the average value of H_T calculated with eq. (3). $\bar{G}_{ST}$ is the average value of $\hat{G}_{ST}$ calculated with eq. (5). The weighting, $w = n_1/(n_1 + n_2)$ was used to calculate the statistic K_S. For K_S^*, Z, Z^*, and H_S, the weighting $w_1 = (n_1-2)/(n_1+n_2-4)$ was used. If the observed value of the statistic had $P \leq 0.05$, the null hypothesis of no genetic differentiation was rejected.

[a] Data are fractions of samples for which null hypothesis was rejected when indicated statistic was used.

[b] Obtained using critical values of χ^2 distribution.

w, with much reduced power when suboptimal values of w are used. As expected, when $n_1 = n_2$, $w = 0.5$ is optimal for all statistics. For K_S it was found that the most powerful test was obtained with a weighting of $w = n_1/(n_1 + n_2)$. For the other statistics, $w = (n_1-2)/(n_1+n_2-4)$ was close to optimal for the parameter values in table 1. These are the weightings used in table 1. This issue of optimal w values needs more thorough investigation for different values of n_1 and n_2. For a given total sample size, the power of the test is largest when $n_1 = n_2$, as is also shown in table 1.

The χ^2 statistics were calculated by lumping haplotypes so that the expected number in each locality is ≥ 1. In all samples generated for table 1, there were at least two haplotypes present in each sample after this lumping, and thus the test statistic could be calculated. With smaller mutation rates, such as some of the cases in table 3, there remained only one haplotype after lumping, and no χ^2 test was possible. In these cases the power reported in the table is the proportion of all 1,000 samples for which the test (a) was possible and (b) resulted in rejection of the null hypothesis.

Results

The results are shown in tables 1–3. Although there is no single test that is best for every parameter combination considered, there are certain patterns that are discernible in tables 1–3. In every case, the power of K_S is less than or equal to that of K_S^*, and the power of Z is less than or equal to that of Z^*. Thus K_S and Z are inferior statistics for our purpose and will not be discussed further. The test based on χ^2, using the critical values of the χ^2 distribution [indicated by χ^2 (table)] is always less powerful than the test based on χ^2 with the P values estimated by the permutation-based method. At either low sample sizes or high mutation rates (when most haplotypes are in very low frequency and when lumping is important) the difference in power can be substantial, but in other cases the differences are not great. The test using the critical values of the χ^2 distribution will not be considered further. This leaves us with the following four test statistics to discuss: K_S^*, Z^*, H_S, and χ^2.

For almost all cases without recombination, the test based on χ^2 is most powerful. The only exceptions are either with very large mutation rate ($4Nu = 15$ and $n_1 = n_2 = 25$) or with rather small sample size in one or both localities (with $4Nu = 5$ and n_1

Table 2
Power of Tests: Effects of Sample Size, $4Nm$, and $4Nr$

		$4Nr = 0$[a]								$4Nr = 20$[b]							
			Power[c] of								Power[c] of						
$n_1 (=n_2)$	$4Nm$	$\bar{G}_{ST}$	K_S^*	Z^*	H_S	χ^2	χ^2 (table)	K_S	Z	$\bar{G}_{ST}$	K_S^*	Z^*	H_S	χ^2	χ^2 (table)	K_S	Z
10	5.0	0.030	0.31	0.33	0.29	0.29	0.18	0.24	0.25	0.020	0.50	0.45	0.24	0.14	0.08	0.39	0.34
15	5.0	0.033	0.48	0.53	0.51	0.55	0.41	0.37	0.39	0.019	0.68	0.66	0.54	0.38	0.23	0.57	0.51
25	1.0	0.069	0.99	1.00	0.99	1.00	0.99	0.93	0.96	0.029	1.00	1.00	1.00	0.97	0.91	1.00	1.00
25	2.0	0.054	0.94	0.97	0.95	0.99	0.96	0.81	0.85	0.025	0.99	0.99	0.97	0.93	0.83	0.98	0.95
25	5.0	0.032	0.67	0.75	0.77	0.86	0.77	0.51	0.56	0.019	0.92	0.91	0.87	0.82	0.63	0.79	0.73
25	10.0	0.020	0.45	0.49	0.55	0.59	0.48	0.32	0.35	0.013	0.69	0.64	0.70	0.61	0.42	0.50	0.44
50	5.0	0.033	0.92	0.96	0.97	0.99	0.99	0.77	0.83	0.018	1.00	1.00	1.00	1.00	1.00	0.97	0.93
50	10.0	0.019	0.71	0.79	0.85	0.96	0.93	0.52	0.57	0.013	0.94	0.93	0.97	0.99	0.97	0.81	0.71

[a] Average $H_T = 0.91$ or 0.92.
[b] Average $H_T = 0.97$.
[c] Estimated as in table 1, with $4Nu = 5.0$ and either $4Nr = 0.0$ or $4Nr = 20.0$, where r is expected number of recombination events (crossovers) per generation per gamete in region sequenced.

Table 3
Power of Tests: Effects of $4Nu$ and $4Nr$

	$4Nr = 0$								$4Nr = 16Nu$[a]							
		Power[b]								Power[b]						
$4Nu$	$\bar{H}_T$ ($\bar{G}_{ST}$)	K_S^*	Z^*	H_S	χ^2	χ^2 (table)	K_S	Z	$\bar{H}_T$ ($\bar{G}_{ST}$)	K_S^*	Z^*	H_S	χ^2	χ^2 (table)	K_S	Z
0.156 ...	0.24 (0.041)	0.16	0.16	0.16	0.33	0.22	0.16	0.16	0.26 (0.039)	0.19	0.19	0.18	0.34	0.23	0.18	0.19
0.313 ...	0.40 (0.039)	0.31	0.31	0.29	0.39	0.35	0.27	0.30	0.41 (0.040)	0.31	0.30	0.29	0.40	0.34	0.29	0.29
0.625 ...	0.57 (0.040)	0.44	0.44	0.41	0.50	0.47	0.39	0.42	0.62 (0.040)	0.47	0.47	0.46	0.56	0.50	0.44	0.44
1.25	0.72 (0.040)	0.53	0.55	0.54	0.65	0.57	0.47	0.50	0.79 (0.040)	0.61	0.61	0.61	0.74	0.65	0.54	0.54
2.5	0.84 (0.038)	0.61	0.66	0.66	0.78	0.70	0.51	0.54	0.92 (0.029)	0.79	0.80	0.79	0.86	0.74	0.67	0.64
5.0	0.91 (0.031)	0.70	0.75	0.79	0.86	0.78	0.52	0.57	0.97 (0.019)	0.92	0.90	0.88	0.81	0.64	0.80	0.73
10.0	0.95 (0.025)	0.77	0.84	0.86	0.87	0.78	0.53	0.59	0.99 (0.010)	0.98	0.95	0.86	0.54	0.31	0.92	0.84
15.0	0.97 (0.020)	0.78	0.86	0.89	0.84	0.68	0.49	0.55	0.99 (0.007)	0.99	0.98	0.73	0.34	0.17	0.97	0.91

[a] With $4Nr$ equal to four times mutation parameter.
[b] Estimated as in table 1, with $n_1 = n_2 = 25$ and $4Nm = 5.0$.

$= n_2 = 10$, or $n_1 = 35$, $n_2 = 5$, or $n_1 = 30$, $n_2 = 10$). With recombination, χ^2 is most powerful for sample sizes of 50 or for low mutation rates ($4Nu \leq 2.5$). For most other cases with recombination, K_S^* is most powerful, although, with unequal subsample sizes, H_S and Z^* are slightly more powerful than K_S^*, as shown in table 1.

The results in table 2 show that the power of the tests decreases rapidly as $4Nm$ increases. For $n_1 = n_2 = 25$ and $4Nr = 0$, the power of the tests decreases from nearly 100%, at $4Nm = 1.0$, to ~50%, at $4Nm = 10.0$. This decrease in power is not surprising, since it is well known that the island model behaves as a single large panmictic population if the migration rate is large (Hartl and Clark 1989, pp. 310–315). For the parameters in tables 2 and 3, the presence of recombination increases the tests' power to detect subdivision. For example, with $n_1 = n_2 = 10$, $4Nm = 5.0$, $4Nu = 5.0$, and $4Nr = 0$, the power of the test with Z^* is 0.33, whereas, with $4Nr = 20.0$, the power with Z^* is 0.45. These results suggest that, for detecting population differentiation, genes with recombination are more informative than are genes without recombination, everything else being equal. Hence, for making inferences of this sort about genetic differentiation of subpopulations, nuclear genes may be more informative than are mitochondrial genes. One must keep in mind that the neutral mutation rate may be considerably larger for typical mitochondrial genes than for nuclear genes and that the effective population size of mitochondria may be significantly less (perhaps by a factor of 4) than that for nuclear genes.

From table 2, it is clear that larger samples can increase the power of the test substantially. For example, when $4Nm = 5.0$ and $4Nr = 0$, the power of the test with Z^* and χ^2 goes from 0.33 and 0.29, respectively, to 0.96 and 0.99, respectively, as the total sample size goes from 20 to 100.

Table 3 shows that the mutation rate also influences the power considerably. However, note that for $n_1 = n_2 = 25$ and $4Nm = 5$ there seems to be, for substantial increases in $4Nu$ beyond 5.0, only a minor gain in power, when there is no recombination. Thus, for regions with low recombination rates, increased power may be obtained more easily by increasing the sample sizes rather than by increasing the size of the region sequenced. However, with recombination, the power continues to grow as increasingly larger regions are examined.

Applications

Our first example is a data set describing restriction-site and insertion/deletion polymorphism in a 45-kb region of the *white* locus on the X chromosome in *Drosophila melanogaster* (Miyashita and Langley 1988). Sixty-four DNA sequences were sampled from three different locations: 20 from Raleigh, N.C., 27 from Texas, and 17 from Fukuoka, Japan. A total of 109 polymorphic sites were detected by using four- and six-cutter restriction enzymes. For simplicity we considered only the six-cutter results for which there were 16 restriction site and 25 insertion/deletion polymorphisms. In table 4 the estimated P values of the various statistics for the three pairwise comparisons are given. These results for the sequence statistics suggest that the Raleigh and Texas subpopulations are not genetically differentiated but that the Fukuoka subpopulation is different. In view of the geographical distances involved, these results are not surprising. In contrast to the sequence statistics, the haplotype statistics have fairly large P values for all pairs of localities. We note that H_T, the haplotypic diversities are high (>0.99), and consequently we expect the haplotype statistics to lack power to detect genetic differentiation.

Table 4
Analysis of Six-Cutter Data of Miyashita and Langley (1988)

		ESTIMATED P VALUE[a]						
	H_T	K_S^*	Z^*	H_S	χ^2	χ^2 (table)	K_S	Z
Texas vs. Fukuoka ($n_1 = 27$; $n_2 = 17$) ...	0.995	0.00	0.00	0.08	0.10	>0.05	0.00	0.00
Raleigh vs. Fukuoka ($n_1 = 20$; $n_2 = 17$) ...	0.997	0.00	0.00	0.18	0.52	>0.05	0.00	0.00
Raleigh vs. Texas ($n_1 = 20$; $n_2 = 27$) ...	0.994	0.23	0.34	0.10	0.52	>0.05	0.37	0.32

[a] Based on 1,000 random partitions.

For a second example, we consider the data of Kreitman and Aguadé (1986*b*), describing restriction-enzyme and insertion/deletion variation in a 2.7-kb region of the genome surrounding the *Adh* locus in *D. melanogaster*. A total of 87 DNA sequences were sampled: 60 from Raleigh and 27 from Putah Creek, Calif. There are 17 restriction-site and 11 insertion/deletion polymorphisms. In addition, there are two electrophoretically distinct alleles, Adh^S and Adh^F, which differ by a single amino acid. Thus, there are a total of 29 polymorphic sites. Three of the insertion/deletion polymorphisms appear to have more than two sizes in the sample (see Kreitman and Aguadé 1986*b*, fig. 3). To simplify the analysis, these positions will be ignored. The estimated P values of the statistics are shown in table 5. The sequence statistics all have low P values, suggesting that the subpopulations are genetically different. The null hypothesis is not rejected by the haplotype statistics, but again the haplotypic diversity is quite high, and so we do not expect the haplotype statistics to be very powerful.

The frequencies of Adh^F in Raleigh and Putah Creek are ~0.3 and ~0.6, respectively. There is strong evidence that both the *Adh* polymorphism and its geographic pattern of allele frequencies are maintained by natural selection (Oakeshott et al. 1982; Kreitman and Aguadé 1986*a*; Hudson et al. 1987; Hudson and Kaplan 1988). Thus it is possible that the differences between the Raleigh and Putah Creek populations are primarily the result of relatively strong selection acting on the protein polymorphism, rather than just the result of genetic drift with limited migration between the populations (for a theoretical treatment of this situation, see Kaplan et al. 1991). To test this hypothesis, it is useful to examine the two Adh^F allele groups separately, to ascertain whether the Adh^F alleles in Putah Creek are genetically different from the Adh^F alleles of Raleigh. And, similarly, the Adh^S alleles can be tested in this way. In table 5 the estimated P values are given for the two subsamples—one consisting of only Adh^F DNA sequences and one consisting only of Adh^S DNA sequences—and in both cases the results are not statistically significant. Thus, we cannot reject the null hypothesis that the subpopulations are genetically undifferentiated neither for the Adh^F sequences alone nor for the Adh^S sequences alone. This is consistent with the hypothesis that there is a relatively high rate of migration between the subpopulations but that, despite the migration, selection maintains the allele frequency difference (Kaplan et al. 1991).

Table 5
Analysis of Data of Kreitman and Aguadé (1986*b*)

		Estimated *P* Value[a]						
	H_T	K_S^*	Z^*	H_S	χ^2	χ^2 (table)	K_S	Z
All data ($n_1 = 60$; $n_2 = 27$) ...	0.96	0.01	0.01	0.12	0.38	>0.05	0.01	0.00
Adh^S only ($n_1 = 42$; $n_2 = 11$) ...	0.93	0.21	0.22	0.36	0.34	>0.05	0.14	0.17
Adh^F only ($n_1 = 18$; $n_2 = 16$) ...	0.91	0.45	0.42	0.32	0.39	>0.05	0.54	0.61

[a] Based on 1,000 random partitions.

Discussion

We have proposed a permutation-based statistical test for detecting genetic differentiation of subpopulations at different localities and have examined its power to detect differentiation under a selectively neutral Wright-Fisher island model. Test statistics based on haplotype frequencies and on the number of nucleotide differences between sequences were compared. Simulation results indicate how the power of the tests depends on the sample size and on the rates of mutation, migration, and recombination. For high mutation rates (i.e., $H_T \sim 1$) or small sample sizes, the sequence-based statistics are more powerful. For low mutation rates or large sample sizes, tests based on haplotype frequencies are more powerful. The simulation results suggest that, for given sample sizes, when H_T is below a critical value, χ^2 is best—whereas, if H_T is above the critical value, K_S^* is usually best. For example, with $n_1 = n_2 = 25$, if $H_T < 0.95$, then χ^2 appears to be most powerful, and, if $H_T > 0.95$, then K_S^* is usually best. For smaller sample sizes, the critical value of H_T is lower, ~ 0.91 for $n_1 = n_2 = 15$. For $n_1 = n_2 = 50$, the critical value appears to be >0.97. An approximate rule is that the χ^2 statistic is most powerful if $H_T < 1 - [1/\min(n_1, n_2)]$. There are exceptions to this rule, even among the small number of cases that we have examined, as shown in table 1 for the case of $n_1 = 25$ and $n_2 = 15$. This issue merits further investigation.

Since no simulations are needed when the χ^2 statistic is used, if the *P* values are obtained from the tabulated distribution of the χ^2 distribution, this is the easiest procedure to carry out. However, the results in tables 1–3 show that this test [χ^2(table)] is always less powerful than the permutation-based χ^2 test. Taking into account all our results, we suggest the following strategy for testing the null model:

1. Use χ^2(table) first. If the null model is rejected, then quit. If it is not, go to step 2.
2. Carry out the permutation tests using χ^2, K_S^*, and Z^*. If one is forced to choose only one statistic for the permutation test, then the following rules are suggested for choosing the statistic: (*a*) If $H_T < 1 - [1/\min(n_1, n_2)]$, then use χ^2. (*b*) If $H_T > 1 - [1/\min(n_1, n_2)]$, then use K_S^*.

The results in tables 2 and 3 indicate that genes with recombination result in more powerful tests. Two possible reasons for this are that recombination (1) increases H_T and (2) decreases the level of linkage disequilibrium between nucleotide sites. Since mitochondrial DNA has no recombination, nuclear DNA may therefore be

preferable for testing for genetic differentiation of subpopulations. However, results in table 3 suggest that samples of mitochondrial genes and samples of nuclear genes with comparable values of H_T would lead to tests having comparable power. Also, as mentioned earlier, $4Nm$ is likely to be smaller for mitochondrial genes than for nuclear genes, because mitochondrial genomes are haploid and often only transmitted maternally.

Acknowledgments

This work was supported in part by NIH grant GM42447 to R.R.H. Brian Metscher provided expert programming assistance.

LITERATURE CITED

HARTL, D. L., and A. G. CLARK. 1989. Principles of population genetics. Sinauer, Sunderland, Mass.

HUDSON, R. R. 1983. Properties of a neutral allele model with intragenic recombination. Theor. Popul. Biol. **23**:183–201.

HUDSON, R. R., and N. L. KAPLAN. 1988. The coalescent process in models with selection and recombination. Genetics **120**:831–840.

HUDSON, R. R., M. KREITMAN, and M. AGUADÉ. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics **116**:153–159.

KAPLAN, N. L., R. R. HUDSON, and M. IIZUKA. 1991. The coalescent process in models with selection, recombination and geographic subdivision. Genet. Res. **57**:83–91.

KIMURA, M. 1969. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. Genetics **61**:893–903.

KREITMAN, M., and M. AGUADÉ. 1986*a*. Excess polymorphism at the Adh locus in *Drosophila melanogaster*. Genetics **114**:93–110.

———. 1986*b*. Genetic uniformity in two populations of *Drosophila melanogaster* as revealed by filter hybridization of four-nucleotide-recognizing restriction enzyme digests. Proc. Natl. Acad. Sci. USA **83**:3562–3566.

LYNCH, M., and T. J. CREASE. 1990. The analysis of population survey data on DNA sequence variation. Mol. Biol. Evol. 7:377–394.

MIYASHITA, N., and C. H. LANGLEY. 1988. Molecular and phenotypic variation of the white locus region in *Drosophila melanogaster*. Genetics **120**:199–212.

NEI, M. 1982. Evolution of human races at the gene level. Pp. 167–181 *in* B. BONNE-TAMIR, T. COHEN, and R. M. GOODMAN, eds. Human genetics, part A: The unfolding genome. Alan R. Liss, New York.

———. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

NEI, M., and R. K. CHESSER. 1983. Estimation of fixation indices and gene diversities. Ann. Hum. Genet. **47**:253–259.

OAKESHOTT, J. G., J. B. GIBSON, P. R. ANDERSON, W. R. KNIBB, D. G. ANDERSON, and G. K. CHAMBERS. 1982. Alcohol dehydrogenase and glycerol-3-phosphate dehydrogenase clines in *Drosophila melanogaster* on three continents. Evolution **36**:86–96.

SLATKIN, M., and N. H. BARTON. 1989. A comparison of three indirect methods for estimating average level of gene flow. Evolution **43**:1349–1368.

STONEKING, M., L. B. JORDE, K. BHATIA, and A. C. WILSON. 1990. Geographic variation in human mitochondrial DNA from Papua New Guinea. Genetics **124**:717–733.

STROBECK, C. 1987. Average number of nucleotide differences in a sample from a single subpopulation: a test for population subdivision. Genetics **117**:149–153.

WATTERSON, G. A. 1975. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **10**:256–276.

WEIR, B. S. 1990. Genetic data analysis. Sinauer, Sunderland, Mass.
WRIGHT, S. 1951. The genetical structure of populations. Ann. Eugenics **15**:323–354.

MASATOSHI NEI, reviewing editor

Received November 19, 1990; revision received May 20, 1991

Accepted May 21, 1991