

## A Recent Shift from Polygyny to Monogamy in Humans Is Suggested by the Analysis of Worldwide Y-Chromosome Diversity

Isabelle Dupanloup,<sup>1</sup> Luísa Pereira,<sup>2</sup> Giorgio Bertorelle,<sup>1</sup> Francesc Calafell,<sup>3</sup> Maria João Prata,<sup>2</sup> Antonio Amorim,<sup>2</sup> Guido Barbujani<sup>1</sup>

<sup>1</sup> Dipartimento di Biologia, Università di Ferrara, via L. Borsari 46, I-44100 Ferrara, Italy

<sup>2</sup> Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), R. Dr. Roberto Frias, s/n, 4200 Porto, and Faculdade de Ciências da Universidade do Porto, Praça Gomes Teixeira, 4050 Porto, Portugal

<sup>3</sup> Unitat de Biologia Evolutiva, Universitat Pompeu Fabra, Doctor Aiguader 80, 08003 Barcelona, Spain

Received: 12 August 2002 / Accepted: 11 February 2003

**Abstract.** Molecular genetic data contain information on the history of populations. Evidence of prehistoric demographic expansions has been detected in the mitochondrial diversity of most human populations and in a Y-chromosome STR analysis, but not in a previous study of 11 Y-chromosome SNPs in Europeans. In this paper, we show that mismatch distributions and tests of mutation/drift equilibrium based on up to 166 Y-chromosome SNPs, in 46 samples from all continents, also fail to support an increase of the male effective population size. Computer simulations show that the low nuclear versus mitochondrial mutation rates cannot explain these results. However, ascertainment bias, i.e., when only highly variable SNP sites are typed, may be concealing any Y SNPs evidence for a recent, but not an ancient, increase in male effective population sizes. The results of our SNP analyses can be reconciled with the expansion of male effective population sizes inferred from STR loci, and with mitochondrial evidence, by admitting that humans were essentially polygynous during much of their history. As a consequence, until recently only a few men may have contributed a large fraction of the Y-chromosome pool at every generation. The number of breeding males may have increased, and the variance of their reproductive success

may have decreased, through a recent shift from polygyny to monogamy, which is supported by ethnological data and possibly accompanied the shift from mobile to sedentary communities.

**Key words:** Human populations — Y chromosome — Single-nucleotide polymorphisms — Mismatch distributions — Demographic expansions — Polygyny

### Introduction

Patterns of DNA diversity in contemporary populations offer insight into the populations' past. Processes such as migration, geographic dispersal, and admixture leave recognizable marks at the DNA level (von Haeseler et al. 1996; Cavalli-Sforza 1998; Bertranpetit 2000; De Knijff 2000), and rapid changes in population size can be inferred from the distributions of pairwise sequence differences in nonrecombining DNA segments. In particular, theory predicts that demographic changes affect the shape of the gene genealogies. Expansions result in star-like genealogies, and most mutations occurring on those genealogies do not tend to be shared among lineages (Donnelly 1996). The resulting plots of differences between pairs of individuals, or mismatch distributions, are smooth and unimodal; their mode depends on the time passed since the expansion.

sion (Rogers and Harpending 1992). Most human mitochondrial mismatch distributions agree with expansion expectations, and the few exceptions have been explained as a result of demographic crises in hunting-gathering communities (Excoffier and Schneider 1999).

For the Y chromosome, the results are not equally straightforward. Two studies (Pritchard et al. 1999; Shen et al. 2000) concluded that Y-chromosome diversity is more compatible with exponentially increasing than with constant population sizes, suggesting a recent population growth. On the contrary, in a previous study we found no evidence of growth in the mismatch distributions inferred from 11 Y-chromosome SNPs in Europe (Pereira et al. 2001). In agreement with what would be expected for populations of constant size, all mismatch distributions had multiple peaks, and statistics sensitive to demographic changes (Tajima's  $D$  and Fu's  $F_S$ ) were insignificant. The small number of polymorphic sites available might have reduced the sensitivity of the tests, although we observed that mitochondrial mismatch distributions maintain their unimodal shape even when small subsets of sites are considered (Pereira et al. 2001). We interpreted our results as reflecting either a combination of selective and demographic processes or the fact that the effective population sizes of European males ( $N_m$ ) have really remained low until recently, while female population sizes ( $N_f$ ) increased sharply in prehistoric times. Three problems remained open, namely, (1) whether a greater number of SNPs could have led to differently shaped distributions for Europe; (2) whether the lack of Y SNPs evidence for expansion extends beyond Europe; and (3) whether ascertainment bias and/or low mutation rates typical of Y SNPs but not of STRs (Pritchard et al. 1999) could have concealed an existing signal of expansion.

To address the first two questions, we analyzed two sets of data, comprising, respectively, 1007 Y chromosomes from Europe (Semino et al. 2000) and 1062 Y chromosomes from all continents (Underhill et al. 2000). Mismatch distributions were calculated, and tests of mutation-drift equilibrium were carried out. As for the third question, we simulated the effects of factors such as different mutation rates and different probabilities of ascertainment of polymorphic sites, in both stationary and expanding populations, and we compared the resulting mismatch distributions with those observed in the empirical analyses.

The results we obtained suggest that male population sizes increased substantially later than female population sizes. This observation raises the possibility that a polygamous mating system might have been widespread in prehistoric human populations.

## Materials and Methods

### *The Data*

We analyzed the data sets of Y-chromosome single-nucleotide polymorphisms published by Semino et al. (2000) (Europe [EU] data set) and Underhill et al. (2000) (world [WO] data set). They comprised, respectively, 1007 individuals from 25 European populations, typed at 22 polymorphic sites, and 1062 individuals from 21 populations of all continents, who had been typed at 166 polymorphic sites or whose genotype could be inferred with a high degree of confidence, assuming that all SNPs result from mutations that occurred only once in human evolution (Underhill et al. 2000). The only known violation of the assumption that no site mutated more than once is the M116 polymorphism, where three different alleles have been recorded. We chose to disregard that site, and therefore we considered as identical two haplotypes that differ only by a substitution at M116, namely, haplotype 19 (which had been observed only once, in an African individual) and haplotype 22. To evaluate the consequences of the lumping of different populations, in both cases we also jointly analyzed all chromosomes of either data set, regardless of their origin. The European and Near Eastern samples of the WO data set include part of the chromosomes of the EU data set. Therefore, the two data sets are not fully independent.

### *Mismatch Distributions*

Allele genealogies tend to have long internal branches in stationary populations, so that many mutations are shared by several individuals. Rapidly expanding populations, conversely, will show long terminal branches in their gene trees (or *star-like* genealogies); the mutations occurring along those branches will often be unique to single individuals (Donnelly 1996). These different patterns of substitutions are reflected in the shape of the distribution of pairwise differences between sequences, or mismatch distribution. Population subdivision (Marjoram and Donnelly 1994) and admixture (Bertorelle and Slatkin 1995) may act as confounding factors. As a rule, however, irregular and multimodal mismatch distributions are expected in stationary or shrinking populations, whereas a smooth, unimodal shape is typical of expanding populations (Rogers and Harpending 1992; Rogers et al. 1996; Excoffier and Schneider 1999).

Mismatch distributions were estimated 48 times, namely, for each of the 25 populations of the EU data set, for each of the 21 populations of the WO data set, and for the two entire data sets. They were also estimated for a number of data sets generated by computer simulation, to represent a broad spectrum of evolutionary and demographic scenarios.

In all cases, mismatch distributions and gene diversity, i.e., the probability that two randomly sampled chromosomes differ from each other (Nei 1987), were estimated by ARLEQUIN 2.0 (Schneider et al. 2000), and the observed distribution of mismatches was fitted to expectations relative to an expanding population, by Monte Carlo randomization (Schneider and Excoffier 1999). The null hypothesis was one of expansion, because there is no established expectation for the mismatch distribution in a stationary population (Harpending 1994). The age of the expansion,  $\tau$ , was also estimated from the data (Rogers and Jorde 1995) when the expansion hypothesis was not rejected and the distribution was unimodal.

### *Tests of Mutation-Drift Equilibrium*

Departures from mutation-drift or mutation-selection equilibrium were tested, in each population and in the pooled samples, by Tajima's  $D$  and Fu's  $F_S$ . In Tajima's (1989a, b) test, the parameter

$\theta = 2N\mu$  (where  $N$  is the population size, and  $\mu$  is the mutation rate) is independently estimated twice, from the number of polymorphic sites and from the average number of pairwise differences (or average mismatch) in the sample. Under equilibrium, the two  $\theta$  estimates should overlap. Differences between them, measured by the statistic  $D$ , may be caused by changes in the population size, or selection, or both. Fu's (1997)  $F_S$  statistic compares the observed number of alleles in a sample with the number of alleles expected in a population of constant size on the basis of the observed average mismatch. Both  $D$  and  $F_S$  take negative values when the population expands and positive values when it shrinks. Different selective regimes may affect the shape of the underlying gene tree and, hence, mimic the effects of demographic changes.

The significance of  $D$  and  $F_S$  was tested by randomization, in agreement with Simonsen and co-workers' (1995) observation that critical values of the former test based on the beta distribution are too conservative. By the coalescent simulation program implemented in the ARLEQUIN package (Schneider et al. 2000), random samples were repeatedly generated from hypothetical stationary populations whose parameter  $\theta$  was equal to the average number of pairwise differences observed in the population of interest (Tajima 1989a). In this way, empirical null distributions of the relevant statistics were generated by repeating the simulations 1000 times, each time recording the values of  $D$  and  $F_S$ . It was straightforward to obtain empirical estimates of the probability of the observed  $D$  and  $F_S$  values from these distributions, under the hypothesis of neutrality and constant population size.

### Monte Carlo Simulations

Computer simulations were run to evaluate the effects of mutation rates and ascertainment bias on the probability of detecting an expansion, once it has occurred. Biallelic Y-chromosome markers mutate slowly; estimated mutation rates per site and per year,  $\mu$ , range between  $1.2 \times 10^{-9}$  (Thomson et al. 2000) and  $2.5 \times 10^{-8}$  (Hammer 1995; Jobling et al. 1997). On the contrary, for the hypervariable region of the mitochondrial genome  $\mu$  estimates can be as high as  $3.2 \times 10^{-7}$  (Soodyall et al. 1997; Jazin et al. 1998; Siggurdottir et al. 2000). It is conceivable that low rates of mutation may reduce the possibility to identify population growth in studies based on Y-chromosome SNPs.

To evaluate the effects of different mutation rates on the possibility to detect an expansion, we generated samples from stationary and expanding populations by Monte Carlo simulation using the SIMCOAL program (Excoffier et al. 2000). The simulation algorithm was based on the coalescent process with superimposed mutations, as described by Hudson (1990). Each sample was obtained by first generating its genealogy. Mutations were then randomly placed on the genealogy, assuming that they occur according to a Poisson process. More details are given by Pereira et al. (2001).

For four mutation rates (from  $5 \times 10^{-9}$  to  $1 \times 10^{-7}$  per site and per year), we simulated 1000 samples of 80 chromosomes under the assumption of a large and constant effective population size ( $N_m = 5000$  haploid individuals) and random mating. Each chromosome had 1000 potentially variable SNP sites, and each site could mutate only once. For the same mutation rates we also simulated exponential population expansions using the same coalescent approach. The simulated expansion started 50,000 years (or 2500 generations) ago, a figure commonly estimated in mitochondrial studies (Rogers and Harpending 1992; Excoffier and Schneider 1999). Population size increased by a factor of 100 to a final effective size  $N_0 = 100,000$ , corresponding to a rate of exponential growth  $r = 0.0018$ . Depending on the mutation rates and on the shapes of population genealogies, variable numbers of sites (in practice, never exceeding 199) mutated and became polymorphic.

Tajima's  $D$  and Fu's  $F_S$  statistics were estimated in each of the simulated samples. The fit of the observed distribution of mismatches to a model of population expansion was tested by the bootstrap approach implemented in ARLEQUIN (Schneider et al. 2000). The statistic  $SSD$ , a sum of squared deviations from expansion expectations, was estimated and its empirical probability was computed by performing sets of 100 coalescent simulations of expansions. Finally, we defined three basic shapes of the mismatch distribution, namely, unimodal with a peak at zero differences (Type 0), unimodal with a maximum  $>0$  (Type 1), and bi- or multimodal (Type 2), and counted the number of occurrences of each type in each set of 1000 simulations.

While complete sequences of hundreds or thousands of base pairs are analyzed in mitochondrial studies, and in Shen and co-workers' (2000) Y-chromosome study, in most SNP analyses only sites known in advance to be variable are typed. In this way, some rare or private polymorphisms are likely to be missed, possibly affecting the inferred mismatch distributions. That phenomenon is called selection of sites (a term we prefer to avoid because of the possible confusion with natural selection), or ascertainment bias. A second set of Monte Carlo simulations was run to evaluate the effects of ascertainment bias on the power to detect population growth. We simulated 1000 samples from populations that underwent an ancient (50,000 years ago) or a recent (10,000 years ago) demographic expansion. Each sample was composed of 100 individuals, i.e., 100 sets of 30,000 potentially variable SNP sites. Because of the low mutation rate ( $\mu = 1 \times 10^{-8}$ , a value close to the estimates for Y-chromosome biallelic markers [Hammer 1995; Jobling et al. 1997]), we brought to 30,000 the number of sites considered, so that a sufficiently large number of them would become polymorphic in the course of each simulation.

The effect of ascertainment bias was reproduced by excluding from the analysis the least polymorphic sites, i.e., those that have a higher chance to escape detection in the phase of polymorphism discovery (Underhill et al. 1997). After each run of the simulation, we computed Tajima's  $D$  statistic, by considering both all sites that mutated and so became polymorphic or (to represent the loss of sites that escape ascertainment) only the sites whose rarer allele had a frequency  $p > 1\%$  (mutations shared by at least two chromosomes), or  $p > 2\%$  (mutations shared by at least three chromosomes), or  $p > 3\%$  (mutations shared by at least four chromosomes).

## Results

### Mismatch Distributions

A ragged pattern, either bi- or trimodal, is observed in the analysis of all populations in the EU data set (Fig. 1). Despite considering twice as many polymorphic sites as in the previous study of the same continent (Pereira et al. 2001), mismatch distributions with a peak at zero differences are still the most common. Predictably, by doubling the number of sites considered, the average distance between peaks increased. For instance, in the Iberian populations, mismatch distributions based on 11 sites (Pereira et al. 2001) displayed peaks at zero, three, and five differences, whereas, in this study, the peaks are located at zero, four, and seven differences.

Doubling the number of polymorphic sites analyzed (Table 1; compare with Table 1 of Pereira et al. 2001), the number of different haplotypes increased

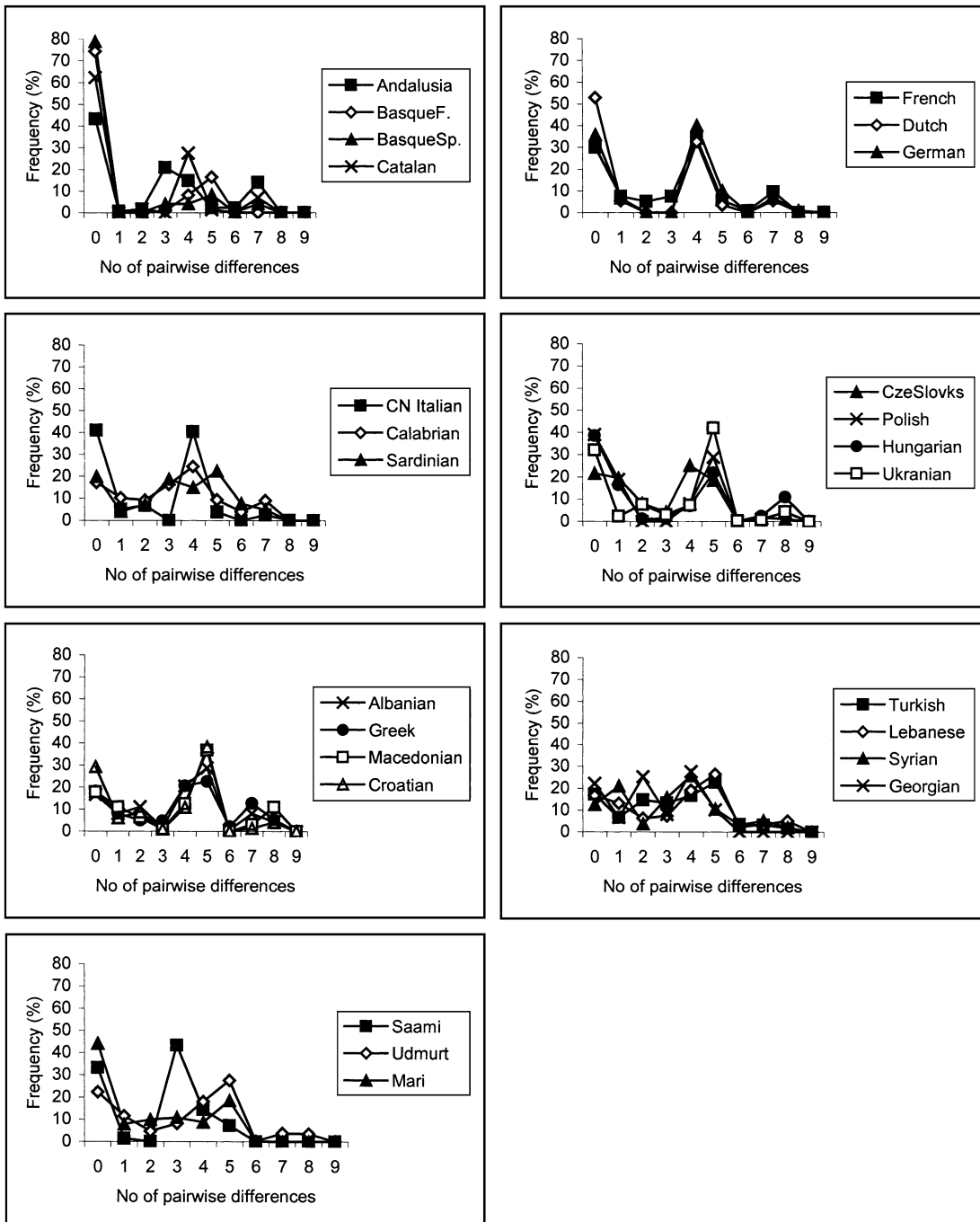


Fig. 1. Mismatch distributions in the EU data set (data from Semino et al. 2000).

(from 2–8 per population using 11 SNPs to 3–13 using 22 SNPs). However, gene diversities did not increase as much, in agreement with Semino et al.'s (2000) observation that more than 95% of the chromosomes they typed could be assigned to clades of haplotypes defined by just 10 key mutations. Tajima's and Fu's statistics were insignificant, except for a negative  $F_S$  for Turks, who also showed a rather smooth distribution. However, this result was no longer significant after Bonferroni's correction for multiple tests (Sokal and Rohlf 1995).

Even when 166 biallelic markers were studied (WO data set) most distributions were multimodal (Fig. 2). The sub-Saharan samples were the ones that displayed the most irregular distributions, with peaks at 16 (Sudan), 17 (Ethiopia), or 18 (Khoisan) differences, confirming extensive divergence of African Y chromosomes. In the European samples there were minor differences between the results of the analysis of 22 (EU data set) and 166 (WO data set) sites, in terms of both the shape of the mismatch distributions and the related statistics (Table 2). Sardinia shows a

**Table 1.** Measures of genetic diversity estimated from the EU data set (Semino et al. 2000)<sup>a</sup>

Population	<i>n</i>	Hp	Average mismatch	<i>P</i> (exp)	<i>H</i>	<i>D</i>	<i>P</i> ( <i>D</i> )	<i>F</i> <sub>S</sub>	<i>P</i> ( <i>F</i> <sub>S</sub> )	$\tau$
Andalusia	29	7	2.49 ± 1.38	0.080	0.567	-0.363	0.420	0.182	0.581	— <sup>b</sup>
Basque-Spanish	45	5	1.00 ± 0.69	0.090	0.211	-1.471	0.055	-0.221	0.450	—
Basque-French	22	3	1.18 ± 0.79	<b>0.040</b>	0.255	-0.882	0.218	1.730	0.821	—
Catalan	24	5	1.69 ± 1.03	0.070	0.377	-1.228	0.107	0.502	0.657	—
French	23	6	2.75 ± 1.51	<b>0.030</b>	0.700	-0.336	0.681	0.923	0.682	—
Dutch	27	4	1.93 ± 1.13	0.050	0.470	-0.558	0.324	2.147	0.855	—
German	16	4	2.70 ± 1.51	0.060	0.642	-0.017	0.529	2.309	0.887	—
Czech + Slovakian	45	9	2.63 ± 1.43	0.120	0.784	-0.550	0.336	-0.330	0.464	—
Centr. North Italian	50	6	2.19 ± 1.23	<b>0.030</b>	0.589	-0.313	0.442	1.476	0.789	—
Calabrian	37	9	3.11 ± 1.65	0.053	0.829	-0.191	0.844	-0.104	0.517	—
Sardinian	77	11	3.27 ± 1.70	<b>0.040</b>	0.801	-0.404	0.658	0.224	0.618	—
Croatian	58	7	3.05 ± 1.61	<b>0.020</b>	0.707	0.804	0.817	2.224	0.837	—
Albanian	51	8	3.57 ± 1.85	<b>0.010</b>	0.833	1.339	0.915	1.844	0.826	—
Greek	76	11	3.75 ± 1.91	<b>0.020</b>	0.818	0.893	0.839	0.807	0.691	—
Macedonian	20	6	3.72 ± 1.96	<b>0.010</b>	0.821	1.120	0.880	1.597	0.810	—
Polish	55	4	2.37 ± 1.31	0.080	0.609	0.558	0.746	4.253	0.945	—
Hungarian	45	7	2.67 ± 1.45	0.180	0.614	-0.085	0.539	1.196	0.742	—
Ukrainian	50	10	3.08 ± 1.63	<b>0.000</b>	0.678	-0.243	0.447	-0.211	0.521	—
Georgian	63	9	2.43 ± 1.34	0.050	0.778	0.732	0.803	-0.058	0.554	—
Turkish	30	13	3.15 ± 1.68	0.310	0.828	-0.567	0.310	-3.969	<b>0.042</b>	4.519
Lebanese	31	9	3.36 ± 1.77	0.111	0.832	0.106	0.588	-0.211	0.503	—
Syrian	20	8	3.04 ± 1.65	0.190	0.874	-0.067	0.517	-0.808	0.348	—
Saami	24	4	2.26 ± 1.29	<b>0.000</b>	0.667	1.231	0.884	2.504	0.889	—
Udmurt	43	8	3.12 ± 1.65	0.080	0.776	0.681	0.786	0.933	0.703	—
Mari	46	6	1.88 ± 1.09	0.350	0.558	-0.313	0.722	0.807	0.729	—
EU data set	1007	19	3.29 ± 1.70	0.050	0.847	0.544	0.748	0.304	0.644	—

<sup>a</sup> *n*, sample size; Hp, number of different haplotypes observed; *P*(exp), probability (estimated by bootstrap) of the observed mismatch distribution under the hypothesis of population expansion for a time  $\tau$  estimated from the data; *H*, Gene diversity; *D*, Tajima's *D*; *F*<sub>S</sub>, Fu's *F*<sub>S</sub>;  $\tau$ , time since expansion in units of mutation rate. Initial and final  $\theta$  in the Turkish population, the only one showing evidence of expansion: 0.003–8.115.

<sup>b</sup> Not calculated due to absence of signal of expansion.

more irregular shape in the WO data set, but that might reflect the small sample size, 22, in the study by Underhill et al. (2000), presumably a subset of the 77 individuals described by Semino et al. (2000).

Unimodal mismatch distributions were observed in three Central and Eastern Asian populations. Fu's *F*<sub>S</sub> was negative and significant (as is the case for mitochondrial data) in these samples, and Tajima's *D* in one of them, but these significances did not withstand Bonferroni's correction. A unimodal distribution was also observed in the American sample, but the peak is at zero differences (which is what we defined as the Type 0 distribution), reflecting the fact that 78% of the Y chromosomes belong to haplotype 115, a haplotype not found in other continents (Underhill et al. 2000). Tajima's *D* and Fu's *F*<sub>S</sub> are negative but, once again, both insignificant after Bonferroni's correction. We do not know how well the sample considered represents the whole continent. However, based on the evidence available, it seems that Y-chromosome diversity in this American sample reflects a severe bottleneck (Bonatto and Salzano 1997), with most Y-chromosome variation presumably restricted to the more rapidly evolving STR sites (Ruiz-Linares et al. 1999).

To understand the effects of aggregation of individuals from distant populations, we ran two global analyses of the EU and WO data sets (Fig. 3). For the populations of the EU data set, we observed a trimodal distribution, and insignificant, positive Tajima's *D* and Fu's *F*<sub>S</sub> (Table 1). For the more heterogeneous set of populations in the WO data set, the mismatch distribution was still bimodal, but Tajima's *D* and Fu's *F*<sub>S</sub> statistics were negative, and the former remained significant ( $p = 0.048$ ) even after Bonferroni's correction (Table 2). The simplest explanation of this apparently puzzling result is that, if one picks up chromosomes from different populations, many substitutions become relatively rare. As a consequence, a greater fraction of mutations is likely to appear almost lineage-specific. In turn, that may lead to a mismatch distribution that looks similar to those resulting from expansions, and to values of *D* and *F*<sub>S</sub> compatible with an expansion, even when there is no evidence of expansion in any single population. In agreement with Tajima (1989a), who specified that his test is valid only if applied to a set of chromosomes that evolved together, we interpret as a statistical artifact the significant *D* value observed in the global analyses.

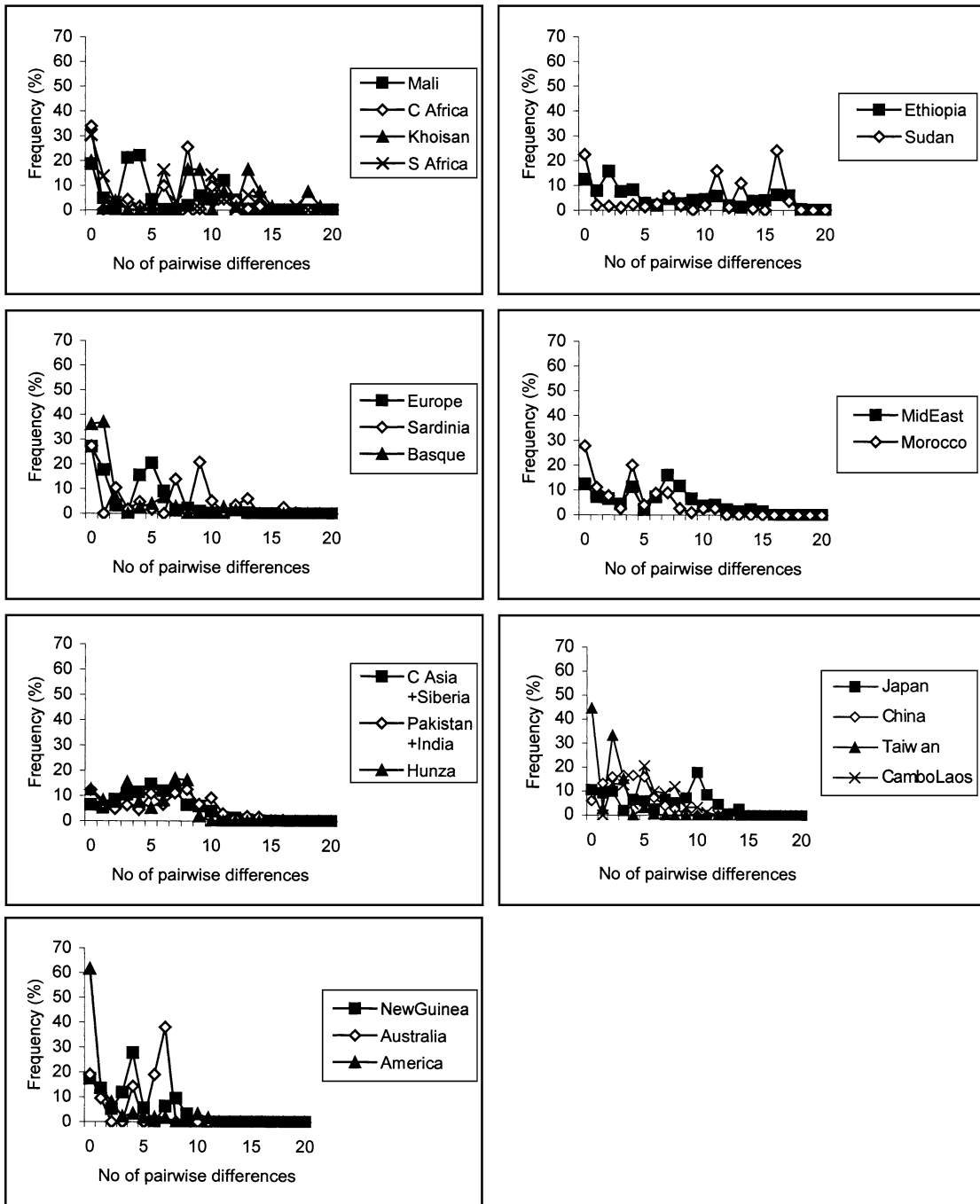


Fig. 2. Mismatch distributions in the WO data set (data from Underhill et al. 2000).

### Simulations: Effects of the Mutation Rates

In simulated stationary populations, when considering the whole set of sites, the average mismatch observed is close to the expected value, i.e., the parameter  $\theta$  used to generate the simulated samples, and thus increased with the mutation rate in the different simulations (Table 3). As expected, in expanding populations both the average mismatch and its standard deviation are reduced.

For both stationary and expanding populations, multimodality is more frequent as the mutation rate increases, which does not support the view that the low Y-chromosome mutation rate increases the probability of observing multimodal mismatch distributions. Under stationarity, for example, for  $\mu = 5 \times 10^{-9}$  per site and per year, about 60% of the mismatch distributions have a single peak (Types 0 and 1) and 40% of them show a maximum at zero difference. But when  $\mu$  is  $10^{-7}$  all but one simulations

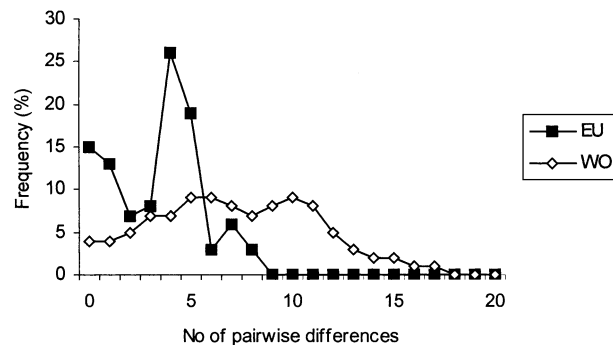
**Table 2.** Measures of genetic diversity estimated from the WO data set (Underhill et al. 2000)<sup>a</sup>

Population	<i>n</i>	Hp	Average mismatch	<i>P</i> (exp)	<i>H</i>	<i>D</i>	<i>P</i> ( <i>D</i> )	<i>F</i> <sub>S</sub>	<i>P</i> ( <i>F</i> <sub>S</sub> )	$\tau$
Sudan	40	9	9.05 ± 4.25	<b>0.000</b>	0.776	1.283	0.912	5.673	0.952	— <sup>b</sup>
Ethiopia	88	15	6.62 ± 3.16	0.930	0.876	-0.053	0.551	1.791	0.775	—
Mali	44	11	4.76 ± 2.37	<b>0.030</b>	0.814	-0.880	0.184	0.605	0.634	—
Morocco	28	8	3.36 ± 1.78	0.470	0.722	-0.799	0.235	0.313	0.608	—
C. Africa	36	9	5.21 ± 2.58	<b>0.020</b>	0.660	-0.343	0.417	1.869	0.771	—
Khoisan	39	6	8.86 ± 4.17	<b>0.000</b>	0.802	1.692	0.961	9.876	0.993	—
S. Africa	54	9	5.22 ± 2.56	0.210	0.697	-0.276	0.472	3.115	0.870	—
Europe	60	12	3.05 ± 1.61	0.300	0.731	-1.201	0.102	-1.174	0.342	—
Sardinia	22	7	5.75 ± 2.86	<b>0.030</b>	0.727	-0.959	0.164	2.513	0.877	—
Basque	45	8	1.89 ± 1.10	<b>0.010</b>	0.638	-1.523	<b>0.041</b>	-0.779	0.358	0.882
MidEast	24	11	5.73 ± 2.84	0.680	0.877	-0.886	0.200	-0.505	0.422	—
C. Asia + Siberia	185	33	5.11 ± 2.49	0.940	0.936	-1.220	0.070	-8.009	<b>0.047</b>	5.536
Pakistan + India	88	24	5.82 ± 2.81	0.600	0.877	-1.203	0.088	-3.811	0.157	—
Hunza	38	13	4.42 ± 2.23	0.280	0.875	-0.529	0.351	-1.397	0.319	—
Japan	23	12	6.28 ± 3.09	0.530	0.893	0.202	0.635	-1.090	0.303	—
China	20	12	3.41 ± 1.82	0.850	0.937	-1.233	0.105	-4.507	<b>0.010</b>	3.749
Taiwan	74	5	1.37 ± 0.86	0.090	0.554	-0.110	0.483	1.213	0.765	—
Cambo + Laos	18	12	4.80 ± 2.46	0.280	0.895	-1.728	0.027	-3.496	<b>0.043</b>	6.101
NewGuinea	23	6	3.46 ± 1.83	0.240	0.826	0.937	0.857	1.689	0.809	—
Australia	7	4	4.48 ± 2.51	<b>0.030</b>	0.810	1.161	0.883	1.486	0.760	—
America	106	12	1.38 ± 0.86	0.360	0.384	-1.766	<b>0.015</b>	-4.082	<b>0.049</b>	3.027
WO data set	1062	115	7.33 ± 3.43	0.730	0.964	-1.895	<b>0.001</b>	-23.937	<b>0.007</b>	6.230

<sup>a</sup> *n*, sample size; Hp, number of different haplotypes observed; *P*(exp), probability (estimated by bootstrap) of the observed mismatch distribution under the hypothesis of population expansion for a time  $\tau$  estimated from the data; *H*, gene diversity; *D*, Tajima's *D*; *F*<sub>S</sub>, Fu's *F*<sub>S</sub>;  $\tau$ , time since expansion in units of mutation rate. Initial and final  $\theta$  in the populations showing evidence of expansion

are as follows: C. Asia + Siberia, 0.567–16.595; China, 0.170–16.432; and Cambodia + Laos, 0.034–13.963. For the Americas, the peak is at zero differences, and the estimated  $\theta$  is the same, 0.560, both before and after the expansion.

<sup>b</sup> Not calculated, owing to absence of signals of expansion.



**Fig. 3.** Mismatch distributions for the pooling of all populations in the EU and WO data sets.

yield distributions with multiple peaks (Type 2). In expanding populations, multimodality is rare, especially at low mutation rates. When  $\mu < 10^{-8}$ , more than 90% of the mismatch distributions are either Type 0 or Type 1. Therefore, the low Y-chromosome mutation rate does not seem to affect the shape of the mismatch distribution much; our results show that, if anything, it may enhance an existing signal of expansion, reducing the frequency of Type 2 distributions.

In the simulated stationary populations, both *D* and *F*<sub>S</sub> show wide distributions centered on zero, regardless of mutation rates. The fraction of signifi-

cant *D* values is very close to the nominal level  $\alpha = 0.05$  (in fact, lower than that). On the contrary, the *F* test seems permissive, at least for low mutation rates. After expansions, *D* and *F*<sub>S</sub> are always negative and, in more than 78% of the cases, significant. As expected, the power of these statistics increases as the mutation rate, and therefore the number of polymorphic sites, increases.

#### Simulations: Effects of the Ascertainment Bias

We report in Fig. 4 the Tajima *D* values and the number of times they achieved significance over 1000 simulations of expanding populations. A strong effect of the ascertainment bias on the values (Fig. 4a) and the significance (Fig. 4b) of Tajima's *D* is evident, especially when the expansion event is recent. In the populations that underwent an ancient expansion, when passing from the total number of polymorphic sites to the analysis of mutations that are shared by at least two chromosomes, the number of significant Tajima *D* values drops from 1000 to 511, while in the populations that expanded recently it drops drastically, from 940 to 61.

When only mutations that are shared by at least three chromosomes are analyzed, Tajima's *D* fails to show any evidence of population growth. For old

**Table 3.** Simulation results

Simulated population	$\mu$	Average mismatch <sup>a</sup>	Type 0 <sup>b</sup>	Type 1 <sup>b</sup>	Type 2 <sup>b</sup>	SSD signif. <sup>c</sup>	$D^d$	Signif. <sup>e</sup>	$F_S^f$	Signif. <sup>e</sup>
Stationary	$5 \times 10^{-9}$	1.005 (0.778)	0.407	0.183	0.410	0.356	-0.027 (0.949)	0.065	-0.042 (2.203)	0.120
	$1 \times 10^{-8}$	2.064 (1.243)	0.132	0.211	0.657	0.311	[-1.939, 3.282]	0.044	[-6.871, 12.659]	0.081
	$5 \times 10^{-8}$	9.715 (4.904)	0.000	0.027	0.973	0.269	[-2.016, 3.277]	0.046	[-9.404, 14.086]	0.088
	$1 \times 10^{-7}$	19.280 (9.536)	0.000	0.001	0.999	0.243	[-0.073 (0.892)]	0.042	[-0.526 (4.103)]	0.073
Expanding	$5 \times 10^{-9}$	0.522 (0.312)	0.790	0.172	0.038	0.197	[-2.175, 3.010]	0.788	[-16.896, 20.196]	0.926
	$1 \times 10^{-8}$	1.045 (0.442)	0.436	0.499	0.065	0.164	[-0.043 (0.931)]	0.939	[-0.843 (4.747)]	0.994
	$5 \times 10^{-8}$	5.187 (1.191)	0.000	0.825	0.175	0.049	[-12.248, 3.248]	0.999	[-23.947, 23.828]	1.000
	$1 \times 10^{-7}$	10.402 (1.936)	0.000	0.634	0.366	0.028	[-1.740 (0.385)]	1.000	[-25.209, -24.059]	1.000

<sup>a</sup> Mean of the mismatch distribution (and standard deviation) across 1000 simulations.

<sup>b</sup> Fraction of the 1000 simulations showing the shapes of the mismatch distribution: Type 0 (unimodal with a maximum at 0), Type 1 (unimodal with a maximum  $> 0$ ), and Type 2 (bimodal/multimodal).

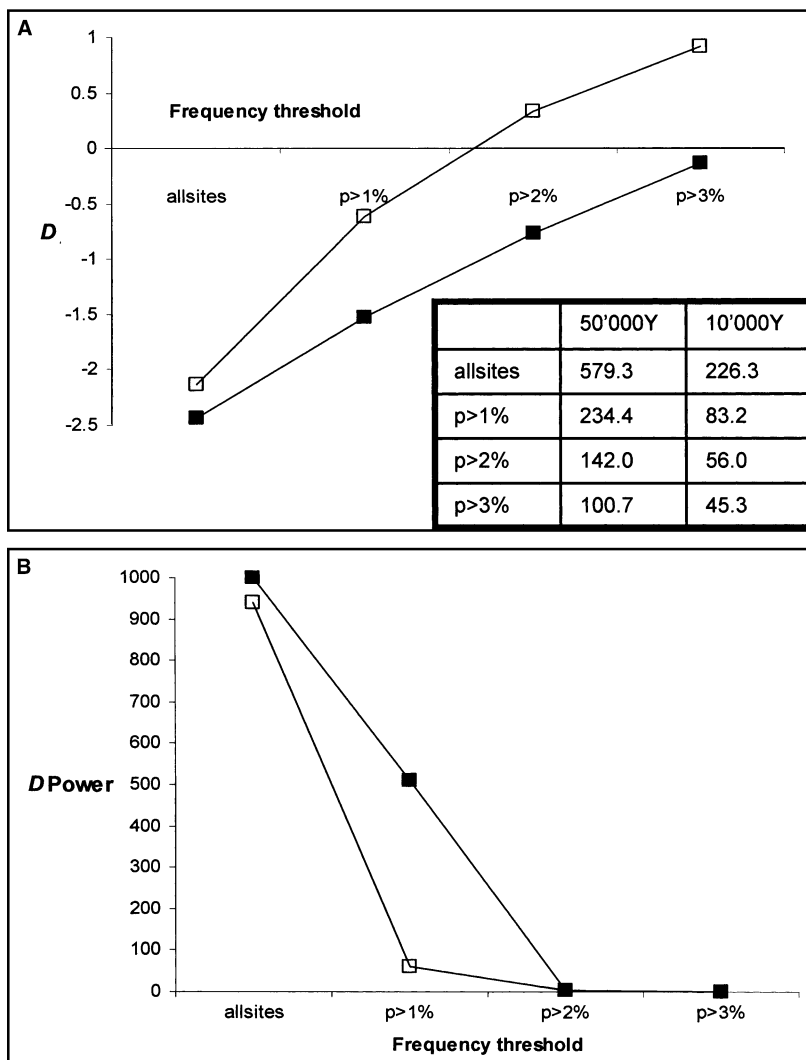
<sup>c</sup> Fraction of the cases where the  $p$  value of the SSD statistic is lower than 5%.

<sup>d</sup> Average  $D$  (Tajima 1989a). Standard deviation is in parentheses and range of observed values is in brackets.

<sup>e</sup> Fraction of significant cases at the 5% level, of 1000 simulations.

<sup>f</sup> Average  $F_S$  (Fu 1997) over 1000 simulations. Standard deviation is shown in parentheses and range of observed values is in brackets.





**Fig. 4.** Tajima's  $D$  values (**A**) and number of significant cases (**B**) for 1000 simulations of expanding populations. Open and filled symbols correspond, respectively, to recent (10,000 years) and ancient (50,000 years) expansion events. Tajima's  $D$  statistics were estimated from different subsets of SNP sites, selected on the basis of their polymorphism: all sites,  $p > 1\%$  (mutations shared by at least two chromosomes),  $p > 2\%$  (mutations shared by at least three chromosomes), and  $p > 3\%$  (mutations shared by at least four chromosomes). The average number of sites analyzed within each frequency class is indicated in the inset.

expansions, the  $D$  values are insignificant but mostly negative on average, and for recent expansions, Tajima's  $D$  often takes positive values.

Therefore, recent expansions are more likely to go undetected than ancient expansions, if there is an ascertainment bias. How large is the ascertainment bias in the SNPs of this study? The biallelic polymorphisms considered here were discovered by comparing the Y chromosomes of different individuals by DHPLC and looking for heteroduplexes (Underhill et al. 1997). If  $p$  is the frequency of the most common allele at a site, the probability that all  $n$  screened chromosomes share that allele is  $p^n$ , and so the possibility to detect the polymorphism is  $1 - p^n$ . The individuals screened varied from 53 (Underhill et al. 1997, 2000) to 72 (Shen et al. 2000). If we use these values to define a range of  $n$ , the fraction of sites failing to be ascertained is between 2.5 and 6.6% for  $p = 0.95$  and between 48.0 and 58.7% for  $p = 0.99$ . In other words, most moderately polymorphic sites have probably been identified as such, whereas one-

half or more of the sites whose rarer allele has a frequency  $< 0.01$  are likely to appear monomorphic.

## Discussion

Analyses of Y-chromosome SNPs (Pereira et al. 2001; this paper) do not suggest a rapid growth of the males' effective population. Despite the fact that we considered many polymorphic sites, nearly all mismatch distributions are still multimodal, and there is no statistical support for departures from mutation-drift equilibrium. Y-chromosome SNP diversity provides no evidence for male population expansion in populations around the world, particularly in Africa. Thus, as for the questions listed in the Introduction, it is clear that (1) a greater number of SNPs has not changed the shape of the mismatch distributions, and (2) the apparently nonexpansion pattern is not a European, but a worldwide, feature.

Although there is only one human Y-chromosome tree, and so, in principle, all Y-chromosome markers

should lead to the same demographic inferences, two previous studies reached different conclusions. Shen et al. (2000) inferred a rapid growth of the male population from a negative Tajima's test and from the distribution of mutants at independent sites. For these calculations, however, 72 individuals from 46 populations were considered, and that violates the assumptions of Tajima's (1989a, p. 593) test. As confirmed by our mismatch distributions and related statistics, joint analysis of individuals of different origins tends to render  $D$  and  $F_S$  compatible with an expansion (Table 2), even though no population shows evidence for growth when separately analyzed.

A better fit of a model of rapid growth than of constant  $N_m$  was also found by Pritchard et al. (1999), who studied eight microsatellite loci in 445 individuals. The excess of rare haplotypes they observed suggests an expansion 18,000 years ago for the whole human population (95% confidence interval between 7000 and 41,000; see also Table 4).

There are doubtless more humans now than in the Pleistocene (Biraben 1979; Weiss 1984; see also Zietkiewicz et al. 1998; Harpending et al. 1998), and so the idea that  $N_m$  stayed constant is counterintuitive. However, expansions may be difficult to recognize if poorly polymorphic sites are not efficiently ascertained (Nielsen 2000; Wakeley et al. 2001), so that most mutations considered will be ancient and shared by several chromosomes. Our simulations show that an ascertainment bias, in our case leading to lumping rare ( $p < 0.01$ ) haplotypes with their nearest evolutionary neighbors, may actually conceal the effects of an expansion (question 3 in the Introduction). However, that effect (Fig. 4) was strong for simulations of recent expansions (10,000 years ago), and much less so for older phenomena (50,000 years ago). Therefore, if Pritchard et al.'s (1999) time estimates are approximately correct, the male effective population size might have increased too recently for the expansion to be detected at the SNP level. Because the likely dates of female and male population growth estimated in previous studies do not overlap (see Table 4) (Excoffier and Schneider 1999; Pritchard et al. 1999), the obvious implication is that the two genders had different demographic histories.

In principle, apparent differences between the demographic history of males and that of females may reflect different selective regimes. A mitochondrial selective sweep (Excoffier 1990; Harris and Hey 1999b; Wall and Przeworski 2000) may have led investigators to reject constant  $N_f$  erroneously. Alternatively, stabilizing selection upon the Y chromosome (Jobling and Tyler-Smith 2000) may have determined patterns compatible with constant  $N_m$ , when populations, in fact, expanded (Tajima 1989a; Wall and Przeworski 2000; see also Pereira et al. 2001). The effects of selective pressures and

demographic changes cannot be discriminated *a posteriori* from population data (Takahata 1996), and hence the present study does not provide evidence relevant to this question. However, unless selection has really been strong (which would force us to reconsider crucial aspects of human evolution inferred from DNA evidence under the assumption of quasineutrality, such as the age of the most recent human ancestors), the available data suggest at least that the human demographic past cannot be envisaged as a process of linear growth, to which females and males contributed equally and in parallel. Variation in other genome regions does not help clarify the picture. Some nuclear loci seem to support rapid population growth (Reich and Goldstein 1998; Zhao et al. 2000; Alonso and Armour 2001), but others (Takahata et al. 1995; Harding et al. 1997; Harris and Hey 1999a; Beaumont, 1999) do not.

#### *A Recent Shift from Polygyny to Monogamy?*

In this section we explore the consequences of a model in which the effective populations of females expanded earlier than those of males. The main genetic consequence would be that the terminal branches in the Y-chromosome tree would be shorter than those in the mitochondrial tree, because of the shorter time elapsed from the expansion. In addition, as we showed, an ascertainment bias causes rare ( $p < 0.01$ ) mutations in the terminal branches of the tree to be largely missed. In this way, very few haplotypes will differ for mutations in the terminal branches, both because there are not many such mutations and because a fraction of them would not be discovered. Ultimately, that would result in ragged and multimodal Y-chromosome mismatch distributions and in insignificant values of the  $D$  and  $F_S$  statistics (Rogers and Jorde 1995). That is not the case for the STR sites of Pritchard et al. (1999), both because they mutate more rapidly and because all their alleles are identified without any bias.

As a consequence, we propose the following.

- (1) Female effective population sizes, in agreement with mitochondrial studies, increased relatively early, at the moments at which archaeological evidence places the human expansions from Africa into the various continents (see Table 4) (Excoffier and Schneider 1999).
- (2) Male effective population sizes increased later (see Table 4), and therefore, over much of human prehistory, polygyny was the rule rather than the exception; a high variance in the males' offspring numbers is a necessary consequence of that mating system.

**Table 4.** Estimated times of expansion for mitochondrial and Y-chromosome data (Excoffier and Schneider 1999; Pritchard et al. 1999)

Region	$t$ (KY)	$t$ (range) <sup>b</sup>
mtDNA estimate <sup>a</sup>		
Africa	69.750 (18.607)	48–110
East Africa	86.667 (20.306)	73–110
West Africa	59.600 (7.503)	48–68
Europe	37.000 (13.468)	14–63
West Asia	52.667 (6.658)	47–60
East Asia	62.000 (16.793)	40–77
Oceania	78.800 (18.539)	48–97
America	57.000 (2.828)	55–59
Y-chromosome estimate <sup>c</sup>		
World	18	7–41
Africa	15	5–37
South Africa	13	3–32
East Africa	15	5–38
West Africa	12	0.2–49
Europe	13	4–32
West Asia	17	5–43
East Asia	15	4–40
Oceania	16	1–45
America	14	3–37

<sup>a</sup> Estimates obtained by Excoffier and Schneider (1999): The mean and standard deviation (in parentheses) are computed for every region as the mean and standard deviation of the estimates obtained for each sample.

<sup>b</sup> Range defined as minimum and maximum expansion times estimated for the samples within each region.

<sup>c</sup> Estimates obtained by Pritchard et al. (1999).

(3) Such recent expansions of male population sizes can be identified at the STR (Pritchard et al. 1999) but not at the SNP (Pereira et al. 2001; this study) level, because the sample sizes used to discover SNP polymorphisms likely prevent the identification of variable sites whose rarer allele has a frequency  $p \leq 0.01$ , i.e., the ones where the expansion has had the greatest chance of leaving a mark.

Under the hypothesis of a delayed increasing in  $N_m$ , most apparent inconsistencies among the mtDNA and Y-chromosome evidence would disappear. It would no longer be necessary to imagine different selection regimes on female- and male-transmitted traits (although these regimes may well have differed) or a different tendency to migrate (although male and female migration rates may well have differed). The greater genetic variances observed for the Y chromosome (Seielstad et al. 1998) could be explained by the greater impact of genetic drift over the smaller populations of males.

In turn, the resulting increased differentiation of the male populations prior to the shift from polygyny to monogamy is also expected to result in a stronger impact of admixture on the males' genetic diversity, which would also contribute to determining multimodal mismatch distributions. Indeed, hybrid

populations may be more or less internally heterogeneous, depending on the degree of differentiation between the parental groups from which they derive. If the parental groups differed substantially and the contact is recent, hybrid populations fail to show signatures of past expansions (Marjoram and Donnelly 1994). There is empirical evidence of this phenomenon in some bimodal mitochondrial mismatch distributions observed in Africa, probably reflecting the presence of genes from two different gene pools (Bandelt and Forster 1997; Brakez et al. 2001). If  $N_m$  increased later than  $N_f$ , admixture has had a greater chance to induce multimodality in the Y chromosome than in the mitochondrial mismatch distributions, making male population growth difficult to identify.

Is there any evidence of a relatively recent shift from polygyny to monogamy in humans? In the ethnological literature, there is ample consensus that humans evolved in multimale polygynous bands (Lee and DeVore 1968; Flannery 1972b; Keen 1982; Badcock 1991), like gorillas (Pussey 2001). Under those conditions, although the sex ratio at birth is close to one in humans (James 1987),  $N_m$  was smaller than  $N_f$ , because some men made a large contribution to the next generation's gene pool, and some did not contribute at all. Note that if the number of offspring per male is highly variable, the  $N_m$  estimated from genetic data is even lower than the actual number of reproducing males (Crow 1958; Donnelly et al. 1996).

A likely moment for the shift to monogamy is difficult to define exactly, and there is no reason to imagine that it happened at the same moment everywhere. The dates Pritchard et al. (1999) estimated from their different continental and intercontinental groups of samples (Table 4) are clearly averages of processes that probably occurred at different times but affected most human populations. Slightly later than these average dates, but within their confidence interval, namely, between 10,000 and 5000 years ago in Europe and Asia and more recently in Africa and in the Americas, a major change is documented in the archeological record, i.e., the development of technologies for farming and animal breeding, or Neolithic transition (Cavalli-Sforza et al. 1994; Bellwood 2001). Although the shift to food production may have caused an initial decline in nutrition and health (Cohen 1989), there is no doubt that its long-term effect was an increase in population density and growth rates (Weiss 1984; Cavalli-Sforza et al. 1994). Sociobiological studies do suggest that the development of extensive farming resulted in a decrease in the levels of polygyny, although those levels did not appear to be very high among hunters and gatherers either (van den Berghe 1979).

Especially with the shift to farming in the Neolithic period, although not necessarily then, sedentary, more structured communities developed.

Nuclear families replaced the polygamous, extended-family compounds typical of hunting–gathering populations, and the household, rather than the band, became the main socioeconomic unit (Flannery 1972a). If so, monogamy may have become widespread, and  $N_m$  may have started increasing. Nowadays there are examples of polygyny both in farming and in hunting–gathering societies. For the above model to be generally correct, however, it seems necessary only that polygyny (and the related small  $N_m$ ) became increasingly uncommon as time passed.

Then, unless so far unspecified selective pressures have not been the main factor shaping human gene genealogies, Pritchard and co-workers' (1999) time estimates reflect a change from a typically polygynic social structure to one in which more males had access to reproduction. Our empirical results and our simulations suggest that that change happened too recently to leave a trace in the variation observed at most Y-chromosome SNPs. Starting at approximately the same time, populations that did not turn to food production began to suffer from competition with farming communities; the resulting bottlenecks are reflected in their current mitochondrial diversity (Excoffier and Schneider 1999). This model seems able to reconcile the results of several, apparently contradictory, analyses of genes inherited through the female and male lines. One could test the model using comparisons of Y-chromosome diversity in contemporary polygynous and monogamous populations. This model may provide a framework for addressing specific questions concerning human demography of the past.

*Acknowledgments.* This paper was supported by grants from the European Science Foundation (Eurocores Programme: The Origin of Man, Language, and Languages) through the Italian CNR, the University of Ferrara, and Programa Operacional Ciência, Tecnologia e Inovação (POCTI), Quadro Comunitário de Apoio III. I.D. was supported by a grant from the Swiss National Science Foundation (FNRS) for advanced researchers; L.P., by a research grant (PRAXIS XXI BD/13632/97) from Fundação para a Ciência e a Tecnologia. We thank Laurent Excoffier for critical discussion and for providing us with a copy of the SIMCOAL program to perform the coalescent simulations and Steven LeBlanc for many suggestions and for his guidance through the maze of the ethnographic literature.

## References

- Alonso S, Armour JA (2001) A highly variable segment of human subterminal 16p reveals a history of population growth for modern humans outside Africa. *Proc Natl Acad Sci USA* 98:864–869
- Badcock C (1991) *Evolution and individual behavior: An introduction to human sociobiology*. Blackwell, Oxford
- Bandelt HJ, Forster P (1997) The myth of bumpy hunter-gatherer mismatch distributions. *Am J Hum Genet* 61:980–983
- Beaumont MA (1999) Detecting population expansion and decline using microsatellites. *Genetics* 153:2013–2019
- Bellwood P (2001) Early agricultural diasporas? *Annu Rev Anthropol* 30:181–207
- Bertorelle G, Slatkin M (1995) The number of segregating sites in expanding human populations, with implications for estimates of demographic parameters. *Mol Biol Evol* 12:887–892
- Bertranpetit J (2000) Genome, diversity, and origins: The Y chromosome as a storyteller. *Proc Natl Acad Sci USA* 97:6927–6929
- Biraben JN (1979) Essai sur l'évolution du nombre des hommes. *Population* 34:13–25
- Bonato SL, Salzano FM (1997) A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. *Proc Natl Acad Sci USA* 94:1866–1871
- Brakez Z, Bosch E, Izaabel H, Akhayat O, Comas D, Bertranpetit J, Calafell F (2001) Human mitochondrial DNA sequence variation in the Moroccan population of the Souss area. *Ann Hum Biol* 28:295–307
- Cavalli-Sforza LL (1998) The DNA revolution in population genetics. *Trends Genet* 14:60–65
- Cavalli-Sforza LL, Menozzi P, Piazza A (1994) *The history and geography of human genes*. Princeton University Press, Princeton, NJ
- Cohen MN (1989) *Health and the rise of civilization*. Yale University Press, New Haven, CT
- Crow JF (1958) Some possibilities for measuring selection intensities in man. *Hum Biol* 30:1–13
- De Knijff P (2000) Messages through bottlenecks: On the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. *Am J Hum Genet* 67:1055–1061
- Donnelly P (1996) Interpreting genetic variability: The effects of shared evolutionary history. In: Weiss K (ed) *Variation in the human genome*. Wiley, Chichester, UK, pp. 25–50
- Donnelly P, Tavaré S, Balding DJ, Griffiths RC (1996) Estimating the age of the common ancestor of men from the ZFY intron. *Science* 272:1357–1359
- Excoffier L (1990) Evolution of human mitochondrial DNA: Evidence for departure from a pure neutral model of populations at equilibrium. *J Mol Evol* 30:125–139
- Excoffier L, Schneider S (1999) Why hunter-gatherer populations do not show signs of Pleistocene demographic expansions. *Proc Natl Acad Sci USA* 96:10597–10602
- Excoffier L, Novembre J, Schneider S (2000) SIMCOAL: A general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. *J Hered* 91:506–509
- Flannery K (1972a) The origins of the village as a settlement type in Mesoamerica and the Near East. In: Uckho P, Tringham R, Dimpleby GW (eds) *Man, settlement, and urbanism*. Schenckman, Cambridge, MA, pp. 23–53
- Flannery K (1972b) The cultural evolution of civilizations. *Annu Rev Ecol Syst* 3:399–426
- Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147:915–925
- Hammer MF (1995) A recent common ancestry for human Y chromosomes. *Nature* 378:376–378
- Harding RM, Fullerton SM, Griffiths RC, Bond J, Cox MJ, Schneider JA, Moulin DS, Clegg JB (1997) Archaic African and Asian lineages in the genetic ancestry of modern humans. *Am J Hum Genet* 60:772–789
- Harpending HC (1994) Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum Biol* 66:591–600
- Harpending HC, Batzer MA, Gurven M, Jorde JB, Rogers AR, Sherry ST (1998) Genetic traces of ancient demography. *Proc Natl Acad Sci USA* 95:1961–1967
- Harris EE, Hey J (1999a) X chromosome evidence for ancient human histories. *Proc Natl Acad Sci USA* 96:3320–3324

- Harris EE, Hey J (1999b) Human demography in the Pleistocene: Do mitochondrial and nuclear genes tell the same story? *Evol Anthropol* 8:81–86
- Hudson RR (1990) Gene genealogies and the coalescent process. In: Futuyma D, Antonovics J (eds) *Oxford surveys in evolutionary biology*. Oxford University Press, Oxford, pp. 1–44
- James WH (1987) The human sex ratio, part I: A review of the literature. *Hum Biol* 59:721–752
- Jazin E, Soodyall H, Jalonen P, Lindholm E, Stoneking M, Gyllenstein U (1998) Mitochondrial mutation rate revisited: Hot spots and polymorphism. *Nat Genet* 18:109–110
- Jobling MA, Tyler-Smith C (2000) New uses for new haplotypes. The human Y chromosome, disease and selection. *Trends Genet* 16:356–362
- Jobling MA, Pandya A, Tyler-Smith C (1997) The Y chromosome in forensic analysis and paternity testing. *Int J Legal Med* 110:118–124
- Keen I (1982) How some Murngin men marry ten wives. *Man* 17:620–642
- Lee RB, DeVore I (1968) *Man the hunter*. Aldine, Chicago
- Marjoram P, Donnelly P (1994) Pairwise comparisons of mitochondrial DNA sequences in subdivided populations and implications for early human evolution. *Genetics* 136:673–683
- Nei M (1987) *Molecular evolutionary genetics*. Columbia University Press, New York
- Nielsen R (2000) Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics* 154:931–942
- Pereira L, Dupanloup I, Rosser Z, Jobling MA, Barbujani G (2001) Y-chromosome mismatch distributions in Europe. *Mol Biol Evol* 18:1259–1271
- Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW (1999) Population growth of human Y chromosomes: A study of Y chromosome microsatellites. *Mol Biol Evol* 16:1791–1798
- Pussey AE (2001) Of genes and apes: Chimpanzee social organization and reproduction. In: de Waal FBM (ed) *Tree of origin: What primate behavior can tell us about human social evolution*. Harvard University Press, Cambridge, MA, pp. 9–37
- Reich DE, Goldstein DB (1998) Genetic evidence for a Paleolithic human population expansion in Africa. *Proc Natl Acad Sci USA* 95:8119–8123
- Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol* 9:552–569
- Rogers AR, Jorde LB (1995) Genetic evidence on modern human origins. *Hum Biol* 67:1–36
- Rogers AR, Fraley AE, Bamshad MJ, Watkins WS, Jorde LB (1996) Mitochondrial mismatch analysis is insensitive to the mutational process. *Mol Biol Evol* 13:895–902
- Ruiz-Linares A, Ortiz-Barrientos D, Figueroa M, et al. (1999) Microsatellites provide evidence for Y-chromosome diversity among the founders of the New World. *Proc Natl Acad Sci USA* 96:6312–6317
- Schneider S, Excoffier L (1999) Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: Application to human mitochondrial DNA. *Genetics* 152:1079–1089
- Schneider S, Roessli D, Excoffier L (2000) *Arlequin ver. 2.000: A software for population genetics data analysis*. Genetics and Biometry Laboratory, University of Geneva, Geneva, Switzerland
- Seielstad M, Minch E, Cavalli-Sforza LL (1998) Genetic evidence for a higher female migration rate in humans. *Nat Genet* 20:278–280
- Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, et al. (2000) The genetic legacy of Paleolithic *Homo sapiens* in extant Europeans: A Y chromosome perspective. *Science* 290:1155–1159
- Shen P, Wang F, Underhill PA, Franco C, Yang WH, Roxas A, Sung R, et al. (2000) Population genetic implications from sequence variation in four Y chromosome genes. *Proc Natl Acad Sci USA* 97:7354–7359
- Sigurgardottir S, Helgason A, Gulcher JR, Stefansson K, Donnelly P (2000) The mutation rate in the human mtDNA control region. *Am J Hum Genet* 66:1599–1609
- Simonsen KL, Churchill GA, Aquadro CF (1995) Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* 141:413–429
- Sokal RR, Rohlf FJ (1995) *Biometry*. W.H. Freeman, New York
- Soodyall H, Jenkins T, Mukherjee A, du Toit E, Roberts DF, Stoneking M (1997) The founding mitochondrial DNA lineages of Tristan da Cunha Islanders. *Am J Phys Anthropol* 104:157–166
- Tajima F (1989a) Statistical method for testing the neutral mutation hypothesis by DNA polymorphisms. *Genetics* 123:585–595
- Tajima F (1989b) The effect of change in population size on DNA polymorphism. *Genetics* 123:597–601
- Takahata N (1996) Neutral theory of molecular evolution. *Curr Opin Genet Devel* 6:767–772
- Takahata N, Satta Y, Klein J (1995) Divergence time and population size in the lineage leading to modern humans. *Theor Pop Biol* 48:198–221
- Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW (2000) Recent common ancestry of human Y chromosomes: Evidence from DNA sequence data. *Proc Natl Acad Sci USA* 97:7360–7365
- Underhill PA, Jin L, Lin AA, et al. (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res* 7:996–1005
- Underhill PA, Shen P, Lin AA, et al. (2000) Y chromosome sequence variation and the history of human populations. *Nat Genet* 26:358–361
- van den Berghe P (1979) *Human family systems. An evolutionary view*. Waveland Press, Prospect Heights, IL
- von Haeseler A, Sajantila A, Pääbo S (1996) The genetical archaeology of the human genome. *Nat Genet* 14:135–140
- Wakeley J, Nielsen R, Liu-Cordero SN, Ardlie K (2001) The discovery of single-nucleotide-polymorphisms—and inferences about the human demographic history. *Am J Hum Genet* 69:1332–1347
- Wall JD, Przeworski M (2000) When did the human population start increasing? *Genetics* 155:1865–1874
- Weiss KM (1984) On the number of members of the genus *Homo* who have ever lived, and some evolutionary implications. *Hum Biol* 56:637–649
- Zhao Z, Jin L, Fu YX, et al. (2000) Worldwide DNA sequence variation in a 10-kilobase noncoding region of human chromosome 22. *Proc Natl Acad Sci USA* 97:11354–11358
- Zietkiewicz E, Yotova V, Jarnik M, et al. (1998) Genetic structure of the ancestral population of modern humans. *J Mol Evol* 47:146–155