

Larger Genetic Differences Within Africans Than Between Africans and Eurasians

Ning Yu,* Feng-Chi Chen,*[†] Satoshi Ota,* Lynn B. Jorde,[†] Pekka Pamilo,[§] Laszlo Patthy,** Michele Ramsay,^{††} Trefor Jenkins,** Song-Kun Shyue^{††} and Wen-Hsiung Li*¹

*Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, [†]Department of Life Science, National Tsing Hua University, Hsinchu, 300 Taiwan, [‡]Department of Human Genetics, University of Utah, Salt Lake City, Utah 84112, [§]Department of Biology, University of Oulu, 90014 Oulu, Finland, **Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, H-1518 Budapest, Hungary, ^{††}Department of Human Genetics, South African Institute for Medical Research and University of the Witwatersrand, Johannesburg, 2050 South Africa and ^{††}Institute of Biomedical Sciences, Academia Sinica, Taipei, 115 Taiwan

Manuscript received August 30, 2001
Accepted for publication February 19, 2002

ABSTRACT

The worldwide pattern of single nucleotide polymorphism (SNP) variation is of great interest to human geneticists, population geneticists, and evolutionists, but remains incompletely understood. We studied the pattern in noncoding regions, because they are less affected by natural selection than are coding regions. Thus, it can reflect better the history of human evolution and can serve as a baseline for understanding the maintenance of SNPs in human populations. We sequenced 50 noncoding DNA segments each ~500 bp long in 10 Africans, 10 Europeans, and 10 Asians. An analysis of the data suggests that the sampling scheme is adequate for our purpose. The average nucleotide diversity (π) for the 50 segments is only 0.061% \pm 0.010% among Asians and 0.064% \pm 0.011% among Europeans but almost twice as high (0.115% \pm 0.016%) among Africans. The African diversity estimate is even higher than that between Africans and Eurasians (0.096% \pm 0.012%). From available data for noncoding autosomal regions (total length = 47,038 bp) and X-linked regions (47,421 bp), we estimated the π -values for autosomal regions to be 0.105, 0.070, 0.069, and 0.097% for Africans, Asians, Europeans, and between Africans and Eurasians, and the corresponding values for X-linked regions to be 0.088, 0.042, 0.053, and 0.082%. Thus, Africans differ from one another slightly more than from Eurasians, and the genetic diversity in Eurasians is largely a subset of that in Africans, supporting the out of Africa model of human evolution. Clearly, one must specify the geographic origins of the individuals sampled when studying π or SNP density.

THERE has been much interest in single nucleotide polymorphisms (SNPs) in human populations because such data are useful for studying human evolution and the mechanism of maintenance of genetic variability in human populations and for identifying genes associated with complex disease (HARDING *et al.* 1997; NICKERSON *et al.* 1998; ZIETKIEWICZ *et al.* 1998; CHARGILL *et al.* 1999; HALUSHKA *et al.* 1999; HARRIS and HEY 1999; JARUZELSKA *et al.* 1999; KAESSMANN *et al.* 1999; RIEDER *et al.* 1999; NACHMAN and CROWELL 2000). The much interest notwithstanding, the worldwide pattern of SNP variation remains incompletely understood. This is especially true for noncoding regions because such regions are medically less interesting. However, data from noncoding regions are less affected by natural selection than data from coding regions and so can reflect more accurately the history of human evolution. Moreover, the pattern of SNP variation in noncoding regions can

serve as a baseline for understanding the maintenance of SNPs in human populations. For these reasons, we have obtained SNP data from 50 noncoding regions in Africans, Europeans, and Asians and have estimated the levels of nucleotide diversity within and between populations. To strengthen our conclusions, we have also used data from GenBank and the literature.

MATERIALS AND METHODS

DNA samples: The 10 Africans used were 1 Biaka Pygmy, 1 Mbuti Pygmy, 1 Ghanaian, 1 Kikuyu, 1 !Kung, 1 Luo, 2 Nigerians (Yuroba and Rivers), 1 South African Bantu speaker, and 1 Zulu (also a South African Bantu speaker); the 10 Europeans were 1 Finnish, 1 French, 1 German, 1 Hungarian, 1 Italian, 1 Portuguese, 1 Russian, 1 Spanish, 1 Swedish, and 1 Ukrainian; and the 10 Asians were 1 Cambodian, 2 Chinese (North and South), 1 Han Taiwanese, 2 Indians (Punjab and Bengal), 1 Japanese, 1 Mongolian, 1 Vietnamese, and 1 Yakut. As every segment studied is autosomal, the number of sequences studied for each segment is 60 (20 for each continent studied).

Selection of DNA segments: Fifty noncoding, nonrepetitive genomic segments (each ~1 kb), which covered almost all autosomes, were selected randomly with reference to the Ge-

¹Corresponding author: Department of Ecology and Evolution, University of Chicago, 1101 E. 57th St., Chicago, IL 60637.
E-mail: whli@uchicago.edu

nome Channel (<http://genome.ornl.gov/GCat/species.shtml>); see CHEN and LI (2001) for details. All of them were chosen to avoid coding or close linkage to any coding regions. In each segment and its nearby regions there was no registered gene in GenBank and no potential coding region was detected by either GenScan or GRAIL-EXP.

PCR amplification and DNA sequencing: Touchdown PCR (DON *et al.* 1991) was used and the reactions were carried out following the conditions described (ZHAO *et al.* 2000). The PCR products were purified by the Wizard PCR Preps DNA purification resin kit (Promega, Madison, WI). Sequencing reaction was performed according to the protocol of ABI Prism BigDye Terminator sequencing kits (Perkin-Elmer, Norwalk, CT) modified by quarter reaction. The extension products were purified by Sephadex G-50 (DNA grade; Pharmacia, Piscataway, NJ) and run on an ABI 377XL DNA sequencer using 4.25% gels (Sooner Scientific). About 500 bp of each segment was sequenced.

ABI DNA Sequence Analysis 3.0 was used for lane tracking and base calling. The data were then proofread manually and heterozygous sites were detected as double peaks. The forward and reverse sequences were assembled automatically in each individual using SeqMan in DNASTAR. The assembled files were carefully checked by eye. Fluorescent traces for each variant site were rechecked again in all individuals. All singletons, which were variants that appear only once in the total sample, were verified by PCR reamplification and resequencing the PCR products in both directions.

Data analysis: The sequences were aligned by SeqMan in the DNASTAR or the DAMBE package (XIA 2000). Nucleotide diversity values were calculated using DNASP (ROZAS and ROZAS 1999), DAMBE, and our own programs.

RESULTS AND DISCUSSION

Distribution of SNPs: A total of 146 SNPs were found in the total sample; 53 of them were observed only once (*i.e.*, singletons) and 22 only twice (doubletons). The number of variant sites found in the African sample was 118, of which 68 (36 singletons, 15 doubletons, and 17 others) were not found in the Eurasian sequences (*i.e.*, they were unique). In contrast, in the Eurasian sample only 78 variant sites were found and only 28 of them (17 singletons, 4 doubletons, and 7 others) were unique, though the combined sample size was twice the African sample size. Thus, beyond the 50 variants already observed in the African sample, the combined Eurasian sample contains in addition only 17 singletons and 11 nonsingleton variants. The high frequencies of singletons in the African and Eurasian samples are similar to those observed in other studies (KAESSMANN *et al.* 1999; ZHAO *et al.* 2000; YU *et al.* 2001). Note that in a neutral Wright-Fisher population with θ , the expected number of mutations of size i in a random sample of n sequences is θ/i (FU 1995). So the number of singletons should be twice the number of doubletons and thrice the number of tripletons. In our total sample we found 53 singletons, 22 doubletons, and 7 tripletons. Therefore, there is an excess of singletons, which suggests a population expansion in the recent past.

Nucleotide diversity: Nucleotide diversity (π) is defined as the number of nucleotide differences between

two randomly chosen sequences in a population. The π -value fluctuates greatly among the segments studied (Table 1). The range of π is from 0 (5 segments) to 0.27% in the total sample, from 0 (5 segments) to 0.58% in the African sample, from 0 (19 segments) to 0.27% in the Asian sample, and from 0 (18 segments) to 0.29% in the European sample. Such large fluctuations are not surprising because the nucleotide diversity in a short DNA region is subject to strong stochastic effects. In addition, variation in π may also arise from different mutation rates among different segments, although we found no correlation between π and the divergence between human and ape sequences. The average π -values are only 0.061% \pm 0.010% among Asians and 0.064% \pm 0.011% among Europeans, but almost twice as high among Africans (0.115% \pm 0.016%); π is 0.088% \pm 0.011% for the total sample. The average π -value within Africans is actually somewhat higher than that between Africans and Eurasians (0.096% \pm 0.012%). In other words, Africans differ on average more among themselves than from Eurasians.

Adequacy of sampling scheme and sample size: We now consider the adequacy of our sampling schedule. The fact that the individuals used were chosen to cover various geographic areas and ethnic backgrounds in each of the three continents studied may tend to overestimate π , whereas the inclusion of the two sequences within each individual may tend to underestimate π . The two tendencies should be reflected in between-individual π -values (π_b) and in within-individual π -values (π_w), respectively, and the average π_b and π_w can be taken as an upper and a lower bound of the true π -value. To simplify the analysis, we concatenate the segments in an individual in a random manner into two continuous sequences. For the African sequences the distribution of π_b -values, which ranges from 0.059 to 0.187%, is only somewhat wider than that of the 10 π_w -values, which ranges from 0.059 to 0.152%. Therefore, the average π_b (0.115%) is only slightly higher than the average π_w (0.108%), implying that the geographic locations of the individuals sampled have little effect on the average π -value. For the Asian sample there are two very low π_w -values (0.023 for the North Chinese and the Bengal) and the average π_w -value (0.051%) is substantially lower than the average π_b (0.063%). The average π_w without the two outliers becomes 0.058%, which is similar to the between-individual average π_b (0.063%). So, our sampling schedule in Asia may cause at most only a minor overestimate of the true average π -value. For the European sample, the average π_w -value is only 0.049% and, after excluding the two lowest values (0.027% for the Ukrainian and 0.031% for the Russian), it becomes 0.054%, which is still not close to the between-individual average of 0.066%. This comparison suggests that our sampling schedule may have inflated somewhat the average π -value for the Europeans. However, as the final estimate of 0.063% should have been

TABLE 1
Nucleotide diversity in each segment within and between continents

Region (segment length, bp)	Nucleotide diversity (%)						
	All	Afr	Asia	Eur	Afr/Asia	Afr/Eur	Asia/Eur
NT2041 (461)	0	0	0	0	0	0	0
NA1364 (417)	0	0	0	0	0	0	0
NT0953 (644)	0	0	0	0	0	0	0
NT2012 (539)	0	0	0	0	0	0	0
NT2064 (522)	0	0	0	0	0	0	0
NT1584 (563)	0.006	0.018	0	0	0.009	0.009	0
NT2563 (526)	0.012	0.036	0	0	0.019	0.019	0
NT0946 (450)	0.015	0.042	0	0	0.022	0.022	0
NT2191 (415)	0.016	0.046	0	0	0.023	0.023	0
NT2659 (497)	0.020	0.040	0.020	0	0.030	0.020	0.010
NT1419 (485)	0.020	0.060	0	0	0.030	0.030	0
NT1506 (476)	0.020	0	0.021	0.040	0.011	0.020	0.030
NT24894 (476)	0.021	0.063	0	0	0.032	0.032	0
NT2472 (456)	0.021	0.059	0	0	0.031	0.031	0
NT2984 (478)	0.034	0.098	0	0	0.051	0.051	0
NT2609 (452)	0.036	0.066	0	0.042	0.033	0.054	0.022
NT812 (960)	0.038	0.041	0.038	0.038	0.039	0.039	0.038
NT2906 (485)	0.040	0.082	0	0.017	0.051	0.056	0.009
NT2986 (492)	0.043	0.020	0.055	0.043	0.038	0.037	0.045
NT2265 (519)	0.044	0.075	0.039	0.019	0.056	0.048	0.029
NT866 (493)	0.046	0.020	0.059	0.054	0.040	0.039	0.057
NT2019 (462)	0.047	0.041	0.063	0.041	0.051	0.040	0.051
NT2266 (477)	0.056	0.127	0	0	0.079	0.079	0
NT2018 (548)	0.063	0.099	0.018	0.049	0.066	0.081	0.035
NT1482 (479)	0.064	0.110	0.042	0.04	0.078	0.075	0.040
NT2963 (541)	0.068	0.118	0	0.081	0.062	0.098	0.043
NT2020 (488)	0.081	0.090	0.081	0.081	0.084	0.084	0.079
NT2568 (444)	0.091	0.060	0.154	0.042	0.115	0.054	0.100
NT813 (528)	0.094	0.010	0.100	0.075	0.097	0.091	0.091
NT1469 (468)	0.099	0.040	0.094	0.112	0.077	0.103	0.105
NT2294 (748)	0.106	0.115	0.154	0.051	0.135	0.082	0.103
NT10604 (511)	0.109	0.118	0.097	0.102	0.108	0.110	0.106
NT787 (442)	0.114	0.076	0.118	0.100	0.102	0.115	0.113
NT2560 (507)	0.114	0.107	0.053	0.103	0.090	0.142	0.092
NT2987 (465)	0.116	0.106	0.103	0.095	0.120	0.121	0.097
NT0151 (426)	0.121	0.103	0.104	0.112	0.102	0.129	0.12
NT2085 (577)	0.122	0.263	0	0	0.171	0.171	0
NT2352 (550)	0.123	0.226	0.061	0.067	0.15	0.153	0.063
NT1251 (529)	0.142	0.162	0	0.254	0.083	0.206	0.135
NT2021 (485)	0.145	0.153	0.148	0.125	0.153	0.146	0.135
NT2558 (509)	0.151	0.236	0.155	0.037	0.196	0.143	0.122
NT1386 (418)	0.157	0.292	0.045	0.106	0.183	0.202	0.079
NT1412 (555)	0.179	0.202	0.123	0.199	0.171	0.204	0.159
NA2920 (508)	0.193	0.219	0.174	0	0.221	0.129	0.184
NT784 (541)	0.203	0.110	0.128	0.161	0.217	0.184	0.157
NT2557 (431)	0.212	0.305	0.067	0.195	0.222	0.255	0.141
NT864 (493)	0.212	0.245	0.216	0.158	0.237	0.205	0.190
NT2988 (540)	0.248	0.301	0.194	0.246	0.252	0.268	0.223
NT1636 (506)	0.250	0.582	0.037	0	0.355	0.345	0.019
NT2924 (525)	0.266	0.251	0.274	0.289	0.260	0.268	0.275
Average	0.088	0.115	0.061	0.064	0.095	0.096	0.066
SE	0.011	0.016	0.010	0.011	0.012	0.012	0.010

Afr, Africa; Eur, Europe.

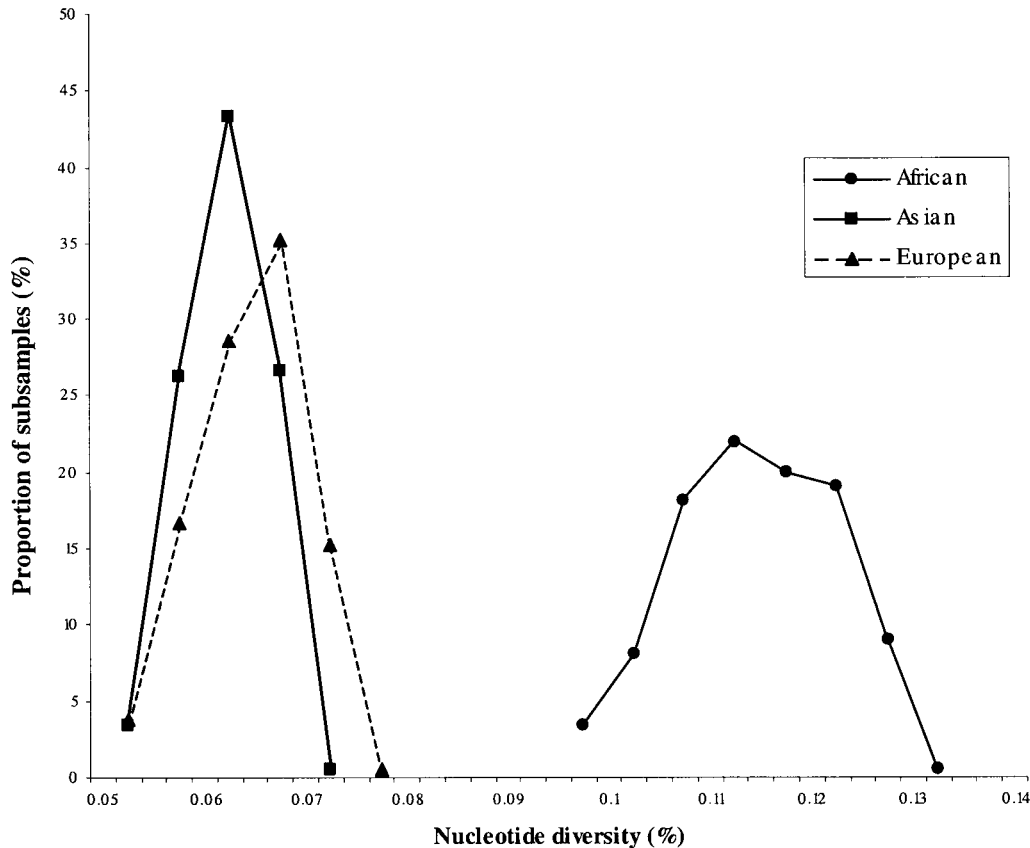


FIGURE 1.—Distributions of average π -values in subsamples of 6 individuals from the sample of 10 individuals in each continent studied. The distribution is based on 1000 replicates.

compensated to some extent by the π_w -values, which tend to be low, it should not differ much from the true value.

Next let us consider whether the sample size of 10 individuals from each continent studied is sufficiently large for our purpose. To answer this question, let us consider subsamples of the original samples; we consider concatenated sequences. Figure 1 shows the distribution of the average π -values when each subsample in a continent contains only 6 individuals from the original sample of 10; there are 210 such possible subsamples. It is seen that the three distributions are rather concentrated. For example, in each distribution none of the average π -values deviate significantly from the mean π -value of the total sample and the proportions of average π -values in subsamples that deviate more than 10% from the mean are 18.1, 7.6, and 20.4% for the African, Asian, and European samples, respectively. This analysis suggests that even a sample of 6 independent individuals would usually give an estimate of π in a continent reasonably close to the true value.

Worldwide pattern of SNP variation and implications for human evolution: A more general pattern of worldwide nucleotide diversity is shown in Table 2, which includes the present data and data from the literature and our unpublished studies. For the autosomal regions included (47,038 bp) the π -values are 0.105% for Africans,

0.070% for Asians, 0.069% for Europeans, and 0.097% for between Africans and Eurasians. For X-linked regions (47,421 bp) the corresponding values are 0.088, 0.042, 0.053, and 0.082%. As in many previous studies (VIGILANT *et al.* 1991; HAMMER *et al.* 1997; NICKERSON *et al.* 1998; HALUSHKA *et al.* 1999; KAESSMANN *et al.* 1999; INGMAN *et al.* 2000; JORDE *et al.* 2000; UNDERHILL *et al.* 2000; ZHAO *et al.* 2000; YU *et al.* 2001), the nucleotide diversity is considerably higher in Africans than in non-Africans, probably because of a larger effective population size or longer evolutionary history (STONEKING *et al.* 1997; INGMAN *et al.* 2000). Asians and Europeans have very similar π -values and are considerably closer to each other than to Africans.

The genome-wide nucleotide diversity has been estimated to be 0.075% for the National Institutes of Health (NIH) diversity panel (INTERNATIONAL SNP WORKING GROUP 2001; see also VENTER *et al.* 2001). The NIH panel consists of 90 individuals from European-Americans, African-Americans, Hispanic Americans, Native Americans, and Asian Americans, while our estimate of 0.089% is from a worldwide sample (Table 2) and is slightly higher. The SNP density is 1 SNP/1331 bp for the former data and 1 SNP/1123 bp for our data. For X-linked sequences the SNP density is 1 SNP/1408 bp (Table 2). It is clear from Table 2 that the geographic origins of the individuals sampled should be specified

TABLE 2
Nucleotide diversity within and between populations in noncoding regions

Regions	Type of sequence	Total no. of sequences	Length (bp)	Nucleotide diversity (%)							References
				Total	Afr	Asia	Eur	Afr/Asia	Afr/Eur	Asia/Eur	
Autosomal											
50 segments	Noncoding	60	25,607	0.088	0.115	0.061	0.064	0.095	0.096	0.066	
1q24	Introns	122	9,221	0.057	0.074	0.046	0.044	0.067	0.064	0.045	YU <i>et al.</i> (2001)
22q11	Noncoding	128	9,834	0.088	0.085	0.075	0.077	0.083	0.108	0.091	ZHAO <i>et al.</i> (2000)
β-Globin	5' flanking	349	814	0.318	0.323	0.331	0.232	0.393	0.309	0.331	HARDING <i>et al.</i> (1997)
	Introns	349	980	0.140	0.072	0.142	0.116	0.171	0.096	0.182	HARDING <i>et al.</i> (1997)
	3' flanking	349	582	0.264	0.217	0.273	0.255	0.304	0.242	0.319	HARDING <i>et al.</i> (1997)
Total			47,038	0.089	0.105	0.070	0.069	0.096	0.098	0.077	
X-linked											
Xq21	Noncoding	63	10,158	0.072	0.071	0.073	0.070	0.073	0.071	0.072	Our unpublished data
Dmd9	Intron	35	10,835	0.040	0.076	0.009	0.036	0.042	0.053	0.024	Our unpublished data
Dmd7	Intron	31	2,383	0.037	0.080	0	0.008	0.048	0.054	0.004	NACHMAN and CROWELL (2000)
Dmd44	Intron	31	2,984	0.150	0.174	0.101	0.144	0.139	0.173	0.132	NACHMAN and CROWELL (2000)
Xq13.3	Noncoding	69	9,564	0.034	0.035	0.025	0.034	0.038	0.034	0.035	KAESSMANN <i>et al.</i> (1999)
PDHA1	Introns	35	3,530	0.169	0.216	0	0.009	0.248	0.251	0.005	HARRIS and HEY (1999)
Dys44	Introns	847	7,185	0.102	0.109	0.085	0.071	0.107	0.110	0.084	ZIETKIEWICZ <i>et al.</i> (1998)
ZFX	Intron	336	782	0.021	0.053	0.004	0	0.030	0.028	0.002	JARUZELSKA <i>et al.</i> (1999)
Total			47,421	0.071	0.088	0.042	0.053	0.079	0.084	0.050	

when considering nucleotide diversity or SNP density. For example, the SNP density for autosomal regions is 1 SNP/952 bp for Africans but only 1 SNP/1430 bp for Asians and Europeans.

Interestingly, both autosomal and X-linked sequence data show higher DNA variation within Africans than between Africans and Eurasians (Table 2), contrary to the general observation of lower within-population than between-population differences in population genetics. This finding implies that Africans differ on average more among themselves than from Eurasians. Thus, with the exception of many minor unique variants, the nucleotide diversity in Eurasians is essentially a subset of that in Africans, as suggested by the observation that both Y-linked and autosomal haplotypes found outside of Africa were often a subset of the collection of haplotypes found in Africa (ARMOUR *et al.* 1996; TISHKOFF *et al.* 1996, 2000; HAMMER *et al.* 1997; UNDERHILL *et al.* 2000). Our finding is more in agreement with the out of Africa model of human evolution than with the multi-regional model because it is consistent with the view that modern humans originated in Africa and that a smaller subset of this population later migrated to other parts of the world (see STONEKING *et al.* 1997 and references therein). During and after the migration some variants would have been lost and, as the separation time is still short, non-Africans have not yet acquired many high-frequency variants, though they might have derived some variants from indigenous archaic popula-

tions in Asia and Europe. For these reasons, the genetic differences between non-Africans and Africans are on average smaller than the genetic differences within Africans.

This work was supported by National Institutes of Health grants GM55759, HD38287, GM30998, and GM59290.

LITERATURE CITED

- ARMOUR, J. A. L., T. ANTTINEN, C. A. MAY, E. E. VEGA, A. SAJANTILA *et al.*, 1996 Minisatellite diversity supports a recent African origin for modern humans. *Nat. Genet.* **13**: 154–160.
- CARGILL, M., D. ALTSHULER, J. IRELAND, P. SKLAR, K. ARDLIE *et al.*, 1999 Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.* **22**: 231–238.
- CHEN, F.-C., and W.-H. LI, 2001 Genomic divergences between human and other hominoids and the effective population size of the common ancestor of human and chimpanzee. *Am. J. Hum. Genet.* **68**: 444–456.
- DON, R. H., P. T. COX, B. J. WAINWRIGHT, K. BAKER and J. S. MATTICK, 1991 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.* **19**: 4008.
- FU, Y. X., 1995 Statistical properties of segregating sites. *Theor. Popul. Biol.* **48**: 172–197.
- HALUSHKA, M. K., J. B. FAN, K. BENTLEY, L. HSIE, N. SHEN *et al.*, 1999 Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat. Genet.* **22**: 239–247.
- HAMMER, M. F., A. B. SPURDLE, T. KARAFET, M. R. BONNER, E. T. WOOD *et al.*, 1997 The geographic distribution of Y chromosome variation. *Genetics* **145**: 787–805.
- HARDING, R. M., S. M. FULLERTON, R. C. GRIFFITHS, J. BOND, M. J. COX *et al.*, 1997 Archaic African and Asian lineages in the genetic ancestry of modern humans. *Am. J. Hum. Genet.* **60**: 772–789.
- HARRIS, E. E., and J. HEY, 1999 X chromosome evidence for ancient human histories. *Proc. Natl. Acad. Sci. USA* **96**: 3320–3324.

- INGMAN, M., H. KAESMANN, S. PÄÄBO and U. GYLLENSTEN, 2000 Mitochondrial genome variation and the origin of modern humans. *Nature* **408**: 708–713.
- INTERNATIONAL SNP WORKING GROUP, 2001 A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**: 928–933.
- JARUZELSKA, J., E. ZIETKIEWICZ, M. BATZER, D. E. C. COLE, J.-P. MOISAN *et al.*, 1999 Spatial and temporal distribution of the neutral polymorphisms in the last ZFX intron: analysis of the haplotype structure and genealogy. *Genetics* **152**: 1091–1101.
- JORDE, L. B., W. S. WATKINS, M. J. BAMSHAD, M. E. DIXON, C. E. RICKER *et al.*, 2000 The distribution of human genetic diversity: a comparison of *mitochondrial*, autosomal, and Y chromosome data. *Am. J. Hum. Genet.* **66**: 979–988.
- KAESMANN, H., F. HEIßIG, A. VON HAESLER and S. PÄÄBO, 1999 DNA sequence variation in a non-coding region of low recombination on the human X chromosome. *Nat. Genet.* **22**: 78–81.
- NACHMAN, M. W., and S. L. CROWELL, 2000 Contrasting evolutionary histories of two introns of Duchenne muscular dystrophy gene, *Dmd*, in humans. *Genetics* **155**: 1855–1864.
- NICKERSON, D. A., S. L. TAYLOR, K. M. WEISS, A. G. CLARK, R. G. HUTCHINSON *et al.*, 1998 DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nat. Genet.* **19**: 233–240.
- RIEDER, M. J., S. L. TAYLOR, A. G. CLARK and D. A. NICKERSON, 1999 Sequence variation in the human angiotensin converting enzyme. *Nat. Genet.* **22**: 59–62.
- ROZAS, J., and R. ROZAS, 1999 DnaSP version 3.0: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**: 174–175.
- STONEKING, M., J. J. FONTIUS, S. L. CLIFFORD, H. SOODYALL, S. S. ARCOT *et al.*, 1997 Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Res.* **7**: 1061–1071.
- TISHKOFF, S. A., E. DIETZSCH, W. SPEED, A. J. PAKSTIS, J. R. KIDD *et al.*, 1996 Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* **271**: 1380–1387.
- TISHKOFF, S. A., A. J. PAKSTIS, M. STONEKING, J. R. KIDD, G. DESTRO-BISOL *et al.*, 2000 Short tandem-repeat Polymorphism/Alu haplotype variation at the PLAT locus: implications for modern human origins. *Am. J. Hum. Genet.* **67**: 901–925.
- UNDERHILL, P. A., P. SHEN, A. A. LIN, L. JIN, G. PASSARINO *et al.*, 2000 Y chromosome sequence variation and the history of human populations. *Nat. Genet.* **26**: 358–361.
- VENTER, J. C., M. D. ADAMS, E. W. MYERS, P. W. LI, R. J. MURAL *et al.*, 2001 The sequence of the human genome. *Science* **291**: 1304–1351.
- VIGILANT, L., M. STONEKING, H. HARPENDING, K. HAWKES and A. C. WILSON, 1991 African populations and the evolution of human mitochondrial DNA. *Science* **253**: 1503–1507.
- XIA, X., 2000 *Data Analysis in Molecular Biology and Evolution*. Kluwer Academic Publishers, Boston.
- YU, N., Z. ZHAO, Y.-X. FU, N. SAMBUUGHIN, M. RAMSAY *et al.*, 2001 Global patterns of human DNA sequence variation in a 10-kb region on chromosome 1. *Mol. Biol. Evol.* **18**: 214–222.
- ZHAO, Z., L. JIN, Y.-X. FU, M. RAMSAY, T. JENKINS *et al.*, 2000 World-wide DNA sequence variation in a 10-kilobase noncoding region on human chromosome 22. *Proc. Natl. Acad. Sci. USA* **97**: 11354–11358.
- ZIETKIEWICZ, E., V. YOTOVA, M. JARNIK, M. KORAB-LASKOWSKA, K. K. KIDD *et al.*, 1998 Nuclear DNA diversity in worldwide distributed human populations. *Gene* **205**: 161–171.

Communicating editor: Y.-X. FU