Y-Chromosome Mismatch Distributions in Europe

Luísa Pereira,† Isabelle Dupanloup,† Zoë H. Rosser,‡ Mark A. Jobling,‡ and Guido Barbujani†*

*Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP) and Faculdade de Ciências da Universidade do Porto, Porto, Portugal; †Dipartimento di Biologia, Università di Ferrara, Ferrara, Italy; and ‡Department of Genetics, University of Leicester, Leicester, England

Ancient demographic events can be inferred from the distribution of pairwise sequence differences (or mismatches) among individuals. We analyzed a database of 3,677 Y chromosomes typed for 11 biallelic markers in 48 human populations from Europe and the Mediterranean area. Contrary to what is observed in the analysis of mitochondrial polymorphisms, Tajima's test was insignificant for most Y-chromosome samples, and in 47 populations the mismatch distributions had multiple peaks. Taken at face value, these results would suggest either (1) that the size of the male population stayed essentially constant over time, while the female population size increased, or (2) that different selective regimes have shaped mitochondrial and Y-chromosome diversity, leading to an excess of rare alleles only in the mitochondrial genome. An alternative explanation would be that the 11 variable sites of the Y chromosome do not provide sufficient statistical power, so a comparison with mitochondrial data (where more than 200 variable sites are studied in Europe) is impossible at present. To discriminate between these possibilities, we repeatedly analyzed a European mitochondrial database, each time considering only 11 variable sites, and we estimated mismatch distributions in stable and growing populations, generated by simulating coalescent processes. Along with theoretical considerations, these tests suggest that the difference between the mismatch distributions inferred from mitochondrial and Y-chromosome data are not a statistical artifact. Therefore, the observed mismatch distributions appear to reflect different underlying demographic histories and/or selective pressures for maternally and paternally transmitted loci.

Introduction

Analyses of mitochondrial and Y-chromosome polymorphisms in humans tend to suggest that the female- and the male-transmitted gene pools evolved under somewhat different conditions. Although some studies (notably, Poloni et al. 1997; see Bertranpetit 2000) found congruent results, Y-chromosome data seem to be characterized by lower diversity within populations and more significant spatial structure than mitochondrial data (see, e.g., Jorde et al. 2000). It is unclear why this is so. Selection (Excoffier 1990; Wise et al. 1998; Wyckoff, Wang, and Wo 2000) and different demographic histories for females and males (Sajantila et al. 1996; Seielstad, Minch, and Cavalli-Sforza 1998; Perez-Lezaun et al. 1999) are two popular types of explanations.

Changes in population size tend to leave recognizable signatures in the patterns of nucleotide diversity. Therefore, the distribution of pairwise sequence differences in a sample (or, simply, the mismatch distribution) contains information on the population's history (Rogers and Harpending 1992). The genealogy of a population of constant size is expected to have long deep branches (Donnelly 1996); mutations occurring along these branches will be shared by several lineages, which will result in an irregular or ragged distribution of pairwise sequence differences. Conversely, the genealogy of a population that has substantially grown in size has long terminal branches, and the mutations that have occurred along these branches, i.e., most mutations, will be specific to a single lineage (Donnelly 1996). Under these conditions, one expects unimodal mismatch distributions (Harpending et al. 1993; Harpending 1994; Marjoram and Donnelly 1994), whose means, under an infinite-sites mutation model, increase as a function of the time elapsed after population growth (Sherry et al. 1994). However, different selective regimes may mimic the effects of changes in population size.

In addition, recombination acts as a confounding factor, for it brings together chromosome regions that evolved independently. For that reason, with only one exception (Alonso and Armour 2001), human mismatch distributions have only been studied at the mitochondrial level so far, based on both restriction fragment length polymorphisms (RFLPs) (Rogers and Harpending 1992; Harpending 1994) and hypervariable region I (HVRI) sequences (Excoffier and Schneider 1999). Almost all of these distributions are unimodal. In general, both the mean and the variance are highest (and therefore the curve is smoothest) in African populations, lower among Asians and Americans, and lowest among Europeans. The mean mismatch is related to the time of the expansion through the mutation rate (Rogers and Harpending 1992; Rogers and Jorde 1995), so the dates of the main demographic expansions are estimated at around 110,000 years ago in East Africa, 70,000 years ago in the rest of Africa and Asia, 55,000 years ago in America, and 40,000 years ago in Europe and in the Middle East (Excoffier and Schneider 1999). The few exceptions are represented by populations which may have undergone recent bottlenecks, thus presumably losing the typical genetic features of expanding populations (Excoffier and Schneider 1999).

Abbreviations: HVRI, hypervariable region I; RFLP, restriction fragment length polymorphism.

Key words: Y chromosome, polymorphism, human, mismatch distribution, population expansion.

Address for correspondence and reprints: Guido Barbujani, Dipartimento di Biologia, Università di Ferrara, via L. Borsari 46, I-44100 Ferrara, Italy. E-mail: bjg@unife.it.

Mol. Biol. Evol. 18(7):1259–1271. 2001

Because only mitochondrial mismatch distributions have been studied so far, it is not yet known whether the inferred expansions affected the entire populations or only their female components. In this study, we calculated the mismatch distributions in a data set of Y-chromosome biallelic polymorphisms in 48 population samples from Europe and the Mediterranean area (data from Rosser et al. 2000). The results obtained differ sharply from those observed for mitochondrial data. To understand the causes of that discrepancy, a database of European mitochondrial sequences was reanalyzed and patterns in the mismatch distributions of computer-simulated populations were studied with the aim of determining the effects of expansions versus constant population sizes.

Materials and Methods

Databases

The database of Y-chromosome biallelic markers we considered (Rosser et al. 2000) comprised data on 48 populations (listed in table 1), for a total of 3,677 individuals. Most were European, but populations from the eastern and southern shores of the Mediterranean sea, the Caucasus region, and Greenland were also included.

The 11 biallelic markers considered, namely, 2 insertion/deletion and 9 single-nucleotide (SNP) polymorphisms, defined 10 alleles, 6 of them polymorphic or subpolymorphic (frequency > 0.01) and 4 of them rare in Europe. For the sake of consistency with other studies, and because micro- and minisatellite variation has been observed within most such alleles (Jobling and Tyler-Smith 1995, 2000; Karafet et al. 1999), we refer to each of them as a ''haplogroup'' (De Knijff 2000). A single minimum-spanning tree could be constructed on the basis of those 10 haplogroups (fig. 1). Therefore, there was reason to believe that each of the 11 polymorphisms of interest represented the effect of a unique mutational event, without ambiguities. Full haplogroup descriptions and frequencies are in Rosser et al. (2000).

A database of mitochondrial DNA HVRI sequences in Europe (updated from Simoni et al. 2000) was used for comparisons of patterns of genetic diversity in maternally transmitted genes. Of the 48 population samples available there, 21 were selected that approximately matched the geographic location of the samples considered in this analysis of Y-chromosome diversity.

Mismatch Distributions and Neutrality Tests

Mismatch distributions and gene diversity, i.e., the probability that two randomly sampled chromosomes differ from each other (Nei 1987), were estimated for both databases by ARLEQUIN, version 2.0 (Schneider, Roessli, and Excoffier 2000). The fit of the observed distribution of mismatches to a model of population expansion was tested by a bootstrap approach, also implemented in ARLEQUIN. Note that for this method, the null hypothesis is one of population expansion, due to the fact that there is no quantitative expectation as for the shape of the mismatch distribution in a stationary population (Harpending 1994), whereas under the hypothesis of expansion, a parameter τ estimated from the data allows one to predict the average mismatch.

Within each population sample (including simulated samples; see below), departures from mutation-drift or mutation-selection equilibrium were tested by means of Tajima's D and Fu's F_S. In Tajima's (1989*a*, 1989*b*) test, the parameter $\theta = 2N\mu$ (where N is the population size and μ is the mutation rate) is independently estimated twice, once from the number of polymorphic sites and once from the average mismatch in the sample. Differences between the two estimates are then attributed to selection or to the demographic history of the population studied. Similarly, Fu's (1997) F_S statistic compares the observed number of alleles in a sample with the number of alleles expected if the population has kept a constant size.

The significance of D and F_S was tested by randomization. By the coalescent simulation program implemented in the ARLEQUIN package (Schneider, Roessli, and Excoffier 2000), separately for each sample studied, we generated random samples from a hypothetical stationary population whose parameter $\theta = 2N\mu$ was equal to the average number of observed pairwise differences. For each sample, the procedure was repeated 1,000 times, in every case recomputing the D and F_S statistics so as to obtain empirical null distributions of these statistics and hence the probability of the observed D and F_S values under the hypothesis of demographic stationarity.

Reanalysis of Reduced Mitochondrial Data Sets

To understand whether the Y-chromosome mismatch distributions could reflect an insufficient number of sites considered, we reanalyzed the mtDNA database in two ways.

1. From one randomly selected European population sample, we repeatedly calculated the mismatch distribution, each time removing 10 nucleotide sites from the initial 360 (starting from positions 16024–16033 of the Cambridge reference sequence; Anderson et al. 1981), until 10 sites were left.
2. In the 21 European samples, four sets of 11 polymorphic sites were selected so as to analyze the same number of sites for mtDNA and for the Y chromosome. The criteria of selection were as follows. In two runs of the analysis (data sets A and B), 11 sites were selected at random; in one run (data set C), 10 highly variable sites were used; and in one run (data set D), selection was among poorly polymorphic sites.

Coalescent Simulations

We generated samples from stationary and expanding populations by Monte Carlo simulation to see whether the shapes of the Y-chromosome mismatch distributions described in this study were compatible with some form of demographic expansion. The simulation algorithm was based on the coalescent process with su-

Table 1
Measures of Genetic Diversity Estimated from Y-Chromosome Data

Population	*N*	Hp	Average Mismatch	*P*(exp)	*H*	*D*	*P*(*D*)	F_S	$P(F_S)$
Northern Central Europe	939	9	1.48 ± 0.90	0.100	0.514	−0.003	0.505	0.863	0.682
Bavaria	80	6	2.07 ± 1.17	0.460	0.701	−0.187	0.442	−1.684	0.263
Belgium	92	7	1.60 ± 0.96	0.009	0.551	−0.688	0.244	−2.661	0.120
Cornwall	51	2	0.89 ± 0.63	0.030	0.297	−1.846	0.014	−7.573	0.000
Holland	84	6	2.02 ± 1.15	0.210	0.698	−0.224	0.380	−1.715	0.253
East Anglia	172	7	1.65 ± 0.98	0.110	0.585	−0.340	0.358	−1.550	0.314
France	40	7	2.19 ± 1.24	0.280	0.691	−0.467	0.281	−2.908	0.103
Ireland	257	6	1.05 ± 0.70	0.030	0.329	−0.966	0.152	−3.306	0.110
Scotland	43	4	0.84 ± 0.61	0.090	0.364	−2.010	0.004	−8.576	0.000
West Scotland	120	4	1.19 ± 0.77	0.130	0.437	−1.069	0.111	−3.754	0.072
Iberia	537	7	1.87 ± 1.07	0.210	0.562	1.198	0.875	2.993	0.879
Basque	26	3	0.77 ± 0.58	0.210	0.440	−2.440	0.000	−11.317	0.000
North Portugal	328	6	1.91 ± 1.09	0.130	0.575	0.244	0.598	−0.134	0.580
South Portugal	57	6	2.27 ± 1.27	0.140	0.637	−0.138	0.422	−1.906	0.238
Spain	126	7	1.69 ± 0.10	0.310	0.509	−0.430	0.328	−1.889	0.255
South-Central Europe	384	9	2.44 ± 1.33	0.600	0.795	1.288	0.876	2.530	0.881
Bulgaria	24	5	2.33 ± 1.32	0.900	0.772	−0.708	0.217	−4.325	0.011
Greece	36	6	2.40 ± 1.33	0.680	0.798	−0.301	0.376	−2.768	0.092
Italy	99	6	2.34 ± 1.29	0.300	0.728	0.254	0.560	−0.799	0.447
Romania	45	7	2.49 ± 1.37	0.320	0.810	−0.036	0.458	−2.022	0.216
Sardinia	10	4	2.53 ± 1.49	0.130	0.778	−1.560	0.040	−27.76	0.000
Slovenia	70	6	2.34 ± 1.29	0.280	0.745	0.071	0.508	−1.370	0.357
Yugoslavia	100	7	2.14 ± 1.20	0.580	0.706	0.016	0.481	−1.169	0.364
Eastern Europe	835	8	2.73 ± 1.45	0.100	0.773	2.384	0.976	5.556	0.991
Belarus	41	7	2.53 ± 1.39	0.200	0.728	−0.049	0.475	−2.170	0.172
Chuvash	17	7	2.65 ± 1.49	0.670	0.882	−0.692	0.244	−5.486	0.001
Czech Republic	53	6	2.56 ± 1.40	0.300	0.779	0.159	0.521	−1.532	0.317
Estonia	207	8	2.71 ± 1.44	0.080	0.762	1.076	0.816	0.923	0.760
Germany	30	6	1.87 ± 1.10	0.320	0.731	−1.047	0.120	−4.621	0.009
Hungary	36	5	2.52 ± 1.39	0.680	0.773	−0.154	0.423	−2.530	0.123
Latvia	34	4	2.63 ± 1.44	0.030	0.711	−0.069	0.407	−2.497	0.134
Lithuania	38	4	2.70 ± 1.47	0.020	0.656	0.099	0.537	−2.070	0.200
Mari	48	6	2.67 ± 1.45	0.050	0.776	0.231	0.574	−1.555	0.293
Poland	112	7	1.99 ± 1.13	0.330	0.645	−0.114	0.437	−1.318	0.315
Russia	122	8	2.78 ± 1.48	0.050	0.727	0.908	0.776	0.274	0.678
Slovakia	70	8	2.42 ± 1.33	0.330	0.717	0.171	0.560	−1.220	0.376
Ukraine	27	5	2.55 ± 1.41	0.030	0.681	−0.354	0.332	−3.402	0.050
Scandinavia and Finland	325	8	2.45 ± 1.33	0.010	0.744	1.575	0.916	3.110	0.930
Denmark	56	6	1.90 ± 1.10	0.060	0.647	−0.586	0.267	−2.769	0.103
Finland	57	6	2.03 ± 1.16	0.030	0.569	−0.420	0.315	−2.411	0.142
Gotland	64	5	1.89 ± 1.09	0.050	0.599	−0.522	0.299	−2.515	0.144
North Sweden	48	6	2.25 ± 1.26	0.050	0.709	−0.268	0.372	−2.318	0.155
Norway	52	6	2.16 ± 1.22	0.110	0.727	−0.331	0.369	−2.339	0.153
Saami	48	4	2.58 ± 1.41	0.000	0.696	0.115	0.529	−1.716	0.257
Southern Mediterranean	156	5	1.27 ± 0.80	0.160	0.446	0.041	0.566	1.663	0.793
Algeria	27	4	1.60 ± 0.98	0.000	0.584	−1.446	0.038	−5.959	0.000
North Africa	129	5	1.13 ± 0.74	0.150	0.396	−1.111	0.112	−3.909	0.066
Turkey	212	8	2.18 ± 1.21	0.640	0.782	0.978	0.823	1.841	0.824
North Cyprus	45	7	2.05 ± 1.17	0.430	0.775	−0.554	0.284	−2.929	0.086
Anatolia	167	8	2.17 ± 1.21	0.490	0.779	0.303	0.620	−0.370	0.495
Caucasus	200	7	2.04 ± 1.15	0.230	0.743	0.750	0.784	2.228	0.865
Armenia	89	7	2.11 ± 1.19	0.240	0.757	−0.078	0.479	−1.413	0.318
Georgia	64	6	1.70 ± 1.01	0.390	0.682	−0.747	0.200	−3.031	0.086
Ossetia	47	6	2.27 ± 1.27	0.120	0.700	−0.265	0.370	−2.337	0.180
Northwest Atlantic	89	5	1.50 ± 0.91	0.010	0.534	0.207	0.630	1.734	0.819
Greenland	61	4	1.41 ± 0.87	0.020	0.455	−1.130	0.108	−4.170	0.027
Iceland	28	3	1.71 ± 1.03	0.130	0.659	−1.295	0.068	−5.425	0.002

NOTE.—*N* = sample size; Hp = number of different haplogroups observed; average mismatch = mean and standard deviation of mismatch distribution; *P*(exp) = probability (estimated by bootstrap) of the observed mismatch distribution under the hypothesis of population expansion for a time estimated from the data; *H* = gene diversity; *D* = Tajima's *D*; *P*(*D*) = *P* value for *D*; F_S = Fu's F_S; $P(F_S)$ = *P* value for F_S.

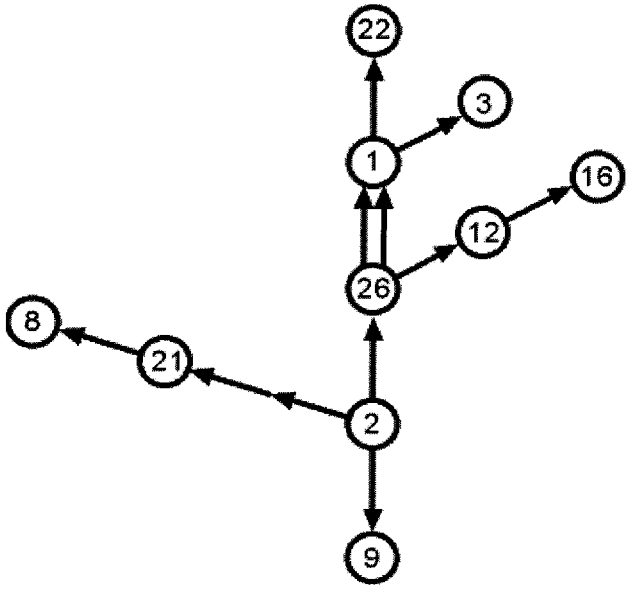

FIG. 1.—Network summarizing the evolutionary relationships among the 10 haplogroups observed in Europe. Each arrow represents one mutational event, whose probable direction is indicated by the arrow (from Rosser et al. 2000).

perimposed mutations, as described by Hudson (1990). Each sample was obtained by first generating its genealogy. Mutations were then randomly placed on the genealogy assuming they occurred according to a uniform and constant Poisson process.

First, we simulated 1,000 samples of genes under the assumption of a large and constant population size from a single panmictic deme. Each sample was composed of 80 individuals, i.e., 80 sets of 300 potentially variable sites. The size of the deme was 5,000 haploid individuals, and the mutation rate was 2×10^{-4} per generation for the whole sequence, i.e., 6.7×10^{-7} for each site. Although a plausible mutation rate for Y-chromosome biallelic polymorphisms seems to be around 5×10^{-7} (Hammer 1995; Jobling, Pandya, and Tyler-Smith 1997) or less (Thomson et al. 2000), we chose this higher rate so as to obtain similar mean pairwise differences in the simulated and in the real samples. The initial number of sites, 300, was also chosen because with those mutation rates, most sites (>97% in stationary populations) were monomorphic at the end of each simulation.

Second, four processes of exponential population expansion were simulated using the same coalescent approach, namely, (1) expansion from 50,000 years ago until now, with a 100-fold increase in population size, final effective population size $N_0 = 50{,}000$; (2) expansion from 50,000 years ago until now, with a 100-fold increase in population size, $N_0 = 100{,}000$; (3) expansion from 50,000 years ago until now, with a 100-fold increase in population size, $N_0 = 200{,}000$; and (4) expansion from 100,000 years ago until now, with a 100-fold increase in population size, $N_0 = 100{,}000$. The mutation rate was the same as that considered for the stationary populations. One thousand samples of 80 individuals were generated in this way for each of the four processes.

To more easily compare the simulation results, we defined three basic shapes of the mismatch distribution, namely, unimodal with a maximum at 0 (type 0), unimodal with a maximum >0 (type 1), and bimodal (type 2); examples are shown in figure 5.

Results

Mismatch Distributions and Neutrality Tests

Mismatch distributions obtained for Y chromosome biallelic markers were multimodal with one exception: the Chuvash population from Russia (fig. 2). Each distribution had at least two peaks, one at 0 and the other at a number of differences that varied among populations. These shapes reflect the fact that in many populations, most Y chromosomes belong to two frequent haplogroups, whereas other haplogroups occur at lower frequencies. Therefore, the peak at 0 differences corresponded to the comparisons between individuals that share the same allele, and the second peak was located at the mismatch representing the number of mutational steps separating the most frequent haplogroups. Where more than two haplogroups occurred at intermediate or high frequencies, there was a third, and sometimes a fourth, peak.

For 13 populations, the hypothesis of expansion could be rejected at the $P < 0.05$ level (fifth column of table 1). Although these probabilities were only nominal, Bonferroni's correction for multiple tests (Sokal and Rohlf 1995) confirmed significant overall departure from expansion expectations for the samples in this study. Conversely, all 21 mismatch distributions of the mtDNA samples appeared compatible with the effects of a population increase (fifth column of table 2). One parameter of the mismatch distribution, τ, estimates the time elapsed since population expansion (Rogers and Jorde 1995; Rogers 1995). This parameter is not reported in table 1 because we found little or no evidence for expansions in the shape of the Y-chromosome mismatch distributions.

It is possible to lump together Y-chromosome distributions based on their shapes; the clusters obtained in this way corresponded to sets of geographically near populations (fig. 3). This seems to be a consequence of the clinal variation shown by most nuclear markers in Europe (Chikhi et al. 1998; Casalotti et al. 1999; Quintana-Murci et al. 1999; Rosser et al. 2000; Barbujani and Bertorelle 2001). Most Western and Central European populations (British Isles, France, Belgium, and the Netherlands) showed a peak at three differences, i.e., the mutational distance between haplogroups 1 and 2, which represented the largest fraction of haplogroups there. Iberian populations (except Basques) showed an additional peak at five differences, resulting from the presence of a substantial number of haplogroup 21 chromosomes, which differ from those of haplogroup 1 by five mutational steps. The Southern-Central European and Turkish samples showed smoother distributions, reflecting larger numbers of different haplogroups. The

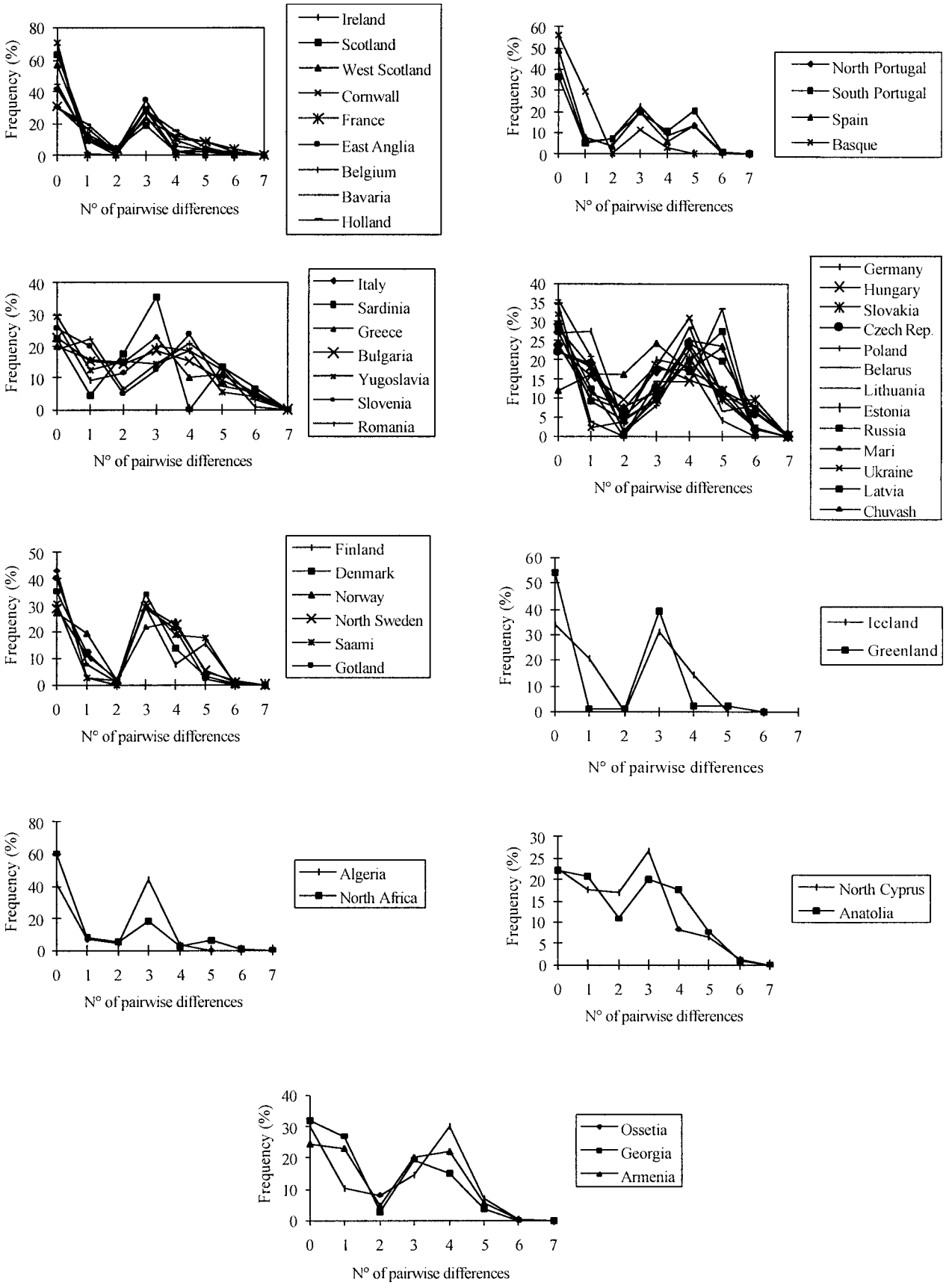

FIG. 2.—Mismatch distributions for Y-chromosome biallelic markers in 48 populations.

Table 2
Measures of Genetic Diversity Estimated from mtDNA Data

Population	*N*	Hp	Mean Mismatch	*P*(exp)	*H*	*D*	*P*(*D*)	F_S	*P*(F_S)
Basque	106	52	2.95 ± 1.56	0.430	0.936	−2.226	0.001	−26.498	0.000
Cornwall.....	69	45	3.89 ± 1.89	0.540	0.965	−2.127	0.002	−25.967	0.000
Sardinia	73	50	4.25 ± 2.13	0.790	0.956	−2.116	0.002	−25.824	0.000
Turkey	96	79	5.45 ± 2.65	0.850	0.988	−2.159	0.000	−25.345	0.000
Greece	48	37	4.67 ± 2.33	0.700	0.991	−2.001	0.005	−25.647	0.000
Bulgaria	30	22	4.55 ± 2.30	0.430	0.977	−1.878	0.016	−14.397	0.000
Denmark.....	32	20	3.41 ± 1.79	0.810	0.934	−1.586	0.041	−12.986	0.000
Sweden.....	32	27	4.58 ± 2.31	0.840	0.988	−1.871	0.014	−24.394	0.000
Georgia.....	45	28	4.57 ± 2.29	0.930	0.964	−1.774	0.013	−18.602	0.000
Germany.....	108	70	3.92 ± 1.98	0.630	0.973	−2.115	0.000	−25.912	0.000
Iceland	53	38	4.83 ± 2.40	0.300	0.979	−1.574	0.031	−25.597	0.000
Italy.....	115	95	6.14 ± 2.94	0.910	0.993	−2.139	0.000	−25.085	0.000
Norway.....	30	20	3.26 ± 1.73	0.860	0.954	−1.798	0.017	−14.418	0.000
France.....	111	73	3.83 ± 1.94	0.820	0.961	−2.217	0.000	−25.955	0.000
Belgium	33	25	3.35 ± 1.77	—[a]	0.974	−2.196	0.003	−26.498	0.000
Saami	240	37	3.60 ± 1.83	0.300	0.799	−1.172	0.107	−16.668	0.000
Estonia	28	23	4.36 ± 2.22	0.790	0.979	−1.728	0.018	−18.773	0.000
Portugal	54	38	3.60 ± 1.85	1.000	0.934	−1.988	0.007	−26.077	0.000
Finland	79	46	3.74 ± 1.91	0.720	0.970	−1.909	0.005	−26.048	0.000
Spain.....	74	61	5.25 ± 2.57	0.680	0.987	−2.043	0.001	−25.454	0.000
Russia.....	103	64	4.22 ± 2.11	0.960	0.965	−2.010	0.002	−25.784	0.000

NOTE.—*N* = sample size; Hp = number of different haplogroups observed; mean mismatch = mean and standard deviation of mismatch distribution; *P*(exp) = probability (estimated by bootstrap) of the observed mismatch distribution under the hypothesis of population expansion for a time estimated from the data; *H* = gene diversity; *D* = Tajima's *D*; *P*(*D*) = *P* value for *D*; F_S = Fu's F_S; *P*(F_S) = *P* value for F_S.

[a] This value is missing because the last-squares procedure to fit model mismatch distribution and observed distribution did not converge after 1,800 steps.

increased within-population diversity seems to be largely due to the fact that differently oriented clines, from the southeast into the northwest and from south to north, converge in that area. As a consequence, haplogroups that were very rare or absent elsewhere tended to reach substantial frequencies in these populations. For northern-eastern samples, other peaks at three, four, and five differences were evident, resulting from differences between haplogroups 2 and 16, haplogroups 2 and 3, and haplogroups 3 and 16, respectively.

Mismatch distributions appeared to be bi- or multimodal, regardless of whether single samples (fig. 2) or groups thereof (fig. 3) were analyzed. Accordingly, insufficient sample size does not seem to be a plausible explanation for that finding. Bimodal distributions were also observed in the few Asian and North African samples available, and among Greenlanders. We do not know of any suitable Y-chromosome data set which could allow comparison with other continents.

Mitochondrial and Y-chromosome variation also differed when summarized by means of Tajima's *D* and Fu's F_S (seventh and ninth columns of tables 1 and 2). For mitochondrial data, both statistics were negative and significant (with the exception only of Saami), and all F_S values were significant at the 0.001 level. Conversely, when estimated from Y-chromosome data, most values of F_S, and especially of *D*, were insignificant, and the latter were even positive in 12 cases. Such positive *D* values were not associated with any spatial pattern that we could recognize. On the contrary, the four negative and significant values occurred in linguistic (Basques) or geographic isolates (Sardinians, Scots, Cornish; note, however, the small sample size in Sardinia), also showing low gene diversity. Gene diversity seems to be patterned in space (sixth column of table 1), with comparatively high values in the south and in the east, as also observed for mitochondrial variation (Comas et al. 1997).

By and large, taken at face value, mismatch distributions, Tajima's *D*, and Fu's F_S would suggest that the European male population has had a different history than the female population and that only the latter has increased substantially in numbers. Before drawing any conclusions, however, it is better to ask whether those apparent differences between sexes may simply be some sort of statistical artifact. Only 11 Y-chromosome polymorphic sites were studied, versus more than 200 for mtDNA. Might that have biased the results?

Reanalysis of Reduced Mitochondrial Data Sets

Initially, we repeatedly estimated the mitochondrial mismatch distribution and related statistics in one randomly chosen sample, the Cornish sample, each time considering a decreasing number of sites, from 360 to 10. Figure 4 shows that the characteristic, unimodal pattern of the mismatch distribution is always maintained through repeated reductions of molecular information in the 69 HVRI sequences considered. As expected, the mean moved left and the variance decreased as fewer and fewer nucleotide positions were considered. The reduction in the diversity was not linear; it was slow in the first steps, and it accelerated later, probably reflecting the fact that the 5′ and 3′ extremes of the HVRI are less variable than is the intermediate segment.

As the number of sites considered decreased, *D* always remained negative and lost significance when 50 sites were left, whereas F_S was significant even when

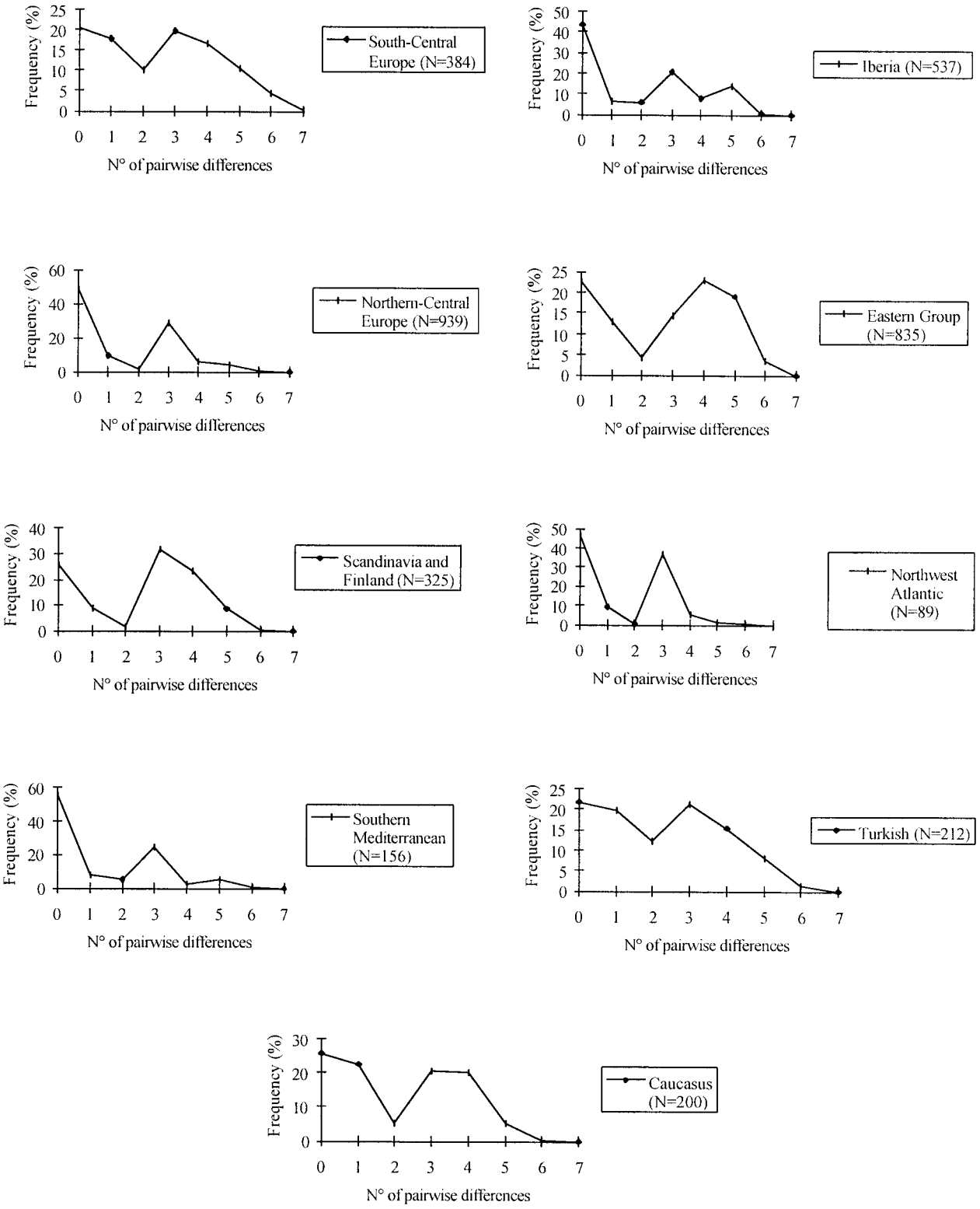

FIG. 3.—Mismatch distributions for Y-chromosome biallelic markers in the various population groups.

30 nucleotide sites were analyzed (the last 20 sites were monomorphic; table 3). This may indicate that although D is more robust than F_S for small sample sizes, F_S is more sensitive when few polymorphic positions are considered.

As a second test, we estimated the mismatch distributions from four different sets of 11 mtDNA polymorphic sites in the 21 populations for which both mtDNA and Y-chromosome data were available (table 4). The shape of the mismatch distribution was almost

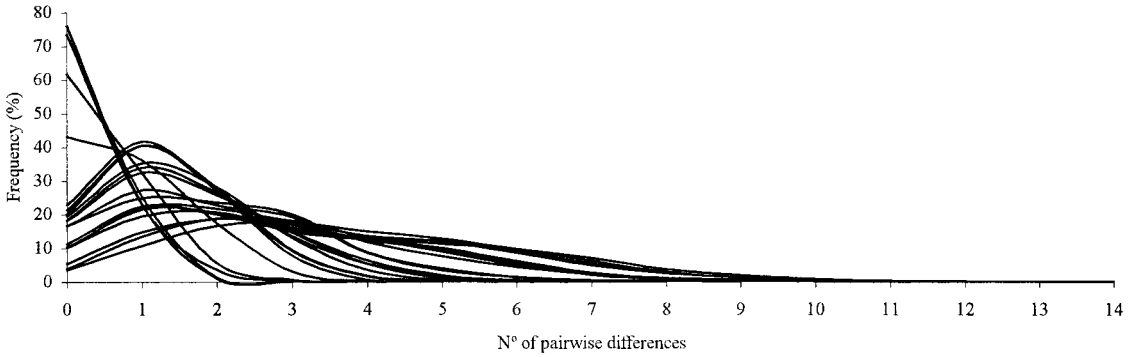

FIG. 4.—Mismatch distributions observed through analysis of subsets of data obtained by progressively removing 10 (constant or polymorphic) sites from the mitochondrial hypervariable region I of the Cornish population.

always unimodal, with only three vaguely bimodal shapes in a total of 81 mismatch distributions simulated (in the other three cases, all in data set D, no site was polymorphic among sequences, and therefore no statistic could be estimated). These results did not depend on the way polymorphic sites were selected, i.e., at random or on the basis of their levels of variation. Predictably, analyses of highly variable sites (data set C) yielded the highest means and variances.

Once again, Fu's F_S and Tajima's D values were negative with one exception (Saami), Also, confirming data obtained in the previous simulation, F_S negative values were statistically significant in the vast majority of the cases. In contrast, Tajima's D negative values did not often reach significance, although they tended to do so in data set D.

It is evident that even when the number of sites considered is the same as that available for the Y chromosome, mtDNA mismatch distributions are unimodal and therefore different from those calculated for the Y chromosome.

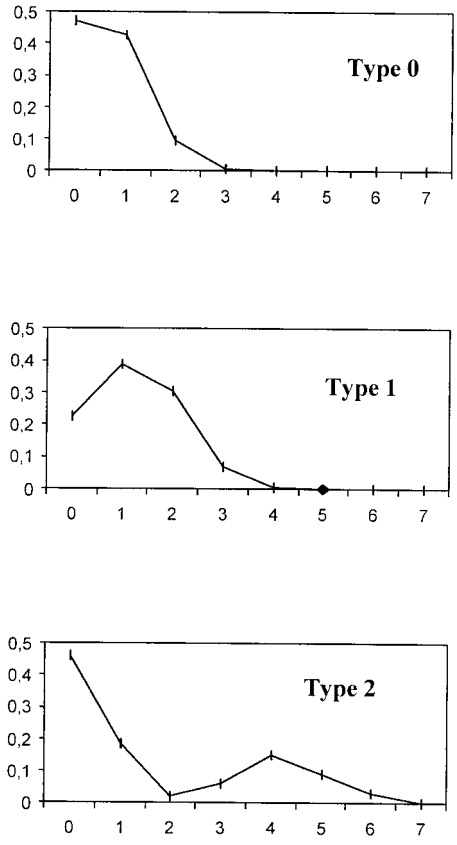

FIG. 5.—Scheme of the most common shapes of the mismatch distribution observed in the simulations.

Simulations

In simulated stationary populations, the average mismatch was higher than that for expanding populations (table 5) and close to the expected value, i.e., the parameter θ used to generate the simulated samples. This is what one expects under mutation-drift equilibrium (Rogers and Harpending 1992; Rogers et al. 1996). Also, the observed standard deviation (1.22) was close to the expectation (1.26) derived by Tajima (1983, eq. 30). In expanding populations, conversely, the average mismatch and its standard deviation were reduced, but both increased with the size of the population after the expansion and with the time since the expansion event.

Less than 12% of the mismatch distributions in the samples generated under stationarity showed a geometric form (type 0), and <20% had a single peak (type 1). Around 70% of the distributions showed multiple peaks (type 2). Conversely, for expanding populations, the number of bell-shaped mismatch distributions increased with the size of the population after the expansion and with the time to the expansion event and, correspondingly, the number of type 0 distributions decreased. Also, the proportion of distributions with multiple (generally two) peaks increased with the time since expansion and with the size of the population after expansion. The largest value was observed for populations that had expanded for 50,000 years, reaching an effective size N_0 of 200,000, where 21.8% of the mismatch distributions had two peaks (type 2).

In synthesis, the bimodal mismatch distributions observed in Y-chromosome European samples can be generated in simulations of both stationary and postexpansion populations. However, depending on the modes of the simulated expansion, bimodality is from 3–10

Table 3
Diversity Parameters Estimated in the Analysis of Mitochondrial Hypervariable Region I in the Cornish Population ($N = 69$), Considering a Decreasing Number of Sites Each Time

No. of Sites	Hp	Mean Mismatch	P(exp)	H	D	$P(D)$	F_S	$P(F_S)$
360	45	3.89 ± 1.98	0.540	0.965	−2.127	0.002	−25.973	0.000
310	44	3.62 ± 1.86	—	0.962	−2.186	0.000	−26.106	0.000
260	42	3.43 ± 1.77	0.640	0.945	−2.172	0.001	−26.209	0.000
210	38	2.55 ± 1.39	0.740	0.888	−2.292	0.000	−26.771	0.000
160	29	1.68 ± 1.00	0.920	0.819	−2.292	0.000	−27.730	0.000
110	23	1.33 ± 0.84	—	0.799	−2.127	0.000	−24.147	0.000
100	22	1.30 ± 0.83	—	0.786	−2.088	0.004	−22.365	0.000
90	20	1.22 ± 0.79	—	0.771	−1.950	0.006	−19.262	0.000
80	13	0.81 ± 0.59	0.700	0.567	−1.656	0.019	−10.700	0.000
70	8	0.50 ± 0.41	0.370	0.384	−1.878	0.005	−6.247	0.000
60	5	0.31 ± 0.32	0.550	0.266	−1.599	0.024	−3.135	0.021
50	5	0.28 ± 0.30	0.510	0.266	−1.410	0.058	−3.437	0.013
40	4	0.25 ± 0.29	0.480	0.240	−1.154	0.100	−2.239	0.042
30	4	0.25 ± 0.29	0.370	0.240	−1.154	0.105	−2.239	0.037
20/10.....	1	—	—	—	—	—	—	—

NOTE.—N = sample size; Hp = number of different haplogroups observed; mean mismatch = mean and standard deviation of mismatch distribution; P(exp) = probability (estimated by bootstrap) of the observed mismatch distribution under the hypothesis of population expansion for a time estimated from the data; H = gene diversity; D = Tajima's D; $P(D) = P$ value for D; F_S = Fu's F_S; $P(F_S) = P$ value for F_S. Some values are missing because the least-squares procedure to fit model mismatch distribution and observed distribution did not converge after 1,800 steps.

times as frequent in stationary populations as in expanding populations.

In the simulated stationary populations, both D and F_S showed a wide distribution centered on 0 (552 negative values out of 1,000 simulations for D, and 577 negative values for F_S). Only in 3.9% and 8.7% of the cases, respectively, were these values significant. Conversely, when expansions were simulated, F_S was always negative and was significant at the 5% level in >95% of the iterations. Tajima's D was also negative in nearly all cases of expansion and reached statistical significance in >85% of the simulations.

Discussion

The mismatch distributions inferred in this study from Y-chromosome biallelic markers were bimodal and did not resemble those inferred in the same populations from mtDNA data, which were unimodal. Statistical tests failed to reject a neutral equilibrium model for European Y-chromosome variation, whereas there was highly significant departure from equilibrium for mtDNA data (Merriwether et al. 1991; Excoffier and Schneider 1999) (table 5).

Models of population expansion do not predict, even transitorily, the presence of multiple peaks in the mismatch distribution (Rogers and Harpending 1992; Rogers and Jorde 1995). In Slatkin and Hudson's (1991) simulations, bimodal distributions with a peak at 0 were observed only for stationary populations. Conversely, expanding populations showed no instance of bimodality, and there were virtually no observations for mismatch = 0 (Slatkin and Hudson 1991, p. 560). Similar results were obtained by Harpending et al. (1998), who also showed that gene trees with a few well-differentiated alleles, much like those described in this study for the Y chromosome, are the rule in populations whose size has stayed constant or contracted.

Three lines of evidence suggest that the results of this study are not simply a statistical artifact:

1. By analyzing biallelic variation as we did in this study, one neglects other possible, but so far unde-

Table 4
Median Observed Values in the Analysis of Reduced mtDNA Data Sets Comprising 11 Sites in the 21 Population Samples of Table 2

Data Set[a]	Mean Mismatch	P(exp)	H	D	$P(D)$	No. significant D	F_S	$P(F_S)$	No. Significant F_S	No. of Unimodal Distributions
A	0.905	0.630	0.566	−1.117	0.132	1	−4.038	0.005	18	20/21
B	0.875	0.740	0.579	−0.914	0.185	1	−4.943	0.011	18	21/21
C	1.778	0.440	0.798	−0.411	0.315	0	−9.977	0.000	19	19/21
D	0.095	0.235	0.091	−1.418	0.039	11	−3.399	0.004	15	18/18[b]

NOTE.—P(exp) = probability (estimated by bootstrap) of the observed mismatch distribution under the hypothesis of population expansion for a time estimated from the data; H = gene diversity; D = Tajima's D; $P(D) = P$ value for D; F_S = Fu's F_S; $P(F_S) = P$ value for F_S.

[a] Sites considered: data set A—16051, 16069, 10693, 16104, 16124, 16126, 16129, 16145, 16163, 16172, and 16182; data set B—16187, 16189, 16192, 16222, 16223, 16224, 16245, 16249, 16256, 16261, and 16264; data set C—16069, 16126, 16183, 16189, 16192, 16223, 16270, 16278, 16294, 16298, and 16311; data set D—16051, 16104, 16142, 16167, 16176, 16215, 16248, 16257, 16265, 16290, and 16327.

[b] In three samples, no mismatch distribution could be calculated because all sequences were identical at the sites considered.

Table 5
Simulation Results

Simulated Population	Average Mismatch[a]	One Mode, Type 0[b]	One Mode, Type 1[b]	Two Modes Type 2[b]	D[c]	Signif.[d]	F_S[e]	Signif.[d]
Stationary.....	1.960 (1.218)	0.113	0.185	0.702	−0.041 (0.945) [−2.348, 2.775]	0.039	−0.159 (2.579) [−11.287, 10.955]	0.087
Expanding 1....	0.820 (0.375)	0.563	0.373	0.063	−1.826 (0.359) [−2.408, −0.019]	0.855	−10.193 (4.176) [−28.159, −0.568]	0.964
Expanding 2....	1.077 (0.457)	0.385	0.516	0.099	−1.959 (0.349) [−2.577, −0.244]	0.921	−14.208 (5.453) [−31.869, −1.039]	0.983
Expanding 3....	1.486 (0.651)	0.197	0.585	0.218	−1.987 (0.369) [−2.611, 0.005]	0.908	−17.245 (6.454) [−33.023, 1.320]	0.989
Expanding 4....	1.658 (0.563)	0.108	0.775	0.117	−2.015 (0.290) [−2.577, −0.746]	0.959	−19.018 (5.643) [−30.639, −4.013]	0.999

[a] Average mean of the mismatch distribution across 1,000 simulations. The standard deviation is shown in parentheses.
[b] Fraction of the 1,000 simulations showing shapes of the mismatch distribution.
[c] Average D (Tajima 1989*a*) over 1,000 simulations. The standard deviation is shown in parentheses, and the range of observed values is in brackets.
[d] Fraction of significant cases at the 5% level (out of 1,000 simulations).
[e] Average F_S (Fu 1997) over 1,000 simulations. The standard deviation is shown in parentheses, and the range of observed values is in brackets.

tected, polymorphism at other sites. However, that also applies to the mitochondrial RFLP studies which, as we have seen, show very different, unimodal mismatch distributions (Harpending 1994).

2. When we reanalyzed subsets of mitochondrial HVRI data, the shape of the distributions remained unimodal, and Tajima's D and Fu's F_S remained negative (although only the latter was always significant).
3. In our simulations, bimodal distributions appeared much more frequently in stationary than in expanding populations. We did not estimate a likelihood ratio because the numerical results depended on admittedly approximate expansion parameters and mutation rates. However, all factors considered, constant population sizes seemed roughly 3–10 times as likely as expansions.

The demographic scenarios we simulated were very simple. It is customary to model expansions as either instantaneous (see, e.g., Rogers and Jorde 1995) or exponential phenomena (this study; see also Excoffier and Schneider 1999), but populations may have grown in other ways. To mention just one, temporary contractions may have punctuated periods of general expansion, and that may have had an impact on levels and patterns of genetic diversity (Excoffier and Schneider 1999). However, in the absence of more sophisticated testable models, this study does not suggest that the Y-chromosome and the mitochondrial mismatch distributions differ from each other because of the limited resolution offered by the available Y-chromosome data.

Nonunimodal distributions (and insignificant Tajima's and Fu's statistics) are regarded as evidence that populations evolved under neutrality, without significantly increasing in size (Harpending et al. 1993). Departures from the expected shape can reflect adaptation (if one assumes constant population size), demographic changes (if one assumes neutrality), or both and can also occur when mutation rates vary across nucleotide sites (Aris-Brosou and Excoffier 1996). In principle, therefore, the different results obtained in the analysis of maternally and paternally transmitted genes in Europe may be due to differences in mutation mechanisms, differences in selective regimes, differences in past demographic history, or combinations thereof.

Might some sort of distorted mutational process have generated spurious multimodal mismatch distributions? Each of the 11 polymorphisms considered in this study probably results from a mutation that occurred only once (Rosser et al. 2000), and therefore these polymorphisms meet the assumptions of the model underlying the theory of mismatch distributions, the infinite-sites model (Rogers and Harpending 1992). In addition, simulations suggest anyway that the shape of the mismatch distribution tends to faithfully reflect the demographic history of a population, despite even substantial violations of the infinite-sites model (Rogers at al. 1996). Finally, recurrent mutation at some sites, reflecting mutation rate heterogeneity, may mimic the effects of population growth (Aris-Brosou and Excoffier 1996), but here the problem is the opposite, i.e., how to explain the apparent constancy of population size suggested by data. In short, we cannot rule out yet-to-be-discovered peculiarities of the Y-chromosome mutation process, but even if they existed, at present it is hard to imagine how such peculiarities could account for the results of this study.

Adaptation is the second factor. Tests based on the comparison of within-species and between-species nucleotide diversity have failed so far to reject the hypothesis of neutrality for Y-chromosome markers in comparisons between humans and mice (Nachman 1998), but not in comparisons between humans and Old World monkeys (Wyckoff, Wang, and Wu 2000). In addition, some loci of the Y chromosome are known to affect male fertility (Vogt 1997), and some detrimental mutations have been shown to occur more frequently on a particular Y-chromosome background (Jobling et al. 2000). Therefore, some role of selective pressures appears probable. However, Nachman's (1998) results and other simple calculations suggest that selection can ex-

plain, per se, only a small fraction of the human Y chromosome variation (Bertranpetit 2000).

The differences described here between mitochondrial and Y-chromosome data seem therefore to reflect, at least in part, the effects of past demographic phenomena. There are a few complications, though. Mismatch distributions from chromosomes subject to recombination contain little unambiguous evolutionary information. However, Tajima's D and Fu's F_S have been estimated within autosomal regions with no apparent recombination (Harding et al. 1997; Hey 1997; Zietkiewicz et al. 1998), and their values do not appear to depart from neutral, stationary expectations. Fay and Wu (1999) proposed that the smaller mitochondrial population size (one fourth that of autosomal genes) has caused a stronger impact of past population bottlenecks on mitochondrial variation. That interpretation seems at odds with the results of this study, because indices of Y-chromosome diversity resemble those estimated at the other nuclear loci, despite the fact that, in principle, Y-chromosome and mitochondrial effective population sizes should be the same.

It thus seems necessary to envisage either different demographic histories for males and females, with the former leaving a stronger mark on autosomal variation, or some combination of demographic changes and selective processes. Schematically, three hypotheses appear compatible with the available data:

1. The European female population increased in size; the male population did not.
2. Neither population increased in size, but there was disruptive selection for mtDNA.
3. Both populations increased in size, and there was purifying selection on the Y chromosome.

The data we analyzed do not allow, at present, discrimination among these hypotheses. However, it is worth noting that a small population size does not necessarily mean small numbers of individuals (of males in the present case). A high variance of reproductive success among individuals reduces the effective population size (Crow 1958). If the number of offspring has been generally more variable among males than among females, the effective population size inferred from Y-chromosome diversity is expected to be less (which is what this study suggests), and the genetic differences between populations tend to be greater (which has been demonstrated by, among others, Seielstad, Minch, and Cavalli-Sforza [1998] and Perez-Lezaun et al. [1999]). In other words, European men may have been approximately as numerous as European women, but a fraction of men may have left many descendants at each generation, and another fraction may have left just a few or none. The correlation in family size across generations, demonstrated in Canadian pedigrees (Austerlitz and Heyer 1998), would increase the evolutionary impact of this effect.

The first two hypotheses appear to contrast with the population expansions inferred from microsatellite (Pritchard et al. 1999) and sequence (Shen et al. 2000) Y-chromosome variation. Those studies, however, considered largely non-European samples and different Y-chromosome polymorphisms; it may be that the demographic history of Europe has been peculiar or that the biallelic polymorphisms we considered offer insight into a different period. Biallelic Y-chromosome markers, with their low mutation rates ($<10^{-8}$ per site per year; Hammer 1995; Jobling, Pandya, and Tyler-Smith 1997; Thomson et al. 2000), may only be able to reveal ancient population growth (see Takahata 1995). Conversely, fast-evolving markers in the mitochondrial genome (estimates of the mutation rate per site per year are 8.6×10^{-5} for the hypervariable region [Stoneking et al. 1992] and about 4.5×10^{-5} for RFLP [Rogers and Harpending 1992]) may contain information on more recent demographic changes.

At any rate, it is not impossible to reconcile the findings of Pritchard et al. (1999) and Shen et al. (2000) with those of the present study. If hypothesis 3 proved correct, the apparent constancy of the European male population size, as inferred from mismatch distributions, Tajima's tests, and Fu's tests, would be due to some form of purifying selection, ultimately concealing the effects of the demographic growth that previous studies of the Y-chromosome have recognized.

Acknowledgments

We are grateful to Giorgio Bertorelle, Laurent Excoffier, Antonio Amorim, and Chris Tyler-Smith, who discussed the results of this study with us and critically read the manuscript. Chris Tyler-Smith and Tatiana Zerjal also gave us access to unpublished data, and we thank them for that. This research was supported by funds from the Italian Ministry of the Universities (MURST COFIN 99) and from the University of Ferrara. L.P. was supported by a Ph.D. grant from Fundação para a Ciência e a Tecnologia (PRAXIS XXI/BD/13632/97), I.D. by a grant from the Swiss National Research Council (FNRS) for Perspective Investigators, Z.H.R. by a BBSRC Studentship, and M.A.J. by a Wellcome Trust Senior Fellowship in Basic Biomedical Science (grant number 057559).

LITERATURE CITED

ALONSO, S., and J. A. ARMOUR. 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.

ANDERSON, S., A. T. BANKIER, B. G. BARRELL et al. (14 coauthors). 1981. Sequence and organization of the human mitochondrial genome. Nature **290**:457–465.

ARIS-BROSOU, S., and L. EXCOFFIER. 1996. The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. Mol. Biol. Evol. **13**:494–504.

AUSTERLITZ, F., and E. HEYER. 1998. Social transmission of reproductive behavior increases frequency of inherited disorders in a young-expanding population. Proc. Natl. Acad. Sci. USA **95**:15140–15144.

BARBUJANI, G., and G. BERTORELLE. 2001. Genetics and the population history of Europe. Proc. Natl. Acad. Sci. USA **98**:22–25.

BERTRANPETIT, J. 2000. Genome, diversity, and origins: the Y chromosome as a storyteller. Proc. Natl. Acad. Sci. USA **97**:6927–6929.

CASALOTTI, R., L. SIMONI, M. BELLEDI, and G. BARBUJANI. 1999. Y-chromosome polymorphisms and the origins of the European gene pool. Proc. R. Soc. Lond. B Biol. Sci. **266**: 1959–1965.

COMAS, D., F. CALAFELL, E. MATEU, A. PEREZ-LEZAUN, E. BOSCH, and J. BERTRANPETIT. 1997. Mitochondrial DNA variation and the origin of the Europeans. Hum. Genet. **99**: 443–449.

CHIKHI, L., G. DESTRO-BISOL, V. PASCALI, V. BARAVELLI, M. DOBOSZ, and G. BARBUJANI. 1998. Clinal variation in the DNA of Europeans. Hum. Biol. **70**:643–657.

CROW, J. F. 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. Pp. 25–50 *in* K. WEISS, ed. Variation in the human genome. Wiley, Chichester, England (CIBA Foundation Symposium).

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., and S. SCHNEIDER. 1999. Why hunter-gatherer populations do not show signs of Pleistocene demographic expansions. Proc. Natl. Acad. Sci. USA **96**:10597–10602.

FAY, J. C., and C. I. WU. 1999. A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. Mol. Biol. Evol. **16**:1003–1005.

FU, Y.-X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics **147**:915–925.

HAMMER, M. F. 1995. A recent common ancestry for human Y chromosomes. Nature **378**:376–378.

HARDING, R. M., S. M. FULLERTON, R. C. GRIFFITHS, J. BOND, M. J. COX, J. A. SCHNEIDER, D. S. MOULIN, and J. B. CLEGG. 1997. Archaic African and Asian lineages in the genetic ancestry of modern humans. Am. J. Hum. Genet. **60**:772–789.

HARPENDING, H. C. 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. Hum. Biol. **66**:591–600.

HARPENDING, H. C., M. A. BATZER, M. A. GRUVEN, L. B. JORDE, A. R. ROGERS, and S. T. SHERRY. 1998. Genetic traces of ancient demography. Proc. Natl. Acad. Sci. USA **95**:1961–1967.

HARPENDING, H. C., S. T. SHERRY, A. R. ROGERS, and M. STONEKING. 1993. The genetic structure of ancient human populations. Curr. Anthropol. **34**:483–496.

HEY, J. 1997. Mitochondrial and nuclear genes present conflicting portraits of human origins. Mol. Biol. Evol. **14**:166–172.

HUDSON, R. R. 1990. Gene genealogies and the coalescent process. Pp. 1–44 *in* D. FUTUYMA and J. ANTONOVICS, eds. Oxford surveys in evolutionary biology. Oxford University Press, Oxford, England.

JOBLING, M. A., A. PANDYA, and C. TYLER-SMITH. 1997. The Y chromosome in forensic analysis and paternity testing. Int. J. Legal Med. **110**:118–124.

JOBLING, M. A., and C. TYLER-SMITH. 1995. Fathers and sons: the Y chromosome and human evolution. Trends Genet. **11**: 449–456.

———. 2000. New uses for new haplotypes: the human Y chromosome, disease, and selection. Trends Genet. **16**:356–362.

JOBLING, M. A., G. WILLIAMS, K. SCHIEBEL, A. PANDYA, K. MCELREAVEY, L. SALAS, G. A. RAPPOLD, N. A. AFFARA, and C. TYLER-SMITH. 2000. A selective difference between human Y-chromosomal DNA haplotypes. Curr. Biol. **8**: 1391–1394.

JORDE, L. B., W. S. WATKINS, M. J. BAMSHAD, M. E. DIXON, C. E. RICKER, M. T. SEIELSTAD, and M. A. BATZER. 2000. The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. Am. J. Hum. Genet. **66**:979–988.

KARAFET, T. M., S. L. ZEGURA, O. POSUKH et al. (14 co-authors). 1999. Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. Am. J. Hum. Genet. **64**: 817–831.

MARJORAM, P., and P. DONNELLY. 1994. Pairwise comparisons of mitochondrial DNA sequences in subdivided populations and implications for early human evolution. Genetics **136**: 673–683.

MERRIWETHER, D. A., A. G. CLARK, S. W. BALLINGER, T. G. SCHURR, H. SOODYALL, T. JENKINS, S. T. SHERRY, and D. C. WALLACE. 1991. The structure of human mitochondrial DNA variation. J. Mol. Evol. **33**:543–555.

NACHMAN, M. W. 1998. Y-chromosome variation of mice and men. Mol. Biol. Evol. **15**:1744–1750.

NEI, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

PEREZ-LEZAUN, A., F. CALAFELL, D. COMAS et al. (12 co-authors). 1999. Sex-specific migration patterns in central Asian populations, revealed by the analysis of Y-chromosome short tandem repeats and mtDNA. Am. J. Hum. Genet. **65**:208–219.

POLONI, E. S., G. PASSARINO, A. S. SANTACHIARA-BENERECETTI, O. SEMINO, A. LANGANEY, and L. EXCOFFIER. 1997. Human genetic affinities for Y chromosome p49a,f/Taq I haplotypes show strong correspondence with linguistics. Am. J. Hum. Genet. **61**:1015–1035.

PRITCHARD, J. K., M. T. SEIELSTAD, A. PEREZ-LEZAUN, and M. W. FELDMAN. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol. Biol. Evol. **16**:1791–1798.

QUINTANA-MURCI, L., O. SEMINO, E. MINCH, G. PASSARINO, A. BREGA, and A. S. SANTACHIARA-BENERECETTI. 1999. Further characteristics of proto-European Y chromosomes. Eur. J. Hum. Genet. **7**:603–608.

ROGERS, A. R. 1995. Genetic evidence for a Pleistocene population explosion. Evolution **49**:608–615.

ROGERS, A. R., A. E. FRALEY, M. J. BAMSHAD, W. S. WATKINS, and L. B. JORDE. 1996. Mitochondrial mismatch analysis is insensitive to the mutational process. Mol. Biol. Evol. **13**:895–902.

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

ROGERS, A. R., and L. B. JORDE. 1995. Genetic evidence on modern human origins. Hum. Biol. **67**:1–36.

ROSSER, Z. H., T. ZERJAL, M. E. HURLES et al. (60 co-authors). 2000. Y-chromosomal diversity within Europe is clinal and influenced primarily by geography rather than language. Am. J. Hum. Genet. **67**:1526–1543.

SAJANTILA, A., A. H. SALEM, P. SAVOLAINEN, K. BAUER, C. GIERIG, and S. PÄÄBO. 1996. Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. Proc. Natl. Acad. Sci. USA **93**:12035–12039.

SCHNEIDER, S., D. ROESSLI, and L. EXCOFFIER. 2000. ARLEQUIN: a software for population genetics data analysis. Version 2.0. Department of Anthropology, University of Geneva, Switzerland.

SEIELSTAD, M., E. MINCH, and L. L. CAVALLI-SFORZA. 1998. Genetic evidence for a higher female migration rate in humans. Nat. Genet. **20**:278–280.

SHEN, P., F. WANG, P. A. UNDERHILL et al. (13 co-authors). 2000. Population genetics implications from sequence variation in four Y chromosome genes. Proc. Natl. Acad. Sci. USA **97**:7354–7359.

SHERRY, S. T., A. R. ROGERS, H. C. HARPENDING, H. SOODYALL, T. JENKINS, and M. STONEKING. 1994. Pairwise differences of mtDNA reveal recent human population expansions. Hum. Biol. **66**:761–776.

SIMONI, L., F. CALAFELL, D. PETTENER, J. BERTRANPETIT, and G. BARBUJANI. 2000. Geographic patterns of mtDNA diversity in Europe. Am. J. Hum. Genet. **66**:262–278.

SLATKIN, M., and R. R. HUDSON. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics **129**:555–562.

SOKAL, R. R., and F. J. ROHLF. 1995. Biometry. 3rd edition. Freeman, San Francisco.

STONEKING, M., S. SHERRY, and L. VIGILANT. 1992. Geographic origin of human mtDNA revisited. Syst. Biol. **41**:384–391.

TAJIMA, F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics **105**:437–460.

———. 1989*a*. Statistical method for testing the neutral mutation hypothesis by DNA polymorphisms. Genetics **123**: 585–595.

———. 1989*b*. The effect of change in population size on DNA polymorphism. Genetics **123**:597–601.

TAKAHATA, N. 1995. A genetic perspective on the origin and history of humans. Annu. Rev. Ecol. Syst. **26**:343–372.

THOMSON, R., J. K. PRITCHARD, P. SHEN, P. J. OEFNER, and M. W. FELDMAN. 2000. Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. Proc. Natl. Acad. Sci. USA **97**:7360–7365.

VOGT, P. H. 1997. Molecular basis of male (in)fertility. Int. J. Androl. **20**(Suppl. 3):2–10.

WISE, C. A., M. SRAML, D. C. RUBINSZTEIN, and S. EASTEAL. 1998. Comparative nuclear and mitochondrial genome diversity in humans and chimpanzees. Mol. Biol. Evol. **14**: 707–716.

WYCKOFF, G. J., W. WANG, and C. WU. 2000. Rapid evolution of male reproductive genes in the descent of man. Nature **403**:304–308.

ZIETKIEWICZ, E., V. YOTOVA, M. JARNIK et al. (11 co-authors). 1998. Genetic structure of the ancestral population of modern humans. J. Mol. Evol. **47**:146–155.

JEFFREY LONG, reviewing editor

Accepted March 15, 2001