Statistical Properties of Two Tests that Use Multilocus Data Sets to Detect Population Expansions

David E. Reich, Marcus W. Feldman,† and David B. Goldstein**

*Department of Zoology, University of Oxford, United Kingdom; and †Department of Biological Sciences, Stanford University

We describe two methods for detecting population expansions based on variation at unlinked microsatellite loci. The tests were first used in a study of human demographic history that showed evidence for a Paleolithic human population expansion in Africa. Here, we provide a simple recipe for applying the tests to other data sets and describe the power of the tests as a function of the sample size, number of loci, mutation rate, diploid population size N_0, and time since expansion. An important property of the tests is that as long as the population doubles at least once every $0.1N_0$ generations, where N_0 now represents the pre-expansion population size, and the overall factor of expansion is sufficiently large, the signal of growth will be nearly identical to one generated by a sudden and massive expansion. This greatly simplifies the mathematical modeling necessary to evaluate the test results but also means that many patterns of growth will be indistinguishable using the tests. A second conclusion from our analysis is that the tests show different sensitivities to specific deviations from the biological and demographic models. Hence, more information can be garnered from the two tests taken together than from either alone.

Introduction

We describe two tests that use patterns of diversity at unlinked microsatellite loci to study demographic history (Reich and Goldstein 1998). Previous tests for population expansions, applied mainly to DNA sequence variation at the control region of mitochondria (Rogers and Harpending 1992), have yielded ambiguous results because they utilize information from only a single locus. The problem with single-locus studies is that the effect of a selective sweep on a linked gene cannot be distinguished from the effect of a population expansion. Multilocus data, however, can distinguish between demographic effects and selection because population growth produces a signal at all loci, while selection only influences linked loci.

The basis of our approach is that gene genealogies—the trees of ancestral relationships among sampled alleles—reflect a population's history of growth. For a constant-sized population, a gene genealogical tree tends to be dominated by only a few ancient bifurcations, and mutations occurring in the oldest parts of the tree tend to divide the alleles into a few distinct clusters (Donnelly 1996). On the other hand, for an expanding population, allele types do not tend to be clustered (Donnelly 1996). By taking advantage of the systematic differences in gene genealogies for the two scenarios and by assuming a simple stepwise mutation model for microsatellites whereby allele lengths change by a single unit and the direction of change is unbiased, we have developed within-locus and interlocus tests of demographic history (Reich and Goldstein 1998).

Two Tests for Population Expansion

Within-Locus k *Test*

The within-locus test (Reich and Goldstein 1998) is based on the assumption that microsatellite loci mutate according to a simple stepwise model, as well as on expectations about the systematic differences in the shapes of allele-length distributions for constant-sized and expanding populations. In the case of a constant-sized population, the typical allele-length distribution at a locus is expected to have several modes, corresponding to the small number of ancient bifurcations in the gene genealogy at that locus, while for an expanding population, the distribution is expected to have a single mode and to be more peaked (Reich and Goldstein 1998).

Peakedness is classically quantified using the kurtosis, γ_4/σ^4-3, where σ^4 is the variance squared and γ_4 is the fourth central moment of a frequency distribution. The kurtosis is expected to be positive if a distribution is more peaked than Gaussian and is expected to be negative otherwise. Reich and Goldstein (1998) defined a statistic, denoted *k*, which measures peakedness and is centered at 0 for the allele length distributions expected from a constant sized population. The combination of statistical moments was selected by trial and error using coalescent-based simulations (Appendix 2), and the value of *k* tends to decrease with the greater peakedness and the higher value of the kurtosis that occurs after an expansion.

Specifically, *k* is a linear combination of unbiased estimators—S^2 for the variance, Sig^4 for the variance squared, and Gam_4 for the fourth central moment (Appendix 1)—and an adjustment for the sample size, *n*. We used computer simulations (Appendix 2) to assign weights to these terms and adjusted the weights empirically, so that for a constant-sized population, *k* is positive about 50% of the time for as wide as possible a range of sample sizes and values of $N_0\nu$. (N_0 is the diploid effective population size, and ν is the mutation rate.)

$$k = 2.5*\mathrm{Sig}^4 + 0.28*S^2 - 0.95/n - \mathrm{Gam}_4 \quad (1)$$

The relationship between *k* and the kurtosis is evident in equation (1). Specifically, when the equation is divided through by $(-\mathrm{Sig}^4)$, the first and last terms form

Key words: demographic history, microsatellites, stepwise mutation model.

Address for correspondence and reprints: David B. Goldstein, The Galton Laboratory, Department of Biology, University College London, Wolfson House, 4 Stephenson Way, London NW1 2HE. E-mail: david.goldstein@zoology.ox.ac.uk.

Mol. Biol. Evol. 16(4):453–466. 1999

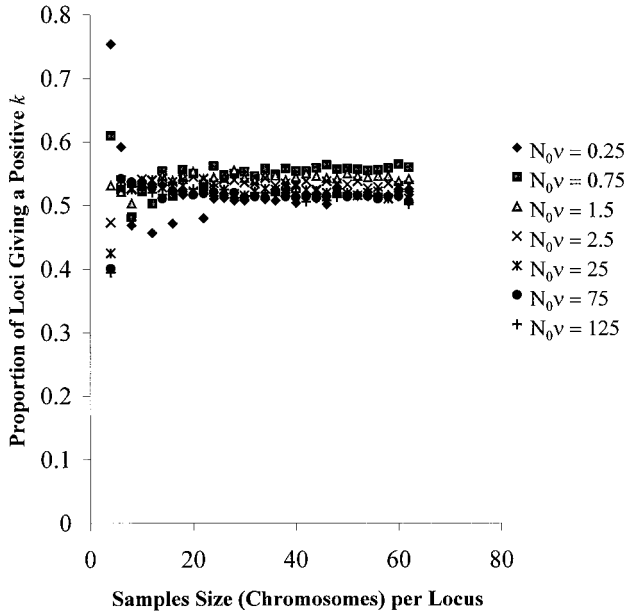


FIG. 1.—The probability of obtaining a positive k as a function of sample size and $N_0\nu$ under an assumption of constant population size. For each data point (i.e., combination of sample size and $N_0\nu$), we perform 10,000 simulations and calculate the probability of a positive k based on the simulations. The probability fluctuates in a narrow range, between 0.515 and 0.55, for sample sizes of at least 10 and $N_0\nu > 0.5$.

an estimator for $\gamma_4/\sigma^4 - 2.5$, and thus the equation has a similar form to the kurtosis. The additional terms of equation (1), although not directly related to the kurtosis, are included to ensure that the probability of a positive k is nearly constant for as wide as possible a range of sample sizes and values of $N_0\nu$. As shown in figure 1, the probability that k is positive is fully constrained between 0.515 and 0.55 for sample sizes of at least 10 and $N_0\nu > 0.5$. To make sure that $N_0\nu$ is at least 0.5, we use the relation that for a population of constant size, $E[\text{Var}] = 2N_0\nu$ (Zhivotovsky and Feldman 1995), implying that if the variance of an allele-length distribution at a locus (*Var*) is greater than 1, it is likely to meet the conditions on $N_0\nu$. Finally, to implement the test, we limit ourselves to loci for which the sample size is at least 10 and the variances of allele-length distributions are at least 1 and count the proportion of loci that give positive k values. In order to assess significance levels, we use a binomial distribution with the number of trials equal to the number of loci and the probability of a positive k set conservatively at its lower boundary of 0.515.

Interlocus g Test

Variance in the widths of allele-length distributions across loci is usually lower in an expanding population than in a constant-sized population (Reich and Goldstein 1998). Thus, by measuring the variance of the allele-length distribution at each locus (V_j) and considering the variance of these variances across loci (Var[V_j]), we obtain a statistic that can be compared with its theoretical expectation to test for size constancy. An interlocus test on this basis can be applied to any type of genetic marker. However, what makes our test specific for microsatellites is the incorporation of the analytical expectation of Var[V_j] for the specific case of the stepwise mutation model.

To implement this approach, we use the analytical result for E[Var[V_j]] in the case when microsatellite mutations are single step and mutation rates are the same across loci and there is no selection (Roe 1992, pp. 137–203; Zhivotovsky and Feldman 1995):

$$E[\text{Var}[V_j]] = (4/3)(\text{E}[\text{V}_\text{j}])^2 + (1/6)\text{E}[\text{V}_\text{j}]. \qquad (2)$$

Here, E[V_j] is the expected value of the variance at a given locus. Kimmel and Chakraborty (1996) have shown this equation to be true even for directionally-biased mutations.

To formulate the interlocus test explicitly, we consider the ratio Var[V_j]/E[Var[V_j]]; that is, the ratio of the observed to the predicted variance of the variance. Substituting the average variance across loci ($\bar{V}$) for the expected variance (E[V_j]) in equation (2), the test statistic we propose is as follows:

$$g = \frac{\text{Var}[V_j]}{\frac{4}{3}\bar{V}^2 + \frac{1}{6}\bar{V}}. \qquad (3)$$

An unusually low value of g may be interpreted as an indication of an expansion.

Significance levels for the interlocus test are obtained empirically, using a modified form of Hudson's (1990) coalescent computer simulation (see Appendix 2). A look-up table of 0.05 significance level cutoffs for a range of numbers of loci and sample sizes—including the $g = 0.333$ cutoff for 30 loci and 40 samples used in many of the examples in this paper—is given in table 1. Importantly, the behavior of g is nearly independent of the mutation rate and population size for all $N_0\nu > 0.5$ (fig. 2), so that the cutoffs provided in table 1 are valid for many values of $N_0\nu$ even though they were derived specifically for $N_0\nu = 0.63$. However, the interlocus test also applies when $N_0\nu < 0.5$, even though cutoffs are not presented for this case. Whatever the underlying value of $N_0\nu$, loci should never be selectively dropped from the analysis, even if they are monomorphic, because doing so would cause bias in the test results. An exception to this rule might be made in cases of a priori evidence for a lower mutation rate, for example due to an interrupted stretch of repeats in the microsatellites.

An alternative multilocus test of demographic history has recently been introduced for use with microsatellite data (Kimmel et al. 1998). This is based on an ''imbalance index,'' $\beta(t)$, which is a ratio of two estimators for $4N_0\nu$: one using the average variance across loci, and the other using the average heterozygosity. The imbalance index changes in response to population expansions and bottlenecks, because its numerator and denominator are differently affected by departures from size constancy. However, the intuitive effects of an expansion on β are not obvious, as in the case of g. In

Table 1
Fifth-Percentile Cutoffs for Interlocus Test

Number of Loci	Sample Size				
	10	20	40	80	160
5....	0.12	0.10	0.08	0.08	0.08
6....	0.13	0.12	0.11	0.10	0.10
7....	0.16	0.15	0.14	0.13	0.12
8....	0.19	0.16	0.15	0.15	0.14
10....	0.23	0.19	0.17	0.18	0.17
12....	0.26	0.21	0.20	0.19	0.19
14....	0.29	0.24	0.22	0.22	0.22
17....	0.31	0.26	0.25	0.24	0.24
21....	0.35	0.29	0.28	0.27	0.27
25....	0.38	0.33	0.31	0.30	0.29
30....	0.39	0.36	0.33[a]	0.33	0.32
37....	0.47	0.39	0.37	0.35	0.35
44....	0.48	0.43	0.40	0.37	0.37
53....	0.53	0.46	0.42	0.41	0.41
64....	0.54	0.47	0.45	0.43	0.44
77....	0.56	0.51	0.47	0.46	0.45
92....	0.60	0.54	0.50	0.49	0.48
110....	0.63	0.56	0.54	0.51	0.51
133....	0.67	0.59	0.56	0.53	0.53
159....	0.70	0.61	0.57	0.57	0.56
191....	0.73	0.65	0.61	0.58	0.58
230....	0.76	0.67	0.64	0.62	0.61

NOTE.—Each entry in the table is obtained on the basis of 2,000 simulations of a constant-sized population, with $N_0\nu = 2.5$. Since g is nearly independent of $N_0\nu$ for $N_0\nu > 0.5$, the cutoffs apply to a range of values of $N_0\nu$ and not just to 2.5. Interestingly, the cutoffs decrease with increasing sample size, the reason being that the variance at each locus is measured less reliably when the sample size is smaller, and hence the variability of the variances across loci is larger.

[a] The fifth-percentile cutoff of $g = 0.33$ is used in many of the examples in this paper.

addition, statistical concerns about $\beta(t)$ remain. For example, due to the statistical difficulties posed by ratios of random variables, Kimmel et al. (1998) construct $\beta(t)$ as the ratio of the estimators averaged separately. However, it may be that variation in the mutation rate causes the ratio of the expectations to have a different value from the expectation of the ratios. Despite these potential problems, g and $\beta(t)$ may turn out to be complementary, with different sensitivities to certain aspects of demographic history.

Power of Tests as a Function of Number of Loci, Sample Size, and Value of $N_0\nu$

In designing studies of demographic history, it is important to know how many loci, what sample size, and what value of $N_0\nu$, where N_0 now represents the *pre*-expansion population size, are necessary for the tests to have a specific level of resolving power. To assess these questions, we consider 100-fold expansions that began N_0 generations ago and that have an approximately 50% chance of producing unusually low test statistics for 30 loci, a sample size of 40, and an $N_0\nu$ value of 0.88. (This set of conditions is evaluated for both the within-locus and interlocus tests.) By varying the number of loci, sample size, and $N_0\nu$ in turn (with the other two parameters fixed at the above values), the power of the tests can be assessed as a function of the three main demographic parameters.

The power of a test is one minus the probability that the test fails to reject size constancy for a particular combination of parameters when an expansion actually occurred (that is, one minus the Type II error rate). To estimate this probability, it is first necessary to find the fifth-percentile lower cutoffs for the proportion of positive k values (within-locus test) and for the value of g (interlocus test). The cutoffs for the within-locus test are obtained from the binomial distribution, and the cutoffs for the interlocus test are taken from table 1. By performing 1,000 computer simulations for each of several combinations of demographic parameters, and then counting the proportion of test statistics that are above the fifth-percentile cutoffs when a 100-fold sudden expansion occurred N_0 generations ago, we derive the probability of not rejecting size constancy and use one minus this probability to calculate the power of the tests (fig. 3).

Figure 3*A* shows that the power of both tests increases with an increasing number of loci. Figure 3*B* shows that the power of the within-locus test increases with increasing sample size, while the power of the interlocus increases very little once the sample size is

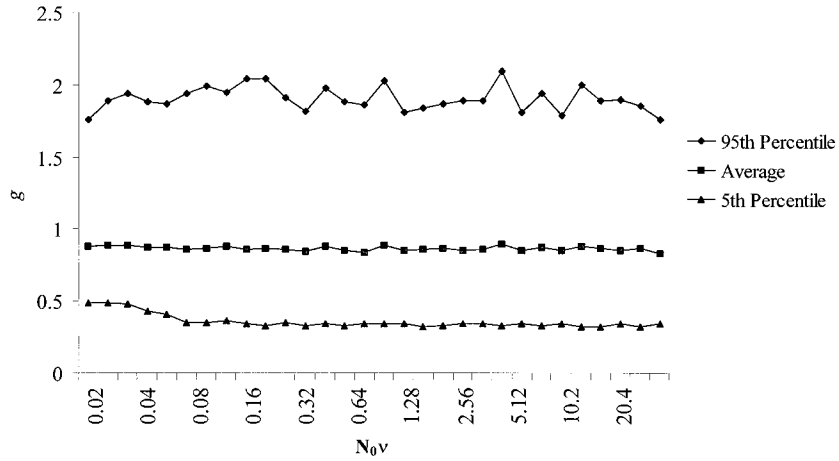
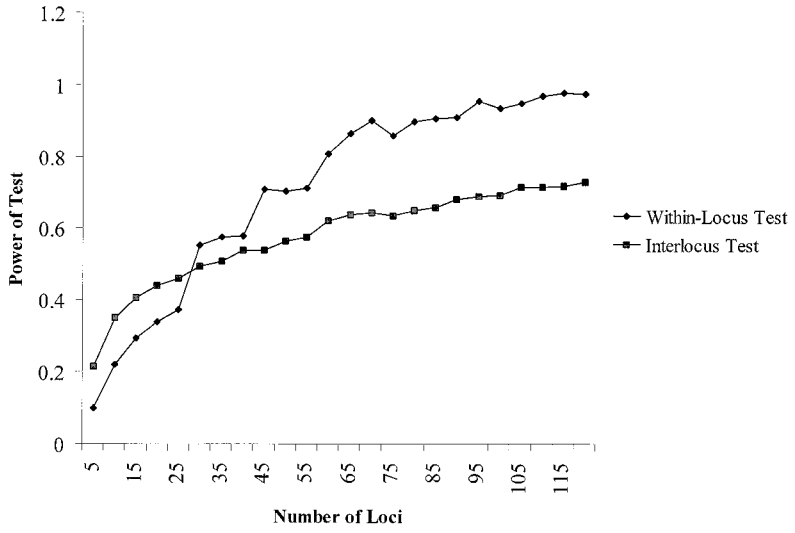
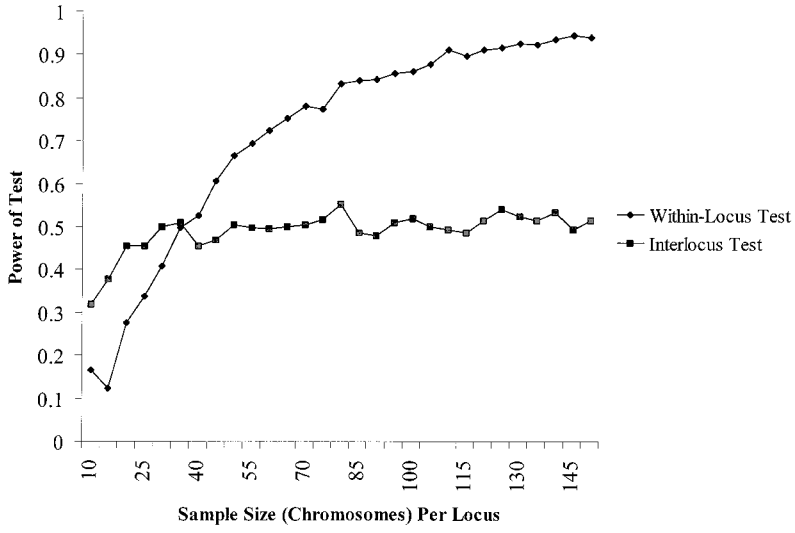


FIG. 2.—Ninety percent central confidence intervals for g and the average value of g over 1,000 simulations, under the assumption of a constant population size. For each value of $N_0\nu$, we use 30 loci and a sample size of 40 and calculate the expectation and confidence interval for g. When $N_0\nu$ is greater than 0.25, g is nearly independent of $N_0\nu$. The near independence of g from $N_0\nu$ also holds for other numbers of loci and sample sizes.

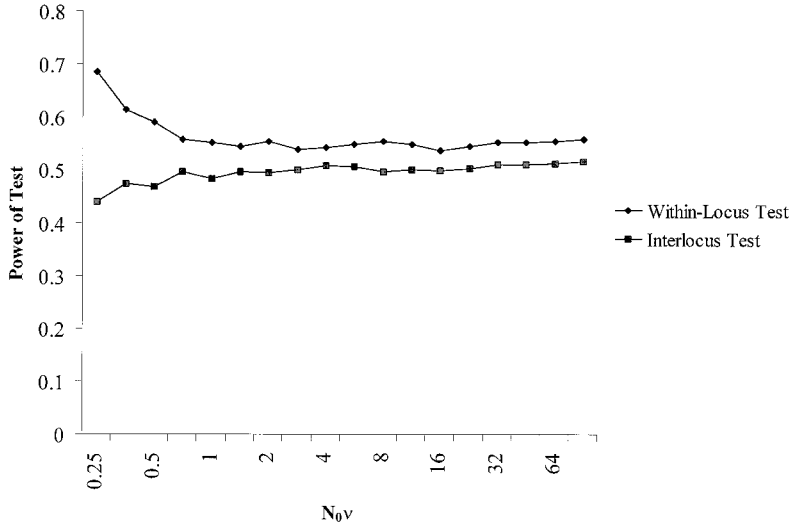
A


1.2
1
0.8
0.6
0.4
0.2
0
Power of Test
5
15
25
35
45
55
65
75
85
95
105
115
Number of Loci
Within-Locus Test
Interlocus Test

B


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Power of Test
10
25
40
55
70
85
100
115
130
145
Sample Size (Chromosomes) Per Locus
Within-Locus Test
Interlocus Test

C


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Power of Test
0.25
0.5
1
2
4
8
16
32
64
N0v
Within-Locus Test
Interlocus Test

above a minimum. The reason for this is that the within-locus test is based on the fourth central moment—best assessed using a large sample size—while the interlocus test is based on the variance, which can be estimated with relatively fewer samples. Finally, figure 3*C* shows that the power of both tests is nearly independent of $N_0\nu$ as long as $N_0\nu \geq 1.0$, as noted elsewhere. However, the test statistics will not be independent of $N_0\nu$ if microsatellites are subject to constraints on the lengths of alleles.

Due to space limitations, only a few combinations of demographic parameters are presented in figure 3. Three general principles, however, are apparent: (1) increasing the number of loci increases the power of both tests, (2) increasing the sample size increases the power of the within-locus test only, and (3) the magnitude of $N_0\nu$ affects neither test when $N_0\nu > 1.0$.

Sensitivity of Within-Locus and Interlocus Tests to the History of Expansion

Statistical signals of expansion take a considerable time to develop and fade gradually as the genetic variation in a population approaches a new state of mutation-drift equilibrium. The time window of sensitivity of the within-locus and interlocus tests—that is, the range of dates for which growth would be detected at a specified significance level—is now explored for a range of demographic models.

Sudden and Gradual Expansions

We begin by simulating 100-fold, sudden population growth from $N_0\nu = 0.88$ to $N_f\nu = 88$, where N_f is the postexpansion population size. For each date of expansion, we perform 1,000 simulations—each involving 30 unlinked microsatellite loci and a sample size of 40—and calculate central confidence intervals for the two tests. The dates of expansion to which the tests are maximally sensitive, that is, that produce the most marked reductions of the expected test statistics, are $5.1N_0$ generations ago for the within-locus test, and $14.6N_0$ generations ago for the interlocus test (fig. 4). The within-locus test is more sensitive to recent expansions than the interlocus test, a fact that is also reflected in the dates of expansion that the tests can detect with greater than 50% probability: $0.87N_0$ to $26N_0$ generations ago for the within-locus test, and $1.02N_0$ to $171N_0$ generations ago for the interlocus test (data not shown). Simulations for a range of $N_0\nu$ values other than 0.88 show that as long as $N_0\nu > 0.5$, the range of dates of expansion to which the tests are most sensitive, scaled in units of N_0, is independent of $N_0\nu$.

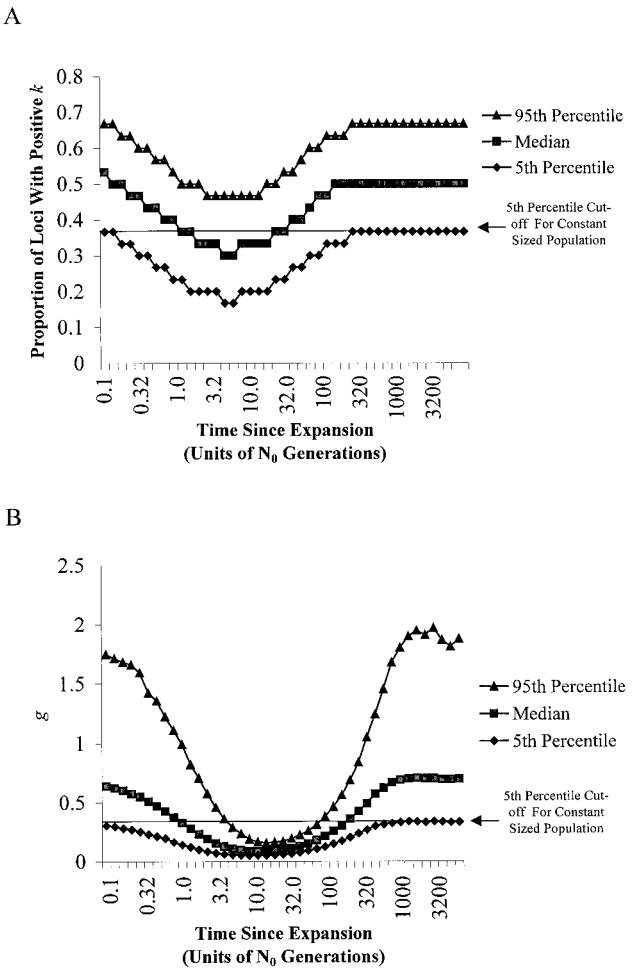


FIG. 4.—(*A*) Power of the within-locus test and the interlocus test as functions of the number of loci, for a sample size of 40, and for population growth from $N_0\nu = 0.88$ to $N_f\nu = 88$. The power is defined as 1 minus the probability of not detecting an expansion at the 5% significance level if one indeed occurred. To obtain this probability, we assume a model of 100-fold growth that occurred $1.0N_0$ generations ago, perform 5,000 simulations for each combination of parameters, and calculate the percentage of results that are below the fifth percentile cutoff for a constant size population. Graph (*B*) shows the power of the two tests as functions of sample size, for 30 loci and population growth from $N_0\nu = 0.88$ to $N_f\nu = 88$. Each data point is calculated on the basis of 1,000 simulations. Graph (*C*) shows the power of the two tests as functions of $N_0\nu$, for 30 loci and a sample size of 40 and with each data point calculated on the basis of 5,000 simulations.

We now plot the proportion of loci with positive *k*, and the average value of *g*, against the time since expansion for 10-fold, 100-fold, and 1,000-fold expansions (fig. 5). The minimum points of the curves in figure 5, corresponding to the dates of expansion to which the

←

FIG. 3.—Confidence intervals for the within-locus and interlocus test statistics as functions of the time since expansion. We perform 10,000 simulations for each date of expansion and consider 100-fold growth from $N_0\nu = 0.88$ to $N_f\nu = 88$ with 30 loci and a sample size of 40. The within-locus test statistic is expected to be lowest for an expansion that occurred $5.1N_0$ generations ago, and the probability of detecting an expansion using the test is expected to be greater than 50% for expansions that occurred $0.87N_0 - 26N_0$ generations ago (*A*). The date range is given by the intersection points of the 50th percentile cutoff for a constant-sized population and the curve. In contrast, the interlocus test statistic (*g*) is expected to be lowest for an expansion that occurred $14.6N_0$ generations ago, and the probability of detecting an expansion is expected to be greater than 50% for expansions that occurred $1.02N_0 - 171N_0$ generations ago (*B*).

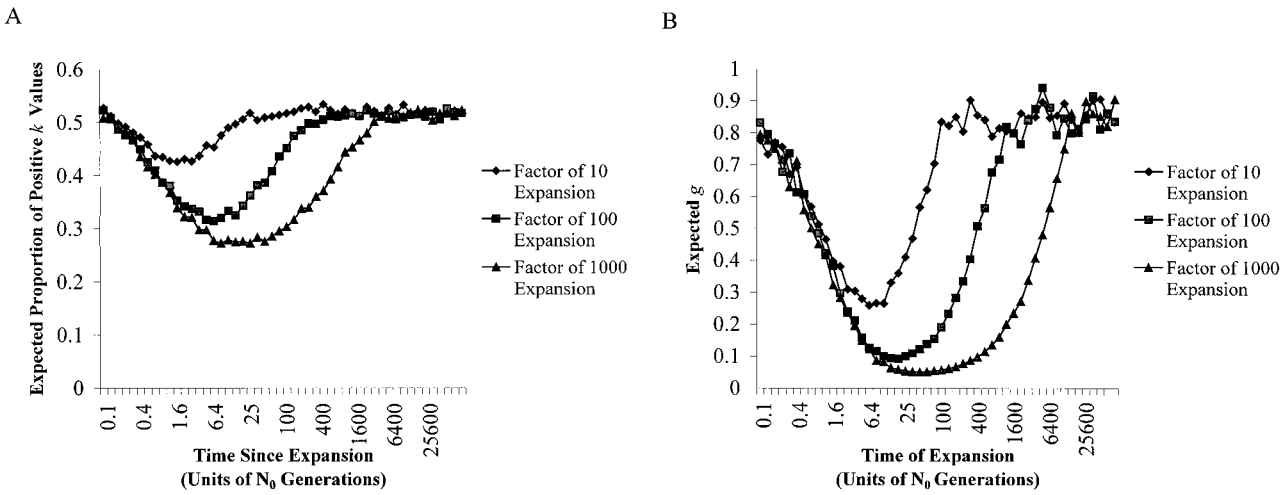


FIG. 5.—(Sudden growth) Average values of the within-locus statistic (*A*), and the interlocus statistic (*B*), as functions of the time since expansion and the factor of sudden growth. For each date of expansion, we use 30 loci, a sample size of 40, and calculate test statistics based on an average of 250 simulations. The contours correspond to 10-fold, 100-fold, and 1,000-fold growth from $N_0\nu = 0.88$. Immediately following the expansion (the left side of the figure), the behavior of both test statistics is nearly independent of the factor of expansion.

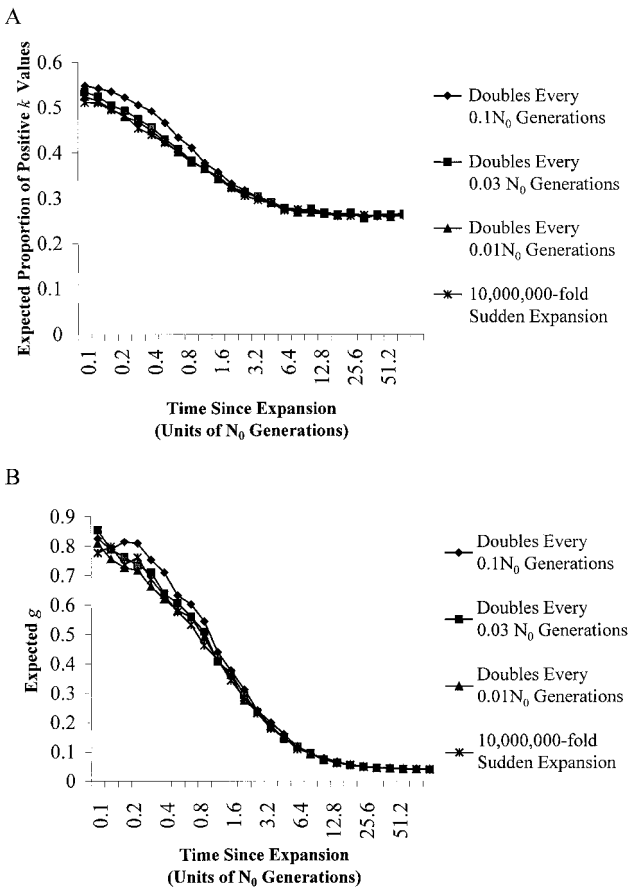


FIG. 6.—(Exponential growth) Average values of the within-locus statistic (*A*) and the interlocus statistic (*B*), as functions of the time since expansion and the rate of exponential growth. For each date, we use 30 loci and a sample size of 40 and calculate test statistics based on an average of 1,000 simulations. The contours correspond to doubling every $0.1N_0$, $0.03N_0$, $0.01N_0$ and $0.001N_0$ generations (the population size is held constant at N_0 before the expansion begins). For comparison, we show a contour corresponding to a massive, 10,000,000-fold sudden expansion. For all contours shown, the test statistic is always within 10% of what is predicted for a massive, sudden expansion.

tests have the greatest sensitivity, occur at older times when the factor of expansion is larger, since it takes a longer time for the population to return to mutation-drift equilibrium. A second important property of the tests—which can also be seen in figure 5—is that the test statistics are independent of the growth factor in the period immediately following expansion. To quantify this observation, we compare the average test statistics for a massive (10,000,000-fold) expansion, to those occurring for smaller factors of growth. Specifically, we assess how long after an expansion the average test statistics stay within 10% of the value expected for a 10,000,000-fold expansion: the answer is $0.7N_0$ generations in the case of a 10-fold expansion, $2.3N_0$ generations in the case of a 100-fold expansion, and $20N_0$ generations in the case of a 1,000-fold expansion. We conclude that if the factor of growth is large enough, and the time since expansion is less than a value determined by the factor of expansion, the tests have almost no power to distinguish among various factors of sudden growth. This is actually a manifestation of a more general property of the tests: whether growth is sudden or more gradual, the tests have almost no power to distinguish among alternative scenarios of expansion. The behavior is likely to be due mainly to the effects of growth on gene genealogies, and is therefore probably a general feature of genetic tests of demographic history.

We now consider the effect of gradual expansions (as opposed to sudden expansions) on the test statistics. Specifically, we consider the case of exponential growth. Figure 6 shows that when the population size doubles at least once every $0.1N_0$ generations, the expected values of the test statistics are always within 10% of the expectation for a massive (10,000,000-fold), sudden expansion (fig. 6). The same holds true for more complex, nonexponential histories of expansion, as long as doubling occurs at least once every $0.1N_0$ generations. The similar behavior of the test statistics for a range of growth rates—evident in figure 6—greatly simplifies the

study of demographic history using these tests. In particular, for many realistic growth scenarios, we can treat gradual expansions as if they had been sudden. For example, in the case of the Paleolithic human population expansion that was recently detected in Africa (Reich and Goldstein 1998), the doubling need only have occurred at least once every 10,000 years—assuming a pre-expansion population size of at least 4,000 and a generation time of 25 years—for the expansion to have been effectively sudden from the perspective of our tests. Since 10,000 years and 4,000 individuals seem to be a rather minimal requirement for a doubling time and pre-expansion population size in humans, the assumption of a sudden expansion (Reich and Goldstein 1998) seems appropriate.

There are also certain cases in which it is not valid to treat expansions as sudden: for example, when the population doubles less frequently than once every $0.1N_0$ generations. In such circumstances, the earliest periods of growth dominate the test statistics, since they influence the times of occurrence of the first few bifurcations in the genealogical trees. These are the critical bifurcations for determining the shapes of allele-length distributions.

Population Growth Interrupted by a Bottleneck

We now examine the response of the test statistics to strong growth interrupted by a bottleneck. For simplicity, we consider a model in which a sudden population expansion, by a factor of F, occurs t_0 generations ago. The population remains constant in size until $t_1 + \Delta t$ generations ago, at which point it contracts to N_C individuals for Δt generations. Finally, the population returns to the prebottleneck size ($N_f = N_0F$) at time t_1, and remains static until the present. Using coalescent simulations of this model, we estimate average values for the test statistics (and also the average value of the variance across loci) as functions of several demographic parameters. Figure 7 graphs simulation results for the average test statistics and average variance when the parameters F, $N_0\nu$, t_0, and Δt are fixed at 100, 0.88, $3.0N_0$ generations, and $0.04N_0$ generations, respectively, and when the population size during the bottleneck, N_C, and the time when the bottleneck ended, t_1, are varied. As shown, the average test statistics tend to be raised due to a bottleneck; however, when the population contraction during the bottleneck is sufficiently severe, the average value can actually be depressed below what is expected, which explains why the curves are nonmonotonic: the tests are most sensitive—that is, the position of the peaks of the curves in (*A*) and (*B*)—is substantially higher for the within-locus than for the interlocus test. Finally, for severe bottlenecks that ended 0 generations ago, the test statistics behave irregularly because the allele-length distributions have extremely low variance (hence, the contours corresponding to these cases, in A and B, are truncated).

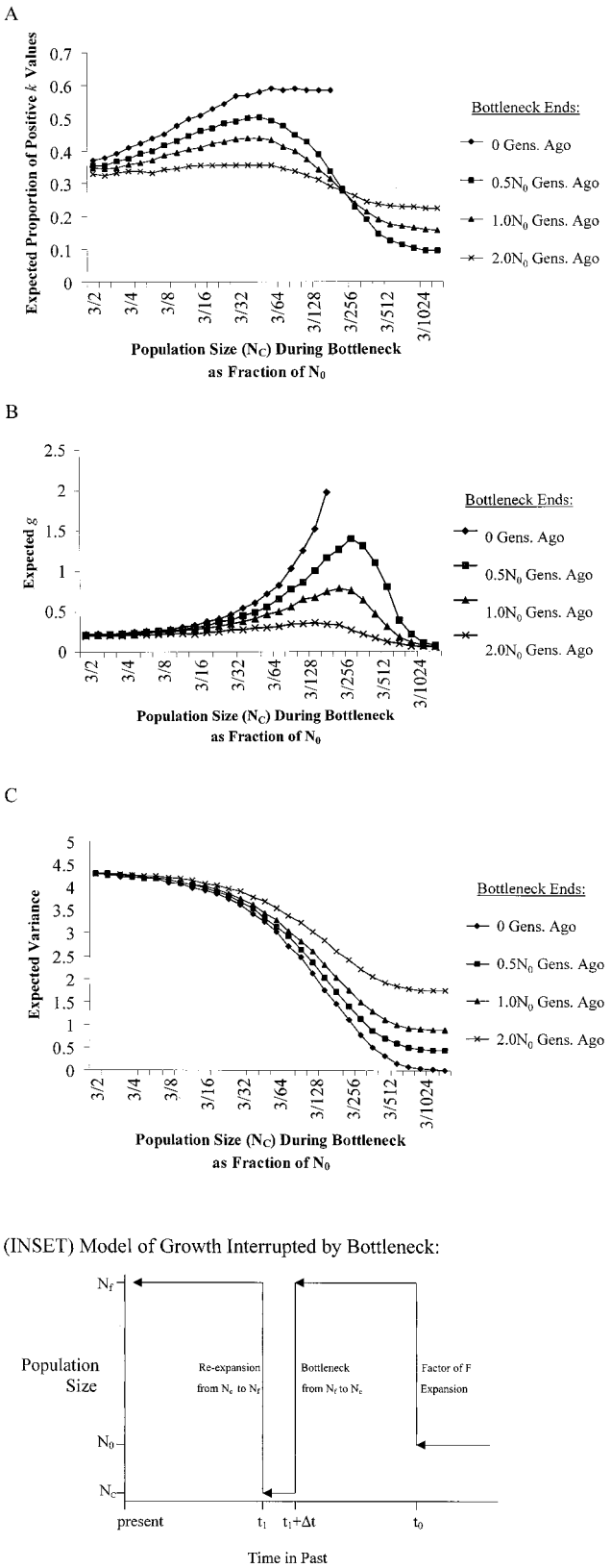


FIG. 7.—Average values of the within-locus statistic (*A*), the interlocus statistic (*B*), and variance of the allele-length distribution (*C*), for sudden growth interrupted by a bottleneck that lasts $0.04N_0$ generations. Inset shows the time course for this scenario of growth interrupted by a bottleneck. For each combination of parameters, we use 30 loci, a sample size of 40, an $N_0\nu$ value of 0.88, and perform 1,000 simulations of a 100-fold sudden expansion that began $3.0N_0$ generations ago and that was interrupted by a bottleneck ending 0, $0.5N_0$, $1.0N_0$ or $2.0N_0$ generations ago. The average value of the test statistic in the absence of a bottleneck can be extrapolated from the far left side of the curves in (*A*) and (*B*). The factor of contraction to which

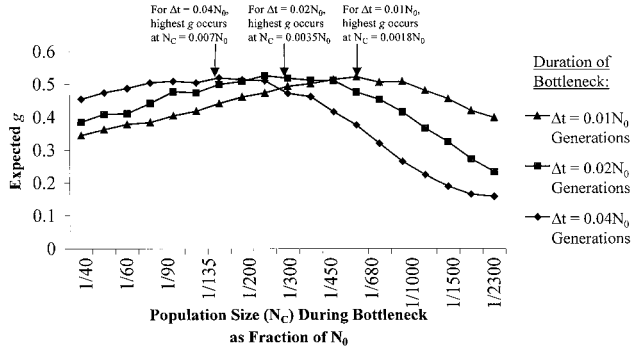


FIG. 8.—The effect of the duration of the bottleneck (Δt) on the interlocus test. Using 30 loci, a sample size of 40, and an $N_0\nu$ value of 0.88, and considering a 100-fold expansion that began $3.0N_0$ generations ago and was interrupted by a bottleneck $1.0N_0$ generations ago, we perform 1,000 simulations for each data point. As shown in the figure, an increase in Δt by a certain factor causes the contours to shift to the left by the same factor. Similar effects occur for the within-locus test.

otonic. Another important feature of figure 7 (*A*–*C*) is that the effects of a bottleneck are more pronounced when the bottleneck is relatively recent (i.e., t_1 is small), than when the bottleneck is older.

To understand the behavior of figure 7, note that a bottleneck increases genetic drift, thereby transforming the starlike genealogies expected for an expansion into the more clustered topologies expected for a constant-sized population. However, the opposite effect can also occur: when a bottleneck is sufficiently severe, genealogies tend to become more starlike, and a powerful signal of expansion is generated because the typical gene genealogy is reduced to only a single ancestral lineage as old as the bottleneck, and this resets the genetic ''clock'' in the sense that the average statistics are below what is expected even in the absence of a bottleneck (cf. the right side of the curves in fig. 7*A* and *B*). Interestingly, the resetting of the genetic clock actually requires a more severe contraction in the case of the interlocus test than in the case of the within-locus test (cf. the peak positions in fig. 7*A* and *B*). For the interlocus test, every one of the sampled loci must be reduced by drift to a single lineage before the genetic clock is reset, while for the within-locus test, the clock is reset locus by locus and the overall requirement for the development of a new signal of expansion is less stringent.

We now consider the effects of varying the parameters that were held fixed in figure 7. Variation in t_0 has a complex effect on the test statistics; however, the factor of expansion to which the tests are maximally sensitive, as well as the overall shape of curves in figure 7*A* and *B*, is independent of t_0. Variation in Δt and its effect on the interlocus test statistic is depicted in figure 8. Figure 8 shows that when Δt is changed by a given factor, the curves in figure 7*B* shift to the left by the same factor and are mostly unchanged in shape after the shift. (Similar effects are observed for the *k* statistic.) Hence, it is possible to extrapolate the quantitative results of figure 7*A* and *B* to bottlenecks of durations other than $0.04N_0$ generations. Indeed, if the ''severity'' of the bottleneck is defined as a simple combination of N_C and Δt—$N_0\Delta t/N_C$—figure 7*A* and *B* can be used to make the very rough estimate that if $\Delta t(N_0/N_C) < 0.04$ (in units of N_0 generations) for the within-locus test, or $\Delta t(N_0/N_C) < 0.08$ (in units of N_0 generations) for the interlocus test, the test statistics will be essentially unaffected by a bottleneck. When N_C is time-dependent, a similar expression (involving an integral over the duration of the bottleneck), is the critical determinant of the effect on the test statistics (not shown).

Application to Human Demography

In connection with an expansion that was recently detected in Africa but not outside of Africa, Reich and Goldstein (1998) speculated that the statistical signal of population expansion may have been generated among the ancestors of all modern humans in the Paleolithic but was then erased in non-African groups due to a population bottleneck that occurred approximately 80,000–100,000 years ago during the emergence of the first humans from Africa. In order to test this idea, we used Rogers and Harpending's (1992) estimation (based on analysis of variability in mitochondrial DNA) of an expansion from $N_0 = 3{,}254$ to $N_f = 547{,}586$ estimated to have occurred 4,800 generations ago (cf. Kimmel et al. [1998]). With an estimated mutation rate for dinucleotide microsatellites of $\nu = 5.6 \times 10^{-4}$ (Weber and Wong 1993), this corresponds to an expansion from $N_0\nu = 1.8$ to $N_f\nu = 306$ occurring $1.48N_0$ generations ago. We then model a bottleneck starting $1.04N_0$ generations ago (85,000 years ago, assuming 25 years per generation) and lasting for $0.04N_0$ generations. Our simulations show that a bottleneck of this type could indeed have obscured an ancient signal of expansion, and would be expected to cause a rise in the average value of *g* above the 0.05 percentile lower cutoff for significance as long as $\Delta t(N_0/N_C)$ is in the approximate range $0.00025 < \Delta t(N_0/N_C) < 0.0045$ (in units of N_0 generations; see e.g., fig. 7*B*). This bottleneck is not expected to cause *g* to rise above the cutoff if the bottleneck occurred more than $1.2N_0$ generations ago. Note that erasure of a signal of population expansion outside of Africa is not necessarily the result of an out-of-Africa bottleneck; it could also be due to more recent bottlenecks, or to population structure (see below) in non-African populations.

Our time-sensitivity results also shed light on the date of the detected human population expansion. In particular, to evaluate our estimate that the expansion was Paleolithic, we ask whether the expansion would have been likely to be detected had it occurred in the Mesolithic or Neolithic time periods. As described above, the most recent population expansion that can be detected with greater than 50% probability is $0.87N_0$ generations ago for the within-locus test and $1.02N_0$ generations ago for the interlocus test. If we now assume that effective African population sizes were at least 50,000 during the Neolithic and Mesolithic, this translates to dates of expansion of 1.1 million years ago and 1.3 million years ago, respectively (assuming a generation time of 25 years). It therefore seems unlikely that the tests would be able to detect a Mesolithic or Neolithic expansion. We conclude that in the case of human demographic history, the two tests are not expected

to be sensitive to any but the most ancient expansions that began from relatively small population sizes. Indeed, before applying these approaches to any data set (not just human data sets), it may be worth making a quick assessment—based on what is known about the history of the species and about the properties of the tests—to anticipate whether the expansions of interest are likely to be detected by the tests or whether the pre-expansion population size and age of the species rule out detection.

Estimating a Date of Expansion and Other Parameters

To estimate a date for a detected expansion, we assume a model of sudden growth by a fixed factor, F, and use computer simulations to find the range of expansion times and pre-expansion $N_0\nu$'s that are consistent with the observed variance and test statistic (Reich and Goldstein 1998). A particular combination of parameters (including a date of expansion) is considered to be ''allowed'' if the observed values of the variance and test statistics are within the specified confidence intervals determined from computer simulation. To calculate the allowed range of dates for the expansion, we then take the full set of dates that, when combined with an appropriate value of $N_0\nu$, are consistent with the observed variance and test statistic in the sense that they comprise an allowed parameter combination. These allowed ranges of dates are specific to a particular factor of expansion, F. To find allowed ranges of dates for other factors of expansion and more complex models of growth, it is necessary to perform further analysis.

We now show how the same simulations can be used to construct a likelihood surface for the observed results as a function of the date of expansion and value of $N_0\nu$. By counting the proportion of results—for each combination of parameters—that are within a narrow window around the observed variance and test statistic, we obtain a surface that is generally bimodal, with peaks at positions that vary with the factor of expansion F. To understand this, we consider a graph showing the average value of the test statistic as a function of the time since expansion (fig. 5). By drawing a horizontal line through the graph at the height of the observed value of the within-locus or interlocus test statistic, we understand immediately why the likelihood surface has two peaks, and we can roughly identify the positions of the peaks as a function of the time since expansion. Specifically, the two dates when the horizontal line and the curve intersect (that is, the dates for which the expected value of the test statistic is equal to the observed value), correspond roughly to the most likely dates of expansion and the peaks in the likelihood surface. The more recent of the two peaks corresponds to a population that is just beginning to register a signal of expansion, while the older peak corresponds to a population that is returning to equilibrium after an older expansion. Note that the points of intersection do not give the best estimates of the positions of the peaks. To obtain the best estimates, it is necessary to take into account the observed average variance across loci.

We now consider how date estimates change in response to different factors of growth. Figure 5 shows growth factors of 10, 100, and 1,000; in all three curves, the expected value of both test statistics shrinks in the period immediately following the expansion, then rises again to its equilibrium levels when the expansion is sufficiently old. Thus, a typical horizontal line through figure 5 intersects the curves twice, with the date of the more recent intersection being the same regardless of the factor of expansion and the date of the older intersection varying with the factor of expansion. To take a specific example, consider a horizontal line through figure 5*B* at the level $g = 0.4$, and note that it first intersects all three curves at a time lag of about $1.2N_0$ generations but intersects the other part of the curves at time lags of $20N_0$, $200N_0$, and $2{,}000N_0$ generations for $F = 10$, 100, and 1,000, respectively. The similar values of the curves shortly after expansion are due to the inability of the tests to distinguish between different factors of growth in the case of recent expansions. The different values of the curves at the right side of the figures are due to the fact that the rate of return to mutation-drift equilibrium depends on population size.

The independence of the more recent intersection—and hence the date of the more recent peak in the likelihood surface—from the assumed model of population growth is especially useful when it is possible to use other evidence, for example from the fossil record, to eliminate the possibility of the ancient peak. If we cannot eliminate an ancient peak, however, our estimate for the date of expansion will be subject to error depending on our assumptions about growth history. There are also other complications that can affect date estimation. For example, if a population has undergone a series of small expansions over an extended period of time or if a severe bottleneck has occurred, a date estimate would be obtained that is not likely to correspond to a specific historical expansion.

Finally, it is possible to estimate other parameters: specifically, the maximum pre-expansion population size, N_0, and the minimum post-expansion population size, N_f. To obtain these values, we assume a variety of factors of expansion, and for each F, use the $N_0\nu$ values associated with the most recent and oldest "allowed dates" to find a maximum pre-expansion population size (N_0), and minimum post-expansion population size (N_f). In principle, the estimates obtained in this way could be different for each factor of expansion. However, we find that the maximum N_0 and minimum N_f (but *not* the minimum N_0 and maximum N_f), are generally extremely robust to changes in the assumed F if it is sufficiently large (Reich and Goldstein, unpublished data).

Complications in the Model

We have assumed single-step and unbiased mutations that are constant in rate across loci, and we have also assumed unstructured populations, an absence of selection, and independent sampling. In what follows, we explore the effects of departures from these assumptions on the test statistics (cf. tables 2 and 3).

Table 2
Quantitative Effects of Deviations from Assumed Model on the Fifth-Percentile Cutoffs for the Within-Locus and Interlocus Tests

	Average Step Size $\bar{s}$							
(A) Multi-step mutations	1.0	1.05	1.1	1.15	1.2	1.35	1.5	2.0
Within-locus test								
Ratio of observed cutoff to no-deviations cutoff	1	0.92	0.92	0.92	0.83	0.75	0.75	0.75
Probability of false rejection of size constancy	0.05	0.08	0.12	0.14	0.17	0.25	0.28	0.35
Interlocus test								
Ratio of observed cutoff to no-deviations cutoff	1.0	1.03	1.07	1.09	1.12	1.15	1.15	1.17
Probability of false rejection of size constancy	0.05	0.04	0.03	0.03	0.03	0.02	0.02	0.02
	Variance of $N_0\nu$ in Units of $(N_0\nu)^2$							
(B) Interlocus variability in mutation rate	0.0	0.05	0.1	0.2	0.3	0.4	0.8	1.6
Interlocus test								
Ratio of observed cutoff to no-deviations cutoff	1.0	1.06	1.11	1.2	1.36	1.42	1.88	2.59
Probability of false rejection of size constancy	0.05	0.04	0.03	0.01	0.007	0.005	0.002	0.000
	Variance of $\bar{s}$ (using $E[\bar{s}] = 1.2$)							
(C) Interlocus variability in step size	0.0	0.01	0.02	0.04	0.08	0.16	0.32	0.64
Interlocus test								
Ratio of observed cutoff to $Var[\bar{s}] = 0.0$ cutoff	1.0	1.03	1.05	1.12	1.20	1.29	1.37	1.34
Probability of false rejection of size constancy	0.03	0.02	0.02	0.02	0.006	0.008	0.006	0.009
	Migration Rate ($M = 4N_0m$)							
(D) Three-island model with migration	12.8	6.4	3.2	1.6	0.8	0.4	0.2	0.1
Within-locus test								
Ratio of observed cutoff to no-deviations cutoff	1.0	1.0	1.08	1.08	1.17	1.25	1.25	1.33
Probability of false rejection of size constancy	0.05	0.04	0.03	0.03	0.01	0.009	0.006	0.008
Interlocus test								
Ratio of observed cutoff to no-deviations cutoff	1.0	1.02	1.03	1.08	1.18	1.35	1.61	1.84
Probability of false rejection of size constancy	0.05	0.05	0.04	0.03	0.02	0.007	0.002	0.001
	No. of Populations ($t_{sep} = 1.0N_0$ generations)							
(E) Multichotomy model	2	3	4	5	7	10	15	25
Within-locus test								
Ratio of observed cutoff to no-deviations cutoff	1.08	1.17	1.17	1.17	1.08	1.08	1.0	0.92
Probability of false rejection of size constancy	0.02	0.02	0.01	0.01	0.02	0.03	0.05	0.08
Interlocus test								
Ratio of observed cutoff to no-deviations cutoff	0.96	0.92	0.88	0.86	0.78	0.71	0.65	0.61
Probability of false rejection of size constancy	0.06	0.08	0.09	0.11	0.14	0.2	0.24	0.31
	Average Multiplicity of Samples							
(F) Double-counting of samples	1	1.2	1.4	1.7	2	4	8	16
Within-locus test								
Ratio of observed cutoff to no-deviations cutoff	1.0	1.0	1.08	1.08	1.08	1.25	1.42	1.42
Probability of false rejection of size constancy	0.05	0.04	0.03	0.02	0.02	0.003	0.000	0.000
Interlocus test								
Ratio of observed cutoff to no-deviations cutoff	1.0	1.01	1.03	1.06	1.10	1.26	1.58	2.30
Probability of false rejection of size constancy	0.05	0.05	0.04	0.04	0.03	0.01	0.001	0.000

NOTE.—To calculate each entry in tables 2A–F, we assume a constant sized population and perform 5,000 coalescent simulations for 30 loci, a sample size of 40, and an average $N_0\nu$ of 0.88 (deviations are modeled as described in the text). For each combination of parameters, we provide the ratio of the fifth-percentile cutoff calculated for the deviation, divided by the cutoff expected in the case of no deviations as well as by the probability of false rejection of size constancy (probabilities less than 0.05 indicate a conservative test).

Deviations from the Mutation Model

Although the stepwise mutation model seems to describe changes in microsatellite allele lengths better than alternative models, it is not completely accurate (Goldstein and Pollock 1997). We begin our study of deviations by considering multistep mutations. In order to assess this effect, we assume 30 loci, a sample size of 40, and an $N_0\nu$ value of 0.88 and simulate genealogical trees for constant-sized populations and with mutations distributed along the branches of the trees according to the method described in Appendix 2. Each mutation has an equal chance of increasing or decreasing the length of an allele, and the number of steps is determined by adding 1 to a random draw from a Poisson distribution with parameter λ. The average step size for an event drawn from this distribution is $\bar{s} = \lambda + 1$. To determine the probability of obtaining a false-positive signal of expansion (Type I error rate), we then use the fifth-percentile cutoffs for the case of a single-step mutation model. Table 2*A* shows that the probability of a false positive

Table 3
Qualitative Effects of Deviations from the Assumed Model on the Test Statistics, with Specific Application to Date Estimation

Class of Deviation	Description of Deviation	Effect on the Within-Locus Test Statistic	Effect on Interlocus Test Statistic	Effect on Rate of Rise of the Variance
Deviations from assumed mutation model	Multistep mutations	Falls	Rises	Quickened
	Range constraints	Depends on model	Depends on model	Slowed
	Directional asymmetry	Slight rise	None	None
Variability across loci	Variable mutation rate	None	Rises	None
	Variable average step size	None	Rises	None
	Variable directional asymmetry	None	None	None
Population structure	Island model with migration	Rises	Rises	None
	Multichotomy model	Rises	Depends on model	None
Miscellaneous	Natural selection	Falls	Rises	Slowed
	Pseudoreplication	Rises	Rises	None

NOTE.—The test statistics and the rate of rise of the variance after an expansion both have effects on date estimation. ''Rises'' indicates a conservative effect on the test statistics, ''Falls'' indicates a nonconservative effect on the test statistics, and ''Depends on model'' indicates that the effect on the test statistics depends on the specific type of range constraints that are in place.

is elevated above 5% in the case of the within-locus test (a nonconservative effect) and reduced in the case of the interlocus test (a conservative effect).

Second, we assess the effect of range constraints on the test statistics and, specifically, a model of strict upper and lower boundaries on the allowed allele lengths (Feldman et al. 1997). The within-locus test statistic is expected to rise for this ''reflecting boundaries'' model because of the flattening of an allele-length distribution between the boundaries (a conservative effect). In contrast, the interlocus test statistic is expected to fall because the boundaries tend to restrict the variation of variances across loci (a nonconservative effect). For more realistic models of range constraints, involving multistep mutations and length-dependent mutation rates, the overall effect on the interlocus test may be conservative. Indeed, it is difficult to predict the effect of realistic range constraints on the test statistics.

Finally, we consider asymmetry in the direction of mutations. Our simulations reveal that this has almost no effect on the test statistics, although the within-locus test is slightly conservative to asymmetric mutations when the deviation is extreme (data not shown).

Variation in the Mutation Process Across Loci

We now consider sources of interlocus variation and their effects on the interlocus test (for obvious reasons, we do not describe effects on the within-locus test). We begin by considering the effect of variation in the mutation rate across loci on the interlocus test. For this purpose, we assume the same set of parameters noted earlier, except that to determine $N_0\nu$ for each locus, we sample from a gamma distribution with a mean of $N_0\nu = 0.88$ and a variance that is between $0.05(N_0\nu)^2$ and $1.6(N_0\nu)^2$. Table 2*B* shows that the fifth-percentile cutoffs for g rise dramatically as variation in the mutation rate across loci increases. A method for estimating variation in the mutation rate across loci—which can then be incorporated into adjusted fifth-percentile cutoffs for the interlocus test—is described elsewhere (Reich and Goldstein 1998).

A second source of variation across loci is variability in the likelihood that a microsatellite will undergo multistep mutations. In order to assess this deviation, we use the usual parameters, except that multistep mutations occur with an average step size of $\bar{s}$ ($\bar{s} = \lambda + 1$, as described above). To determine λ at an individual locus, we then take a random draw from a gamma distribution with a mean of 0.2 and a variance ranging from 0 to 0.64 (the average step size across loci is then $E[\bar{s}] = E[\lambda] + 1 = 1.2$). As shown in table 2*C*, variability in step-size across loci causes a rise in the interlocus test statistic, which is a conservative effect.

Finally, we consider variation in the degree of mutational asymmetry across loci, which our simulations show to have very little effect on the test statistics (and hence we do not present quantitative results for it in the paper).

Deviations Due to Population Structure

To assess the effect of population structure on the tests, we first consider an island model with migration, in which there are j islands, each with a constant-sized population of N_0/j, and a probability of an individual migrating away from its current island of m per generation (note that migration is assumed to occur with equal probability to each of the other islands). To simulate this model, we use Hudson's (1990) coalescent algorithm and consider a constant-sized population with 30 loci, a sample size of 40, $N_0\nu = 0.88$, and a migration parameter of $M = 4N_0\text{m}$. The geographic distribution of sampled alleles at each locus is assumed to be the same for each locus, and each individual contributes a full complement of alleles (two from each locus) to the data set. Our simulations using three populations (table 2*D*) show that both statistics are elevated above the expectation for a panmictic population—a conservative effect that is more pronounced when migration rates are low. The effect occurs because population structure causes genealogical trees to be dominated by a few deep splits and to be more variable from locus to locus (Donnelly 1996). When the number of islands is different from 3, the test statistics are also higher than the expectation for a panmictic population (although the deviation from expectation is less pronounced for larger numbers of islands).

As an alternative model of population structure, we consider a multichotomy, in which a population is assumed to be panmictic until a fixed time in the past (t_{sep}), at which point it suddenly breaks up into j isolated subpopulations, each of size N_0/j. To simulate a multichotomy, we use the same procedure as for the island model, but this time set the migration rate to be 0 and force all subpopulations to be lumped together into a single panmictic population at time t_{sep}. Table 2*E* shows the results of the simulations for $t_{sep} = N_0$ generations ago, for 30 loci, a sample size of 40, and $N_0\nu$ of 0.88 and for various numbers of subpopulations that together comprise a constant-sized population. We observe that the within-locus test is conservative (the statistic is elevated above the expectation for a panmictic population) unless the number of subpopulations is 15 or larger. In contrast, the interlocus test statistic is conservative for two subpopulations, approximately equal to expectation for three subpopulations, and suppressed below expectation (a nonconservative effect) for four or more subpopulations. The same qualitative effects are observed for both tests for larger values of t—although the effects are more pronounced when the value of t is larger.

Selection and Pseudoreplication

We now consider two additional deviations from the assumed model: selection and pseudo-replication. Selection is a major problem for single-locus studies because the effect of selection at a given locus cannot be distinguished from a population expansion that would affect all loci. If many loci are available, however, it is possible to make such distinctions. In particular, for the within-locus test, most loci will reflect the population's history, and it is unlikely that more than a few will contribute to a misleading signal of expansion. In the case of the interlocus test, selection is likely to affect some loci and not others, increasing the variability across loci and thereby causing a rise in g. The conservativeness of the interlocus test in response to selection means that the test can be used to confirm a departure from size constancy that is detected by the within-locus test.

We also consider pseudoreplication, or double counting of samples, which can result from experimental error or accidental sampling of multiple individuals from the same extended family. To assess the robustness of our two tests to pseudoreplication, we perform simulations of constant-sized populations with 30 loci, a sample size of 40, and $N_0\nu$ of 0.88 and resample the alleles to mimic pseudoreplication. For each allele that is produced in a run (beginning with allele $i = 1$), we assume that the allele is double-counted D_i times, with D_i obtained by adding 1 to a random draw from a Poisson process with parameter κ. To obtain the resampled set of 40 alleles, we then take D_1 alleles of type $i = 1$, D_2 alleles of type $i = 2$, etc., until we obtain the full resampled set of 40 (the average amount of double counting in this set is about $D \approx 1 + \kappa$). Table 2*F* shows that an increase in the amount of pseudoreplication increases the expected values of the test statistics—a conservative effect that applies to both the within-locus and interlocus tests.

Date Estimation in the Presence of Deviation from the Assumed Model

To describe how estimates for the dates of an expansion change when there are deviations from the assumed model, we return to the illustration of a horizontal line crossing the curves in figure 5 at the level of the observed test statistic. The intersections of the line with the curves roughly determine the positions of the peaks of the likelihood surface, and if the curves are shifted up (or down) due to a deviation from the assumed model, the maximum-likelihood peaks will also shift. For example, if the deviation causes a systematic rise in the test statistics, the estimate for the date of the more recent likelihood peak will be too recent, the estimate for the date of the older likelihood peak will be too old, and the allowed dates of expansion will cover too wide a range. Opposite effects are expected when the deviations cause a fall in the expected test statistic.

The other parameter that affects date estimation is the rate of rise of the variance after an expansion. A fast rise in the test statistics due to deviations from the assumed model will cause a systematic overestimate of the date of expansion, while a slow rise will result in an underestimate of the date. Table 3 summarizes the effects of each of the deviations discussed above on this parameter.

Conclusions

Our two tests of demographic history are generally conservative to deviations from the assumptions used to design the tests (table 3). This means that the tests are not likely to generate a false-positive signal of expansion. On the other hand, the conservativeness also weakens the tests, and it is tempting to improve their power by estimating the quantitative extent of the deviations and incorporating the estimates directly into the tests. For example, Di Rienzo et al. (1998) studied colorectal cancer cells displaying microsatellite instability and, by observing the patterns of mutations in the tumors, inferred parameters for a generalized stepwise mutation model. While this approach is very interesting, it also seems problematic because such estimates depend on the assumption that the mutational process in germline cells is the same as in somatic cancer cells where mismatch repair enzymes may be defective. In light of the fact that the microsatellite mutation process is currently so poorly understood, we prefer to perform tests of demographic history based on a simple mutation model rather than on mutation models that are parameterized in more detailed (and possible nonconservative) ways. In this case, however, it is always necessary to evaluate the sensitivity of the methods to a variety of departures from these simple assumptions.

We now consider what types of data are appropriate for use with the tests. From figure 3*B*, a sample size of at least 30 seems appropriate for the within-locus test, and a sample size of at least 15 seems appropriate for the interlocus test. Figure 3*A* shows that for both tests, at least 25 loci should be included, and of course the tests are more powerful when mutation rates are similar across loci. Finally, samples should be collected from as

many isolated populations as possible, since it is possible to learn more from a comparison of populations than from individual populations. An example of the usefulness of a multipopulation data set is provided by our study of human data (Reich and Goldstein 1998), in which a significant signal of expansion was detected in some African populations—but in no populations outside of Africa—resulting in an inference that a dramatic demographic event must have occurred to separate Africans from non-Africans. This insight would not have emerged had our study focused on African or non-African groups exclusively. Indeed, a multipopulation data set can sometimes give an indication of an expansion even when no significant signal is observed. For example, the signal of expansion that was recently detected in Africa was corroborated using a 30 tetranucleotide microsatellite data set for which the g values observed in Africa were consistently lower than g values observed elsewhere in the world—even though no single population gave a significant signal.

We have described the properties of two tests that use genetic data from multiple unlinked loci in order to assess demographic history. By considering the behavior of the tests in response to a number of demographic scenarios, we have shown that the tests are sensitive in different ways to various deviations from the assumed mutation and demographic models and that they can be used in conjunction to garner more information about demographic history than could be obtained from either test alone. In addition, the approaches to studying demographic history described in this paper, as well as the results concerning the behavior of the within-locus and interlocus tests in response to different growth models, are not restricted in principle to microsatellite variation. For example, it should be possible to use DNA sequence variation, as well as single nucleotide polymorphisms, to test hypotheses about demographic history. Work with this type of genetic data can be complementary to work with microsatellites because of the different mutation processes that are involved.

Acknowledgments

We would like to thank Peter Donnelly, Jonathan K. Pritchard, and Daniel S. Reich for their helpful comments and discussions with us during various stages of this project.

APPENDIX 1:
Derivation of statistical estimators

The within-locus test is based on the k statistic shown in equation (1). We construct the statistic empirically, using computer simulations and a linear combination of unbiased estimators for the variance (σ^2), the variance squared (σ^4), and the fourth central moment (γ_4). To estimate the variance (σ^2), we use the usual sample variance,

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2, \tag{6}$$

where the X_i's represent individual allele lengths and $\bar{X}$ represents the average allele length in the sample of n chromosomes. To estimate the variance squared and the fourth central moment, we use the following expressions:

$$\mathrm{Sig}^4 = \frac{(n^2-3n+3)}{n(n-1)(n-2)(n-3)}\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right]^2 - \frac{1}{(n-2)(n-3)}\sum_{i=1}^{n}(X_i-\bar{X})^4 \tag{7}$$

$$\mathrm{Gam}_4 = \frac{(n^2-2n+3)}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n}(X_i-\bar{X})^4 - \frac{(6n-9)}{n(n-1)(n-2)(n-3)}\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right]^2. \tag{8}$$

In order to derive the estimator equations (7) and (8), we begin by finding expectations for $E(X_i^3\bar{X})$, $E(X_i^2\bar{X}^2)$, and $E(\bar{X}^4)$. Some simple manipulation produces the following, with $\mu \equiv E(X_i)$:

$$E(X_i^3\bar{X}) = \frac{1}{n}[E(X_i^4) + (n-1)\mu E(X_i^3)] \tag{9}$$

$$E(X_i^2\bar{X}^2) = \frac{1}{n^2}[E(X_i^4) + 2(n-1)\mu E(X_i^3) + (n-1)[E(X_i^2)]^2 + (n^2-3n+2)E(X_i^2)] \tag{10}$$

$$E(\bar{X}^4) = \frac{1}{n^3}[E(X_i^4) + 4(n-1)\mu E(X_i^3) + 3(n-1)[E(X_i^2)]^2 + 6(n^2-3n+2)E(X_i^2) + (n^3-6n^2+11n-6)\mu]. \tag{11}$$

In order to develop unbiased estimators for σ^4 and γ_4, we now find expectations for the expressions $[\Sigma_{i=1}^{n}(X_i-\bar{X})^2]^2$ and $\Sigma_{i=1}^{n}(X_i-\bar{X})^4$. In performing the algebra for these calculations, the identities (9), (10), and (11) are used:

$$\frac{n}{n-1}E\left[\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right]^2\right] = (n-1)^2E(X_i^4) - 4(n-1)\mu E(X_i^3) + (n^2-2n+3)[E(X_i^2)]^2 - 2(n^2-5n+6)\mu^2E(X_i^2) + (n^2-5n+6)\mu^4 \tag{12}$$

$$\frac{n^2}{n-1}E\left[\sum_{i=1}^{n}(X_i-\bar{X})^4\right] = (n^2-3n+3)E(X_i^4) - 4(n^2-3n+3)\mu E(X_i^3) + 3(2n-3)[E(X_i^2)]^2 + 6(n^2-5n+6)\mu^2E(X_i^2) - 3(n^2-5n+6)\mu^4. \tag{13}$$

Finally, to simplify equations (7) and (8), we use the identities $\gamma_4 = E(X_i^4) - 4\mu E(X_i^3) + 6\mu E(X_i^2) - 3\mu^4$ and $\sigma^2 = (X_i^2) - \mu^2$ and obtain the following results:

$$\frac{n}{n-1}E\left[\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right]^2\right] = (n-1)\gamma_4 + (n^2 - 2n + 3)\sigma^2 \quad (14)$$

$$\frac{n^2}{n-1}E\left[\sum_{i=1}^{n}(X_i - \bar{X})^4\right] = (n^3 - 3n + 3)\gamma_4 + 3(2n-3)\sigma^4. \quad (15)$$

With this simple system of equations, we now find the desired expressions for σ^4 and γ_4:

$$\sigma^4 = \frac{(n^2 - 3n + 3)}{n(n-1)(n-2)(n-3)}E\left[\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right]^2\right] - \frac{1}{(n-2)(n-3)}E\left[\sum_{i=1}^{n}(X_i - \bar{X})^4\right] \quad (16)$$

$$\gamma_4 = \frac{(n^2 - 2n + 3)}{(n-1)(n-2)(n-3)}E\left[\sum_{i=1}^{n}(X_i - \bar{X})^4\right] - \frac{(6n-9)}{n(n-1)(n-2)(n-3)}E\left[\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right]^2\right] \quad (17)$$

Equations (16) and (17) are the basis of the statistical estimators Sig^4 and Gam_4 shown in equations (7) and (8). As desired, $E[\text{Sig}^4] = \sigma^4$ and $E[\text{Gam}_4] = \gamma_4$.

To verify empirically that these estimators are unbiased, we applied them to uniform and Gaussian distributions generated by computer. In addition, we tested the statistics against the distributions expected from a simple single-step mutation model and used simulations to show that in the case of a constant-sized population, the estimators are independent of sample size and agree with the analytical predictions for σ^4 and γ_4 (Roe 1992; Zhivotovsky and Feldman 1995).

APPENDIX 2: Computer simulation

Our simulation is based on a coalescent algorithm by Hudson (1990), which was modified to reflect the stepwise mutation model as well as various deviations from the assumed mutation and demographic models. In a coalescent simulation, the genealogical tree is traced backward in time from the sampled individuals to their most recent common ancestor, and a demographic expansion or contraction has the effect of shortening or lengthening of the branches of the tree in proportion to the change in population size (Hudson 1990). Once the genealogical tree is generated, mutations are distributed along the tree according to a Poisson process. To check the accuracy of our coalescent results, we used a conventional forward simulation—a simple Wright-Fisher model—in which members of a parent generation all have equal probability of producing progeny, and stepwise mutations have a fixed probability of occurring at every generation. Computer code for the coalescent simulation (in the C programming language), which takes into account a variety of deviations from the stepwise mutation model, is included on the Goldstein lab web page.

LITERATURE CITED

Di Rienzo, A., P. Donnelly, C. Toomajian, B. Sisk, A. Hill, M. L. Petzl-Erler, G. K. Haines, and D. H. Barch. 1998. Heterogeneity of microsatellite mutations within and between loci, and implications for human demographic histories. Genetics **148**:1269–1284.

Donnelly, P. 1996. Interpreting genetic-variability—the effects of shared evolutionary history. Ciba Found. Symp. **197**:25–40.

Feldman, M. W., A. Bergman, D. D. Pollock, and D. B. Goldstein. 1997. Microsatellite genetic distances with range constraints: analytic description and problems of estimation. Genetics **145**:207–216.

Goldstein, D. B., and D. D. Pollock. 1997. Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. J. Hered. **88**:335–342.

Hudson, R. R. 1990. Gene genealogies and the coalescent process. Oxf. Surv. Evol. Biol. **7**:1–44.

Kimmel, M., and R. Chakraborty. 1996. Measures of variation at DNA repeat loci under a general stepwise mutation model. Theor. Popul. Biol. **50**:345–367.

Kimmel, M., R. Chakraborty, J. P. King, M. Bamshad, W. S. Watkins, and L. B. Jorde. 1998. Signatures of population expansion in microsatellite repeat data. Genetics **148**: 1921–1930.

Reich, D. E., and D. B. Goldstein. 1998. Genetic evidence for a Paleolithic human population expansion in Africa. Proc. Natl. Acad. Sci. USA **95**:8119–8123.

Roe, A. 1992. Correlations and interactions in random walks and population genetics. Ph.D. thesis, University of London, London.

Rogers, A. R., and H. Harpending. 1992. Population growth makes waves in the distribution of pairwise genetic differences. Mol. Biol. Evol. **9**:552–569.

Weber, J. L., and C. Wong. 1993. Mutation of human short tandem repeats. Hum. Mol. Genet. **2**:1123–1128.

Zhivotovsky, L. A., and M. W. Feldman. 1995. Microsatellite variability and genetic distances. Proc. Natl. Acad. Sci. USA **92**:11549–11552.

Wolfgang Stephan, reviewing editor

Accepted December 9, 1998