Ann. Hum. Genet. (2000), **64**, 117–134
Printed in Great Britain

Microsatellite evolution in modern humans: a comparison of two data sets from the same populations

L. JIN[1], M. L. BASKETT[2], L. L. CAVALLI-SFORZA[3], L. A. ZHIVOTOVSKY[4], M. W. FELDMAN[2] AND N. A. ROSENBERG[2]

[1]*Human Genetics Center, University of Texas-Houston, P. O. Box 20334, Houston, TX. USA 77225*
[2]*Department of Biological Sciences and* [3]*Department of Genetics, Stanford University, Stanford, CA. USA 94305*
[4]*Vavilov Institute of General Genetics, Russian Academy of Sciences, 3 Gubkin St., Moscow 117809 Russia*

(Received 8.12.99. Accepted 25.2.00)

SUMMARY

We genotyped 64 dinucleotide microsatellite repeats in individuals from populations that represent all inhabited continents. Microsatellite summary statistics are reported for these data, as well as for a data set that includes 28 out of 30 loci studied by Bowcock *et al.* (1994) in the same individuals. For both data sets, diversity statistics such as heterozygosity, number of alleles per locus, and number of private alleles per locus produced the highest values in Africans, intermediate values in Europeans and Asians, and low values in Americans. Evolutionary trees of populations based on genetic distances separated groups from different continents. Corresponding trees were topologically similar for the two data sets, with the exception that the $(\delta\mu)^2$ genetic distance reliably distinguished groups from different continents for the larger data set, but not for the smaller one. Consistent with our results from diversity statistics and from evolutionary trees, population growth statistics S_k and β, which seem particularly useful for indicating recent and ancient population size changes, confirm a model of human evolution in which human populations expand in size and through space following the departure of a small group from Africa.

INTRODUCTION

Variation at microsatellite loci is extremely useful for the study of the history and genetic structure of individual species. In addition to earlier uses in DNA fingerprinting (Weir, 1996), paternity and relatedness testing (Queller *et al.* 1993), within-species clustering (Bowcock *et al.* 1994), and genetic distance (Goldstein *et al.* 1995*a*, *b*; Slatkin, 1995; Shriver *et al.* 1995), inferences based on microsatellites have proven useful for the analysis of plant (Bowers *et al.* 1999) and animal (MacHugh *et al.* 1998) domestications, for determination of island migration histories (Goldstein *et al.* 1999), for identification of individuals of unknown origin (Shriver *et al.* 1997; Davies *et al.* 1999), and for detection of hidden population structure (Pritchard & Rosenberg, 1999).

Microsatellites, also known as short tandem repeats (STRs), have been particularly informative for human evolutionary inference. Their high variability and rapid mutation have produced sufficient allele frequency differences between groups that it is possible to distinguish human populations with phylogenetic methods using relatively few loci (Bowcock *et al.* 1994). In order to refine inference methods, recent work on human microsatellite evolution has focused on the properties of various statistics and the impacts of these different approaches on historical

Correspondence: Noah A. Rosenberg, Department of Biological Sciences, Stanford, CA. USA 94305. Tel: (650) 723-4952; Fax: (650) 725-0180; E-mail: noah@charles.stanford.edu

inference. Takezaki & Nei (1996), Perez-Lezaun *et al.* (1997) and Calafell *et al.* (1998) considered phylogenetic trees of human populations based on several genetic distances. They found that metrics originally developed for use with the infinite allele model produced human evolutionary trees with higher bootstrap confidence values than those derived from the stepwise mutation model (SMM) of microsatellite evolution (Ohta & Kimura, 1973; Moran, 1975). Using 213 markers in four populations, Cooper *et al.* (1999) studied the properties of the distance $(\delta\mu)^2$ (Goldstein *et al.* 1995*b*). They confirmed the consistency of the distance in distinguishing distant human populations, provided that appropriate procedures are used to average $(\delta\mu)^2$ values across loci.

In addition to their uses in the assessment of population similarity, microsatellites have also been developed as indicators of past population growth and decline. Growth indices based on the allele frequency distributions at microsatellite loci are statistics that have a particular expectation when population size is constant. If, for a given data set, the value of the statistic deviates from that expectation, the direction of the deviation is taken to provide information about population growth or decline. Reich & Goldstein (1998) and Reich *et al.* (1999) have used two statistics, an interlocus test (g) and a within-locus test (k) to consider past population growth. The interlocus test compares the observed variance of allele-size variances across multiple microsatellite loci to its expectation under constant population size. The within-locus test utilizes a polynomial combination of moment estimators, counting the fraction of loci in which this polynomial has a positive value and comparing this fraction to null distributions simulated under constant population size. When they performed these tests on the 30 dinucleotide loci reported by Bowcock *et al.* (1994), Reich & Goldstein (1998) detected substantial evidence for population size expansion within Africa sometime between 49 000 and 64 000 years ago, but no evidence for similar expansions among non-African populations.

Kimmel *et al.* (1998) introduced an 'imbalance index' (β), designed to measure population size changes. Unlike g and k, the imbalance index is based on a microsatellite mutation model that takes asymmetry of the mutation process into account. It is calculated as the ratio of two separate estimators of the mutation parameter θ. When they analysed tetranucleotide loci collected by Jorde *et al.* (1995, 1997), Kimmel *et al.* (1998) found the greatest evidence for size change among Asians and the least evidence among Africans.

Most recently, Zhivotovsky *et al.* (2000) studied an indicator of population growth and decline based on the expectation of the fourth moment of microsatellite allele size distributions. This statistic, the 'expansion index' (S_k), uses the generalized symmetric stepwise mutation model of Zhivotovsky & Feldman (1995), which allows for allele size changes of more than one repeat unit. The expansion index is calculated using the across-locus averages of the variance and kurtosis of allele sizes. Like k and β, S_k is a within-locus statistic, which combines statistics from multiple loci into a single estimate. When they applied S_k to di-, tri-, and tetranucleotide data, Zhivotovsky *et al.* (2000) found evidence for recent population growth on all continents except the Americas and Oceania.

In this article, in order to test the consistency of microsatellite statistics, we compare diversity and growth measures for two sets of autosomal dinucleotide CA repeat microsatellite loci genotyped in the same individuals from the same populations. The older and smaller data set derives from the loci studied by Bowcock *et al.* (1994), and it includes 28 loci on chromosomes 13 and 15. The new data set, which consists of 64 loci spread over 17 chromosomes, is large compared to most human microsatellite studies. Both data sets represent individuals on all inhabited continents. The fact that both data sets were genotyped on the same collection of individuals allows for the assessment of the consistency of microsatellite statistics as data sets increase in size. Since the same individuals were tested, we do not introduce sampling differences between the two data sets. Thus, we can attribute differ-

ent inferences from the two data sets to different properties of the two collections of loci.

For these data sets, we perform three types of analysis: diversity statistics, phylogenetic trees, and growth statistics. In the assessment of diversity, population relationships, and growth, we utilize several alternative measures. Additionally, we consider the variance of the microsatellite mutational jump size distribution in estimating divergence times and effective population sizes. We discuss the consistencies and inconsistencies between the two data sets. Finally, we interpret results of the analyses in terms of their implications for human history.

MATERIALS AND METHODS

Loci

64 dinucleotide microsatellite loci were selected for genotyping from the ABI Prism Linkage Mapping Set (version 1). Each locus was genotyped in at least 10 chromosomes in each of 11 populations, and these genotypes were included as our new data set (abbreviated DS64). These loci include *APOC2* (chromosome 19), D2S222, D2S434, D3S1768, D4S2939, D4S427, D6S1009, D7S460, D7S486, D7S488, D7S493, D7S513, D7S516, D7S519, D7S531, D7S550, D7S657, D7S820, D8S277, D8S279, D8S284, D8S514, D10S537, D12S101, D12S1670, D12S373, D12S43, D12S97, D13S317, D15S120, D15S127, D15S165, D15S205, D15S230, D16S401, D16S405, D16S411, D16S415, D16S420, D16S503, D16S508, D16S667, D18S465, D18S59, D18S844, D19S152, D19S180, D19S210, D19S414, D19S420, D19S49, D1S213, D1S484, D20S100, D20S115, D20S117, D20S118, D20S171, D20S471, D21S1435, D22S1158, D22S445, HUMFABP (*FABP2*, chromosome 4), and *IGF1* (chromosome 12).

Statistics based on DS64 were compared to corresponding statistics for a subset of the Bowcock *et al.* (1994) data set. Bowcock *et al.* analysed 30 dinucleotide loci on chromosomes 13 and 15, a list of which can be found in Table 1 of Barbujani *et al.* (1997). We excluded from analysis two loci genotyped by Bowcock *et al.*: D13S133, due to its unusually high allele size variance; and FES since it is a tetranucleotide rather than a dinucleotide. We refer to this older data set based on 28 dinucleotide loci as DS28. In this data set, in almost all cases, populations were genotyped in at least 10 chromosomes at every locus. The exceptions were: D15S98 in Surui, 0 chromosomes; D15S101 in New Guineans, 2 chromosomes; *GABRB3* in Italians, 4 chromosomes; D13S125 in Karitiana, D13S126 in Japanese, PCR21 in Mayans, and UTSW1523 in New Guineans, 8 chromosomes.

The major methodological differences between the older DS28 and the newer DS64 are the number of loci in the data set, the distribution of the loci across the genome, and the fact that the data sets were genotyped by different methods. The average sample sizes of DS28 and DS64 are similar (19.5 chromosomes per locus per population in DS28; 21.2 in DS64). The individuals genotyped in the two data sets overlap to a great extent.

Populations

The populations we sampled are described by Bowcock *et al.* (1994) and in the map in figure 1 of Barbujani *et al.* (1997). These populations include three African groups – Biaka Pygmy from the Central African Republic (CAR), Lisongo (LIS), and Mbuti Pygmy from Zaire (ZAI); three Asian groups – Cambodian (CAM), Chinese (CHI), and Japanese (JAP); two European groups – Northern European (NEU) and Northern Italian (ITA); three American groups – Karitiana (KAR), Mayan (MAY), and Surui (SUR); one Melanesian (MEL) group; and two groups from Oceania – Australian (AUS) and New Guinean (NGN). All fourteen populations are included in DS28. DS64 includes eleven populations, namely: AUS, CAM, CAR, CHI, ITA, JAP, KAR, LIS, MAY, NGN, ZAI.

Some of these populations are believed to result from recent admixture with other groups. In particular, the Australian sample likely includes genetic contributions from Europeans. The Lisongo population has been considerably admixed with a Pygmy population different from the Biaka and the Mbuti.

Additionally, the Karitiana are all closely related, and they mostly descend from a few individuals who lived about five generations ago.

Genotyping

The 64 markers were labeled with fluorescent dyes and they were typed using multiplex PCR with about 10 markers per panel. Multiplex PCR was achieved with a touch-down PCR regime (Don *et al.* 1991), in order to accommodate primers with different annealing temperatures. Relative intensity among loci was adjusted by allowing different primer concentrations. PCR products were loaded with the loading buffer and size standard directly to the gel, after denaturation, without further post-PCR dilution and pooling. Electrophoresis was performed using ABA 373A sequencers. The gels were then processed by GeneScan™ (PE Biosystems) to track lanes and to measure the size of unknown fragments using the size standards (GS350 TAMRA). Genotyper™ (PE Biosystems) was used to call allele sizes.

Statistical analysis

For each population, expected heterozygosity was estimated as one minus the sum of squares of the allele frequencies at a locus. Note that the expectation of this estimator is $(n-1)/n$ times the true heterozygosity, where n is the total number of chromosomes sampled. With the sample sizes in this study, our estimator has a downward bias of a few percent.

We term alleles found only in one population among our sampled populations as 'private alleles' for that population, as described by Calafell *et al.* (1998). Since the privacy of an allele to a specific population depends on which populations are under consideration, an allele private to one population in a data set with fewer populations might no longer be private when more populations are included. We determined the number of private alleles for each population with respect to data sets DS28 and DS64.

Four genetic distances were computed separately for DS28 and DS64 using the *microsat* program of Minch *et al.* (1995). The distance measures used were the negative logarithm of the proportion of shared alleles (Bowcock *et al.* 1994), Nei's distance (Nei, 1972), Reynolds F_{st} distance (Reynolds *et al.* 1983 as in Weir, 1996), and $(\delta\mu)^2$ (Goldstein *et al.* 1995*b*). The $(\delta\mu)^2$ distance was adjusted for small sample size (Goldstein & Pollock, 1997). Significance levels for normalized Mantel correlations, which indicate the level of similarity between two distance measures (Sokal & Rohlf, 1995), were determined from 100 000 permutations. Consensus trees were obtained from 10 000 bootstraps across loci. All phylogenetic trees were constructed by the neighbor-joining algorithm (Saitou & Nei, 1987).

Divergence time estimates were calculated using $(\delta\mu)^2$ as follows: for the split time between Africans and non-Africans, $(\delta\mu)^2$ was determined between each African and each non-African group. These values were averaged to obtain an overall estimate of $(\delta\mu)^2$ between African and non-African populations. This method of computing $(\delta\mu)^2$ between two composite groups, first used by Goldstein *et al.* (1995*b*) is denoted 'method 1' by Cooper *et al.* (1999). Divergence time measured in years was computed as $[(\delta\mu)^2/(2\mu\sigma^2)]G$, where G is the generation time in years, μ is the per generation mutation rate, σ^2 is the variance of the jump size distribution (Zhivotovsky & Feldman, 1995). We employed a generation time of 27 years (Weiss, 1973). For the mutation rate, we used 6.0×10^{-4} (Feldman *et al.* 1999 based on data from Dib *et al.* 1996), and for the jump size variance, we used 2.5 (Feldman *et al.* 1999).

Effective population sizes were obtained via the estimator of diploid population size, $\overline{V}/(2\mu\sigma^2)$, where $\overline{V}$ denotes the across-locus average variance, μ is the mutation rate, and σ^2 is the jump size variance (Zhivotovsky & Feldman, 1995). Note that this estimator presumes a constant population size.

We measured several statistics that indicate features of population growth and decline. The growth statistic k was calculated as in Reich & Goldstein (1998), the expansion index S_k was determined as in Zhivotovsky *et al.* (2000), and β

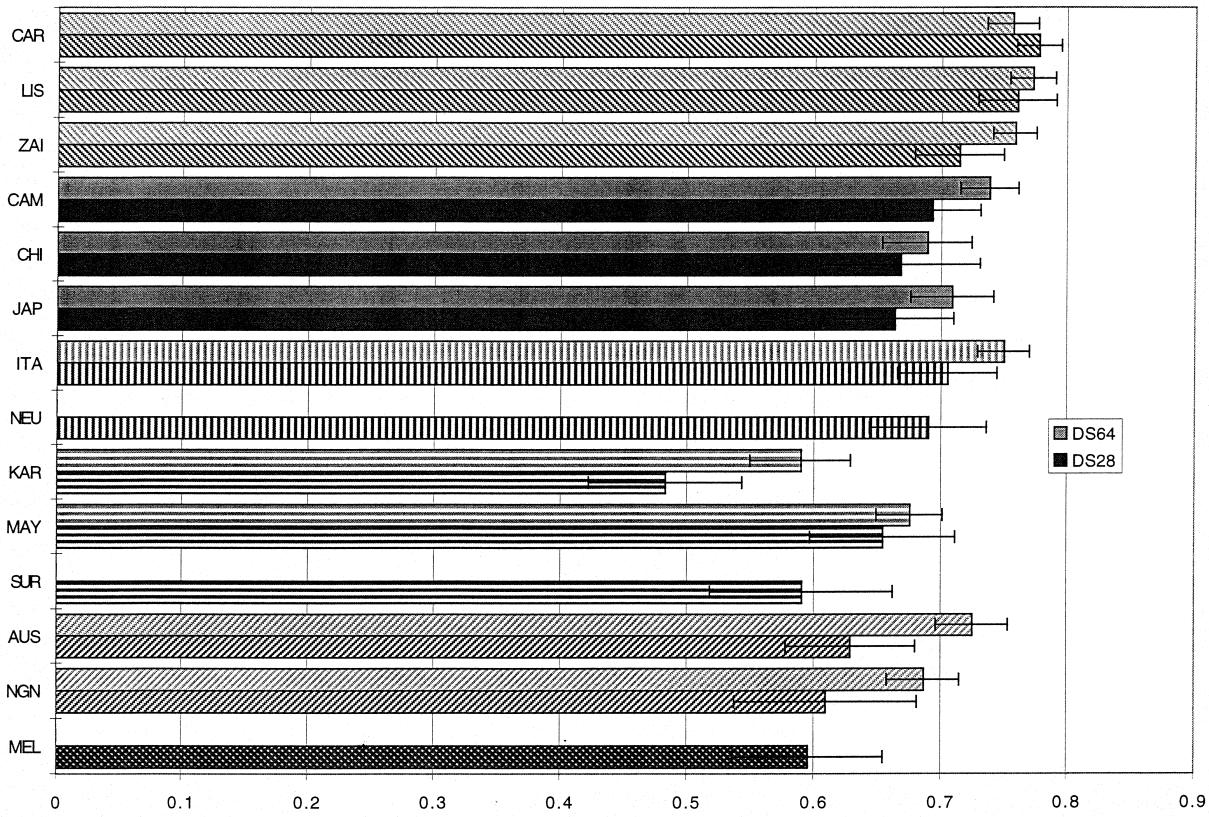

Fig. 1. Mean expected heterozygosity per locus, with 95% confidence intervals based on 1000 bootstraps of the set of loci. Populations are pattern coded by continent. The upper bar for each population is based on DS64; the lower bar is based on DS28.

was computed from the formula of Kimmel *et al.* (1998, p. 1927).

When paired tests were performed between data sets, using corresponding values of statistics in individual populations, groups that were included in DS28 but not in DS64 were excluded from the comparison. The Wilcoxon signed-ranks test was performed as described by Sokal & Rohlf (1995).

To test the hypothesis that two sets of loci had the same mutation rates, we used the fact that for a given locus, $\hat{V}/2N$ estimates the effective mutation rate $\mu\sigma^2$ (Zhivotovsky & Feldman, 1995). $\hat{V}$ is the observed variance of allele sizes at that locus over all chromosomes in the data set. Since the effective population size should be the same for every locus, the Wilcoxon two-sample test (Sokal & Rohlf, 1995), comparing two collections of locus-specific worldwide variances, tests the desired hypothesis.

RESULTS

Genetic diversity

Statistics of variation showed the well-documented patterns of higher microsatellite variation in African groups compared to non-Africans (Bowcock *et al.* 1994; Jorde *et al.* 1997; Calafell *et al.* 1998). For both DS28 and DS64, expected heterozygosities (Fig. 1) were highest in the African groups and lowest in the Native American groups. Expected heterozygosities were higher in DS64 compared to DS28 (Wilcoxon signed-rank test, two-tailed $p = 0.003$).

Other statistics of variation showed similar trends. For DS28, the number of alleles per locus was highest in Lisongo and Central African Republic Biaka Pygmies; while for DS64 the highest numbers of alleles were in Italians, Lisongo, Australians, and Central African Republic Biaka

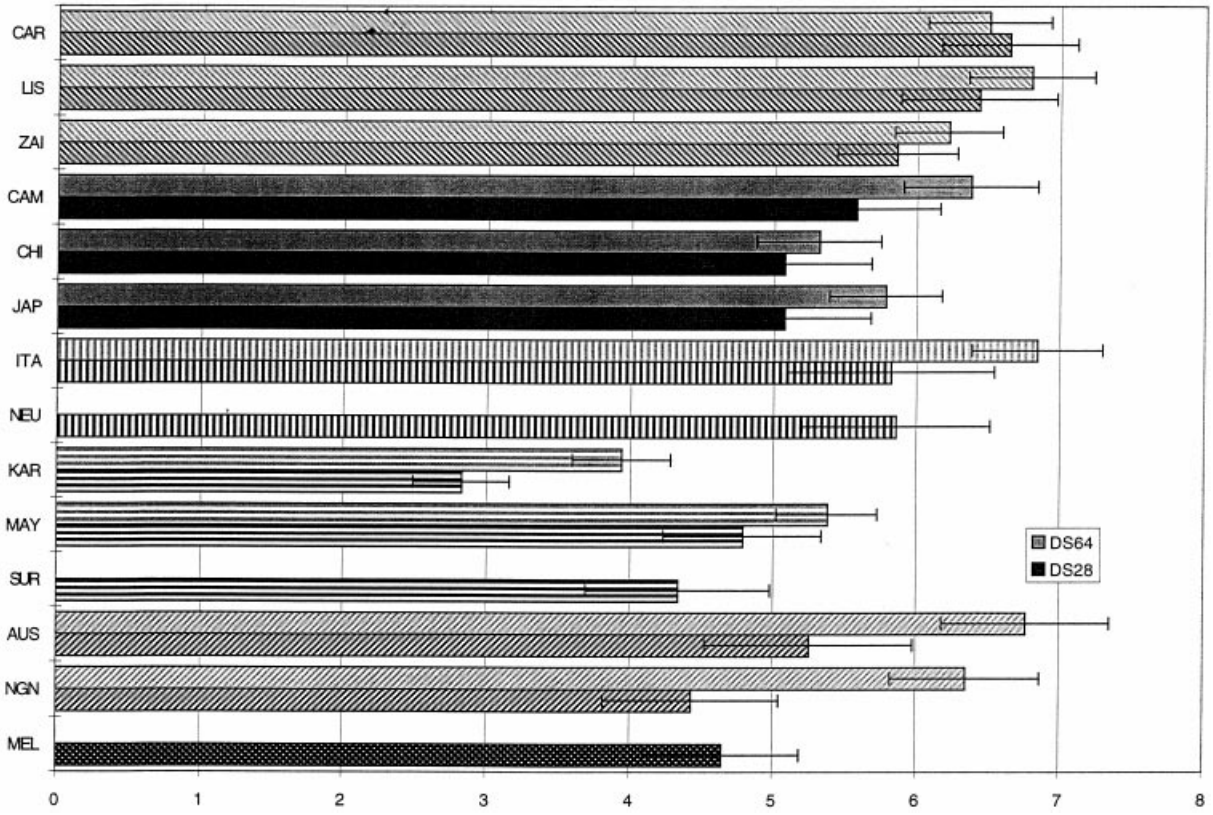

Fig. 2. Mean number of alleles per locus, with 95% confidence intervals based on 1000 bootstraps of the set of loci. Populations are pattern coded by continent. The upper bar for each population is based on DS64; the lower bar is based on DS28.

Pygmies (Fig. 2). In both data sets, Karitiana had the fewest alleles. As with the expected heterozygosity, numbers of alleles per locus were higher in DS64 than in DS28 (Wilcoxon signed-rank test, two-tailed $p = 0.002$).

For DS28, the number of private alleles per locus was highest in the three African groups, in agreement with the findings of Calafell *et al.* (1998). In DS64, however, New Guineans had the most private alleles per locus, followed by Central African Republic Biaka Pygmies and Lisongo (Table 1). DS64 had more private alleles per locus than DS28 (Wilcoxon signed-rank test, two-tailed $p = 0.042$).

For both DS28 and DS64, the mean allele size averaged across loci did not differ at all among populations (results not shown). In fact, for no pair of populations in either data set was a difference in mean allele size noticeable. This result contrasts with those of Cooper *et al.* (1999), who found slight but statistically significant differences in mean allele sizes between pairs of populations. We did not compare the grand mean allele size between data sets: since mean allele sizes depend on the

Table 1. *Mean number of private alleles per locus*

Population	DS28	DS64
CAR	0.2857 ± 0.0842	0.2656 ± 0.0818
LIS	0.2500 ± 0.0923	0.2344 ± 0.0554
ZAI	0.1786 ± 0.0702	0.1719 ± 0.0524
CAM	0.0357 ± 0.0351	0.2188 ± 0.0607
CHI	0.1071 ± 0.0798	0.1250 ± 0.0471
JAP	0.0357 ± 0.0360	0.1094 ± 0.0380
ITA	0.1071 ± 0.0558	0.2031 ± 0.0660
NEU	0.0714 ± 0.0467	
KAR	0.1071 ± 0.0762	0.1094 ± 0.0379
MAY	0.0357 ± 0.0349	0.1563 ± 0.0502
SUR	0.0741 ± 0.0506	
AUS	0.0714 ± 0.0489	0.2188 ± 0.0610
NGN	0.0357 ± 0.0359	0.3281 ± 0.0772
MEL	0.1071 ± 0.0599	

Standard errors are based on 1000 bootstraps of the set of loci.

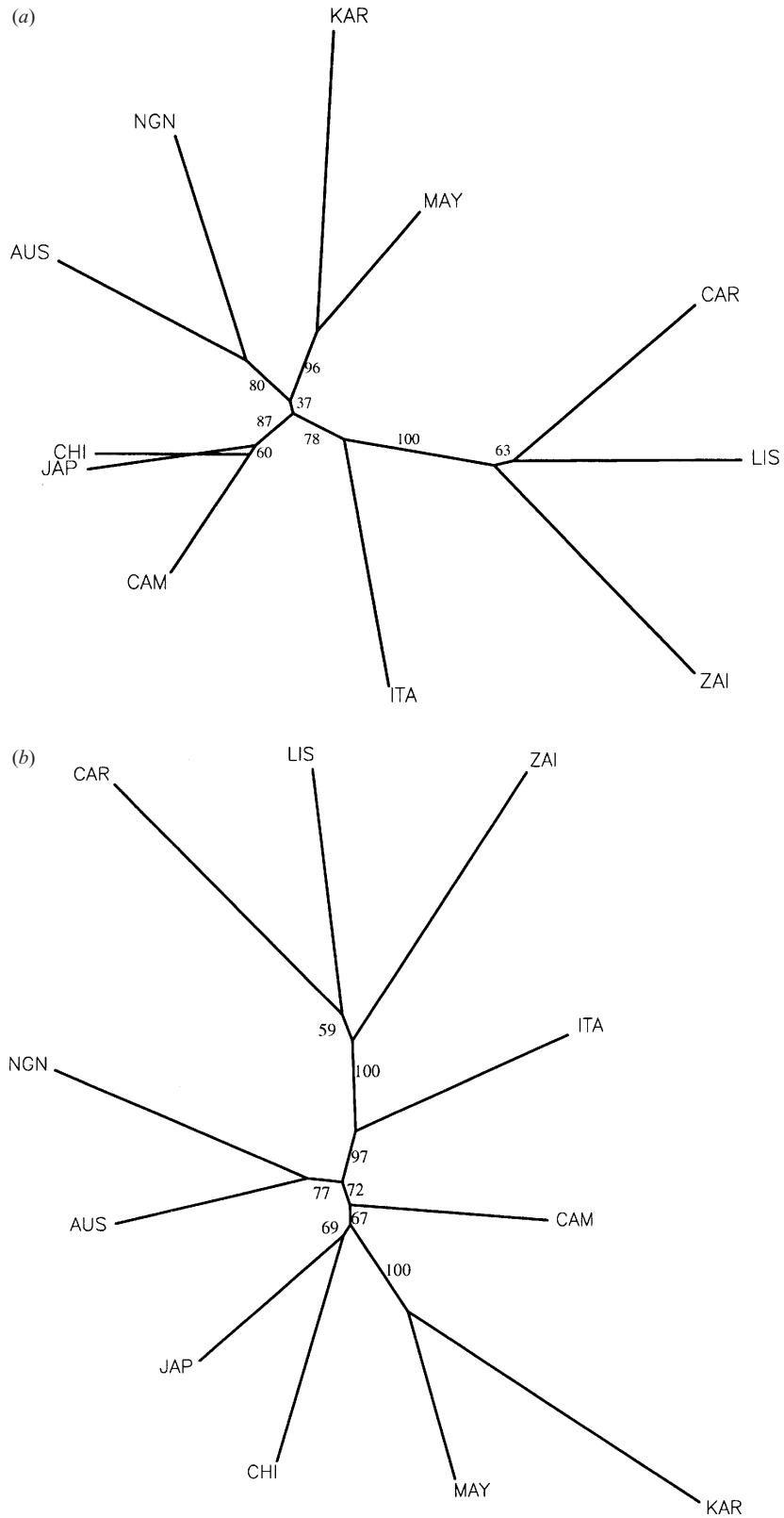

Fig. 3. Consensus neighbour-joining trees based on proportion of shared alleles. Confidence values on edges are obtained from 10 000 bootstrap replicates. (*a*) DS28, (*b*) DS64.

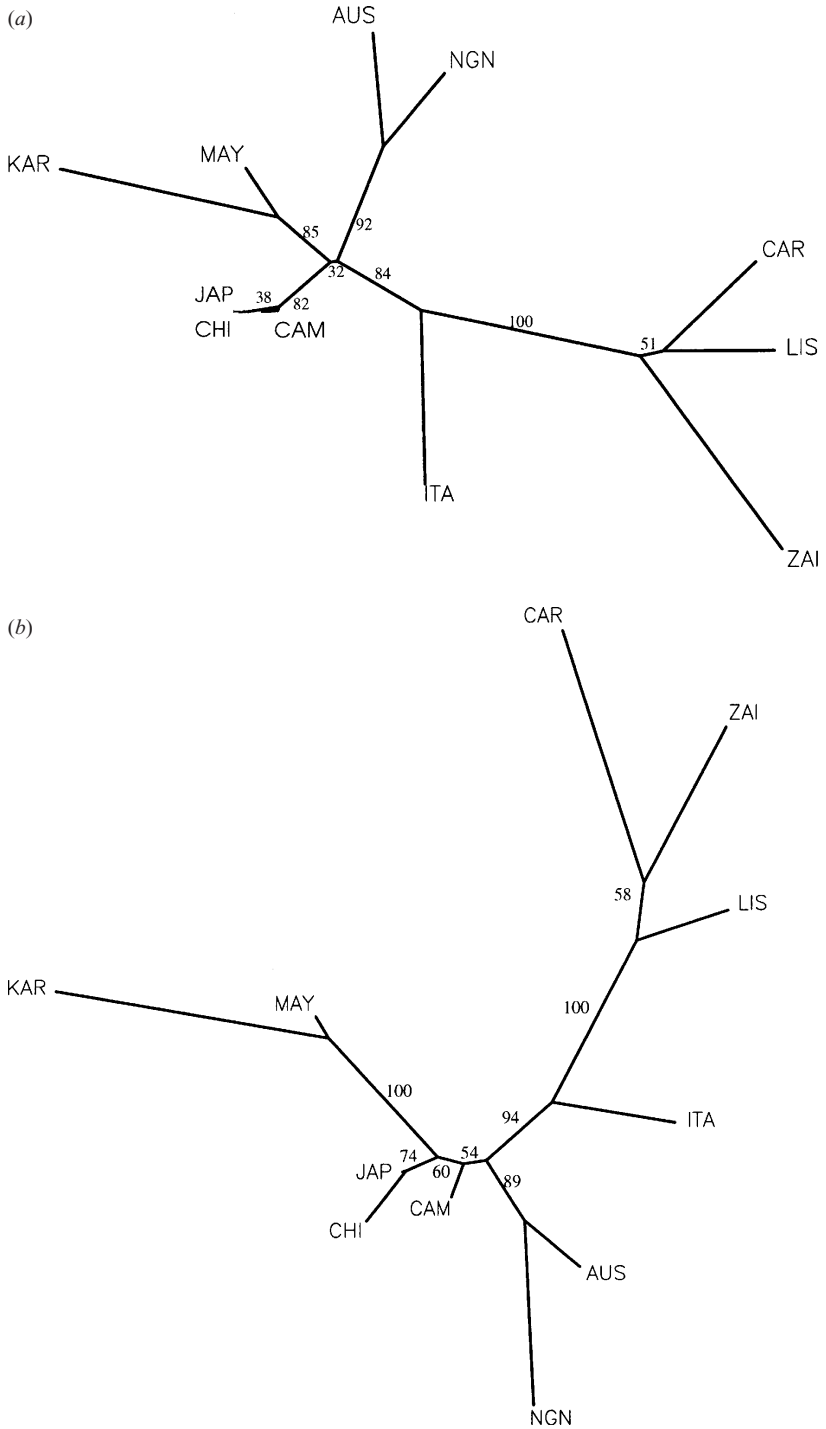

Fig. 4. Consensus neighbour-joining trees based on Nei's distance. Confidence values on edges are obtained from 10 000 bootstrap replicates. (*a*) DS28, (*b*) DS64.

locations of PCR primers used for genotyping, and since genotyping technology changed during the years between the analysis of the two data sets, this comparison of grand mean allele size would most likely reflect the difference in technique and we would be unable to test for differences due to the loci studied.

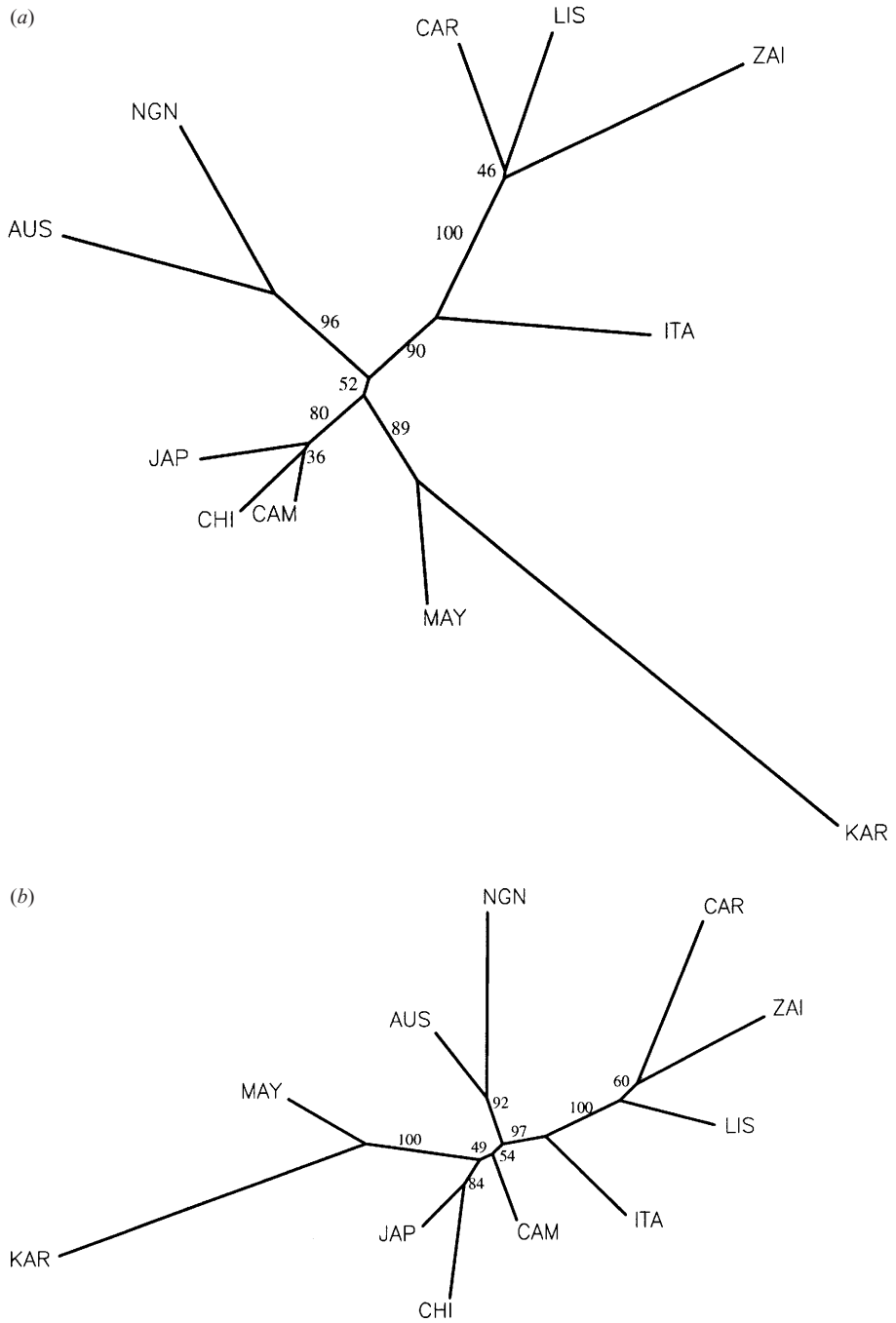

Fig. 5. Consensus neighbour-joining trees based on F_{st}. Confidence values on edges are obtained from 10 000 bootstrap replicates. (*a*) DS28, (*b*) DS64.

Historical patterns

Neighbour-joining population trees based on four genetic distance measures grouped populations from the same continent into clades (Figs 3–6). Bootstrap confidence values on tree nodes were very high for proportion of shared alleles, Nei's distance, and F_{st}. For all four distances, bootstrap confidences were higher with the larger data set.

For proportion of shared alleles, Nei's distance, and F_{st}, tree topologies between DS28 and DS64

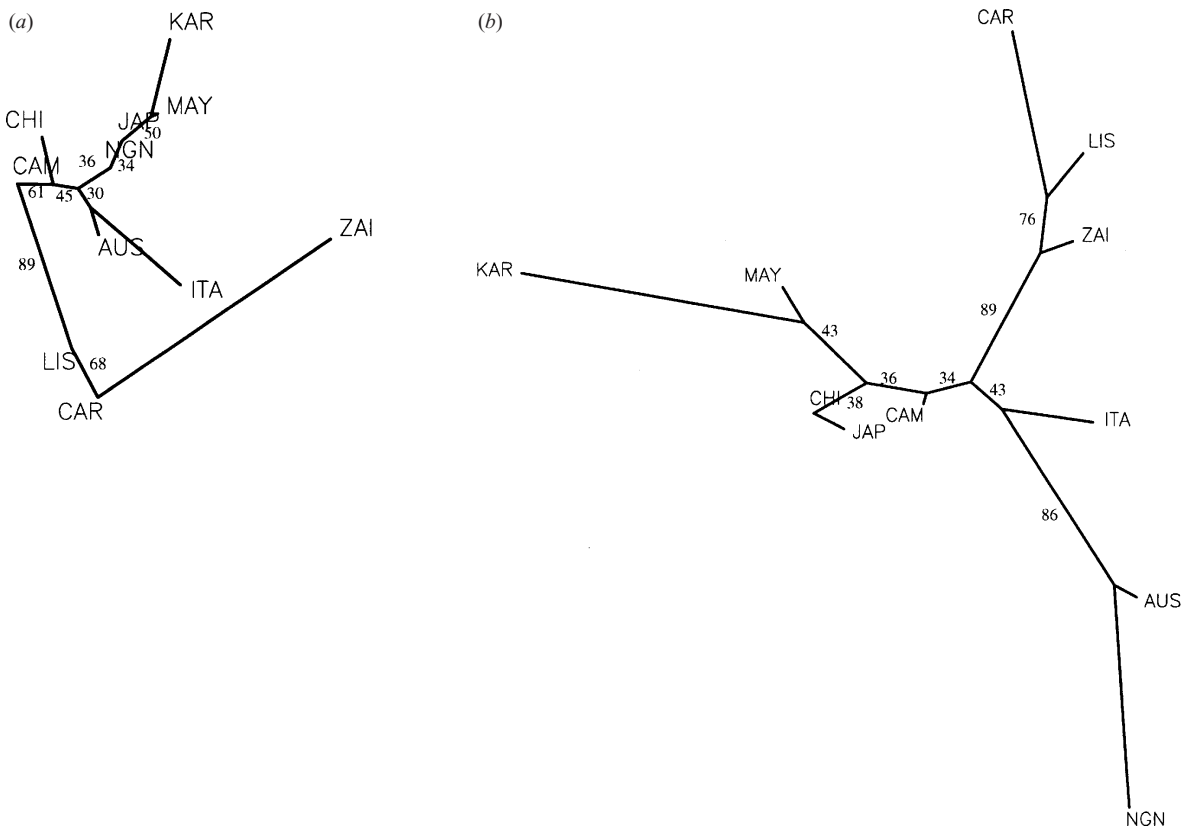

Fig. 6. Consensus neighbour-joining trees based on $(\delta\mu)^2$. Confidence values on edges are obtained from 10 000 bootstrap replicates. (*a*) DS28, (*b*) DS64.

were extremely similar. The main difference across data sets was the placement of the Cambodians; while Cambodians grouped together with Chinese and Japanese in DS28, Chinese and Japanese instead clustered with Karitiana and Maya in DS64. These three distances disagreed among each other and between data sets about the relationships among the three African populations. Additionally, for proportion of shared alleles, the branch containing Australians and New Guineans differed between data sets.

The topologies of trees based on $(\delta\mu)^2$ did not agree between the two data sets, except that for both DS28 and DS64, the African groups separated from the non-Africans (Fig. 6). Interestingly, the topology of the $(\delta\mu)^2$ tree for DS64 agreed well with the topologies of corresponding trees for the other distances, especially that of proportion of shared alleles. For DS28, the $(\delta\mu)^2$ tree was extremely different from trees based on the other distances.

The normalized Mantel correlations among proportion of shared alleles, Nei's distance, and F_{st} were extremely high, larger than 0.8 in most cases (Tables 2 and 3). Correlations between $(\delta\mu)^2$ and the other distances were also high, but smaller than 0.8. For DS64, all Mantel correlations were statistically significant at the 10^{-4} level. Except in the case of the correlation between F_{st} and $(\delta\mu)^2$, the same was true for DS28.

Population age, size, and growth

Assuming a generation time of 27 years (Weiss 1973), we obtained divergence times between Africans and non-Africans of 42 705 ± 12 239 years for DS28 and 23 213±4270 for DS64. Standard errors were obtained using 1000 bootstraps across loci, but these errors did not take into account uncertainty in the generation time, the mutation rate, or the variance of the jump size distribution. Our divergence time estimate for

Table 2. *Mantel correlation coefficients between distance matrices of 11 populations using DS28, and p-values*

	Proportion of shared alleles		$(\delta\mu)^2$		Reynolds F_{st}	
$(\delta\mu)^2$	0.7528	$< 10^{-5}$				
Reynolds F_{st}	0.8348	$< 10^{-5}$	0.5631	0.0177		
Nei's distance	0.9691	$< 10^{-5}$	0.7804	$< 10^{-5}$	0.7929	$< 10^{-5}$

Significance levels were computed from 100 000 matrix permutations.

Table 3. *Mantel correlation coefficients between distance matrices of 11 populations using DS64, and p-values*

	Proportion of shared alleles		$(\delta\mu)^2$		Reynolds F_{st}	
$(\delta\mu)^2$	0.7062	2×10^{-5}				
Reynolds F_{st}	0.9069	$< 10^{-5}$	0.7707	1×10^{-5}		
Nei's distance	0.9750	$< 10^{-5}$	0.7544	$< 10^{-5}$	0.9005	$< 10^{-5}$

Significance levels were computed from 100 000 matrix permutations.

Table 4. *Population sizes based on variances*

Population	DS28	DS64
CAR	2159 ± 260.7	2312 ± 351.0
LIS	2545 ± 362.8	2704 ± 354.6
ZAI	2291 ± 416.8	2792 ± 411.7
CAM	2323 ± 414.9	2821 ± 373.3
CHI	2150 ± 349.8	2526 ± 396.0
JAP	2000 ± 382.6	2319 ± 303.2
ITA	1702 ± 295.6	2696 ± 341.7
NEU	1815 ± 291.4	
KAR	1504 ± 391.7	2046 ± 374.2
MAY	2028 ± 392.6	2598 ± 391.4
SUR	1440 ± 272.2	
AUS	1843 ± 280.9	2386 ± 327.5
NGN	1890 ± 362.4	2187 ± 333.0
MEL	1816 ± 331.0	
World	2301 ± 262.0	3275 ± 343.2

Population sizes were calculated from $\overline{V}/(2\mu\sigma^2)$, where $\overline{V}$ denotes the across-locus average variance, μ is the mutation rate (6.0×10^{-4}), and σ^2 is the jump size variance (2.5). Standard errors are based on 1000 bootstraps of the set of loci.

DS28 differs substantially from the 156 000 year estimate by Goldstein *et al.* (1995 *b*) based on the same data mainly due to our incorporation of the jump size variance σ^2.

Population size estimates based on a constant population size model are given in Table 4. Since these estimates are directly proportional to average allele size variances, for DS28, the largest population sizes were found in African populations and the smallest in American groups. For DS64, the different population sizes had a smaller range. Estimated population sizes for DS64 were higher than those for DS28 (Wilcoxon signed-rank test, two-tailed $p = 0.001$). Consistently, effective population size estimates for individual groups ranged from 1000 to 3000.

Growth statistics indicated statistically significant evidence of size expansion in some populations. For DS28, S_k was significantly positive in Cambodians, Central African Republic Biaka Pygmies, Chinese, Japanese, Lisongo, Northern Europeans, and Zaire Mbuti Pygmies (Fig. 7), as was found by Zhivotovsky *et al.* (2000). However, with DS64, this statistic was ambiguous in all populations. Values of S_k were substantially higher in DS28 than in DS64 (Wilcoxon signed-rank test, two-tailed $p = 0.001$). The growth statistic k showed evidence of growth in Chinese and New Guineans in DS28 but it was ambiguous in DS64 (Table 5) for all populations. It is noteworthy that neither test provides any evidence of population decline in any group.

The imbalance statistic β showed a strong downward imbalance in the three African populations (Fig. 8). In DS28, Central African Republic Biaka Pygmies and Lisongo, along with Italians, had low values of β. In DS64, Central African Republic Biaka Pygmies, Lisongo, and Zaire Mbuti Pygmies had the downward imbalance. Both DS28 and DS64 indicated an upward imbalance in the Karitiana population.

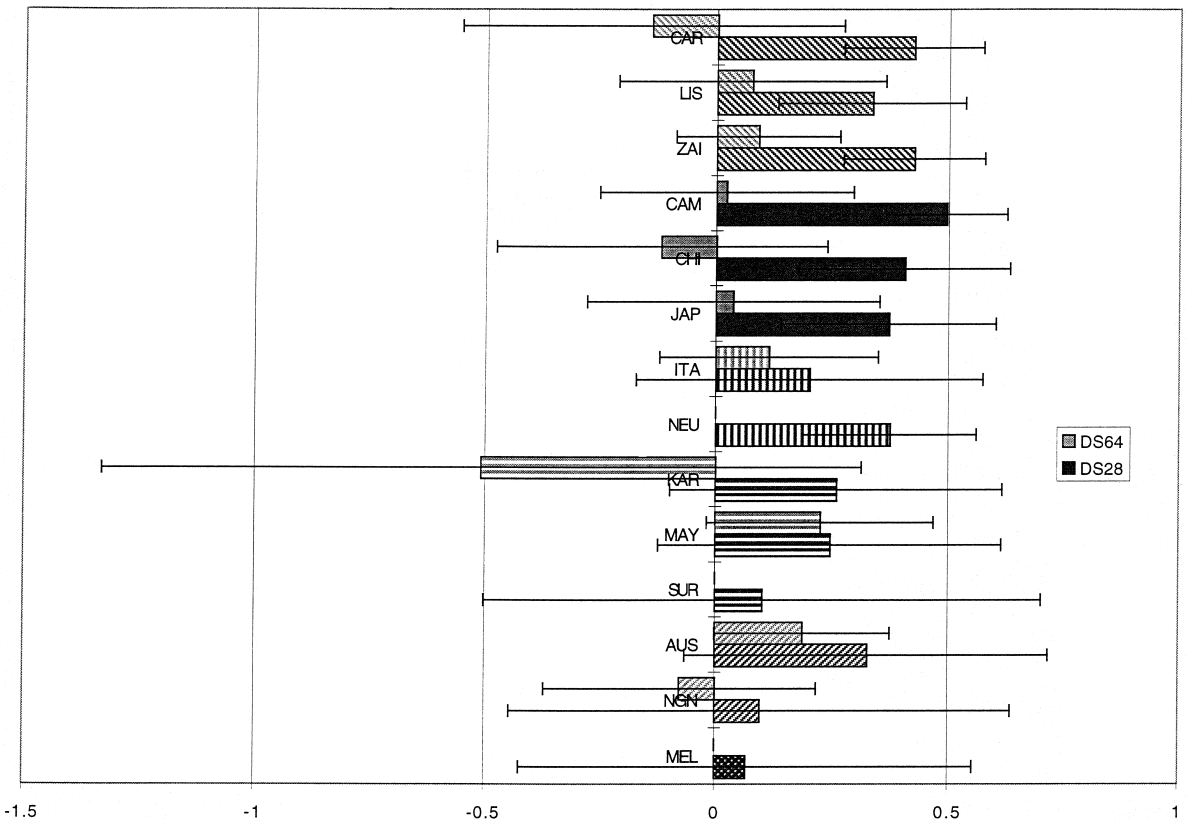

Fig. 7. Expansion index (S_k), with 95% confidence intervals based on 1000 bootstraps of the set of loci. Populations are pattern coded by continent. The upper bar for each population is based on DS64; the lower bar is based on DS28.

Table 5. *Number of positive Reich-Goldstein k statistics over total number of loci, and p-values*

	DS28		DS64	
CAR	18/28	0.9397	31/64	0.3572
LIS	15/28	0.6575	32/64	0.4538
ZAI	15/28	0.6575	35/64	0.737
CAM	18/28	0.9397	30/64	0.2691
CHI	20/28	0.9903	30/64	0.2691
JAP	16/28	0.7837	30/64	0.2691
ITA	17/28	0.8783	30/64	0.2691
NEU	18/28	0.9397		
KAR	16/28	0.7837	32/64	0.4538
MAY	16/28	0.7837	33/64	0.5532
SUR	13/27	0.4375		
AUS	13/28	0.3637	32/64	0.4538
NGN	19/27	0.9857	30/64	0.2691
MEL	13/28	0.3637		

p-values are based on a one-tailed binomial test with a lower boundary at 0.515, as described by Reich & Goldstein (1998).

DISCUSSION

Population age

Our low values for the divergence time between African and non-Africans are consistent with earlier estimates of this value. The difference between our calculation of 23 213 years based on DS64 and the estimate of 53 500 years given by Cooper *et al.* (1999) is largely due to our incorporation of variable mutational jump sizes into the stepwise mutation model through the use of the factor σ^2 (see Feldman *et al.*, 1999). Replacing the mutation rate of 5.6×10^{-4} used by Cooper *et al.* (1999) with the effective mutation rate 1.5×10^{-3} used by Feldman *et al.* (1999), we obtain 20 000 years for the divergence time, using their data. A similar correction of the results of Goldstein *et al.* (1995*b*) produces a divergence time estimate of 38 900 years. These low divergence time estimates are interesting, since recent Y-chromosome studies have placed the most recent common ancestor time of worldwide Y variation around 50 000 years ago (Pritchard *et al.*, 1999).

Unlike studies that utilize $(\delta\mu)^2$, however, Pritchard *et al.* incorporated population growth

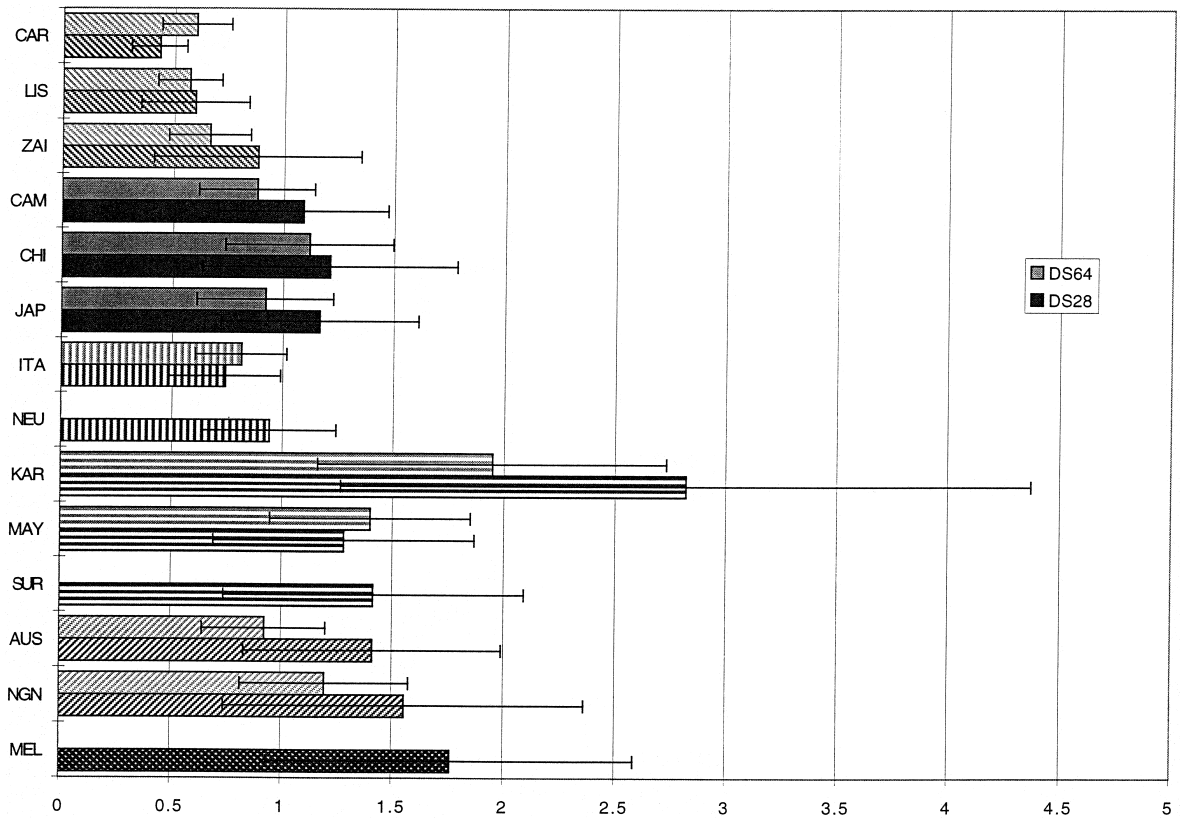

Fig. 8. Imbalance index (β), with 95% confidence intervals based on 1000 bootstraps of the set of loci. Populations are pattern coded by continent. The upper bar for each population is based on DS64; the lower bar is based on DS28.

into the estimation procedure for most recent common ancestor times, and they found that estimates decreased when growth was included in the model. They also found that models that allow for population growth fit the data much better than constant population models. The use of the formula $[(\delta\mu)^2/(2\mu\sigma^2)]G$, which implicitly assumes a constant population size over time for estimating divergence, may introduce a systematic bias, most likely in a downward direction. Corroboration of our low common ancestor times will require further theoretical work to elucidate the nature of the downward bias of $(\delta\mu)^2$ in growing populations.

Population size

Our computed worldwide effective population size of 3275 based on DS64 is reasonable, in light of other genetic studies that estimated sizes of human populations. Pritchard *et al.* (1999) suggest a worldwide effective population size of 1000 for the Y-chromosome. Based on dinucleotide microsatellites, Feldman *et al.* (1999) calculate a worldwide autosomal effective population size of 2900 individuals. Using coalescent analysis of autosomal variation in the β-globin gene, Harding *et al.* (1997) suggest a worldwide effective population size of 11 661. The fact that effective population sizes estimated from different types of genetic data and with different methods of analysis all have fairly low values indicates that human populations were very small for an extremely long time (Harpending *et al.* 1998). Small effective population sizes are consistent with the occurrence of relatively recent major population size expansions.

Population growth

The statistics S_k and β show evidence of size change in several populations, while the statistic k does not. The weak signals of k may be due to the possibility that most populations have

been fairly constant in size. However, given the conflicting results based on S_k and β, as well as evidence from other studies (Pritchard *et al.* 1999), it is more likely that this failure to detect size change reflects the low power of k at sample sizes such as those in our study (Reich *et al.* 1999).

If we assume that populations begin at mutation-drift equilibrium prior to expansion, then positive values of S_k suggest recent population expansions without major bottlenecks. Expansions that have lasted for the longest time produce the highest values of S_k, while recent bottlenecks produce negative values. Similarly, values of β less than one indicate recent population expansions, or bottlenecks in the very distant past. Values of β larger than one arise from recent bottlenecks. Thus, the relative sizes of S_k and of β can be taken as indicators of the relative times at which expansions of populations begain.

Interestingly, the calculated values of β are in agreement for the two data sets: none of the β values in Asia, Melanesia and Oceania deviate substantially from one. The Karitiana group from Brazil has a β value much higher than one, indicative of a recent bottleneck, likely resulting from the fact that this tribe descends from a small number of individuals who lived around five generations ago. African groups have β values substantially less than one, suggesting an expansion that began in the ancient past. For the Italian group, β is also less than one; this also holds for the Northern European group, though not significantly so. These data suggest that the origin of Europeans is more recent than the origin of Africans.

The ancient expansion in Africa suggested by β agrees with the conclusion of Kimmel *et al.* (1998). Using tetranucleotide loci and a different collection of populations, Kimmel *et al.* found evidence of an ancient African expansion. Similar to our results, they also argued that Europeans show earlier evidence of expansion than Asians.

Although for DS64, S_k indicated no strong signals of population expansion or decline, for DS28, S_k is positive and statistically significant in the Asian, African, and European groups. Since S_k is larger in Asians and Africans than in Europeans, Asian and African populations have likely been expanding for the longest time. The two statistics S_k and β are in agreement about the fact that the African expansion is very old; they disagree about the relative order of the the timing of expansion in Asians and Europeans. Both statistics, however, are consistent with the theory that a small group exited Africa, giving rise to Asians and Europeans. Under this scenario, Asian and European groups are younger populations than Africans, and thus they have had less time to grow. Since the differences in the values of S_k and β among Africans, Asians, and Europeans are rather small, we suggest that the early size expansions in all populations were slight and that they did not differ much in magnitude among these three continents. Since both statistics S_k and β detected no evidence of population expansion in the Americas, Melanesia or Oceania, it is likely that population size was roughly constant in these groups for a considerable amount of time.

Trees

In agreement with earlier studies of tetranucleotides (Perez-Lezaun *et al.* 1997; Calafell *et al.* 1998), the trees based on proportion of shared alleles, Nei's distance, and Reynolds F_{st} distance that we produced from DS28 and from DS64 are consistent with a model of human evolution in which Africans are the oldest population. For all four metrics, genetic distances between African and non-African populations are the highest among all interpopulation distances (results not shown), allowing us to root the trees on the branch that separates Africans and non-Africans. In most of the trees, the first groups to separate from the main non-African branch are from Europe, followed by those from Oceania. Asians and Americans are the last to split from the main branch.

It is not surprising that trees based on proportion of shared alleles, Nei's distance, and Reynolds F_{st} produced nearly topologically identical trees, due to their extremely high Mantel

correlations (Tables 2 and 3). Unlike previous studies and in spite of its lower Mantel correlation with the other three distances, however, the tree based on $(\delta\mu)^2$ in DS64 was also very similar. This distance grouped nearly all populations from the various continents with the other populations on those continents. With smaller data sets, such as DS28 and others previously studied, $(\delta\mu)^2$ had only reliably separated Africans from non-Africans. Although we confirm that $(\delta\mu)^2$ produces substantially less consistent trees than the other distances, we suggest that with sufficiently many loci, $(\delta\mu)^2$ may produce reasonable trees.

The fact that $(\delta\mu)^2$ trees do not have bootstrap confidences as high as trees based on other distances is likely due to the high sampling variance of $(\delta\mu)^2$, first noted by Goldstein *et al.* (1995*b*). They suggested that many microsatellite loci would be required in order to make accurate inferences with $(\delta\mu)^2$. Based on a series of simulations, Takezaki and Nei (1996) claimed that $(\delta\mu)^2$ produces less reliable trees than other distances not because the stepwise mutation model is inappropriate, but rather due to the higher sampling variance of $(\delta\mu)^2$ compared to other distances not based on the SMM. The high variance of $(\delta\mu)^2$ was also found by Cooper *et al.* (1999) in a large collection of loci, but it is useful to note that the sampling variance of $(\delta\mu)^2$ from the 213 loci of Cooper *et al.* (1999) is substantially lower than the corresponding variance based on the 30 loci used by Goldstein *et al.* (1995*b*).

Conclusion

The larger data set DS64 confirmed many of the findings of the smaller data set DS28, such as the geographic patterns of variation and most of the phylogenetic trees. Both data sets exhibited high heterozygosity, number of alleles, private alleles, and variance in Africans; intermediate values of these statistics in Europeans and Asians; lower values in Oceanians; and very low values in Americans. For both DS28 and DS64, trees clustered populations into their continental groups with high accuracy. For both data sets, the measures proportion of shared alleles, Nei's distance, and Reynolds F_{st}, were very highly correlated, while $(\delta\mu)^2$ was somewhat less correlated with each of the other three metrics.

For some statistics, inferences became more reliable with the larger data set DS64. Bootstrap confidence intervals for diversity statistics shrank with the larger data set, and bootstrap confidences on trees improved. Additionally, the $(\delta\mu)^2$ tree had higher bootstrap confidences and greater anthropologic plausibility than the corresponding tree for DS28. This result is encouraging: it is possible that this distance, which had never produced a historically reasonable tree until this study, simply requires larger data sets before it becomes useful for phylogenetic trees. The statistic $(\delta\mu)^2$, already uniquely useful for the estimation of divergence times, may acquire a role in tree-building for large data sets.

In some cases, however, the two data sets disagreed. Although imbalance indices were highly consistent across data sets, the growth statistics S_k and k provided stronger results in DS28 than in DS64. The two sets of loci show the same patterns of variability, but DS64 is more variable in all three indices used: heterozygosity, number of alleles, and number of private alleles. Since DS28 and DS64 do share the same individuals from the same populations, differences between statistics calculated for these two data sets may reflect sample size considerations, the different chromosomal distribution of the loci, different mutational properties of the loci themselves, or chance.

It is unlikely that the slightly larger sample size in DS64 compared to DS28 has more than a small impact on the observed difference in variability, since sample size has a different impact on each of the variability statistics used. It is clear that the number of observed alleles per locus increases with the number of individuals typed, but not rapidly enough to explain the entire difference between DS64 and DS28. Estimates of heterozygosity are essentially independent of sample size, the slight bias in the heterozygosity estimator being nearly identical for the two data sets. Lastly, the number of private alleles per locus, can be expected to decrease

slowly with sample size. With extremely small samples, we should expect most populations to have many private alleles; with extremely large samples, we should expect very few private alleles. However, since the number of private alleles also decreases as additional populations are added to a data set, it is conceivable that the fact that DS64 includes fewer populations than DS28 could explain its greater number of private alleles.

No systematic variability in microsatellite statistics across chromosomes has yet been shown, but it would be useful to consider this effect in data sets that contain at least a few loci on most chromosomes. The loci of DS28 lie on chromosomes 13 and 15, whereas those of DS64 are spread across the genome. The low variability of DS28 could be due to a generally lower mutation rate on those chromosomes. However, a comparison of worldwide mutation rates of the set of loci on chromosomes 13 and 15 (all 28 loci from DS28, plus six loci from DS64 on chromosomes 13 and 15), and those on the other chromosomes (58 loci from DS64), was not statistically significant (Wilcoxon two-sample test, $p = 0.7616$).

Since all loci in DS28 derive from the same two chromosomes, it is natural to ask if linkage disequilibrium among the loci affects the average variability per locus. Bowcock *et al.* (1994) note that most pairs of loci in DS28 are sufficiently far apart that no linkage equilibrium should be detectable without a large sample. Recent empirical (Lonjou *et al.* 1999) and theoretical (Kruglyak, 1999) studies, however, claim that linkage equilibrium should be expected between loci over a few hundred kilobases apart. Thus, even for the markers in DS28 closest to each other, it is unlikely that linkage disequilibrium has a sizeable effect.

Lastly, it is conceivable that average mutation rates may differ for the two sets of loci. Due to the fact that the loci in DS28 were selected earlier in time than those in DS64, there might have been a discovery bias, in which extremely polymorphic loci were identified first due to their high levels of variation. However, if this hypothesis were true, we would expect diversity statistics to have been higher in DS28 than in DS64, contrary to our observations.

A hypothesis which could explain the higher variability of DS64 is that the loci in DS64 were known to be highly polymorphic in human populations before the study began, whereas this was not true of DS28. Since the loci of DS64 were genotyped more recently, many polymorphic microsatellites were available at the time of study; since DS28 derives from the first analysis of worldwide human microsatellite variation and since levels of polymorphism were not well known at the time of genotyping, loci with extremely low levels of variability may have been included in that data set. It is notable that that the ten most variable loci in DS64 are more variable than all the loci in DS28. In addition, the ten least variable loci in DS28 are less variable than all except ten of the loci in the larger data set, DS64.

Testing the null hypothesis that the effective mutation rates are the same for the two data sets ($\mu\sigma^2{}_{DS28} = \mu\sigma^2{}_{DS64}$), we obtain $p = 0.1135$ (Wilcoxon two-sample test). We conclude that there are no fundamental differences between the mutational properties of the two sets of loci. Our results support the propositions that no major systematic differences exist between the data sets, that the earlier findings of Bowcock *et al.* (1994) based on DS28 can be corroborated with other microsatellite loci, and that in general, microsatellite inference improves with larger data sets.

The collection of results is consistent with the single origin model of human ancestry. The evidence from this study derives from diversity statistics, distances and trees, and growth statistics. All three types of evidence support a model in which all non-Africans descend from a group that exited Africa recently: diversity statistics show that Africa contains more genetic variation than other regions; genetic distances demonstrate that the separation between African and non-African groups is larger than the distances among non-African groups; and growth statistics indicate that the expansion of African groups

occurred prior to corresponding expansions of other groups.

Our data, based on DS64, suggest a very young date of $23\,213 \pm 4270$ years for the divergence between Africans and non-Africans, too young to match archaeological evidence (Klein, 1999). However, this estimate does not account for uncertainty in the mutation rate μ or the jump size variance σ^2. If we include not only the sampling error but also a range of $5 \times 10^{-4} - 7 \times 10^{-4}$ for μ to include values measured by earlier studies for autosomal microsatellites (e.g. Weber & Wong, 1993) and a range of 2.0–4.0 for σ^2 (Feldman *et al.* 1999), then a 95% confidence interval on the divergence time expands to divergence times expands to 0–60 400 years. This range is consistent with other genetic studies (e.g. Cooper *et al.* 1999; Pritchard *et al.* 1999) as well as anthropologic studies (Klein, 1999). Future calculations that employ $(\delta\mu)^2$ to estimate divergence times should attempt to include the role of population growth.

The authors thank Ken Kidd, Jonathan Pritchard, Federico Stefanini and the reviewers for helpful suggestions. N.A.R. is supported by a National Defense Science and Engineering Graduate Fellowship. This work was funded by NIH grant GM28428.

REFERENCES

Barbujani, G., Magagni, A., Minch, E. *et al.* (1997). An apportionment of human DNA diversity. *Proc. Natl. Acad. Sci. USA* **94**, 4516–4519.

Bowcock, A. M., Ruiz-Linares, A., Tomfohrde, J. *et al.* (1994). High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**, 455–457.

Bowers, J., Boursiquot, J.-M., This, P. *et al.* (1999). Historical genetics: the parentage of chardonnay, gamay, and other wine grapes of northeastern France. *Science* **285**, 1562–1565.

Calafell, F., Shuster, A., Speed, W. C. *et al.* (1998). Short tandem repeat polymorphism evolution in humans. *Europ. J. Hum. Genet.* **6**, 38–49.

Cooper, G., Amos, W., Bellamy, R. *et al.* (1999). An empirical exploration of the $(\delta\mu)^2$ genetic distance for 213 human microsatellite markers. *Am. J. Hum. Genet.* **65**, 1125–1133.

Davies, N., Villablanca, F. X. & Roderick, G. K. (1999). Determining the source of individuals: multilocus genotyping in nonequilibrium population genetics. *Trends Ecol. Evol.* **14**, 17–21.

Dib, C., Faure, S., Fizames, C. *et al.* (1996). A comprehensive genetic-map of the human genome based on 5264 microsatellites. *Nature* **380**, 152–154.

Don, R. H., Cox, P. T., Wainwright, B. J. *et al.* (1991). Touchdown PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.* **19**, 4008.

Feldman, M. W., KUMM, J. & Pritchard, J. K. (1999). Mutation and migration in models of microsatellite evolution. In *Microsatellites: evolution and applications* (eds. D. B. Goldstein & C. Schlotterer), pp. 98–115. Oxford: Oxford University Press.

Goldstein, D. B., Ruiz-Linares, A., Cavalli-Sforza, L. L. *et al.* (1995 a). An evaluation of genetic distances for use with microsatellite loci. *Genetics* **139**, 463–471.

Goldstein, D. B., Ruiz-Linares, A., Cavalli-Sforza, L. L. *et al.* (1995 b). Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc. Natl. Acad. Sci. USA* **92**, 6723–6727.

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

Goldstein, D. B., Roemer, G. W., Smith, D. A. *et al.* (1999). The use of microsatellite variation to infer population structure and demographic history in a natural model system. *Genetics* **151**, 797–801.

Harding, R. M., Fullerton, S. M., Griffiths, R. C. *et al.* (1997). Archaic African *and* Asian lineages in the genetic ancestry of modern humans. *Am. J. Hum. Genet.* **60**, 772–789.

Harpending, H. C., Batzer, M. A., Gurven, M. *et al.* (1998). Genetic traces of ancient demography. *Proc. Natl. Acad. Sci. USA* **95**, 1961–1967.

Jorde, L. B., Bamshad, M. J., Watkins, W. S. *et al.* (1995). Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. *Am. J. Hum. Genet.* **57**, 523–538.

Jorde, L. B., Rogers, A. R., Bamshad, M. *et al.* (1997). Microsatellite diversity and the demographic history of modern humans. *Proc. Natl. Acad. Sci. USA* **94**, 3100–3103.

Kimmel, M., Chakraborty, R., King, J. P. *et al.* (1998). Signatures of population expansion in microsatellite repeat data. *Genetics* **148**, 1921–1930.

Klein, R. G. (1999). *The human career*. Chicago: University of Chicago Press.

Kruglyak, L. (1999). Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat. Genet.* **22**, 139–144.

Lonjou, C., Collins, A. & Morton, N. E. (1999). Allelic association between marker loci. *Proc. Natl. Acad. Sci. USA* **96**, 1621–1626.

MacHugh, D. E., Loftus, R. T., Cunningham, P. *et al.* (1998). Genetic structure of 7 European cattle breeds assessed using 20 microsatellite markers. *Anim. Genet.* **29**, 333–340.

Minch, E., Ruiz-Linares, A., Goldstein, D. B. *et al.* (1995). *Microsat (version 1.5d)*: a program for calculating statistics on microsatellite allele data. *http://lotka.stanford.edu/microsat/microsat.html*.

Moran, P. A. P. (1975). Wandering distributions and the electrophoretic profile. *Theor. Pop. Biol.* **8**, 318–330.

Nei, M. (1972). Genetic distances between populations. *Am. Nat.* **106**, 283–292.

Ohta, T. & Kimura, M. (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genet. Res. Camb.* **22**, 201–204.

Perez-Lezaun, A., Calafell, F., Mateu, E. *et al.* (1997). Microsatellite variation and the differentiation of modern humans. *Hum. Genet.* **99**, 1–7.

Pritchard, J. K. & Rosenberg, N. A. (1999). Use of unlinked genetic markers to detect population stratification in association studies. *Am. J. Hum. Genet.* **65**, 220–228.

Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A. *et al.* (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol. Biol. Evol.* **16**, 1791–1798.

Queller, D. C., Strassmann, J. E. & Hughes, C. R. (1993). Microsatellites and kinship. *Trends Ecol. Evol.* **8**, 285–288.

Reich, D. E., Feldman, M. W. & Goldstein, D. B. (1999). Statistical properties of two tests that use multilocus data sets to detect population expansions. *Mol. Biol. Evol.* **16**, 453–466.

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

Reynolds, J., Weir, B. S. & Cockerham, C. C. (1983). Estimation of the co-ancestry coefficient: basis for a short-term genetic-distance. *Genetics* **105**, 767–779.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.

Shriver, M. D., Jin, L., Boerwinkle, E., *et al.* (1995). A novel measure of genetic distance for highly polymorphic loci. *Mol. Biol. Evol.* **12**, 914–920.

Shriver, M. D., Smith, M. W., Jin, L. *et al.* (1997). Ethnic-affiliation estimation by use of population-specific DNA markers. *Am. J. Hum. Genet.* **60**, 957–964.

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**, 457–462.

Sokal, R. R. & Rohlf, F. J. (1995). *Biometry*, 3 edn. New York: W. H. Freeman and Company.

Takezaki, N. & Nei, M. (1996). Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* **144**, 389–399.

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

Weir, B. S. (1996). *Genetic data analysis II*. Sunderland, MA: Sinauer Associates.

Weiss, K. (1973). Demographic models for anthropology. *Am. Antiquity* **38**, 1–186.

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

Zhivotovsky, L. A., Bennett, L., Bowcock, A. M. *et al.* (2000). Human population expansion and microsatellite variation. *Mol. Biol. Evol.*, in press.