A highly variable segment of human subterminal 16p reveals a history of population growth for modern humans outside Africa

Santos Alonso[†] and John A. L. Armour

Institute of Genetics, University of Nottingham, Queen's Medical Center, Nottingham NG7 2UH, United Kingdom

Edited by Henry C. Harpending, University of Utah, Salt Lake City, UT, and approved November 3, 2000 (received for review May 26, 2000)

We have sequenced a highly polymorphic subterminal noncoding region from human chromosome 16p13.3, flanking the 5′ end of the hypervariable minisatellite MS205, in 100 chromosomes sampled from different African and Euroasiatic populations. Coalescence analysis indicates that the time to the most recent common ancestor (approximately 1 million years) predates the appearance of anatomically modern human forms. The root of the network describing this variability lies in Africa. African populations show a greater level of diversity and deeper branches. Most Euroasiatic variability seems to have been generated after a recent out-of-Africa range expansion. A history of population growth is the most likely scenario for the Euroasiatic populations. This pattern of nuclear variability can be reconciled with inferences based on mitochondrial DNA.

The evolutionary history of a chromosomal locus can be reconstructed under mathematical models including information on its underlying genealogy (1). Ultimately, analyses of independent loci should, in combination, allow us to infer our evolutionary past. Genomic sequences provide unbiased strings of contiguous single nucleotide polymorphisms for this purpose; however, the phase of the linked polymorphisms needs to be resolved. Beyond the mitochondrial microcosm (2), the sex chromosomes provide the opportunity for simple elucidation of haplotypes (3–9), but the autosomes remain the most abundant source of independent genealogies (10–13).

Emerging autosomal sequence data mainly seem so far to conflict with earlier mtDNA and Y chromosome substitutional polymorphism studies, which seem to indicate an expansion in human population size, at approximately 100,000 years ago (14). In some cases, they even fail to reveal an expansion in size that archaeologically seems to be evident; at least in Europe, there is a clear sign of population growth during the Upper Paleolithic (15). This conflicting scenario has been used to support alternative views on human origins and evolution (16).

In investigating human origins, it would be desirable that present patterns of genetic variability could be explained simply by mutation and demography. However, many of the regions sequenced so far map near genes relevant for human health, and inferences on demographic history may be distorted by selection, especially in areas with a very low rate of recombination. On the other hand, recombination within the region under scrutiny can render parsimonious reconstruction of phylogenies doubtful (17) and therefore hinder direct inferences (18). To complicate matters further, the evolutionary pace of some autosomal loci may be insufficient to reveal possible demographic events in the time frame of interest (19), with more recent events requiring faster mutation rates. Thus, the absence of a signal indicating growth might be caused by a low level of polymorphism, rendering a low power to tests devised for that purpose.

The region immediately flanking the 5′ end of minisatellite MS205 at 16p13.3 is assumed to be neutral (because it maps within a large intron approximately 50 kb long) and is G+C rich (65% G+C). G+C-rich regions can contain frequent CpG dinucleotides, which, if subject to methylation-mediated deamination, may reach transition rates five times the background mutation rate (20). This region does contain CpGs methylated in both somatic and sperm DNA (demonstrated by bisulfite mutagenesis; unpublished work). In addition, it maps to a region of high recombination, which may help to shield it further from the distorting effects of genetic hitchhiking or background selection. Consequently, it may constitute a rich source of sequence polymorphism useful for human evolution studies. Therefore, we have sequenced 1.75 kb of this region in a set of different world populations to investigate our demographic history.

Materials and Methods

Genomic DNA from 10 Pygmy (five Biaka and five Mbuti), 10 Kenyan (Mijikenda from the Kilifi district), 10 Japanese (Nagoya), 10 British, and 10 Basque individuals were manually cycle-sequenced for a region encompassing 1.75 kb of the immediately 5′ flanking region of minisatellite MS205 at 16p13.3. The sequencing reactions make use of α^{33}P ddNTP terminators (Amersham Pharmacia). This method results in a more specific labeling, because only properly terminated DNA chains are labeled. "Stop" artifacts and background bands are thus eliminated. Thermosequenase (Amersham Pharmacia) was used as DNA polymerase in the sequencing reactions, because this enzyme has been engineered to efficiently incorporate dideoxynucleotides. In addition, deaza-dGTP was included in the reaction mix to help overcome compression artifacts. A series of primers was designed defining overlapping regions of about 250 bp (primer sequences and cycling conditions are available on request). The presence of a polymorphic position results in two bands of half the intensity of a monomorphic position (if the variant allele is present in a heterozygous state) or in the complete absence of the common allele and presence of an alternative form of the same intensity (if the variant allele is present in a homozygous state). The phase of the polymorphisms was resolved experimentally for all individuals analyzed. Allele-specific PCR (21) and resequencing of the products obtained was performed for that purpose. DNA sequences were processed and assembled by means of the GCG package (22). All 100 haplotype sequences have been submitted to GenBank (accession numbers AJ391838 to AJ391937).

Divergence (K) was estimated by comparison of a random pygmy sequence with one chimp sequence (GenBank accession

---

This paper was submitted directly (Track II) to the PNAS office.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AJ391838 to AJ391937).

See commentary on page 779.

[†]To whom reprint requests should be addressed. E-mail: pdzsaa@granby.nott.ac.uk.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Article published online before print: *Proc. Natl. Acad. Sci. USA*, 10.1073/pnas.011244998. Article and publication date are at www.pnas.org/cgi/doi/10.1073/pnas.011244998

Table 1. Polymorphic positions

Ancestor	ggga+cccgggccgggcccccgacggggtaagctaggggcgt
3P	a.....t........t........a.........a.....
1U	.a.........a..a..................a.....
1J	..c...........aa..................a.....
3B+4J	..c...........a..................a.....
1P	........ca.t.........a....a....at........
1P	............a.a...t...a..........a.....
1J	..............a.................c.a.....
1U	..............a..................a..a...
2P	..............a..................a....c.
1K	..............a..................a.....c
13B+11J+14U+8K+2P	..............a..................a.....
1U	..............a..................a...t..
1B+2UK	..............a.................ca.....
1J	..............a...........t.......a.....
1B+1K	..............a......c...........a.....
1B+2J+4K+4P	..............a....t.............a.....
1P	..............a...t...a..........a.a....
2K	..............a...t...a..........a.....
1K	..............a..t...............a.....
1P+1K	.....
1U	..........t....a.................ca.....
1B	.......t.....
1P	.....t...........t........a.acg..........
3P	....−..t...............a..............
1P	....−.gt...............a..............
2K	...g.....................a....a..........

Dots represent the same state as in the ancestor sequence. + and − in polymorphism number 5 represent presence or absence of a 5-bp motif, respectively. Abbreviations: B (Basques), J (Japanese), K (Kenyans), P (pygmies) and U (U.K.).

nos. AJ252012, AJ252013, and AJ252014) by means of K-ESTIMATOR 5.5 (http://mk-dimension-1.uchicago.edu/) by Josep M. Comeron (23) using a Kimura two-parameter model for multiple hit correction and a transition/transversion rates ratio α:β of 4:1. For this estimate, we assumed a divergence time (t) of 5 million years and an ancestral human-chimp effective population size estimate (N_e) of 10^5 (24). The mutation rate was inferred from divergence by using the formula $\mu = K/(2t + 4N_e)$ (see ref. 25).

To detect departure from a standard neutral model, a series of tests was used on the populations both individually and grouped by continent. ARLEQUIN 2.0 (http://lgb.unige.ch/arlequin and DNASP 3.5 (http://www.bio.ub.es/~julio/DnaSP.html; ref. 26) were used to perform Tajima's D (27), Fu's F_s (28), and Fu and Li's D^* and F^* tests (29). Both ARLEQUIN 2.0 and DNASP 3.5 provide P values based on a coalescent simulation algorithm (10,000 simulations were run). These P values represent the probability that the simulated estimate is less than the observed value in Tajima's D test or less than or equal to the observed value for the rest of the tests. Rejection of these tests may be caused by violation of any of the assumptions in the null hypotheses (neutrality, constant size, panmixia, no recombination). Significant departure of these tests has been explained mainly to be due to an excess of new mutations as results of evolutionary forces, such as selective sweeps or population growth. Processes that produce an excess of old mutations also render significant but positive departures. These processes may include population subdivision and balancing selection (18, 30). Simulations based on a coalescent algorithm with recombination (10,000 simulations, using DNASP 3.5) were performed also to estimate P values of the neutrality tests. A recombination parameter $C = 4N_ec$ (where c is the recombination rate per generation) with values of 1 and 10 were used for this purpose.

Tajima's method uses the difference between the average number of nucleotide differences (k) and an estimate of $\theta = 4N_e\mu$ from the number of segregating sites ($\hat{\theta}_s$). Because under neutrality, equilibrium, and panmixia the expectations of both parameters are θ, we expect $k \approx \hat{\theta}_s$ if these assumptions are correct. Fu's F_s test is based on the probability of having no fewer than k_0 observed alleles in a sample of n sequences, given the estimator of θ based on the average number of pairwise differences $\hat{\theta}_\pi$. Fu's and Li's D^* and F^* tests rely on the difference between two estimates of θ based on the number of mutations in external and internal branches in the genealogy of n sequences (test D^*) or between the average number of nucleotide differences between two sequences in a random sample of n sequences from a population and η_e, the number of mutations in external branches (test F^*).

ARLEQUIN 2.0 was used also to analyze the sequence mismatch distributions. The package fits a distribution to the observations by using a generalized nonlinear least-squares method, from which the parameter $\tau = 2\mu t$ (t being the time since the expansion and μ the mutation rate per sequence) is deduced. Confidence intervals are obtained by parametric bootstrap: this method assumes that the data are distributed according to a sudden expansion model. Thus, a large number of random samples (10,000 in our case) is generated according to the estimated demography with a coalescent algorithm. For each simulated data set, the parameter of interest is reestimated and for a given confidence value α, the approximate limits of the confidence interval are obtained as the $\alpha/2$ and $1 - \alpha/2$ percentile values. Schneider and Excoffier (31) showed that for τ, the true value of the parameter is included in a $100(1 - \alpha)$ confidence interval with a probability very

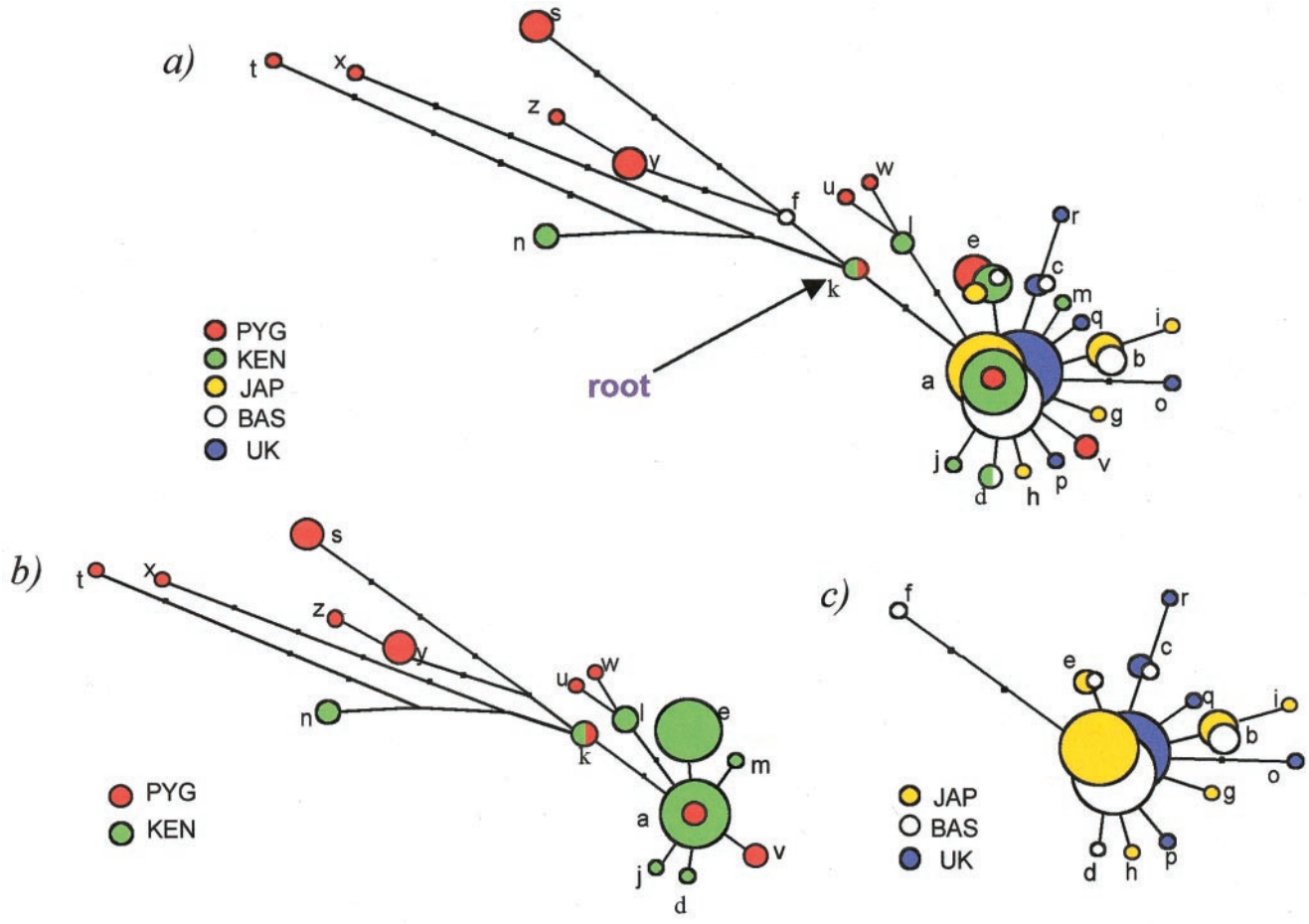

Fig. 1. Median-joining networks depicting the relationships between the haplotypes for all the populations (*a*), for only the African populations (*b*), and for only the non-African populations (*c*).

close to $(1 - \alpha)$. The fit to the expansion model is evaluated by the same parametric bootstrap approach as before, using the sum of square deviation (SSD) between the observed and expected mismatch as a test statistic. In this case, the P value is approximated by P = (number of simulated $SSD_{sim} \geq SSD_{obs}$)/number of simulated samples.

A phylogenetic network (32) describing the genealogical relationships between the different haplotypes was obtained with NETWORK 2.0 (47). To root this tree, Innan and Tajima's method (33) was used to estimate the most recent ancestral states by means of PRANC, a computer program provided by those authors. By using the theory of gene genealogy, this program calculates the probability of ancestry for each polymorphic position, taking into consideration the frequency of each class, the number of segregating sites within each class, and the number of fixed differences between classes. The root of the tree was also estimated by using the GENETREE (http://www.maths.monash.edu.au/~mbahlo/mpg/gtree.html) package (34). In this approach, all possible rooted trees were generated, and the associated likelihood values were obtained by using the coalescent theory. Both approaches in combination were used to deduce the root; in case of ambiguous or conflicting positions, the ancestral state indicated by comparison with the orthologous sequence in other nonhuman primates (GenBank accession nos. AJ252012, AJ252013, and AJ252014) was favored.

Further coalescent analysis was carried out also by using the GENETREE package. Thus, from an initial estimate of $\theta = 8.11 \pm 2.29$ from the number of segregating sites, three rounds of 10 million simulations were run (assuming neutrality, panmixia, and constancy in size). In each round, an initial value of θ was used to obtain a density distribution from which the maximum likelihood estimate ($\hat{\theta}_{mlk}$) was selected and used as a starting value for the next round. After a third round, a $\hat{\theta}_{mlk}$ of 11.06 for all five populations grouped together was obtained. This value was used for further simulations to estimate the time to the most recent common ancestor, for which another 10 million simulations were run.

By using GENETREE, a "quick" exploration for each population was performed independently. Thus, the joint maximum likelihood estimates of θ and the exponential population growth parameter β (growth rate per $2N_e$ generations) were obtained iteratively by fixing a $\hat{\theta}_{mlk}$ as described above and obtaining a likelihood density for β in one round of simulations; after selecting the $\hat{\beta}_{mlk}$, a likelihood density surface for θ in the vicinity of the previous $\hat{\theta}_{mlk}$ was obtained in a further round of simulations. Rounds of simulations in this fashion were performed until both $\hat{\theta}_{mlk}$ and $\hat{\beta}_{mlk}$ stabilized. In this context, quick means 1 million or less simulations in each round. This quick exploration took several weeks on a 400-MHz computer.

Results and Discussion

The sequence region flanking minisatellite MS205 at 16p13.3 is highly polymorphic. We detected 42 substitutions plus one deletion event (involving 5 bp starting at position 219) in 100 human chromosomes. Nucleotide diversity π ranged between 0.3% (SEM 0.2%) for the Pygmies (0.1%, SEM 0.08%, for the

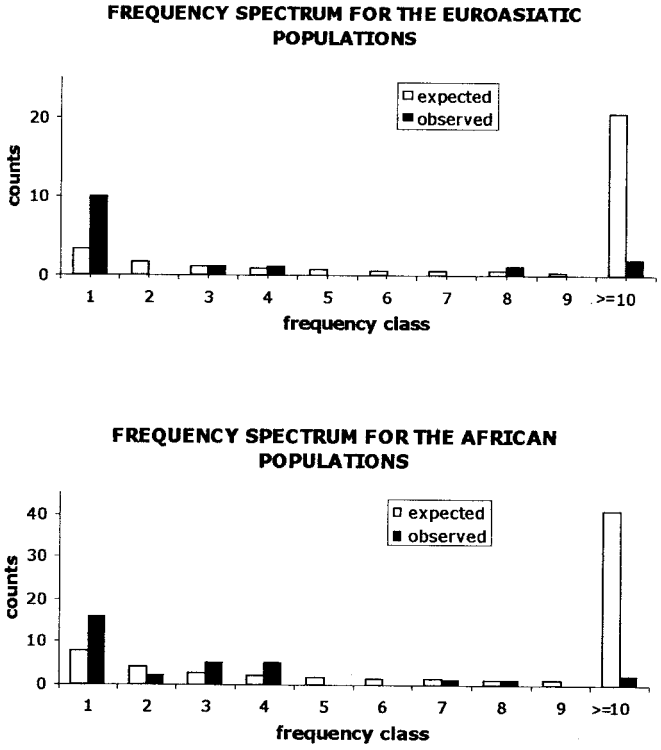

Fig. 2. Frequency spectra for the populations grouped by continent. The frequency class represents the number of segregating sites for which the mutant form is present in i copies and the ancestral estate in $n - i$ copies, with i ranging from 1 to $n - 1$ and n being the total number of sequences. As the ancestral state has been inferred, these frequency spectra are unfolded, that is, classes (i, $n - i$) and ($n - i$, i) can be distinguished. For convenience, frequency classes from $i = 10$ to $n - 1$ have been grouped together. Expected values under neutrality and constant size were obtained by using equation 51 in ref. 27.

Kenyans) and 0.04% (SEM 0.03%) for the U.K. population (0.05%, SEM 0.04%, for both Basques and Japanese). Divergence from the chimpanzee sequence was estimated as 0.0228 (95% confidence interval 0.0157–0.0305). From divergence, the estimate of the average mutation rate per site per year is 2.19×10^{-9}, higher than those described for the PDHA1 locus (8.06×10^{-10}; ref. 3), a ZFX intron (1.34×10^{-9}, ref. 4); a region (5) in Xq13.3 (9.03×10^{-10}), or β-globin (1.1×10^{-9}; ref. 10), (estimates calculated from divergence data in ref. 35 and by using the equation and parameters described in *Materials and Methods*, corrected for X chromosome when necessary) and higher than the average autosomal rate (1.28×10^{-9}; ref. 25). The mutation rate per sequence (1,742 sites) per generation (20 years) was estimated as 7.63×10^{-5}. The abundance of CpG doublets could be an explanation for this high rate, because over 40% of the mutations detected fell within a CpG dinucleotide. However, there is much uncertainty in the estimate of the mutation rate, because, as indicated by ref. 25, allelic (versus species) divergence time and ancestral population size (for instance) cannot be precisely estimated.

During the course of this study, it became clear that the region analyzed lies within a large intron of a low voltage-activated T-type Ca^{2+} channel gene (*CACNA1H*; ref. 36). It is difficult to assess at this stage with what intensity selection on the gene may be affecting the distribution of the polymorphisms in this intron. However, MS205 maps to subterminal 16p (about 1.3 megabases from the telomere), and it is known that genetic recombination increases toward the telomere, particularly in males (37). In fact, a recombinational hot spot has been described (36) in the 85 kb separating the 3′ end of minisatellite MS205 (D16S309) and the 5′ end of minisatellite EKMDA2 (D16S83), situated downstream of MS205. For this region, an enhanced recombination rate of 22-fold above the paternal genome-wide average of 0.9 centimorgans/megabase was reported. Recombination, however, does not seem to disrupt the reconstruction of the evolutionary history of the region immediately flanking the 5′ end of MS205. Whereas all recombination events in the coalescent time of a sequence locus are not likely to be detected by the four-gamete test (17), the assumption of an evenly distributed recombination rate across nucleotides, at least near this region, does not seem to hold. Thus, for instance, between the 85 kb between MS205 and EKMDA2, three of six crossovers could be fine-mapped within a <3-kb interval. This seems to indicate that areas of high recombination may comprise intervals of strong linkage disequilibrium, interspersed with focal regions of more intense recombinational activity. Analysis of a short (1.75 kb) sequence reduces the chance of it containing such a recombinational hot spot.

Table 2. Neutrality tests†

	Fu's		Tajima's		Fu and Li's			
Population	*Fs*	*P*‡	*D*	*P*	*D**	*P*	*F**	*P*
Kenyan	−1.992	0.113	−1.120	0.139	0.116	0.443	−0.279	0.391
Pygmy	−1.198	0.298	−1.047	0.157	−1.127	0.154	−1.284	0.123
All Africans	−3.885	0.184	−1.549	0.035	−1.938	0.053	−2.140	0.034
Japanese	−2.646	0.012	−1.140	0.156	−1.213	0.079	−1.376	0.123
C = 1		0.040		0.121		0.068		0.119
C = 10		0.093		0.089		0.049		0.09
Basque	−2.704	0.013	−1.841	0.016	−2.455	0.007	−2.637	0.018
C = 1		0.076		0.015		0.006		0.016
C = 10		0.210		0.007		0.002		0.007
U.K.	−3.102	0.003	−1.739	0.021	−2.258	0.045	−2.439	0.021
C = 1		0.034		0.024		0.042		0.019
C = 10		0.097		0.014		0.024		0.009
All Eurasians	−10.664	0.0016	−2.184	0.0015	−4.212	0.0015	−4.161	0.0012
C = 1		0.001		0.001		0.0004		0.0009
C = 10		0.007		0.0002		0.0001		0.000

†First *P* value for each population assumes no recombination. Second and third *P* values assume recombination ($C = 4N_eC$) as indicated.
‡The statistic should be considered as significant at the 5% level if the *P* value is below 0.02.

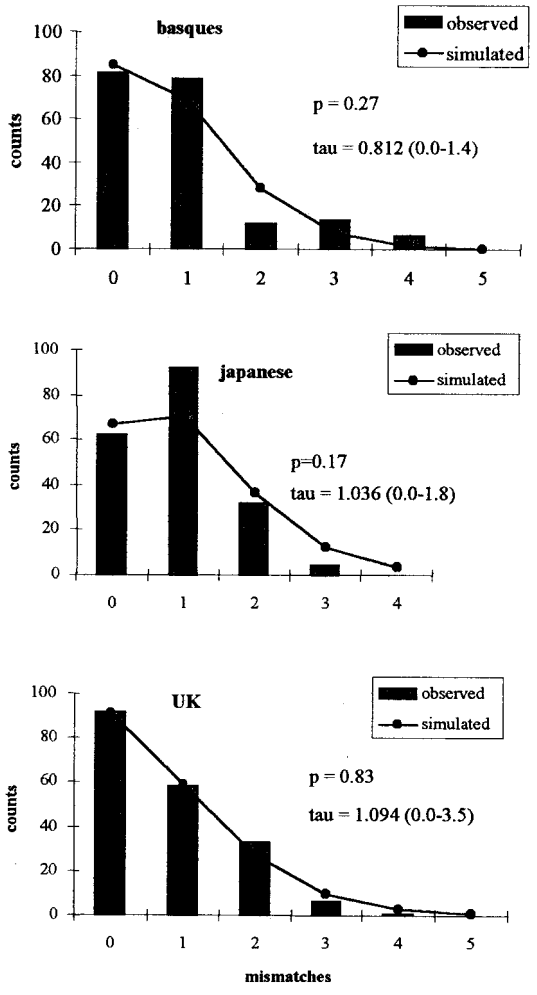

Fig. 3. Mismatch distributions for the Eurasian populations. The *P* value represents the fit to the model of sudden expansion obtained by parametric bootstrap; $\tau = 2\mu t$ (95% confidence interval between parentheses; see *Materials and Methods*).

However, position −8 (a C/T transition 8 bp upstream of the 5′ end of MS205) shows signs of homoplasy, indicated by reticulations in a phylogenetic network (not shown). We favor true homoplasy over any kind of spurious "homoplasy" caused by recombination, because recombination is more likely to involve clusters of homoplasies (38); instead, this position is part of a CpG doublet, which is a well known mutagenic hot spot (by methylation-mediated deamination), which is also polymorphic (C/T) in chimpanzees (39). Gene conversion could also be causing this apparent homoplasy, but if so, it would be preferentially involving the region around −8 (39) into the minisatellite. Pruning of this position and the adjacent 3′ nucleotide position (the last two contiguous positions sequenced) leads to the 1,742 contiguous nucleotides considered for subsequent analyses for all individuals (Table 1). After pruning, there are no incongruent pairs of sites; thus, the minimum number of recombination events [R_M (17)] is 0. Then, the maximum likelihood value of the recombination parameter $C = 4N_e c$ (where c is the recombination rate per generation) is 0 (7, 40). This pruning procedure yields 26 different lineages compatible with an infinite sites model, the genealogical relationships of which are depicted in Fig. 1*a*. For this tree, the inferred root falls within a context of African lineages and is still present in Africa. Assuming that this haplotype truly is the root for the sample, the probability (41) that it is also the ancestral haplotype for the populations analyzed is 0.98.

The coalescent adds a time dimension to the phylogenetic network (tree); thus, assuming neutrality, panmixia, and constancy in population size, the depth of the tree (the time to the most recent common ancestor) is estimated as 0.72 coalescent units, or about 1.04 million years (SEM 0.223 million years). It is not feasible to make an exhaustive exploration of the joint maximum likelihood estimates of parameters such as θ, the array of growth rates (β) for each population, and the matrix of migration rates (m) for each population in all directions that could influence the time to the most recent common ancestor and thus, our estimate should be considered as an approximate time frame for the variability associated with this locus.

Overall, in the African populations, diversity is higher and branches are deeper, whereas in Eurasians, variability seems to have been derived recently from a small subset of African lineages. Contrary to the conclusions of other authors (3–5, 10), we do find evidence of strong population growth for some of the populations, thus reconciling nuclear and mitochondrial inferences. The star-shaped subtree containing both the Euroasiatic variability and some of the African lineages (Fig. 1) immediately suggests significant population growth from a small initial number of lineages (42). Accordingly, for the populations grouped by continent, the frequency spectra show a substantial excess of rare mutations (Fig. 2) compared with the neutral, constant size expectations. This excess is unlikely to be due to sequencing errors because of the robustness of the technique used (see *Materials and Methods*). Furthermore, when establishing the phase of the polymorphisms, resequencing of allele-specific PCR products served as a double check for all initial observations. Finally, neutrality tests in Table 2 show evidence indicating population growth for the Euroasiatic populations. Overall, these tests show negative values, and these results are significant for all tests for the Basques and the U. K. population. Fu's *F*s is also significantly negative for the Japanese. The quick coalescent exploration (see *Materials and Methods*) agreed with this scenario.

Although recombination may decrease the power of neutrality tests, especially *F*s (18), we have argued above that recombination is infrequent enough not to distort the genealogical reconstruction of this region. Under this condition, *F*s has been shown to be considerably more powerful (28) to detect departures from neutrality caused by growth or hitchhiking. The power of this test is correlated also with θ (18, 28); thus, it is likely that the level of polymorphism shown by this locus has provided a good opportunity to detect this pattern. If $\theta = C$, then we would expect in the history of our sample as many recombinant events as segregating sites (17). If, by using the four-gamete test, only approximately 20% (say) of the recombination events are detected (17) for the observed 42 segregating sites, we would expect to detect about eight recombinants. Because we are not detecting any, C must be lower than θ. We have estimated the *P* values of the neutrality tests assuming finite rates of recombination (an additional interesting observation is the opposite effect of recombination on the *P* values of *F*s and the rest of the tests). Thus, we have used a rough upper limit for C of 1, and for comparison we also estimated *P* values assuming a higher value of $C = 10$ (see Table 2). For Europeans, even for $C = 10$, all tests except *F*s still show significant negative values. For $C = 1$ Fs shows *P* values close to α for all Eurasian populations individually; *F*s values are significantly negative when all Eurasian populations are grouped. Overall, this finding indicates that we can be confident that, even assuming undetected recombination, there is a signal of population growth (or genetic hitchhiking) in our data.

On the other hand, we have argued above that a generally high rate of recombination around this region (but not within) may reduce any possible effect of hitchhiking. Therefore, we suggest an explanation for this departure based on population growth in Eurasians.

Given this evidence for population growth in the Euroasiatic populations, mismatch distributions are expected to reflect this process and therefore were used to estimate the time since the expansion. We do not show mismatch distributions for the African populations given the lack of strong signal for population growth in these populations as judged by the neutrality tests applied (note, however, the excess of observed singletons for Africans in Fig. 2). The mismatch distributions for Eurasians (Fig. 3) present strong slopes with peaks at 0–1 differences, indicating a recent origin for this expansion: 106,422 (95% c.i. 0–183,486) years ago for Basques, 143,381 (0–458,715) years for the U.K., and 135,780 (0–253,910) years for the Japanese, respectively. Fine-tuned estimates based on compound haplotypes of a subset of the single nucleotide polymorphisms analyzed here and the diversity accumulated in the linked, highly informative minisatellite MS205 in a larger population sample (39) provide dates for Eurasian-specific lineages that broadly agree with these estimates.

How can this overall pattern be explained? African populations would be expected to show a signature of earlier population growth if we assume the (African) origin for the modern forms of *Homo sapiens* to be a speciation process by cladogenesis within the coalescent time of this sequence region. Although the frequency spectra for Africans shows an excess of the observed number of singletons, suggesting population growth, there is no significant evidence for growth in the two African populations analyzed here. This could be simply a particular characteristic of these populations; alternatively, they could have been growing more slowly, the growth could have been earlier and/or less intense, or this signal may have been overridden by later processes (43). A lack of signal for growth associated to a speciation process could be explained too as speciation by anagenesis, in which physical forms (paleospecies) are generated gradually over time along a single lineage. In any case, historical population numbers (based on θ values) for the African populations analyzed can be considered to be relatively high. Although more African (and other) populations need to be analyzed, in principle, the detected population growth geographically associated with non-African populations would be most likely linked to an out-of-Africa range expansion process. As most of the Euroasiatic variability can be traced back to a single expanding lineage at this locus, the ancestral population that left Africa may have been very small and/or from a geographically localized area.

It is still possible that later migrations also contributed to present-day variability in Europe, as indicated by the presence of a divergent lineage within the Basques (allele F). It is unclear whether this allele represents a later migration (44), a divergent low-frequency allele carried over in the major out-of-Africa migration but sampled only in the Basques, or even a vestige of incomplete population replacement.

The higher substitution rate for this region (and its location in an area of high recombination) may have generated enough variability to recover information on more recent demographic processes. For broadly equivalent effective population sizes, sequenced regions (3–5, 10) with lower evolutionary tempo may not have accumulated enough variability to resolve these processes. In addition, balancing selection (45) may have also played a significant role for some of these regions.

We thank Matthew Stephens and Peter Donnelly for their valuable help. Thanks also to H. Innan and F. Tajima for providing us with PRANC software; to H. J. Bandelt for NETWORK 2.0B, and L. Excoffier for access to the beta version of ARLEQUIN 2.0; to Bob Griffiths and Rosalind Harding for their help with GENETREE; to M. W. Nachman and H. Harpending for providing us with manuscripts before publication; to Emma Rogers for the primate sequences; and to John Brookfield, Paul Sharp, and Jeremy Martinson for their critical reading of the manuscript. We thank Keiji Tamaki, Yoshi Katsumata, and Mark Jobling for sharing DNA samples and Conchi de la Rua, Carmen Manzano, and Neskuts Izagirre for their comments. This work was funded by a grant from the Wellcome Trust (054551).

1. Hudson, R. R. (1990) in *Oxford Surveys in Evolutionary Biology*, eds. Futuyama, D. & Antonovics, J. (Oxford Univ. Press, New York), pp. 1–44.
2. Pääbo, S. (1996) *Am. J. Hum. Genet.* **59,** 493–496.
3. Harris, E. E. & Hey, J. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 3320–3324.
4. Jaruzelska, J., Zietkiewicz, E., Batzer, M., Cole, D. E. C., Moisan, J. P., Scozzari, R., Tavare, S. & Labuda, D. (1999) *Genetics* **152,** 1091–1101.
5. Kaessman, H., Heissig, F., von Haeseler, A. & Pääbo, S. (1999) *Nat. Genet.* **22,** 78–81.
6. Underhill, P. A., Jin, L., Lin, A. A., Mehdi, Q., Jenkins, T., Vollrath, D., Davis, R. W., Cavalli-Sfroza, L. L. & Oefner, P. (1999) *Genome Res.* **7,** 996–1005.
7. Nachman, M. W. & Crowell, S. L. (2000) *Genetics* **155,** 1855–1864.
8. Shen, P., Wang, F., Underhill, P. A., Franco, C., Yang W-H., Roxas, A., Sung, R., Lin, A. A., Hyman R. W., Vollrath, D., *et al.* (2000) *Proc. Natl. Acad. Sci. USA* **97,** 7354–7359.
9. Thomson, R., Pritchard, J. K., Shen, P., Oefner, P. J. & Feldman, M. W. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 7360–7365.
10. Harding, R. M., Fullerton, S. M., Griffiths, R. C., Bond, J., Cox, M. J., Schneider, J. A., Moulin, D. S. & Clegg, J. B. (1997) *Am. J. Hum. Genet.* **60,** 772–789.
11. Clark, A. G., Weiss, K. M., Nickerson, D. A., Taylor, S. L., Buchanan, A., Stengard, J., Salomaa, V., Vartiainen, E., Perola, M., Boerwinkle, E., *et al.* (1998) *Am. J. Hum. Genet.* **63,** 595–612.
12. Rieder, M. J., Taylor, S. L., Clark, A. G. & Nickerson, D. A. (1999) *Nat. Genet.* **22,** 59–62.
13. Ranna, B. K., Hewett-Emmett, D., Jin, L., Chang, B. H.-J., Sambuughin, N., Lin, M., Bamshad, M., Jorde, L. B., Ramsay, M., Jenkins, T. & Li, W.-H. (1999) *Genetics* **151,** 1547–1557.
14. Wall, J. D. & Przeworski, M. (2000) *Genetics* **155,** 1865–1874.
15. Mellars, P. A. (1998) in *Prehistoric Europe: An Illustrated History*, ed. Cunliffe, (Oxford Univ. Press, Oxford), pp. 42–78.
16. Hawks, J., Hunley, K., Lee, S.-H. & Wolpoff, M. (2000) *Mol. Biol. Evol.* **17,** 2–22.
17. Hudson, R. R. & Kaplan, N. L. (1985) *Genetics* **111,** 147–164.
18. Wall, J. D. (1999) *Genet. Res.* **74,** 65–79.
19. Takahata, N. (1995) *Annu. Rev. Ecol. Syst.* **26,** 343–372.
20. Krawczak, M., Ball, E. V. & Cooper, D. N. (1998) *Am. J. Hum. Genet.* **63,** 474–488.
21. Newton, C. R., Graham, A., Heptinstall, L. E., Powell, S. J., Summers, C., Kalsheker, N., Smith, J. C. & Markham, A. F. (1989) *Nucleic Acids Res.* **17,** 2503–2516.
22. Genetics Computer Group (1996) (GCG SEQLAB, Madison, WI).
23. Comeron, J. P. (1995) *J. Mol. Evol.* **41,** 1152–1159.
24. Takahata, N. (1993) *Mol. Biol. Evol.* **10,** 2–22.
25. Nachman, M. W. & Crowell, S. L. (2000) *Genetics* **156,** 297–304.
26. Rozas, J. & Rozas, R. (1999) *Bioinformatics* **15,** 174–175.
27. Tajima, F. (1989) *Genetics* **123,** 585–595.
28. Fu, Y. X. (1997) *Genetics* **147,** 915–925.
29. Fu, Y. X. & Li, W.-H. (1993) *Genetics* **133,** 693–709.
30. Fu, Y. X. (1996) *Genetics* **143,** 557–570.
31. Schneider, S. & Excoffier, L. (1999) *Genetics* **152,** 1079–1089.
32. Bandelt, H. J., Foster, P. & Röhl, A. (1999) *Mol. Biol. Evol.* **16,** 37–48.
33. Innan, H. & Tajima, F. (1997) *Genetics* **147,** 1431–1444.
34. Griffiths, R. C. & Tavaré, S. (1994) *Theor. Popul. Biol.* **46,** 131–159.
35. Przeworski, M., Hudson, R. R. & Di Rienzo, A. (2000) *Trends Genet.* **16,** 296–302.
36. Badge, R. M., Yardley, J., Jeffreys, A. J. & Armour, J. A. L. (2000) *Hum. Mol. Genet.* **9,** 1239–1244.
37. Broman, K. W., Murray, J. C., Steffield, V. C., White, R. L. & Weber, J. L. (1998) *Am. J. Hum. Genet.* **63,** 861–869.
38. Templeton, A., Clark., A. G., Weiss, K. M., Nickerson, D. A., Boerwinkle, E. & Sing, C. F. (2000) *Am. J. Hum. Genet.* **66,** 69–83.
39. Rogers, E. J., Shone, A. C., Alonso, S., May, C. A. & Armour, J. A. L. (2000) *Hum. Mol. Genet.* **9,** 2675–2681.
40. Hey, J. & Wakeley, J. (1997) *Genetics* **145,** 833–846.
41. Waterson, G. A. (1982) *Adv. Appl. Probability* **14,** 206–224.
42. Slatkin, M. & Hudson, R. R. (1991) *Genetics* **129,** 555–562.
43. Excoffier, L. & Schneider, S. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 10597–10602.
44. Jin, L., Underhill, P. A., Doctor, V., Davies, R. W., Shen, P., Cavalli-Sforza, L. L. & Oefner, P. J. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 3796–3800.
45. Harpending, H. & Rogers, A. (2000) *Annu. Rev. Genomics Hum. Genet.* **1,** 361–385.
46. Schneider, S., Roessli, D. & Excoffier, L. (1999) ARLEQUIN 2.0 (Genetics and Biometry Laboratory, University of Geneva, Switzerland).
47. Röhl, A. (1997) NETWORK 2.0 (University of Hamburg, Hamburg, Germany).