Am. J. Hum. Genet. 61:705–718, 1997

Multilocus Genotypes, a Tree of Individuals, and Human Evolutionary History

Joanna L. Mountain[1] and L. Luca Cavalli-Sforza[2]

[1]Department of Integrative Biology, University of California, Berkeley; and [2]Department of Genetics, Stanford University, Stanford

Summary

Our goal is to infer, from human genetic data, general patterns as well as details of human evolutionary history. Here we present the results of an analysis of genetic data at the level of the individual. A tree relating 144 individuals from 12 human groups of Africa, Asia, Europe, and Oceania, inferred from an average of 75 DNA polymorphisms/individual, is remarkable in that most individuals cluster with other members of their regional group. In order to interpret this tree, we consider the factors that influence the tree pattern, including the number of genetic loci examined, the length of population isolation, the sampling process, and the extent of gene flow among groups. Understanding the impact of these factors enables us to infer details of human evolutionary history that might otherwise remain undetected. Our analyses indicate that some recent ancestor(s) of each of a few of the individuals tested may have immigrated. In general, the populations within regional groups appear to have been isolated from one another for <25,000 years. Regional groups may have been isolated for somewhat longer.

Introduction

For decades, population-genetic data have held the promise of providing insight into human evolutionary history. For much of this period, individuals were tested for very few genetic loci. Researchers therefore summarized these data at the population level, inferring trees of populations on the basis of allele frequencies (Edwards and Cavalli-Sforza 1964; Cavalli-Sforza 1967; Cavalli-Sforza and Edwards 1967; Nei and Roychoudhury 1974). They then interpreted such trees in terms of population relationships and major human migrations. Understanding that a strictly bifurcating tree is unlikely to reflect human history very accurately, a few researchers went a step further by inferring population admixture (Wijsman 1984, 1986; Bowcock et al. 1991*b*).

More recently, data for individuals have begun to accrue. These data make possible the consideration of multiple nucleotide sites, linked or unlinked, for many individuals and, hence, the direct comparison of either haplotypes or sets of genotypes. Such data may enable one to obtain greater detail regarding population separations, gene flow, and population substructure. They also make less essential the assignment of individuals to populations prior to analyses. Currently existing data sets for individuals include DNA sequences and haplotypes, microsatellite-polymorphism genotypes (also known as "short tandem repeats" [STRs]), and RFLP genotypes. The latter are the focus of this paper. We examine an inferred individual tree—that is, a tree inferred from the set of genotypes of individuals. We proceed by interpreting this tree in light of a simulation study.

Figure 1 provides a schematic summary of possible relationships between a population history (fig. 1*a*) and a tree inferred from data for individuals (fig. 1*b–d*). Much of this paper is concerned with the extent to which inferred trees for individual data fit into the category of consistency either at a regional level (fig. 1*c*) or at both the regional and the population levels (fig. 1*d*). By the term "consistency" we mean here that the tree inferred from the data for individuals corresponds with the population affiliation of those individuals; that is, all individuals of each group fall into a single cluster in the tree. The relationship between population history and a tree inferred from data for individuals depends on many factors, as discussed below. In a general sense, this approach parallels that of Cockerham (1969, 1973), who (along with authors mentioned therein) considered a hierarchical structure of individuals within isolates within subpopulations and examined the correlations between genes sampled from within each of these levels.

The first segment of DNA sequence to be studied in detail, for samples from individuals of several human populations, was the mitochondrial genome. Aquadro and Greenberg (1983), Johnson et al. (1983), Cann et al. (1987), and Vigilant et al. (1989) were among the earliest to infer relationships among individual mtDNA genomes. Since then, thousands of samples from dozens of populations have been examined. Although some re-

Received July 15, 1996; accepted for publication June 10, 1997.

Address for correspondence and reprints: Dr. Joanna L. Mountain, Department of Integrative Biology, University of California, Berkeley, CA 94720-3140. E-mail: joanna@mws4.biol.berkeley.edu

0002-9297/97/6103-0029$02.00

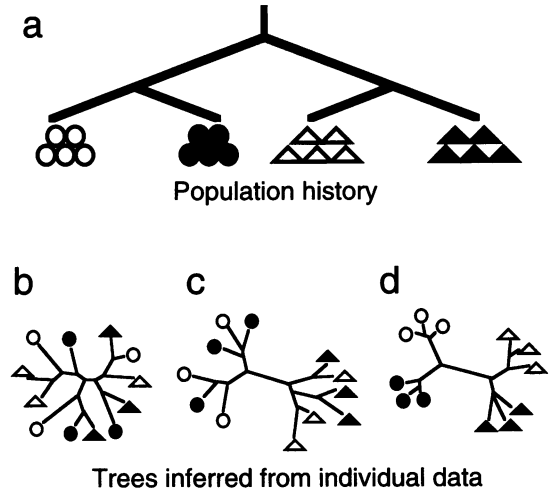

Figure 1 Schematic representation of consistency and inconsistency between regional or population affiliation (*a*) and corresponding gene or individual trees (*b–d*). *a,* Hypothetical population history. A parent population separates into two regional groups (circles vs. triangles), which, in turn, separate into two populations (blackened vs. unblackened symbols). *b–d,* Hypothetical trees inferred from data for individuals of four populations in *a*. *b,* Clustering inconsistent with both regional and population affiliation. *c,* Clustering consistent at the regional level but not at the population level. *d,* Clustering consistent at both the regional level and the population level. A fourth possibility, consistency at the population level but not at the regional level, is not shown.

searchers have found population-specific mutations, trees inferred from these sequences or haplotypes are generally inconsistent with population affiliation; that is, sequences very often cluster most closely with sequences obtained from samples of a different population (fig. 1*b*). One plausible explanation for this inconsistency between the mtDNA gene tree and population history is that much of the polymorphism observed for mtDNA probably predates population separations (Takahata 1989). Furthermore, because the mitochondrial genome undergoes no recombination, the 16,569-bp genome behaves evolutionarily as a single locus. Inferences from any one such locus lack robustness (Pamilo and Nei 1988).

Although the known polymorphisms of the Y chromosome are very few, this chromosome might potentially be studied as has the mitochondrial genome; that is, gene trees might be inferred. Of the known Y-chromosome polymorphisms, at least two appear to be continent specific. A C→T transition is limited to Amerindian males (Underhill et al. 1996), whereas an A→G transition is limited to African males (Seielstad et al. 1994). Other haplotypes, such as ALU+ chromosomes, are found in only a few geographic locations (Hammer 1994; Hammer and Horai 1995). Additional polymorphisms will reveal the extent of agreement between a Y-chromosome gene tree and population affiliation.

For at least two reasons, only a few segments of the autosomal nuclear genome have been sequenced for a large number of individuals. First, long stretches (several kilobases) of nuclear DNA must usually be examined in order to detect more than one or two variable sites. Second, the sequencing of alleles of diploid nuclear genes is more challenging in that cloning is generally required. Among the segments that have been examined closely are the HLA complex of loci on chromosome 6 and the β-globin gene cluster on chromosome 11. These data are possibly more difficult to interpret than mtDNA sequences, because gene conversion and recombination within the loci may have taken place. Although a few haplotypes appear to be population specific, trees inferred from HLA sequences are essentially inconsistent with population affiliation (Belich et al. 1992). In fact, these gene trees are often inconsistent even at the species level (within-species divergence between individuals is sometimes greater than between-species divergence), indicating that natural selection may have played a role in generating and maintaining the diversity that we observe today (Lawlor et al. 1988; Ayala 1995). Haplotypes of the β-globin locus are generally found in multiple populations (Wainscoat et al. 1986), indicating either extensive gene flow among populations, the influence of natural selection (Thompson 1975), that haplotypes predate population separations, or some combination of these factors.

STR, or microsatellite, polymorphisms have provided the opportunity for consideration of a large number of unlinked loci for each of many individuals. Bowcock et al. (1994) typed individuals from 14 populations, for 30 microsatellite markers (polymorphisms). They estimated genetic distances between individuals, considering the level of allele sharing between individuals. From these genetic distances they inferred a tree of individuals. The level of consistency between this tree and population affiliation is relatively high; that is, the tree is close to the pattern shown in figure 1*d*. Most Asian samples clustered together, as did most European samples, most Amerindian samples, and most African samples. A more detailed discussion of the level of consistency is given below.

The current study involves a large number of nuclear-DNA genetic markers: RFLPs. We considered genotypes for ≤ 100 polymorphisms of 12 individuals from each of 12 populations. These polymorphisms are believed to have a low mutation rate, between 10^{-5}/generation (for electrophoretic loci; Neel et al. 1986) and 10^{-7}/generation (for nucleotide sites; Nei 1987). Eighty-four of the markers are biallelic, and most alleles are found in all populations. This is in contrast with microsatellite (i.e., STR) loci, which have a much higher mutation rate (on the order of 10^{-3}/locus/generation; Weber and Wong 1993).

Given such a data set, we can ask a number of questions. First, is an individual tree inferred from these data consistent with regional and population affiliation (as in fig. 1*d*)? What does the extent of consistency between these two trees tell us about the time of separation of the groups studied, the level of gene flow among the groups, and the degree of substructure within these groups? Have sufficiently many genetic markers been tested to allow us draw conclusions? In order to interpret consistency or the lack thereof, we need to consider mode of inheritance, mutation rates, mutational mechanisms, population sizes, sample sizes, and number of loci tested. We have obtained, through simulation, expectations regarding consistency between individual trees and population histories, given the number of loci examined, sample sizes, population sizes, and population-separation times. We use the results of this simulation study to interpret the trees of individuals inferred here.

Subjects and Methods

Twelve individuals from each of the 12 groups described below were included in this study. Selection of individuals for analysis was based on the number of genotypes available. All individuals are unrelated, except for two pairs of Melanesians. Their population affiliation was determined in any of a number of ways: through self-identification, consideration of language or geographic location, the tracing of the individual's genealogy, or a combination of these. The term "population" is used here loosely and includes both broadly defined groups, such as northern Europeans (Bowcock et al. 1987), and more narrowly defined groups, such as the African Pygmies (collected in single villages). The set of populations includes the regions of Africa, Asia, Europe, and Oceania but not the Americas. Genotypes for ≤ 100 RFLPs were obtained on the 144 individuals. Each individual, therefore, is represented by a set of genotypes that we call a "multilocus genotype"; these multilocus genotypes were the basis for the analyses discussed below.

Sources of the Population Samples

The 12 populations considered here are Australians, Cambodians, Chinese, Europeans, Italians, Japanese, Nasioi Melanesians, coastal New Guineans, highland New Guineans, Biaka Pygmies, Mbuti Pygmies, and Senegalese Mandenka. Australian and New Guinean DNA samples were provided by A. Wilson; these have also been described elsewhere (Cann et al. 1987; Stoneking et al. 1990). The New Guineans include 12 individuals from the highland regions and 12 individuals from the coastal areas of Papua New Guinea. Cambodian samples (from Khmer individuals born in Cambodia and living in Santa Ana, CA) were collected by K. Dumars. Samples of Chinese individuals born in mainland China and living in the San Francisco Bay Area were collected by L. Wang and L. L. Cavalli-Sforza. The Europeans sampled were local residents of the Stanford University and Yale University areas and were primarily northern Europeans. The Italian sample was collected by A. Piazza and colleagues, from the town of Trino in northern Italy. This sample has been described in detail by Matullo et al. (1994). The Japanese sample, collected by L. Wang and A. Lin in 1986, consists of individuals born in Japan and living in the San Francisco Bay area. The Melanesian samples, from Bougainville in the Solomon Islands, were collected by J. Friedlander. The Melanesian sample of 12 includes two pairs of related individuals (a parent-offspring pair and an uncle-niece pair), because data for 12 unrelated individuals were unavailable. The Biaka Pygmies from the Central African Republic, sometimes called "western Pygmies," have been shown to be a 70%–75% admixture (of unknown date) with other Africans, mostly of Nilo-Saharan or Bantu origin (Wijsman 1986; Cavalli-Sforza et al. 1994, p.90). These samples were collected by L. L. Cavalli-Sforza and B. Hewlett. The Mbuti Pygmies of the Ituri forest in northeastern Zaire appear to be the least admixed with neighbors among the Pygmy populations. They are also known as the "eastern Pygmies." These samples also were collected by L. L. Cavalli-Sforza and B. Hewlett. The Senegalese sample is from the Niokolonke of the Mandenka population in the eastern part of the Senegal and was collected in 1990 by A. Langaney and colleagues (Tiercy et al. 1992; Poloni et al. 1995).

Sample Processing and DNA Analysis

For all except the Australian and New Guinean samples, blood was drawn and Epstein-Barr virus transformation was performed on the B cells, as described elsewhere (Anderson and Gusella 1984; Bowcock et al. 1987). The extraction of DNA from cell lines was performed as described by Bowcock et al. (1987). Australian and New Guinean DNA samples were extracted from placentas, as discussed by Stoneking et al. (1990). Southern blotting, hybridization, and autoradiography were then performed for all samples (Bowcock et al. 1987). Descriptions of the 100 polymorphisms tested have been given elsewhere (Bowcock et al. 1987, 1991*a*). Of these polymorphisms, 84 are biallelic among the 144 individuals whereas 8 reveal 3 alleles; the remaining 8 polymorphisms reveal 4–10 alleles. Of the 100 polymorphisms, some are very closely linked: a total of 73 independent loci (42 genes and 31 anonymous DNA segments) were considered.

Although not all individuals were tested for all polymorphisms, all calculations involving any particular pair

Table 1

Number of Markers per Pair of Individuals

	Mean ± SD No. of Markers per Pair of Individuals	Mean (above the Diagonal) and SD (below the Diagonal) No. of Markers, for Pair of Populations											
		CAR	ZAI	CHI	MEL	NEU	JPN	AUS	NGh	NGc	CAM	SEN	TRO
CAR	84.9 ± 7.1		88.3	83.4	84.6	57.7	75.1	72.9	74.4	75.8	37.8	70.2	66.8
ZAI	91.9 ± 2.3	5.6		86.9	88.0	59.3	77.8	75.4	77.0	78.4	38.8	72.5	68.9
CHI	83.8 ± 6.5	6.8	5.3		83.5	57.1	74.0	72.1	73.3	74.8	37.1	69.5	66.2
MEL	86.2 ± 4.7	5.9	4.1	6.5		58.8	76.6	74.3	75.9	77.4	38.7	71.6	68.7
NEU	49.3 ± 5.0	4.3	4.1	5.3	4.6		52.6	53.3	54.0	55.3	26.8	52.2	49.6
JPN	76.8 ± 2.8	4.0	2.5	5.1	3.4	4.2		73.5	75.5	76.9	39.6	66.9	66.3
AUS	76.1 ± 4.3	4.9	4.0	5.6	4.9	4.8	3.9		77.1	78.1	36.3	67.1	66.2
NGh	79.4 ± 1.5	3.4	1.8	4.7	3.2	4.1	2.2	3.7		80.4	38.0	68.3	68.6
NGc	81.8 ± .4	3.2	1.5	4.5	3.0	3.9	1.9	3.6	1.2		38.8	69.6	69.8
CAM	40.6 ± 1.4	1.9	1.8	2.9	1.8	2.4	1.4	2.2	.6	.4		36.4	35.9
SEN	73.8 ± 1.3	2.8	1.8	4.0	3.0	4.4	1.9	3.6	1.5	1.0	.9		64.7
TRO	71.7 ± .8	2.6	1.6	4.2	2.3	3.8	1.6	3.4	1.1	.6	.3	1.0	

[a] CAR = Central African Republic Pygmy; ZAI = Zaire Pygmy; CHI = Chinese; MEL = Melanesian; NEU = northern European; JPN = Japanese; AUS = Australian; NGh = New Guinea highland; NGc = New Guinea coastal; CAM = Cambodian; SEN = Senegalese; and TRO = Trino Italian.

of individuals were performed considering all markers that had been tested for that pair of individuals. Table 1 gives both the average number of markers considered for pairs of individuals from within each population and the average number of markers considered for pairs of individuals from two different populations.

Genetic Distance

The genetic difference between each pair of individuals, m and m', was summarized by means of an allele-sharing distance, $D_{(m,m')}$, as follows:

$$D_{(m,m')} = \frac{1}{l}\sum_{j=1}^{l} d_{(m,m')j},$$

where l is the number of loci for which both individuals have been tested, and $d_{(m,m')}j = 0$ if the individuals have identical genotypes at locus j (e.g., AA:AA or AB:AB), .5 if one individual has only a single allele in common with the other individual (e.g., AB:AA or AB:AC), and 1.0 if the individuals have no alleles in common (e.g., AA:BB). In this manner, a 144 × 144 interindividual genetic-distance matrix was generated.

Tree Inference

We inferred trees of individuals for each pair of the 12 populations, from the genetic distances, according to the neighbor-joining algorithm (Saitou and Nei 1987). We also inferred a tree relating all of the 144 individuals. The Jumble option of the NEIGHBOR program (Felsenstein 1989) was invoked; this option randomizes the order in which the individuals in the input file are considered, thereby eliminating any artificial consistency due to input order. We partitioned the latter tree into major clusters by dividing the tree along its longer internal branches. As stated above, a tree is considered consistent, at some level, if all individuals of a particular sample or set of samples form a single cluster (monophyletic group) in the tree and if no other individuals are found in this cluster (see fig. 1). In a perfectly consistent tree, therefore, each population forms a single, separate cluster (fig. 1*d*). We have not attempted to quantify the level of consistency.

We compared the tree inferred from RFLP genotypes to a tree inferred, in a similar manner, from microsatellite loci (Bowcock et al. 1994). Because only a subset of individuals from a subset of populations was considered in both studies, we compared the tree positions of these individuals only. We asked several questions: Which tree has the greatest consistency with population affiliation? Are any individuals outliers in both trees? Are there any pairs of individuals who cluster in both trees?

Very Recent Immigration Events

In order to determine whether some of the recent ancestors of any of the 144 individuals may have immigrated to their current population, we performed two types of tests. We first performed tests for each individual, in order to assess whether that individual or any recent ancestor(s) had immigrated from a particular population. The test compared the probability under the hypothesis that his or her multilocus genotype was derived from the individual's population versus the probability under the hypothesis that the multilocus ge-

notype was derived from another population (Shriver et al. 1997; Rannala and Mountain, in press). Specifically, for each individual m for each locus j (with k_j alleles), we calculated the probability of his or her genotype (X_{ijm}), given the allele frequencies (x_{ji}) in that individual's population i:

$$\Pr(X_{ijm}|x_{ji}) = \begin{cases} x_{hji}^2 & \text{if} \quad X_{ijm} = hh \\ 2x_{hji}x_{gji} & \text{if} \quad X_{ijm} = hg \end{cases},$$

for all $h = 1, 2, \ldots, k_j$ and $g = 1, 2, \ldots, k_j$ where $g \neq h$ and x_{hji} is the frequency of allele h at locus j in population i. The probability of the individual's set of genotypes was then the product of these single-locus probabilities. We similarly calculated the probability of the individual's set of genotypes, given the allele frequencies in one of the other populations i'. Our test statistic was the ratio of these two probabilities. In performing this test, we assumed that all genetic loci under consideration are in linkage equilibrium (Rannala and Mountain, in press).

We approximated the null distribution of the ratios by means of a Monte Carlo procedure (Rannala and Mountain, in press). Specifically, we generated 2,500 "individuals" (sets of genotypes for multiple loci), given the allele frequencies of the population. For each of these sets of genotypes, we calculated a ratio of posterior probabilities. We then used this distribution to estimate the probability of the observed ratio under the null hypothesis of no recent immigration. This procedure enabled us to identify those individuals whose genotypes were significantly more likely (at the 1% level) to have been derived from another population than would be expected on the basis of population-allele frequencies. We also performed power calculations for each individual, to determine whether the genotypes and allele frequencies provide sufficient statistical power to detect individuals with some immigrant ancestors. For further details of the test, see the work of Rannala and Mountain (in press). Note that, for this test, either the source population or a population closely related to the source population must be included.

The second test examined the probability that each individual's genotype was drawn from his or her population, where that population is defined by its set of allele frequencies. In order to assess the significance of this probability, we used a Monte Carlo approach, generating 1,000 random "individuals" by drawing multilocus genotypes based on the population's allele frequencies. For each individual, we calculated the probability that his or her genotype was drawn from the population. We then compared the observed probability with the distribution under the null hypothesis of no recent immigration. This test is more simple but less powerful than the first, because no explicit alternative hypothesis is being tested.

Simulation Study

We performed a simulation study in order to facilitate the interpretation of individual trees such as those inferred from multilocus RFLP genotypes. Specifically, we considered samples of size n that were drawn from two populations, of size N, that had been isolated from one another for time t (measured in units of $2N$ generations). We fixed the number of biallelic genetic loci under consideration. We used a coalescent approach to simulate the evolutionary process: for each of the two populations, for each locus, we generated a coalescent tree for the sample of $2n$ genes (Hudson 1990).

In order to simulate changes in population size, we followed the suggestion of Hudson (1990). He outlines a coalescent approach to simulation of a sudden change in population size. In these cases, time is scaled in terms of $2N$ generations, where N is the current effective population size (at time $t = 0$). We considered 2-fold, 5-fold, and 10-fold population expansions at time t_e in the past.

Having generated a coalescent tree for each population, either with or without population expansions, we truncated each tree at time t in the past, thereby generating a set of genes ancestral to the present-day sample, drawn from the parent population at time t. In order to assign genotypes to the n individuals of each present-day sample, we first chose, randomly (uniform distribution), allele frequencies for the genes of the parent population at time t. All genes present in the parent sample at time t were randomly assigned an allelic type on the basis of these allele frequencies. We assumed the mutation rate at these loci to be so low as to be negligible. This assumption is likely to be valid, given that the great majority of the alleles at the RFLP loci are believed to have arisen prior to the initial divergence among the ancestors of extant, modern humans (Mountain and Cavalli-Sforza 1994). Furthermore, for most polymorphisms, all alleles are found in most populations. Therefore, in the simulation, all descendants of an ancestral gene present at time t received the allelic type of that ancestral gene. Genotypes for n individuals from each of the two populations were generated by random pairing of the $2n$ genes present at time $t = 0$. We retained only those cases for which polymorphism remained in the two samples to the present time ($t = 0$).

From the set of genotypes for these $4n$ simulated individuals, we estimated allele-sharing genetic distances between individuals. From these distances we then inferred a tree. Finally, we examined the consistency of this tree; a tree with all individuals of each sample falling into a single cluster was considered consistent (fig. 1*d*); any other configuration was considered inconsistent. For each set of parameters, we performed 1,000 simulation

runs; for example, for the case of 100 loci with a given separation time t, we generated 100,000 total pairs of coalescent trees. We then determined the proportion of these 1,000 runs that resulted in consistent trees (as in fig. 1*d*).

Results of Simulation Study

We found that the number of loci tested had a strong impact on consistency. Consideration of only 50 loci led to a probability of consistency <.35, for samples of size 10 from two populations that had been isolated for time, $t = .1$ (corresponding to 2,000 generations, for populations of effective size 10,000). Consideration of 100 loci, however, increased this probability to ~.75, whereas consideration of 1,000 loci increased the probability to ~.99 (fig. 2*a*). If we assume that human generations are of length 25 years, this result implies that, even if two populations are isolated from one another for as long as 50,000 years, 50 loci—and even 100 loci—are too few to allow us to expect perfect consistency between the tree and population affiliation. For shorter separation times, we are even less likely to observe consistency. Once the time of separation is as large as $t = .2$ (in terms of $2N$ generations), however, even as few as 50 loci are very likely ($P > .9$) to lead to consistent trees. The number of individuals sampled per population plays a role as well, but, although increasing the number of individuals from 10 to 25 does reduce consistency somewhat, this effect is less dramatic than that of the number of loci (J. L. Mountain, unpublished simulation results).

We also explored the impact that population expansions have on the consistency between an inferred individual tree and the population affiliation of individuals (fig. 2*b*). This impact depends highly on the time, t_e, as well as on the size of the expansion. Recent expansion (wherein both populations have reached size N only recently) has the greatest impact. Consider the case of 75 loci examined for each of 10 individuals from each of two populations that separated at time $t = .05$ in the past. Without any expansion, the probability of consistency is very low (~6%). If instead the populations both doubled in size at time $t_e = .04$, the probability increases to nearly 50% (fig. 2*b*). If the sizes increased 10-fold more recently than time $t_e = .025$ ago, the probability of consistency is essentially 1.0 (fig. 2*b*). Thus, even if populations have not been isolated from one another for long, if they have only recently reached their current size, then trees inferred for individuals from these populations are likely to be consistent.

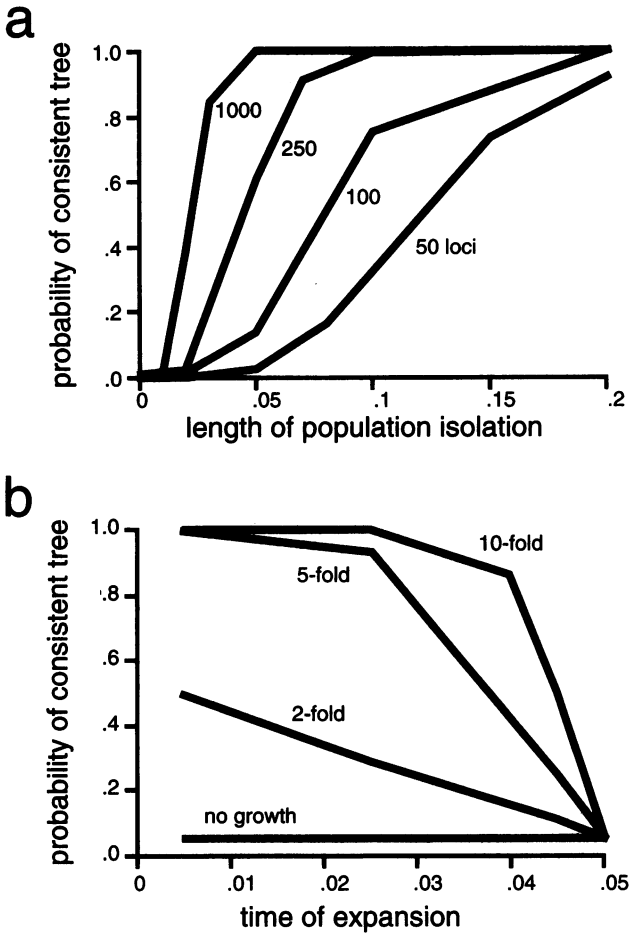

Figure 2 Probability of consistency of tree of 10 individuals from each of two populations, obtained through simulation. Probabilities are given for simulations considering genotypes for 50, 100, 250, or 1,000 polymorphic loci/individual. *a*, Constant population sizes over time. Populations are assumed to have been completely isolated from one another for a time period scaled in terms of $2N$ generations, where N is the current effective population size of each of the two populations. $t = .1$, for example, corresponds to 50,000 years, under the assumptions that two populations have effective sizes of 10,000 individuals and that generations are of length 25 years. *b*, Expanding populations. Each population is assumed to have reached its current effective size N after a 2-fold, 5-fold, or 10-fold expansion at time t_e in the past. Simulations were performed considering 75 loci and under the assumption that the two populations separated at time $t = .05$ in the past. For further details, see text.

Results

Genetic Distances from Multilocus Genotypes

For each pair of the 144 individuals, we estimated a genetic distance. These 10,296 distances are summarized in figure 3 and table 2. The Melanesian and New Guinea highland samples have the smallest average between-individual distances, whereas the two European samples (northern European and Trino) have the largest average between-individual distances (table 2). The Australian sample has the largest range of between-individual distances (fig. 3). The smallest average distances are found for individuals from within each of two clusters of populations: individuals from the three African samples and individuals from the three Australian/New Guinean

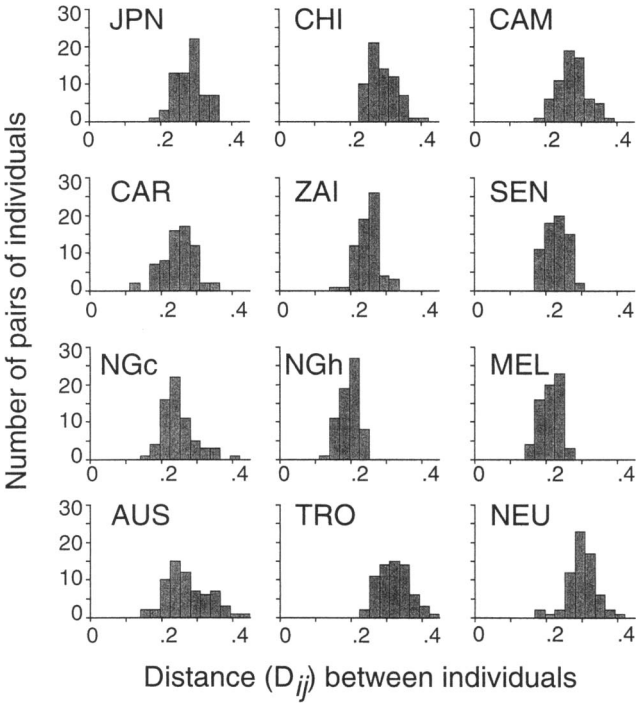

Figure 3 Histograms of genetic distances, $D_{(m,m')}$, between pairs of individuals (m and m') in each sample. Abbreviations are as in table 1.

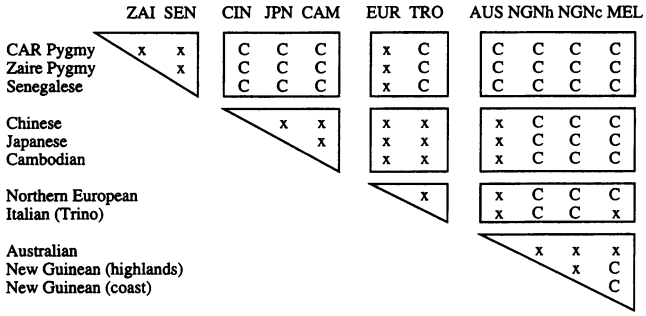

Figure 4 Summary of trees inferred for individuals from each pair of populations. C = tree of 24 individuals, which is consistent with population affiliation (fig. 1*d*); and x = tree of 24 individuals, which is inconsistent with population affiliation. Triangles include within-region results; and rectangles include between-region results.

samples are most genetically similar to one another (table 2). The smallest average distance (.238 ± .045), for instance, is for the comparison of the two New Guinea samples, as might be expected. The largest average distances are found for comparisons of individuals of African versus Oceanic, African versus Asian, and Oceanic versus Asian samples (table 2). The largest average distance (.386 ± .045), for example, is for the comparison of the Mbuti Pygmy sample from Zaire with the sample from Cambodia.

Inferred Individual Trees

Two populations/tree.—For each pair of populations, we inferred a tree of the 24 individuals, from the genetic distances. Results of this analysis are provided in figure 4. For 11 of 13 within-region cases, the trees are inconsistent. The exceptions are the trees of Melanesian and New Guinean individuals. Of the 53 between-region cases, only 15 are inconsistent. All of these cases involve European and/or Australian samples.

All populations.—From the genetic distances between individuals, we inferred a tree according to the neighbor-joining algorithm (fig. 5). We have labeled nine clusters in the tree, each defined by an internal branch (*A-I*). Although these clusters are somewhat arbitrary, other

Table 2

$D_{(m,m')}$ between Individuals

	Mean ± SD $D_{(m,m')}$ between Individuals	Mean (above the Diagonal) and SD (below the Diagonal) $D_{(m,m')}$, for Pair of Populations											
		CAR	ZAI	CHI	MEL	NEU	JPN	AUS	NGh	NGc	CAM	SEN	TRO
CAR	.249 ± .045		.265	.371	.333	.318	.349	.364	.358	.361	.378	.259	.347
ZAI	.253 ± .032	.032		.373	.343	.321	.353	.365	.352	.355	.386	.267	.346
CHI	.295 ± .039	.040	.031		.326	.354	.296	.363	.359	.345	.305	.363	.349
MEL	.215 ± .029	.037	.033	.034		.325	.313	.293	.274	.267	.329	.330	.335
NEU	.303 ± .040	.044	.045	.046	.038		.339	.353	.353	.344	.361	.314	.319
JPN	.284 ± .038	.045	.037	.045	.039	.050		.345	.342	.335	.304	.327	.342
AUS	.274 ± .061	.039	.036	.037	.043	.052	.046		.248	.269	.368	.362	.359
NGh	.207 ± .029	.042	.035	.040	.047	.054	.043	.053		.238	.358	.339	.368
NGc	.245 ± .043	.037	.032	.042	.041	.048	.043	.048	.045		.332	.355	.354
CAM	.296 ± .063	.053	.045	.059	.058	.062	.060	.051	.054	.052		.365	.361
SEN	.235 ± .028	.036	.036	.036	.034	.047	.040	.040	.036	.029	.055		.332
TRO	.325 ± .044	.050	.039	.048	.051	.054	.056	.048	.046	.046	.058	.043	

Note.—Abbreviations are as defined in footnote to table 1.

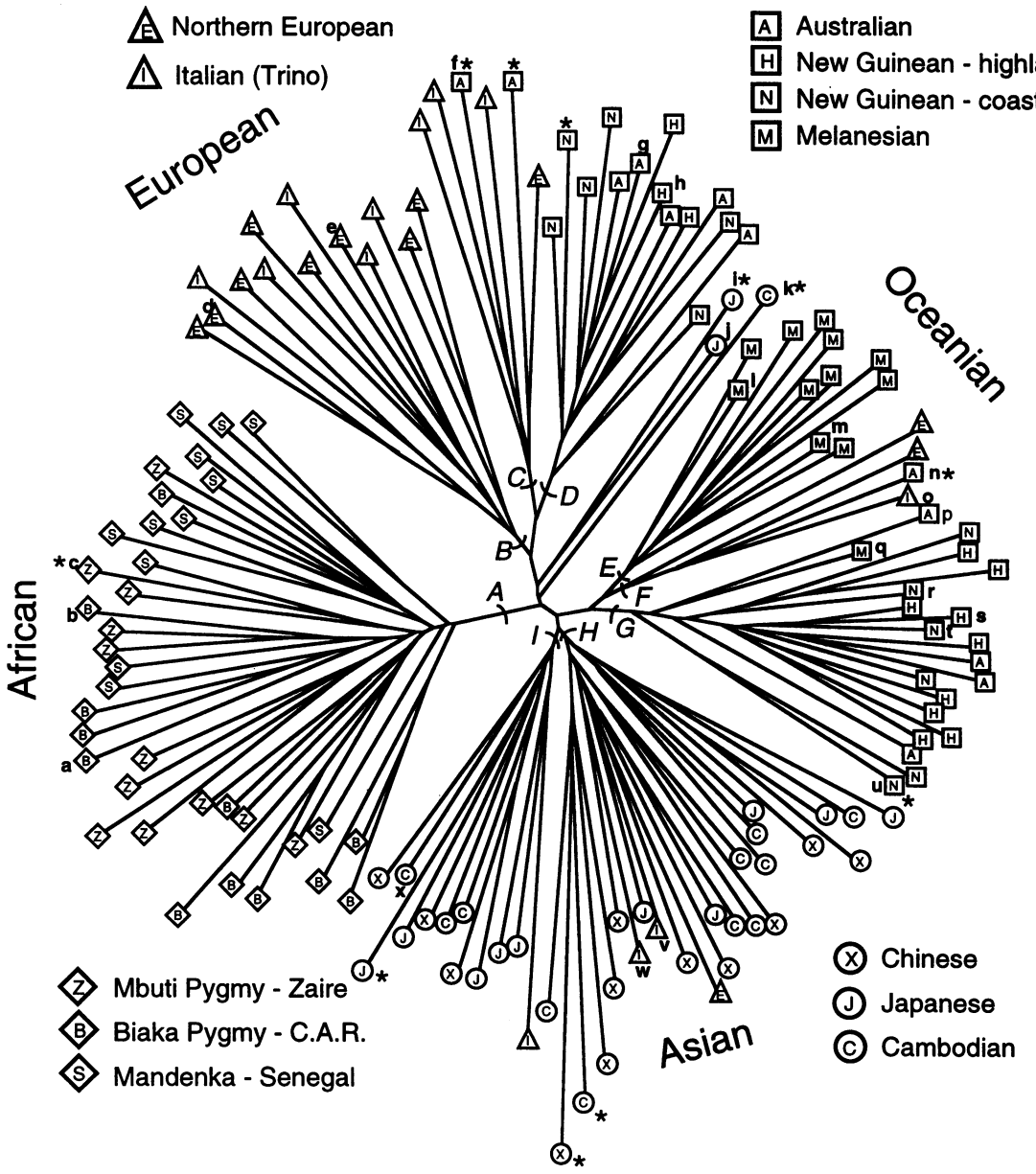

Figure 5 Tree inferred from between-individual genetic distances, according to the neighbor-joining algorithm (Saitou and Nei 1987). A total of 144 individuals from 12 samples of four world regions are represented. Small curved bars crossing interior branches partition the tree into nine clusters (A-I, excluding three outliers), corresponding to those of table 4. Lowercase letters (a–x) indicate those individuals whose genotypes appear likely (at the 1% significance level) to have been derived from another population. An asterisk (*) indicates that an individual's multilocus genotype is significantly improbable (at the 5% level), given the allele frequencies of its own population.

partitions are likely to generate similar conclusions, given the small number of internal branches. In general, each sampled individual falls within a cluster with other members of his or her regional group. All individuals sampled in Africa, for instance, fall into a single cluster (*A*), as do 33 of 36 individuals of Asian origin (cluster *H* + *I*). Although 13 of 24 European individuals form a single cluster (*B*), the European samples are somewhat scattered. European samples fall into two clusters with Oceanic individuals and into two clusters with Asian individuals. The Oceanic samples as a group fall into several clusters. One of these (*D*) is associated with a European cluster (*B*) and a mixed European/Australian cluster (*C*). The others (*E* and *G*) form a larger cluster with a mixed European/Australian cluster (*F*). The composition of these roughly defined clusters is summarized in table 3.

The consistency of the 12 population tree (fig. 5) re-

Table 3

Composition of Clusters of Tree Shown in Figure 5

	CLUSTER										
REGION	A	B	C	D	E	F	G	H	I	OUTLIERS	TOTAL
Africa	36										36
Europe		13	4			3		3	1		24
Oceania			2	14	11	1	20				48
Asia								22	11	3	36
Total	36	13	6	14	11	4	20	25	12	3	144

NOTE.—Population samples have been grouped according to their regions of origin.

flects that of the two population trees (fig. 4); although the tree is roughly consistent at the regional level (as shown schematically in fig. 1*c*), within-region consistency at the population level is rare. The Melanesian sample, with 11 of 12 individuals falling into a single cluster (*E*), is an exception. For seven of the remaining population samples (three African, two New Guinean, Japanese, and two northern European), at least half of the individuals form clusters, of two or three each, with other members of their population sample. The Cambodian, Chinese, Australian, and Italian individuals, on the other hand, fall more often into small clusters with members of other samples. Overall, 65 of 144 individuals are most closely associated, in this tree, with members of their own sample.

Previous analyses indicated that genotype frequencies for the combined New Guinean sample (coastal plus highland individuals) deviated from those expected under the assumption of Hardy-Weinberg equilibrium (Lin et al. 1994). We therefore chose to consider the set of New Guineans as two samples (those from the highland regions and those from coastal areas). Although these two groups do not form distinct clusters in the tree shown in figure 5, highland individuals form pairs in 8 of 12 cases, and coastal individuals form pairs in 6 of 12 cases. In no instance does a highland individual form a pair with a coastal individual.

Very Recent Immigration Events

By performing 1,584 (144 individuals × 11 populations) ratio tests, we were able to identify those individuals some of whose recent ancestors appear to have immigrated from another of the populations. Although, of the 1,584 tests, we would expect 1%, or ~16, to give significant results by chance, many more (45) showed significance. We identified 24 individuals (3 African, 5 European, 4 Asian, and 12 Oceanic) whose genotypes are significantly different (at the 1% level) from the expectation under a null hypothesis of no recent immigration (fig. 5 and table 4). The genotypes of eight of these individuals appeared to be at least partially derived from a population other than their own but from within their own region. Seven of these eight individuals fall into clusters with other members of their region, as is expected. The exception is the Australian, "n," with a set of genotypes that appears to be derived from several populations (table 4). In the remaining 16 cases, the genotypes appeared to be consistent with affiliation with at least one population from another region of the world. Of these 16, 7 fall into clusters consisting primarily of individuals of another region—that is, are outliers. These individuals are the most likely to be of mixed ancestry. Power calculations indicated that, in all but 10 of the 1,584 tests, the power (probability of rejecting the null hypothesis when it is false) is >.95. All of the exceptions are for comparisons of northern European individuals versus the Italian (Trino) sample. The allele frequencies in these two samples are so similar that there is insufficient power, with the available number of loci, for detection of immigration between the two populations.

We located each of the 24 individuals who may have mixed ancestry in the tree shown in figure 5, and in 9 cases we found them to be outliers—that is, clustered with members of other regional groups. For instance, the three Asian individuals (i–k) not included in any of the nine clusters each have genotypes consistent with allele frequencies in non-Asian populations. Of the four European individuals, two (v and w in clusters *H* and *I*) who fall into the Asian clusters have genotypes consistent with allele frequencies in one or more Asian populations. An Italian individual (o) with significant ratios falls into a mixed cluster (*F*). The only Melanesian sample (q) falling outside of the Melanesian cluster appears similar to the coastal New Guinean sample. Two Australians fall into small clusters with European individuals. The multilocus genotype of one of these individuals (f in cluster *C*) appears similar to those of the two European samples and the Chinese sample, whereas that of the other (n in cluster *F*), remarkably, appears similar to

Table 4

Twenty-Four Individuals Identified as Possibly Having Mixed Ancestry, on the Basis of the Ratio Test, Given Individual's Set of Genotypes for Multiple Loci

	AFRICA			EUROPE		OCEANIA				ASIA		
	CAR	ZAI	SEN	NEU	TRO	AUS	MEL	NGh	NGc	CAM	CHI	JPN
CAR		b**	b		a							
ZAI						c**						
NEU	d		d**		e							
TRO				o			o, w*			v**		v**, w
AUS	n**	n**	n*	f*, n*	f*		n*	g**		n*, p	f**, n**, p**	n*, p**
MEL					l**			m	q**		l	
NGh						h	s					
NGc							t	t**		r		u**
CAM	k			k**			x*		x			x**
JPN			i		j*							

NOTE. For details of the test, see the text and the paper by Rannala and Mountain (in press). For these tests, the significance level obtained through the Monte Carlo approach is $P < .01$ (specific P values for some entries are given below). Designations of individuals (a–x) are as defined in figure 5. Each individual is located within the row of his or her own sample and in the column(s) of the sample(s) to which his or her multilocus genotype appears to be similar. Underlining denotes that the individual was also considered in the STR tree (Bowcock et al. 1994); and a full box denotes that the individual falls outside his or her regional cluster in the STR tree.

* $P < .001$.

** $.001 < P < .005$.

those of eight different samples, including African, Asian, and European population samples.

Examining the probability that each individual's genotype was drawn from his or her population, we detected no individuals whose genotypes appear improbable (at the 1% level) under a hypothesis of no recent immigration. The genotypes of 11 individuals appeared improbable at the 5% level. These individuals are indicated by an asterisk (*) in figure 5. Because of the large number of tests performed (144), we expect to see seven significant cases simply by chance. Our detection of 11 individuals is somewhat higher than this, and we therefore conclude that a subset of these individuals is likely to have immigrant ancestry. Of the 11 individuals, 6 are among the 24 individuals whose genotypes showed significance in the ratio tests described above. Several of the remaining five individuals are found at the tips of the longer branches in the tree. These individuals may have an immigrant ancestor from a population not considered in this study.

Comparison with STR Tree of Individuals

In a previous study of 30 STR (i.e., microsatellite) markers, a tree relating 148 individuals was inferred. Of these 148, 129 fell into a clustered pair with an individual from their own global region (Bowcock et al. 1994). Within the regional groups, some population samples, such as the Karitiana, fell into a single cluster. Close to half of the individuals considered in this microsatellite study have also been examined in the RFLP genotype analysis. We were able, therefore, to compare the trees inferred from the two types of markers. The STR study included 148 individuals (including five pairs related) from 14 populations, whereas the RFLP study included 144 individuals (including two pairs related) from 12 populations. Sixty-seven individuals from eight populations were considered in both studies. Both trees show consistency at the regional level, in that African individuals tend to cluster, as do Asian, Oceanic, and European individuals. There are 16 exceptions in the STR tree (dividing the tree into five regional clusters), and there are 14 exceptions in the RFLP tree (dividing the tree into one African cluster [*A*], one Asian cluster [*H* + *I*], one European cluster [*B*], and two Oceanic clusters [*C* + *D* and *E* + *F* + *G*]). At the population level, the Melanesians cluster in both trees. Other populations (e.g., Central African Republic Pygmy and Zaire Pygmy) cluster more consistently in the STR tree. This may be due, in part, to the smaller number of individuals considered per population, the inclusion of more related pairs, and the consideration of a different set of populations. At the lowest possible level of clustering, only twice do two individuals who are paired in the RFLP tree appear as a clustered pair in the STR tree.

Although the two data sets lead to trees with a similar degree of consistency at the regional level, they are both less consistent at the population level, and they differ dramatically in the details. Nonetheless, some individuals appear to be outliers in both data sets. Of the 16 outlying individuals in the STR tree, 10 are also outliers in the RFLP study. We find that, of these 10, 7 (h, i, k, m, n, p, and x; fig. 5 and table 4) have significantly low ratios, given their multilocus RFLP genotypes. Given that 24 of the 144 individuals in the RFLP study show significance, we expect <2 of the 10 STR outlying individuals to show significance by chance. There is evidence from two independent studies, therefore, that many, if not all, of these seven individuals are of mixed ancestry.

Discussion

We have examined the multilocus RFLP genotypes of 144 individuals from 12 populations of Africa, Asia, Europe, and Oceania. The fraction of these polymorphisms that have been strongly influenced by natural selection appears to be small (Bowcock et al. 1991*b*): estimates based on this set of markers are therefore likely to reflect patterns from neutral loci. We summarized these data graphically by inferring trees relating the individuals (figs. 4 and 5). The trees inferred for two populations of the same region are almost always inconsistent, whereas most of the trees inferred for two populations of different regions are consistent. Exceptions in the latter cases always involve the European and Australian samples (fig. 4). In the 12 population tree, individuals tend to fall into clusters with other individuals from the same region (table 3). There is less consistency, however, at the population level: for only one sample, the Melanesian, do the majority (11 of 12) individuals form a cluster that includes no individuals of another population, and this sample includes two pairs of related individuals. Nonetheless, nearly half of the 144 individuals form a cluster, of at least two individuals, with members of their own population sample.

In order to interpret these graphic summaries of the data, we need expectations under various models of population history. We must also consider how the number of individuals per population and the number of loci per individual influence the pattern of the tree. As has been summarized above (see Subjects and Methods), we performed a simulation study designed to provide such expectations, at least under a set of simplified models. In that study, we considered the effects of the length of time that populations have been isolated from one another, of population expansions, of the number of individuals per population included, and of the number of loci tested. Results of those simulations enable us to begin to interpret the tree of individuals.

Implications of Simulation Study for Present Analysis

The tree relating 144 individuals, inferred from an average of 75 RFLP genotypes/individual, is roughly consistent at the regional level. Similarly, trees inferred for two populations of different regions are most often consistent (70% of cases). The latter trees most closely parallel those examined in the simulation study described above. These simulations suggest two models of human evolution (without and with a population size increase) that would lead to consistency between an individual tree and population history. Considering the results in figure 2*a*, we would conclude that the regional groups have been effectively isolated from one another for $t = .10-.15$: these are the values for which 100 and 50 loci, respectively (we considered an average of 75 loci), lead to an ~70% chance of consistency. If it is assumed that populations have maintained a roughly constant effective size of ~10,000 individuals during the course of recent human evolution, these values correspond to 2,000–3,000 generations, or 50,000–75,000 years (25-year generations are assumed). Thus, given a constant population size model, the RFLP data indicate that the African, Asian, and European plus Oceanic groups are likely to have been isolated from one another for approximately this length of time. Populations within the regional groups, showing less consistency, appear to have been isolated from one another for less time.

It is very unlikely, however, that human populations have maintained a constant size during this period of time. Although the overall long-term effective population size for humans has been estimated to be ~10,000 individuals, it is likely that the individual ancestral populations within the global regions were initially small and subsequently expanded (Shields et al. 1993). Incorporating even a moderate size increase (5–10 fold) into the simulation model, we find consistency between the tree clustering and population affiliation, even if populations have been separated for $t = <.05$, or <1,000 generations (25,000 years), if $N = 10{,}000$ (fig. 2*b*). We conclude that populations within regional groups, which do not show consistency in the tree shown in figure 5, have been isolated from one another for <25,000 years. If some of the regional ancestral populations also underwent expansion from initial sizes of 2,000–5,000 individuals, these groups need not have been separated from one another for as long as 50,000 years.

Admixture and Gene Flow among Populations

The lengths of separation times suggested above were obtained on the assumption that there is complete isolation of populations, after separation. They therefore might be termed "effective" separation times, analogous to effective population sizes. Such estimated effective times would be shorter than actual separation times, if

gene flow had taken place among populations after they had separated. It may well be that the populations within regional groups separated as long as 50,000–75,000 years ago but that gene flow among the populations continued thereafter. At this point we have incorporated neither gene flow nor population admixture into the simulation study, and so we have yet to determine the magnitude of the effect of these factors on the pattern of consistency of the individual tree. Models incorporating admixture and gene flow are certainly necessary to consider; although strictly bifurcating models are unlikely to represent the evolutionary history of modern humans, bifurcating models that incorporate the two factors might be reasonably realistic. There is evidence, for example, that the European population arose as an admixture, having originated through direct or indirect genetic contributions from neighboring Asian and African populations (Bowcock et al. 1991*b*).

Trees and Recent Immigration Events

We find strong evidence, from these data as well as from an independent set of DNA markers, that a number of the 144 individuals are of mixed ancestry (table 4). Three of these individuals (i–k) appear as the outliers in the tree shown in figure 5. Several others (f, n, o, v, and w) fall into clusters with members of other regions. Individuals i, k, and n not only are found to have significantly low ratios in the immigration test but also appear as outliers in both the STR tree and the RFLP tree. These individuals are very likely to have immigrant ancestry. Eleven individuals (denoted by an asterisks in fig. 5) have improbable genotypes, if it is assumed that there has been no recent immigration and in view of the allele frequencies of their populations. Of these 11, some have genotypes that appear to have been drawn from other populations, whereas others are peripheral in the tree shown in figure 5. The latter may have ancestors from populations not considered in this study.

Only the Senegalese and Chinese samples appear to include none of these possibly mixed individuals; other samples include several. Four of the 12 Australian individuals, for instance, have multilocus genotypes that might easily have been drawn from other populations. The histograms shown in figure 3 reveal a similar pattern, in that the Senegalese and Chinese genetic distances fall within a much narrower range than do those of the Australian sample. This finding is consistent with the conclusion, reached elsewhere (Lin et al. 1994), that the Australian sample is mixed. Such mixture is likely, considering that samples (placental tissue from individuals at a hospital) were initially obtained for the purpose of studying mtDNA; sample collection may have been conducted without extensive information on paternal ancestry. A separate study, which considered α-globin–locus haplotypes of Oceanic populations, indicated that gene flow from Southeast Asia to northwestern Australia has had a major genetic impact (Roberts-Thomson et al. 1996). According to that study, populations from the central part of the continent appear to have received less immigration from outside Australia.

Population Samples and Tree

The tree shown in figure 5 and the trees summarized in figure 4 are possibly the consequence not only of the history of these populations but also of the particular nature of these samples. As indicated above, for example, the Australian sample may include some individuals of mixed ancestry. This may explain why clusters of European and Oceanic individuals appear in the tree. More generally, these samples were certainly not selected as a random global sample of 144 individuals. Nor were the 12 populations chosen at random; instead, they were selected, at least in some cases, because the populations were believed to have been relatively free of recent admixture. We would therefore expect a tree inferred for samples chosen at random to show much less consistency. An additional complicating factor is the bias in the ascertainment of these polymorphisms (Mountain and Cavalli-Sforza 1994; Rogers and Jorde 1996). Most were included for study because they were found to be polymorphic in a small European sample. The heterozygosity values for the European samples, therefore, are higher than for those for other samples. This may have reduced the consistency of the European sample.

In summary, we have inferred a tree relating 144 individuals sampled from 12 populations of four world regions. Such a tree enables us to summarize the genotype data at the level of the individual, eliminating the usually necessary assumption that all individuals are equally representative of their populations. This tree is consistent at the regional level, with exceptions, but is inconsistent at the population level. Simulations indicate that the extent of consistency at the regional level may have resulted either from isolation of these regions for $\geq$50,000 years or from a shorter isolation period with subsequent population expansion. The lack of consistency at the population level may be the result of relatively short separation times among populations within regions (with no gene flow or admixture). Other possible explanations for the lack of consistency include the formation of populations through admixture and the intermixing of populations through gene flow. We have identified a subset of individuals some of whose ancestors may have recently immigrated to the current population. These individuals too have probably reduced the consistency of the tree. We conclude that the data are consistent with the hypothesis that, although regional groups may have been effectively isolated from one another for as many as $\geq$50,000 years, populations within regions

have probably been isolated from one another for a much shorter length of time.

In the future, larger studies, including ≥1,000 polymorphisms tested in hundreds of individuals from numerous groups around the world, should provide more-robust results. With such data we may also begin to compare individual trees and gene trees inferred from different segments of the human nuclear and mitochondrial genomes. Inferred individual trees should nonetheless be interpreted with caution, given the many factors that influence the pattern of the tree. Consideration of the roles of sampling strategy, mutation, admixture, and gene flow warrant further study.

Acknowledgments

We thank L. Jin, A. Lin, E. Minch, R. Nielsen, B. Rannala, M. Slatkin, Z. Yang, and two anonymous reviewers for valuable suggestions and/or discussion; B. Rannala for suggesting the ratio test for immigration; R. Nielsen for suggesting a second test for immigration; and E. Minch for providing subroutines. We also thank those who provided access to genotype data: R. Griffo, G. Matullo, and A. Piazza (Italians of Trino); E. Poloni and L. Excoffier (Senegalese Mandenka); and K. Ha (Cambodians). This research was funded in part by National Institutes of Health grants GM40282 to M. Slatkin and GM20467 to L.L.C.-S.

References

Anderson MA, Gusella J (1984) Use of cyclosporin A in establishing Epstein-Barr virus transformed human lymphoblastoid cell lines. In Vitro 20:856–858

Aquadro CF, Greenberg BD (1983) Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. Genetics 103:287–312

Ayala FJ (1995) The myth of Eve: molecular biology and human origins. Science 270:1930–1936

Belich MP, Madrigal JA, Hildebrand WH, Zemmour J, Williams RC, Luz R, Petzi-Erier ML, et al (1992) Unusual HLA-B alleles in two tribes of Brazilian Indians. Nature 357:326–329

Bowcock AM, Bucci C, Hebert JM, Kidd JR, Kidd KK, Friedlaender JS, Cavalli-Sforza LL (1987) Study of 47 DNA markers in five populations from four continents. Gene Geogr 1:47–64

Bowcock AM, Hebert JM, Mountain JL, Kidd JR, Rogers J, Kidd KK, Cavalli-Sforza LL (1991*a*) Study of an additional 58 DNA markers in five human populations from four continents. Gene Geogr 5:151–173

Bowcock AM, Kidd JR, Mountain JL, Hebert JM, Carotenuto L, Kidd KK, Cavalli-Sforza LL (1991*b*) Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. Proc Natl Acad Sci USA 88:839–843

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites despite a constraint in allele length. Nature 368:455–457

Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. Nature 325:31–35

Cavalli-Sforza LL (1967) Human populations. In: Alexander R (ed) Heritage from Mendel. University of Wisconsin Press, Madison, pp 309–331

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. Am J Hum Genet 19:233–257

Cavalli-Sforza LL, Piazza A, Menozzi P (1994) History and geography of human genes. Princeton University Press, Princeton

Cockerham C (1969) Variance of gene frequencies. Evolution 23:72–84

——— (1973) Analyses of gene frequencies. Genetics 74:679–700

Edwards AWF, Cavalli-Sforza LL (1964) Reconstruction of evolutionary trees. Syst Assoc Publ 6:67–76

Felsenstein J (1989) PHYLIP—phylogeny inference package (version 3.2). Cladistics 5:164–166

Hammer MF (1994) A recent insertion of an Alu element on the Y chromosome is a useful marker for human population studies. Mol Biol Evol 11:749–761

Hammer MF, Horai S (1995) Y chromosomal DNA variation and the peopling of Japan. Am J Hum Genet 56:951–962

Hudson RR (1990) Gene genealogies and the coalescent process. In: Futuyma D, Antonovics J (eds) Oxford surveys in evolutionary biology. Vol. 7. Oxford University Press, Oxford, pp 1–44

Johnson MJ, Wallace DC, Ferris SD, Rattazzi MC, Cavalli-Sforza LL (1983) Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns. J Mol Evol 19:255–271

Lawlor DA, Ward FE, Ennis PD, Jackson AP, Parham P (1988) HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. Nature 335:268–271

Lin AA, Hebert JM, Mountain JL, Cavalli-Sforza LL (1994) Comparison of 79 DNA polymorphisms tested in Australians, Japanese, and Papua New Guineans with those of five other human populations. Gene Geogr 8:191–214

Matullo G, Griffo RM, Mountain JL, Piazza A, Cavalli-Sforza LL (1994) RFLP analysis on a sample from northern Italy. Gene Geogr 8:25–34

Mountain JL, Cavalli-Sforza LL (1994) Inference of human evolutionary history through cladistic analysis of nuclear DNA restriction polymorphisms. Proc Natl Acad Sci USA 91:6515–6519

Neel JV, Satoh C, Goriki K, Fujita M, Takahashi N, Asakawa J-I, Hazama R (1986) The rate with which spontaneous mutation alters the electrophoretic mobility of polypeptides. Proc Natl Acad Sci USA 83:389–393

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nei M, Roychoudhury AK (1974) Genic variation within and between the three major races of man, caucasoids, negroids, and mongoloids. Am J Hum Genet 26:421–443

Pamilo P, Nei M (1988) Relationships between gene trees and species trees. Mol Biol Evol 5:568–583

Poloni ES, Excoffier L, Mountain JL, Langaney A, Cavalli-Sforza LL (1995) Nuclear DNA polymorphism in a Man-

denka population from Senegal: comparison with eight other human populations. Ann Hum Genet 59:43–61

Rannala B, Mountain JL. Detecting immigration using multilocus genotypes. Proc Natl Acad Sci USA (in press)

Roberts-Thomson JM, Martinson JJ, Norwich JT, Harding RM, Clegg JB, Boettcher B (1996) An ancient common origin of aboriginal Australians and New Guinea Highlanders is supported by α-globin haplotype analysis. Am J Hum Genet 58:1017–1024

Rogers AR, Jorde LB (1996) Ascertainment bias in estimates of average heterozygosity. Am J Hum Genet 58:1033–1041

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

Seielstad MT, Hebert JM, Lin AA, Underhill PA, Ibrahim M, Vollrath D, Cavalli-Sforza LL (1994) Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition. Hum Mol Genet 3:2159–2161

Shields GF, Schmiechen AM, Frazier BL, Redd A, Voevoda MI, Reed JK, Ward RH (1993) mtDNA sequences suggest a recent evolutionary divergence for Beringian and northern North American populations. Am J Hum Genet 53:549–562

Shriver MD, Smith MW, Jin L, Marcini A, Akey JM, Deka R, Ferrell RE (1997) Ethnic-affiliation estimation by use of population-specific DNA markers. Am J Hum Genet 60: 957–964

Stoneking M, Jorde LB, Bhatia K, Wilson AC (1990) Geographic variation in human mitochondrial DNA from Papua New Guinea. Genetics 124:717–733

Takahata N (1989) Gene genealogy in three related populations: consistency probability between gene and population trees. Genetics 122:957–966

Thompson EA (1975) Human evolutionary trees. Cambridge University Press, Cambridge

Tiercy J-M, Sanchez-Mazas A, Excoffier L, Shi-Isaac X, Jeannet M, Mach B, Langaney A (1992) HLA-DR polymorphism in a Senegalese Mandenka population: DNA oligotyping and population genetics of DRB1 specificities. Am J Hum Genet 51:592–608

Underhill PA, Li J, Zemans R, Oefner PJ, Cavalli-Sforza LL (1996) A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. Proc Natl Acad Sci USA 93:196–200

Vigilant L, Pennington R, Harpending H, Wilson AC (1989) Mitochondrial DNA sequences in single hairs from a southern African population. Proc Natl Acad Sci USA 86:9350–9354

Wainscoat JS, Hill AVS, Boyce AL, Flint J, Hernandez M, Thein SL, Old JM, et al (1986) Evolutionary relationships of human populations from an analysis of polymorphisms. Nature 319:491–493

Weber JL, Wong C (1993) Mutation of human short tandem repeats. Hum Mol Genet 2:1123–1128

Wijsman EM (1984) Techniques for estimating genetic admixture and applications to the problem of the origin of the Icelanders and the Ashkenazi Jews. Hum Genet 67:441–448

——— (1986) Estimation of genetic admixture in Pygmies. In: Cavalli-Sforza LL (ed) African pygmies. Academic Press, Orlando, pp 347–358