Am. J. Hum. Genet. 71:528–542, 2002

DNA Sequence Variation in a 3.7-kb Noncoding Sequence 5′ of the *CYP1A2* Gene: Implications for Human Population History and Natural Selection

S. P. Wooding,[1] W. S. Watkins,[1] M. J. Bamshad,[2] D. M. Dunn,[1] R. B. Weiss,[1] and L. B. Jorde[1]

Departments of [1]Human Genetics and [2]Pediatrics, University of Utah, Salt Lake City

***CYP1A2* is a cytochrome P450 gene that is involved in human physiological responses to a variety of drugs and toxins. To investigate the role of population history and natural selection in shaping genetic diversity in *CYP1A2*, we sequenced a 3.7-kb region 5′ from *CYP1A2* in a diverse collection of 113 individuals from three major continental regions of the Old World (Africa, Asia, and Europe). We also examined sequences in the 90-member National Institutes of Health DNA Polymorphism Discovery Resource (PDR). Eighteen single-nucleotide polymorphisms (SNPs) were found. Most of the high-frequency SNPs found in the Old World sample were also found in the PDR sample. However, six SNPs were detected in the Old World sample but not in the PDR sample, and two SNPs found in the PDR sample were not found in the Old World sample. Most pairs of SNPs were in complete linkage disequilibrium with one another, and there was no indication of a decline of disequilibrium with physical distance in this region. The average ± SD nucleotide diversity in the Old World sample was 0.00043 ± 0.00026. The African population had the highest level of nucleotide diversity and the lowest level of linkage disequilibrium. Two distinct haplotype clusters with broadly overlapping geographical distributions were present. Of the 17 haplotypes found in the Old World sample, 12 were found in the African sample, 8 were found in Indians, 5 were found in non-Indian Asians, and 5 were found in Europeans. Haplotypes found outside Africa were mostly a subset of those found within Africa. These patterns are all consistent with an African origin of modern humans. Seven SNPs were singletons, and the site-frequency spectrum showed a significant departure from neutral expectations, suggesting population expansion and/or natural selection. Comparison with outgroup species showed that four derived SNPs have achieved high (>0.90) frequencies in human populations, a trend consistent with the action of positive natural selection. These patterns have a number of implications for disease-association studies in *CYP1A2* and other genes.**

Introduction

As the Human Genome Project nears its goal of a completed human DNA sequence, much is being learned about both the structure and composition of our genome and the potential for the pinpointing of disease-causing genes. The full significance of the Human Genome Project will be better realized as multiple genomes are sequenced to assess patterns of variation in human populations. This variation has important implications for association-based gene-mapping efforts (Terwilliger and Göring 2000; Jorde et al. 2001), for accurate forensic analysis (Evett and Weir 1998), and for a more thorough understanding of our evolutionary history (Cavalli-Sforza et al. 1994; Harpending et al. 1998; Jorde et al. 1998; Mountain 1998; Owens and King 1999; Harpending and Rogers 2000).

To date, most studies of human genetic variation have focused on mtDNA variation (Vigilant et al. 1991; Ingman et al. 2000); Y-chromosome variation (Hammer et al. 1998; Forster et al. 2000; Underhill et al. 2000); or specific classes of autosomal polymorphisms, such as RFLPs (Bowcock et al. 1991), STRPs (Bowcock et al. 1994; Jorde et al. 1997; Perez-Lezaun et al. 1997; Deka et al. 1999), and *Alu* insertion/deletion polymorphisms (Stoneking 1997; Watkins et al. 2001). These systems have yielded detailed portraits of human genetic variation that emphasize the extent to which ancient population history and natural selection have shaped modern human variability.

Analyses of DNA sequence variation in specific genomic regions have offered particular insight into the relationship between human evolution, linkage disequilibrium patterns, and the potential for mapping disease-causing genes (Cargill et al. 1999; Jorde et al. 2001; Pritchard and Przeworski 2001; Reich et al. 2001; Stephens et al. 2001; Ardlie et al. 2002; Gabriel et al. 2002; Nordberg and Tavaré 2002). However, the

Received June 4, 2002; accepted for publication June 10, 2002; electronically published August 9, 2002.

Address for correspondence and reprints: Dr. L. B. Jorde, Department of Human Genetics, University of Utah, 15 North 2030 East, Salt Lake City, UT 84112-5330. E-mail: lbj@genetics.utah.edu

0002-9297/2002/7103-0008$15.00

number of sequenced regions remains small. In some cases, the sample of sequenced individuals has been limited in number and ethnic diversity. In addition, some analyses have been based on the typing of SNPs that have previously been identified in a small initial sample, a strategy likely to miss some rare polymorphisms (Nickerson et al. 2000). Many questions remain about the roles that population history and natural selection have in the shaping of diversity in human genes.

Here we report an analysis of DNA sequence variation in a noncoding 3.7-kb region immediately 5′ of *CYP1A2,* which is a member of the cytochrome P450 gene family (MIM 124060). The gene is located on 15q22, approximately midway between the centromere and telomere. The surrounding region is relatively gene rich (76 genes in the interval spanning 67–75 Mb), and two other cytochrome P450 genes are located nearby (namely *CYP1A1* [60 kb away], a tandem duplication of the ancestral gene, and *CYP11A* [635 kb away], which is unrelated and involved in steroid metabolism). *CYP1A2* encodes an enzyme that carries out the oxidative metabolism of many toxicologically significant compounds, including carcinogenic arylamines, acetaminophen, and a number of widely prescribed antipsychotic drugs. Constitutive *CYP1A2* mRNA expression levels in human liver samples can vary by as much as 15-fold (Ikeya et al. 1989; Schweikl et al. 1993), and it has been speculated that genetic differences in constitutive and/or inducible *CYP1A2* expression may play an important role in individual variation in cancer susceptibility, responses to environmental toxins, and the effectiveness of prescribed medications.

Studies of genetic variation in *CYP1A2* have identified a number of polymorphisms in intron and 5′ regulatory regions that might covary with *CYP1A2* expression and activity (Nakajima et al. 1999; Sachse et al. 1999; Aitchison et al. 2000). However, there is continued dispute about the functional significance of these variants because results have been difficult to replicate in different study populations (e.g., see Basile et al. 2000; Schulze et al. 2002). Evidence that *CYP1A2* allele frequencies vary both within and among ethnic groups suggests that long-term historical and selective factors may underlie some aspects of this variation (Aitchison et al. 2000; Basile et al. 2000).

To assess the effects that population history and natural selection have on genetic variation in *CYP1A2,* we examined patterns of DNA sequence variation in the 5′ regulatory region. Sequences were obtained from 113 individuals in 31 populations spread across Africa, Asia, and Europe. These sequences were tested for evidence of natural selection and of population growth. In addition, diversity patterns in the 5′ region of *CYP1A2* were compared with those observed in several other loci. To gain insight into the usefulness that the National Institutes of Health DNA Polymorphism Discovery Resource (PDR) has in the study of *CYP1A2* and other genes, we compared diversity found in the sample of Africans, Asians, and Europeans with that found in the 90-member PDR sample (Collins et al. 1998).

Subjects and Methods

Subjects

DNA sequence was obtained from 113 individuals, together referred to herein as the "Old World sample": 33 Africans (3 Biaka Pygmies, 7 Mbuti Pygmies, 3 Alur, 3 Nande, 3 Hema, 2 Kenyans, 3 Nigerians, 3 Sotho, 3 Nguni, and 3 San), 25 non-Indian Asians (hereafter termed "Asians") (3 Cambodians, 5 Han Chinese, 5 Japanese, 3 aboriginal Malayans, 3 Vietnamese, 2 Mongolians, and 4 Daghestanis), 27 Europeans (5 northern Europeans, 5 French, 5 Italians, 5 Poles, and 7 Finns), and 28 South Indians (3 Brahmin, 3 Kshatriya, 3 Yadava, 3 Mala, 3 Madiga, 3 Relli, 3 Irula, 3 Khondadora, 1 Santal, and 3 Chenchu). The Khondadora, Santal, and Chenchu individuals are members of South Indian "tribal" populations, whereas the other Indians are members of South Indian caste populations. Informed consent was obtained from all subjects whose blood was drawn either at the University of Utah or in India.

Sequencing was also performed on the 90-member PDR sample (Collins et al. 1998). This collection of individuals consists of European Americans, African Americans, Hispanic Americans, Native Americans, and Asian Americans. DNA sequencing was also performed for three ape species: five common chimpanzees (*Pan troglodytes*), one bonobo (*P. paniscus*), and one gorilla (*Gorilla gorilla*).

Laboratory Analysis

PCR primer sequences (5′-AACAGGGACTTCTTGGATGC-3′ and 5′-TGTACCAAAGAGTCCCTGCC-3′) were derived from the human *CYP1A2* genomic sequence (GenBank accession U02993), spanning 0.5 kb downstream and 3.2 kb upstream from the *CYP1A2* promoter. PCR amplification was performed in 50-μl reaction volumes by the Expand Long Template PCR System (Roche). Each reaction contained 50 ng genomic DNA, 350 μM deoxyribonucleoside triphosphates (dNTPs), 0.2 μM each PCR primer, 1 × reaction buffer 2, and 2.6 U *Taq*/*Pwo* polymerase mix. Cycling conditions included an initial denaturation at 94°C for 2 min; 10 cycles at 94°C for 10 s, 55°C for 10 s, 68°C for 2 min; and 15 cycles at 94°C for 10 s, 55°C for 10 s, 68°C for 2 min 20 s. Residual primers and dNTPs were removed from PCR products with a Millipore

glass fiber filter. The sequence-ready templates were eluted in 70 μl of sterile H_2O. Five microliters of each template was aliquoted to 12 wells of a 384-well sequence dish and was evaporated to dryness in a speed-vac.

Internal primers used for sequencing included the following: 5′-GGATGCTTATGATGTCTCTTG-3′, 5′-GTGCGTGTCAGGTCTCTTCA-3′, 5′-AGAAGGAGCGTAATCCCCC-3′, 5′-AACCTGTGAAGATGCCAAGG-3′, 5′-ACCGAGCCTAACCTCAAACC-3′, 5′-TGTAGAGAGGAGGTCTTG-3′, 5′-TTGTGGTCCCAGCTACTC-3′, 5′-GTTTATCCTTGCTTGAGGG-3′, 5′-CCCCTCAAGCAAGGATAAAC-3′, 5′-GGGACATGTAAGCACGGACT-3′, 5′-GGGGATCATGACACTTCCAT-3′, 5′-ATCAGATTGGCCTGGTTGTC-3′, and 5′-CATGTACCTTCATCCCCAGG-3′. Cycle sequencing was performed in 5-μl reaction volumes by use of ABI BigDye Terminator chemistry. Cycling conditions included an initial denaturation at 96°C for 30 s; followed by 46 cycles of 96°C for 10 s, 50°C for 5 s, and 60°C for 4 min. On completion of cycle sequencing, 20 μl of 62.5% ethyl alcohol and 1 M potassium acetate (with pH 4.5) was added to each reaction, and the sequence plates were centrifuged at 4,000 rpm at 4°C for 45 min in an Eppendorf centrifuge. The samples were resuspended in 15 μl of sterile H_2O and were electrophoresed on an ABI 3700 DNA analyzer prepared with POP-5 capillary gel matrix (ABI).

Sequence-trace files were evaluated using the Phred, Phrap, and Consed programs (Ewing et al. 1998). Potential heterozygotes were identified using PolyPhred, version 3.5 (Nickerson et al. 1997). Polymorphisms were verified by manual evaluation of the individual sequence traces. For most polymorphisms, it was possible to evaluate both the forward and reverse sequences. Detailed information about the SNPs typed in the PDR sample is available at the Web sites dbSNP Home Page and GeneSNPs Public Internet Resource.

Statistical Analysis

Allele frequencies for each SNP were determined by gene counting, and the significance of deviations from Hardy-Weinberg equilibrium was tested using the random-permutation procedure implemented in the Arlequin package (Schneider et al. 2000). Haplotype frequencies were estimated separately for each major population (Africans, Indians, Asians, and Europeans) by the expectation-maximization (EM) algorithm provided in the Arlequin package. This procedure has been shown to yield reliable estimates of haplotype frequencies (Tishkoff et al. 2000). In particular, the EM approach should work well in small regions that have both a high degree of linkage disequilibrium and a high proportion of homozygotes. The data analyzed here meet both of these conditions. Unambiguous haplotypes were inferred directly from genotypes in which only zero or one sites were polymorphic. The inferred haplotype data were used to estimate the D' linkage-disequilibrium statistic (Lewontin 1964) for all pairs of SNPs.

Haplotypes were used to perform phylogenetic analyses, with the common chimpanzee as an outgroup species. A minimum-spanning tree (MST) relating haplotypes was constructed using the Arlequin package. Trees were also constructed using the neighbor-joining, parsimony, and maximum-likelihood methods implemented in the PHYLIP software package (Felsenstein 1993).

The recombination rate for the genomic region that includes *CYP1A2* was estimated by comparison of genetic and physical distances between polymorphic STR markers in a 5-Mb region flanking the gene. The genetic and physical distances between three independent marker pairs were determined using Marshfield and Généthon maps, from NCBI Map Viewier (build 29).

The nucleotide-substitution rate of the *CYP1A2* 5′ region was calculated in two steps. First, the expected coalescence time of human and chimpanzee lineages was calculated. The expected coalescence time of human and chimpanzee lineages is equal to the sum of the time since human/chimpanzee speciation and the expected coalescence time in the population ancestral to humans and chimpanzees. The time since human/chimpanzee speciation is often estimated at ~5 million years, and the coalescence time in the population ancestral to humans and chimpanzees has been estimated at 4 million years (Takahata et al. 1995; Sherry et al. 1997). Second, the mean nucleotide difference between humans and chimpanzees was divided by twice the expected coalescence time, to yield a nucleotide-substitution rate measured in substitutions per year. To provide a basis for comparison with previous studies, we estimated the nucleotide-substitution rate by calculating the mean nucleotide difference between human and chimpanzee haplotypes and dividing by twice the estimated time of speciation, 5 million years before the present.

Average nucleotide diversity (π) was assessed in the PDR sample, the Old World sample, and each major population (Africans, Indians, Asians, and Europeans). Because of their distinct geographic location and history, the samples from the Indian subcontinent were treated separately from those from the rest of Asia. Genetic diversity under a mutation-drift equilibrium model (θ) was assessed using the methods of Tajima (1983) and Watterson (1975). These values were in turn used to estimate effective population size, N_e, using the standard equation, $\theta = 4N_e\mu$. The μ value is estimated as νgL, where g is generation length, L is the length of the nucleotide sequence (3,669 bp), and ν is the mutation rate per nucleotide per year. We assume a generation time of 20

Table 1

Haplotypes and Their Estimated Frequencies in Each Population

	Nucleotide Position																			Frequency in				
	−																	+						
Haplotype	2 9 9 3	2 9 9 2	2 9 7 3	2 7 6 0	2 7 8 8	2 6 2 1	2 3 0 7	2 2 9 5	2 0 3 5	2 9 1 2	1 8 1 0	1 9 5 0	2 8 2 5	1 9 0 9	8 8 0 8	1 0 0 0	1 1 6 2	2 2 7 0	3 9 9 4	Africans	Asians	Europeans	Indians	Inferred from Genotype
1	G	A	G	A	T	G	T	G	G	G	T	G	A	T	G	T	G	G		.378	.460	.778	.518	1
2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.		.015				2
3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.		.015		.019		3
4	.	.	.	.	.	T	.	.	A	.	C	.	G	C	.	.	.	.		.030				15
5	.	.	.	.	.	T	.	.	A	.	C	.	G	C	.	G	.	.		.076	.060		.071	16
6	.	.	.	.	.	T	G	.	A	.	C	.	G	C	.	G	.	.		.015	.020		.036	…
7	.	.	.	.	.	T	.	.	A	.	C	.	G	.	.	G	.	.					.018	…
8	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.		.045				7
9	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.212	.220	.019	.054	8
10	.	.	A	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.		.015				…
11	.	.	A	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.		.076				22
12	.	.	A	.	.	.	.	.	T	.	.	.	.	.	.	.	.	A		.015				…
13	.	.	A	.	.	.	.	.	T	.	.	.	.	.	G	.	.	.		.015				23
14	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.			.200	.019	.241	11
15	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.				.019		12
16	.	.	.	G	.	.	G	.	.	.	C	.	.	.	.	.	.	.					.018	…
17	.	G	A	.	.	.	G	.	.	.	.	.	.	.	A	.	.	.					.018	…

Note.—Haplotype frequencies for each population do not sum to 1, because haplotypes that contained missing sequence values were omitted. Each genotype listed is the genotype that unambiguously indicates the presence of this haplotype (see table 5). The absence of a genotype in the right-most column indicates that the haplotype could not be inferred unambiguously. Position −2335 was found to be polymorphic in a sample for which some sequence data were missing.

years throughout. This is lower than current human generation times but is probably reasonable for most of human history (Chen and Li 2001).

The selective neutrality of mutations in these sequences was tested using the methods of Tajima (1989), Fu and Li (1993), and Fay and Wu (2000, 2002). Statistical significance was based on 5,000 simulated samples. Hudson/Kreitman/Aguadé (HKA) tests (Hudson et al. 1987) were used to compare diversity patterns in the *CYP1A2 5′* region with diversity patterns found at eight other loci: *HOXB6* (Deinard and Kidd 1999), *ZFY* (Dorit et al. 1995), *β-globin* (Fullerton et al. 1994), *DMD* (Nachman and Crowell 2000*a*), *ND2* (Wise et al. 1998), *CCR5* (Bamshad et al., in press), Xq13.3 (Kaessmann et al. 1999), and three loci on the Y chromosome (Thompson et al. 2000). HKA tests were performed using the HKA computer program (Hey 2001).

Results

Sequencing of the subjects in the Old World sample revealed 18 SNPs, yielding an average density of 1 SNP per 188 bp (table 1). Seven of these SNPs were singletons (one was found in a European, three were found in Africans, and three were found in Indians). Of the 18 SNPs, 7 were transversions, and 11 were transitions (table 1). Sequencing of the PDR sample revealed 14 SNPs, yield-

Table 2

Alleles and Their Frequencies in the Old World and PDR Samples

	Allele		Frequency of Derived Variant in	
Position	Ancestral	Derived	Old World Sample (n = 226)	PDR Sample (n = 180)
−2993	G	A	.004	.012
−2992	A	G	.004	.012
−2963	G	A	.177	.215
−2796	G	A	.000	.006
−2780	A	G	.004	.006
−2728	T	C	.018	.006
−2701	G	T	.086	.086
−2697	T	G	.136	.058
−2335	G	T	.004	.000
−2215	A	G	.916	.903
−2012	G	T	.031	.000
−1950	C	T	.911	.918
−1880	G	A	.004	.000
−946	C	A	.000	.011
−905	A	G	.096	.080
−809	C	T	.916	.916
−108	G	A	.004	.000
160	G	T	.911	.908
272	G	C	.004	.000
394	G	A	.004	.000

Table 3

Ape Haplotypes

	Nucleotide Position																											
														−														
	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1
	0	9	9	9	9	8	7	7	7	7	7	7	7	6	6	5	4	4	4	3	2	2	1	0	0	9	8	8	8
	7	9	9	6	0	9	9	9	8	8	5	2	0	9	3	9	6	6	5	3	2	1	4	6	1	5	8	7	0
	4	3	2	3	7	8	6	5	8	0	9	8	1	7	4	9	2	1	8	5	8	5	4	4	2	0	0	6	7
Homo sapiens[a]	T	G	A	G	A	C	G	G	A	A	C	T	G	T	T	T	C	A	T	G	A	G	A	C	G	T	G	T	G
P. troglodytes 1	C	.	.	.	G	G	T	.	G	.	T	.	.	.	C	C	T	G	C	.	G	A	T	T	.	C	.	C	.
P. troglodytes 2	C	.	.	.	G	G	T	.	G	.	T	.	.	.	C	C	T	G	C	.	G	A	.	T	.	C	.	C	A
P. troglodytes 3	C	.	.	.	G	G	T	.	G	.	T	.	.	.	C	C	T	G	C	.	G	A	.	T	.	C	.	C	A
P. troglodytes 4	C	.	.	.	G	G	T	.	G	.	T	.	.	.	C	C	T	G	C	.	G	A	.	T	.	C	.	C	A
P. troglodytes 5	C	.	.	.	G	G	T	.	G	.	T	.	.	.	C	C	T	G	C	.	G	A	.	T	.	C	.	C	A
P. paniscus	C	.	.	.	G	G	T	.	G	.	T	.	.	.	C	C	T	G	C	.	G	A	T	T	.	C	.	C	.
G. gorilla	.	.	.	.	G	.	.	A	.	.	.	.	.	.	C	C	.	.	C	.	G	A	.	T	.	C	.	.	.

[a] The human haplotype is a consensus.

ing an average density of 1 SNP per 242 bp. Six polymorphisms were found in the Old World sample but not in the PDR sample, and two were found in the PDR sample but not in the Old World sample (table 2). Seven of the eight SNPs not found in both samples had frequencies <1%. The remaining SNP had a frequency of ~3% in the Old World sample, and its frequency in the African portion of the Old World sample was 11%.

Comparisons with the three ape species showed that one allele at each polymorphic site in the Old World sample was shared by all three species. The apes were monomorphic at all of the positions that were variable in humans (table 3). This permitted the assignment of the ancestral and derived states for each human SNP with little ambiguity (table 2). The ape nucleotide matched the human common allele at 14 of the 18 SNPs, and the ape nucleotide matched the human minor allele at 4 positions: −2215, −1950, −809, and 160.

In addition to the single-base differences between human and chimpanzee, the human sequence contained, a tetranucleotide repeat at position −3056, $(ATTT)_3$, that was found in five copies in the bonobo and four copies in the common chimpanzee. At position −2889, a singe-base repeat, $(T)_9$, was found in 10 copies in the bonobo and 12 copies in the common chimpanzee. At position −1706, a single-base repeat, $(A)_7$, was found in 8 copies in both chimpanzee species. Finally, at position −628, a single-base repeat, $(A)_4$, was found in three copies in both chimpanzee species.

All systems but one were in Hardy-Weinberg equilibrium (HWE). The SNP at position −2012 showed a significant departure ($P < .002$) from HWE, but this site was polymorphic in only the African population. When non-Africans were excluded from the calculation, the significance level dropped to $P < .03$, which would not be significant when a Bonferroni correction is applied for 18 comparisons. With little evidence for a departure from HWE, use of the EM algorithm for the construction of haplotypes is justified (Excoffier and Slatkin 1995).

The average nucleotide diversity, π, was higher in the Old World sample than in the PDR sample, although the SDs overlapped (table 4). In the Old World sample, the African population had the highest level of nucleotide diversity, followed by the Indian and Asian populations (table 4). The European sample, with a π value >10 times lower than those of the other populations, was remarkable for its lack of diversity: only four deviations from the consensus sequence were seen in 54 European chromosomes.

Table 1 lists the 17 haplotypes and their frequencies, as estimated by the EM algorithm, for the Old World sample. Of these 17 haplotypes, 11 could also be unambiguously inferred from genotypes in which only zero or one nucleotide positions varied (table 5). Pairs of the 11 unambiguous haplotypes accounted for >95% of all genotypes.

Of the 17 haplotypes, 12 were found in the African population, 8 were found in the Indian population, 5 were found in the Asian population, and 5 were found in the European population. Among haplotypes observed two or more times in the Old World sample, only one, haplotype 14, was not found in Africa. Haplotype 14 was common in the Asian and Indian populations, having frequencies of 0.200 and 0.232 in these two populations, respectively, and a frequency of 0.019 in the European population. This haplotype may have arisen outside Africa, or it may exist in African populations even though it was not detected in the present survey. The other five haplotypes seen outside Africa were also found in the African population. Among chromosomes sampled outside Africa, ~80% contained haplotypes also found in Africa (table 1).

The mean pairwise difference between the human and chimpanzee haplotypes was 45.3 substitutions, or 0.0122

Nucleotide Position																																							
−																												+													
1	1	1	1	1	1	1	1																																	
7	3	2	2	1	0	0	0	9	9	9	8	8	7	6	5	5	3	3	2	1	1	1	1				1	1	1	2	2	2	3	3						
9	5	9	4	6	8	4	2	5	4	0	3	0	0	7	9	1	7	0	2	7	2	1	0	9		2	7	3	6	7	4	5	7	6	9					
5	0	7	8	8	9	9	8	0	6	5	1	9	8	7	9	8	6	9	1	1	7	5	8	8	4	1	6	5	0	1	9	6	2	1	5					
G	C	G	A	C	G	C	C	G	G	A	A	T	T	G	C	G	G	G	C	A	A	C	G	C	A	T	A	G	T	G	C	C	G	C	G					
C	T	.	G	T	A	A	T	T	.	.	G	C	.	A	T	A	A	.	A	.	C	.	.	.	T	C	G	A	G	A	A	G	.	T	.					
C	.	C	G	T	A	A	T	T	.	.	G	C	G	A	T	A	.	C	A	C	C	.	.	.	T	C	G	A	G	.	A	G	.	T	.					
C	.	C	G	T	A	A	T	T	.	.	G	C	G	A	T	A	.	C	A	.	C	.	.	.	T	C	G	A	G	.	A	G	.	T	.					
C	.	C	G	T	.	A	T	T	.	.	G	C	G	A	T	A	.	C	A	.	C	.	.	.	T	C	G	A	G	.	A	G	.	T	.					
C	.	C	G	T	.	A	T	T	.	.	G	C	G	A	T	A	.	C	A	.	C	.	.	.	T	C	G	A	G	.	A	G	.	T	.					
C	T	.	G	T	?	?	?	?	?	?	G	C	.	A	T	A	A	.	A	.	C	.	.	.	T	C	G	A	G	A	A	G	.	T	.					
C	.	.	G	T	.	A	.	.	.	.	.	C	.	A	.	A	.	.	A	.	C	.	.	T	.	C	.	.	G	.	A	.	.	T	.					

substitutions per nucleotide. Under the assumptions that human/chimpanzee speciation occurred 5 million years ago and that the effective population size of the human/chimpanzee ancestral population was 100,000 (Takahata et al. 1995; Chen and Li 2001), we obtain a nucleotide-substitution rate of 6.8×10^{-10}. This rate is approximately one-half the rate obtained by the conventional method of calculation, which, when applied to our data, yields a nucleotide-substitution rate of 1.22×10^{-9} per site per year. The higher rate is similar to that estimated in a number of other systems (Eyre-Walker and Keightley 1999; Halushka et al. 1999; Nachman and Crowell 2000*b*; Zhao et al. 2000; Yu et al. 2001).

Estimates of the effective population size, N_e, were also similar to earlier estimates when calculated under the assumption of a nucleotide-substitution rate of 6.8×10^{-10} (e.g., see Harding et al. 1997; Zietkiewicz et al. 1998; Zhao et al. 2000). For the Old World sample, Watterson's *S,* which is based on the number of segregating sites in the sample, yielded N_e estimates of 15,030 for the sample as a whole, of 13,680 for Africa, and of 11,570 for non-Africans. Tajima's θ, which is based on average nucleotide diversity, yielded lower estimates: 8,270 total, 10,980 for Africans, and 6,810 for non-Africans.

In the analysis of recombination in the *CYP1A2* region, three independent marker combinations produced genetic distance:physical distance ratios of 0.46, 0.52, and 0.53 cM/Mb. On the basis of these estimates, the regional recombination rate that surrounds *CYP1A2* is approximately one-half the observed genome average of 1 cM/Mb. Considering this low rate of recombination and the relatively small size of this region, we do not find it surprising that most SNPs pairs were in complete linkage disequilibrium, with no evidence for a decline of disequilibrium with physical distance between SNP pairs. The African population had the lowest average pairwise D' value (0.0924), followed by the Indian (0.988), Asian (1.000), and European (1.000) populations. These high levels of disequilibrium are consistent with other empirical studies' findings, which have revealed significant linkage disequilibrium, at distances of 10–50 kb or more, in many genomic regions in most human populations (Jorde et al. 1994; K. K. Kidd et al. 1998; Collins et al. 1999; Goddard et al. 2000; Jorde 2000; Reich et al. 2001; Stephens et al. 2001). They indicate that statistical tests that assume absence of recombination—including Tajima's *D,* Fu and Li's *D* and *F,* and phylogenetic analyses—are justified.

The site-frequency spectra for the Old World and PDR samples are shown in figure 1. For four of the SNPs in the Old World and PDR samples, the minor allele in the human population was the fixed allele in the apes, so the common allele in the human sample was inferred to be in the derived state. Each of these four alleles had frequencies of 90%–94%. The seven singletons observed in the Old World sample represent a strong excess relative to the number expected under neutrality (i.e., ~3, on the basis of the estimate of *S*). As seen in table 6, the hypothesis of neutrality is not rejected when Tajima's *D* statistic is used, but the hypothesis is rejected when other statistics, such as those of Fu (1996) and Fay and Wu (2000), are used. This

Table 4

Diversity (π) Values in the Sample Populations

Population	$\pi \pm$ SD
PDR sample	.00039 ± .00023
Old World sample	.00043 ± .00026
Africans	.00057 ± .00032
Asians	.00043 ± .00027
Europeans	.00004 ± .00006
Indians	.00055 ± .00031

probably reflects the latter tests' greater sensitivity to excesses of low- and high-frequency variants, respectively.

The HKA test is designed to detect departures from evolutionary neutrality by the comparison of diversity patterns in a known neutral locus with diversity patterns in a test locus (Hudson et al. 1987). Neutrally evolving control loci are difficult to identify in practice, so we compared diversity patterns in the *CYP1A2 5′* region with those of several other loci to survey similarities and differences by contrasting the *P* values. Diversity patterns in the *CYP1A2 5′* region were similar to those found in some loci, such as *HOXB6, ZFY,* and *β-globin,* but differed significantly from others, such as *DMD* and *CCR5* (table 7).

Figure 2 shows the MST of haplotypes. In the tree, haplotypes are divided into two clusters. One cluster, containing haplotypes 4–7 (referred to here as "cluster B"), differs from all other haplotypes (referred to here as "cluster A") by four nucleotides. These four nucleotide substitutions, found at positions −2215, −1950, −809, and 160, account for all four high-frequency SNPs (compare table 1 and fig. 1). The root of the maximum-likelihood tree (not shown) fell within haplotype cluster B, and the length of the branch that separates the two clusters differed significantly from 0 ($P < .05$), supporting their separation. The same clusters and root position were seen in trees constructed by the neighbor-joining and parsimony methods.

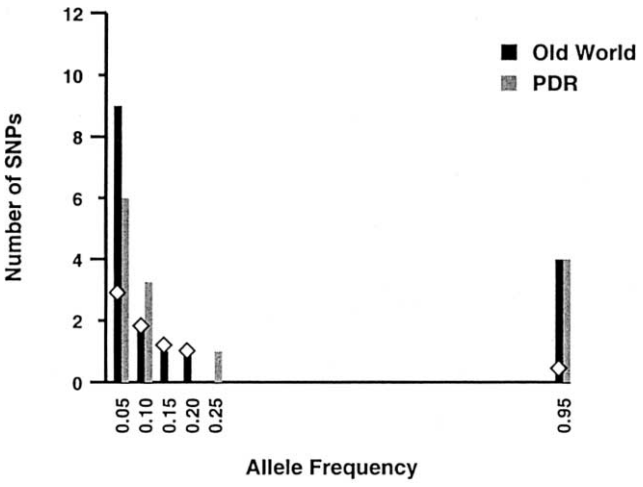

Figure 1 Frequency spectra of derived alleles. The diamonds (◇) show the theoretical expectations under neutrality for the Old World sample.

The two clusters differed appreciably in frequency. Overall, haplotypes from cluster A were found in ~90% of sampled chromosomes. Haplotype cluster B was

Table 5

Genotypes and Their Estimated Frequencies in Each Population

Genotype (Haplotype Pair[s])	Occurrences in				Total Occurrences
	Africans	Indians	Asians	Europeans	
1 (1-1)	5	8	4	19	36
2 (1-2)	1	0	0	0	1
3 (1-3)	1	0	0	1	2
4 (1-5 or 3-4)	3	1	2	0	6
5 (1-6 or 5-14)	1	2	1	0	4
6 (1-7)	0	1	0	0	1
7 (1-8)	2	0	0	0	2
8 (1-9)	5	1	5	1	12
9 (1-10)	1	0	0	0	1
10 (1-12)	1	0	0	0	1
11 (1-14)	0	6	7	1	14
12 (1-15)	0	0	0	1	1
13 (1-16)	0	1	0	0	1
14 (1-17)	0	1	0	0	1
15 (4-4)	1	0	0	0	1
16 (5-5)	0	1	0	0	1
17 (5-9)	2	1	1	0	4
18 (8-9)	1	0	0	0	1
19 (9-9)	2	0	1	0	3
20 (9-11)	2	0	0	0	2
21 (9-14)	0	1	3	0	4
22 (11-11)	1	0	0	0	1
23 (11-13)	1	0	0	0	1
24 (14-14)	0	3	0	0	3
Overall	30	27	24	23	104

Table 6

Results of Neutrality Tests (Tajima's *D*, Fu and Li's *D* and *F*, and Fay and Wu's *H*) in Study Populations, Including the Old World and PDR Samples

Study Population	Neutrality[a] by			*P* Value by
	Tajima's *D*	Fu and Li's *D*	Fu and Li's *F*	Fay and Wu's *H*
Africans	−.54 (−1.40)	−.72 (−1.86)	−.74 (−1.76)	<.05
Non-Africans	−1.00 (−1.42)	−1.76 (−1.77)	−1.71 (−1.72)	<.02
Old World sample (All)	−1.15 (−1.41)	−2.81 (−1.85)	−2.53* (−1.77)	<.05
PDR sample	−.93 (−1.36)	−.36 (−1.76)	−.68 (−1.71)	<.02

[a] Critical values are given in parentheses.
* $P < .02$.

found at its highest frequency in the Indian sample, in which it was found in ~13% of sampled chromosomes. Both clusters were found in the African, Indian, and Asian samples. Haplotype cluster B was absent from the European sample.

The placement of haplotype 17 in the MST was ambiguous, as indicated by a reticulation in the MST. This haplotype differed from haplotypes 9 and 14 by three nucleotides. The ambiguous placement of haplotype 17 may reflect recombination, or it may be the result of convergent nucleotide substitutions.

Discussion

Population History

The nucleotide-diversity values found for the *CYP1A2 5′* region are similar to those found in most other surveys of human autosomal DNA sequences (Halushka et al. 1999; Wall and Przeworski 2000; Zhao et al. 2000; Jorde et al. 2001). In a survey of SNP densities in 200-kb bins spread across the human genome, the International SNP Map Working Group (2001) found a modal π value of ~0.0007, with a range from 0.0000 to 0.0020. The value that we find for the *CYP1A2 5′* region in the Old World sample ($\pi = 0.00043$) falls at approximately the 20th percentile in the International SNP Map Working Group's distribution (see fig. 2*a* in International SNP Map Working Group 2001). The π value that we observe is only slightly lower than 0.00054, which is the average reported by Stephens et al. (2001) in a survey of >300 genes. The low π values generally observed in humans support the conclusion that humans are genetically less diverse than many other species (Li and Sadler 1991; Nachman et al. 1998; Kaessmann et al. 1999). Under the assumption that evolution in the *CYP1A2 5′* region has been neutral, the relatively low level of nucleotide diversity predicts low effective population sizes, estimated here as ~5,000–15,000. These values are similar to values obtained in previous studies of protein polymorphisms (Nei and Graur 1984), mtDNA (Takahata 1993), Y-chromosome DNA (Hammer 1995; Goldstein et al. 1996), X-chromosome polymorphisms (Zietkiewicz et al. 1998; Kaessmann et al. 1999), and autosomal polymorphisms (Harding et al. 1997; Sherry et al. 1997; Halushka et al. 1999; Zhao et al. 2000).

Nucleotide-diversity values were higher in Africa than elsewhere, and Africans had the largest number of haplotypes. Greater African diversity has been observed in many studies of human genetic diversity—including those of protein polymorphisms (Nei et al. 1993), mtDNA (Vigilant et al. 1991), Y-chromosome DNA (Hammer et al. 1997; Seielstad et al. 1999; Jorde et al. 2000), autosomal microsatellites (Bowcock et al. 1994; Deka et al. 1995; Calafell et al. 1997; Jorde et al. 1997), autosomal *Alu* polymorphisms (Jorde et al. 2000), and nuclear SNPs and DNA sequences (Nickerson et al. 1998; Halushka et al. 1999; Kaessmann et al. 1999; Rieder et al. 1999; Labuda et al. 2000; Wall and Przeworski 2000; Zhao et al. 2000; Jorde et al. 2001; Yu et al. 2001). This pattern is consistent with an African origin of modern humans, but it can also be explained by a higher effective population size in Africa (Relethford and Jorde 1999).

Additional evidence for an African origin is provided by the geographical distribution of haplotypes: haplotypes found outside Africa are mostly a subset of those found within Africa. This pattern has been observed in a number of other studies (Tishkoff et al. 1996, 2000;

Table 7

Results of HKA Tests in the Old World and PDR Samples

Locus	*P* Value for		Difference
	Old World Sample	PDR Sample	
HOXB6	.49	.47	.02
ZFY	.11	.13	.02
β-globin	.07	.07	.00
DMD	.05*	.06	.01
ND2syn	.01**	.01**	.00
CCR5	.01**	.01**	.00
Y-SDD	.01**	.01**	.00
Xq13.3	.01**	.01**	.00

* $P < .05$.
** $P < .01$.

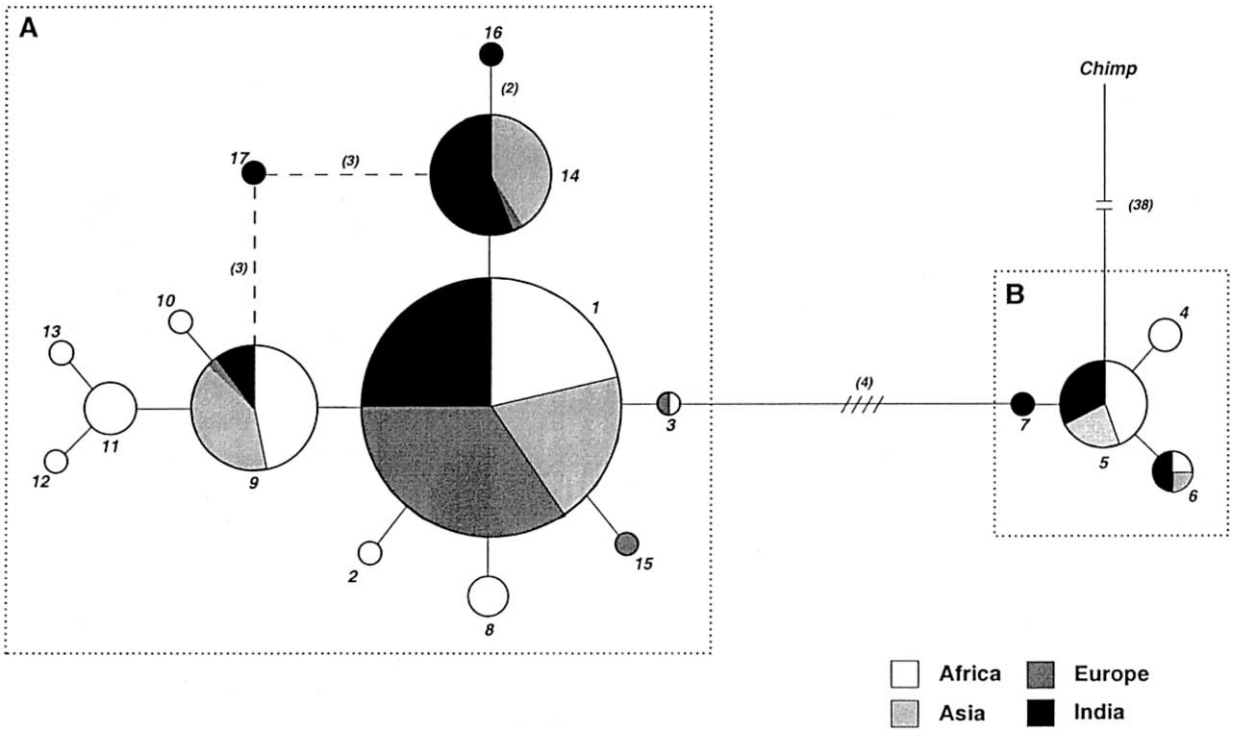

Figure 2 MST of *CYP1A2* haplotypes. Dashed lines indicate alterative topologies of equal length. Node areas are proportional to haplotype frequency. Fully blackened nodes are haplotypes found only in Africa; each node is shaded to indicate the fraction of time it was found in each geographical region. Branches represent one nucleotide substitution except where noted in parentheses. Nucleotide substitutions at which the derived state has high frequency (>90%) are marked with crossbars. Haplotype clusters A and B are shown in the left and right boxes, respectively.

Calafell et al. 1997; Hammer et al. 1997; Watson et al. 1997; Alonso and Armour 2001) and agrees with a scenario in which some African populations left the continent to populate the rest of the Old World (Jorde et al. 1998; Pääbo 1999; Yu et al. 2002).

The linkage-disequilibrium patterns seen in this study are in agreement with a number of other studies that have found less linkage disequilibrium in African than in non-African populations (Tishkoff et al. 1996, 1998, 2000; K. K. Kidd et al. 1998; J. R. Kidd et al. 2000; Reich et al. 2001). Because linkage disequilibrium decreases through time, as a result of recombination, levels of disequilibrium can be correlated with the relative "age" of a population, with older populations having less disequilibrium. Thus, the linkage-disequilibrium results are also consistent with an African origin of modern humans.

Tajima's *D* test failed to reject the hypothesis of neutrality and constant population size. However, the *D* and *F* tests of Fu and Li, which are statistically more powerful under most circumstances (Fu 1996), yielded results that indicate a significant excess of rare variants. This departure could be the result of either positive natural selection or an expansion from a very small population to a much larger one. The latter possibility is expected to produce an excess of rare alleles, because genetic drift, which tends to remove rare alleles from populations, has little effect during and after an expansion (Slatkin and Hudson 1991; Rogers and Jorde 1995).

The results of Fu's *D* and *F* tests are consistent with several lines of evidence for a population expansion in humans. Among genetic systems, support for a human population expansion has been derived from studies of mtDNA (Di Rienzo and Wilson 1991; Rogers and Harpending 1992), Y-chromosome microsatellites (Pritchard et al. 1999) and SNPs (Shen et al. 2000), and autosomal microsatellites (Shriver et al. 1997; Di Rienzo et al. 1998; Kimmel et al. 1998; Reich and Goldstein 1998; Gonser et al. 2000; Jin et al. 2000; Zhivotovsky et al. 2000). Details vary from study to study—for example, the estimated dates of the expansion differ in some cases, and some studies find evidence for expansions in specific populations, such as Africans, whereas others do not. However, analyses of these different types of genetic systems all detect the signature of a population expansion sometime during the past 100,000 years.

Diversity patterns in the *CYP1A2 5′* region combine to form a picture of human genetic diversity that is unremarkable in a number of respects: Africans are more diverse, non-African diversity is largely a subset of Af-

rican diversity, and evidence for population growth is present. However, some aspects of this variation suggest that the history of *CYP1A2* is characterized by more-complex factors as well.

Natural Selection

The failure of Tajima's *D* test to reject the hypothesis of neutrality and the presence of significantly negative *D* and *F* statistics in Fu and Li's test are surprising given the large number of high-frequency variants. More than 20% of sites in the Old World sample have frequencies >90%. Under neutral expectations, <5% of sites should attain such frequencies (fig. 1).

The failure of the *D* and *F* tests to detect an excess of high-frequency variants is probably due to the insensitivity that these tests have to such variants. Tajima's *D* statistic, which compares the number of segregating sites in a sample with the mean pairwise difference (Tajima 1989), is insensitive to the presence of very-high-frequency variants because they contribute little to mean pairwise differences. In a sample of 100 sequences, for example, a derived variant with a frequency of 1/100 will contribute the same to mean pairwise difference as a derived variant with a frequency of 99/100. Fu and Li's statistics, which compare the number of singletons with the number of nonsingletons (Fu and Li 1993), are insensitive to the presence of high-frequency variants because they are scored simply as nonsingletons. Thus, all three of these statistics are probably unaffected by extreme frequency spectra, such as that observed in the *CYP1A2 5′* region.

The application of Fay and Wu's *H* test, which is sensitive to the presence of high-frequency derived variants, rejects the hypothesis of neutrality in our data (table 6) (Fay and Wu 2000; Fay 2002). This result is consistent with the observations that two distinct haplotype clusters are present and that the frequency of the more-derived cluster (A) is higher than expected.

The hypothesis of neutrality in the *CYP1A2 5′* region is placed further in doubt by the HKA tests, which indicate that within- and between-species diversity patterns in the *CYP1A2 5′* region are similar to those of *HOXB6, ZFY,* and *β-globin.* This implies that diversity patterns in the *CYP1A2 5′* region are similar to those in coding regions with known functions. The results of HKA tests do not conclusively exclude either natural selection or evolutionary neutrality in the *CYP1A2 5′* region. Unless *HOXB6, ZFY,* and *β-globin* have been evolving neutrally, however, similarities between the *CYP1A2 5′* region and these regions make neutrality in the *CYP1A2 5′* region seem unlikely.

One factor that can cause the presence of distinct haplotype clusters is population subdivision, which can allow divergent clusters to evolve in different demes. Under these conditions, divergent clusters should be restricted to different geographical areas. This is not the case in the *CYP1A2 5′* region. Haplotype clusters A and B both are widespread geographically (fig. 2). Both are found in several Asian populations (Cambodians, Chinese, Japanese, and Indians) and several African populations (Sotho, Biaka Pygmy, Mbuti Pygmy, Nande, Hema, and Nigerians). The presence of the two clusters does not therefore seem to be due to population subdivision.

Another factor that can cause excesses of high-frequency variants is balancing natural selection. Balancing natural selection can maintain old evolutionary lineages at low frequencies by protecting them from genetic drift (Lewontin and Hubby 1966). Positive selection with low levels of recombination can have similar effects (see fig. 2 in Fay and Wu 2000). Positive selection can have the additional effect of reducing the overall levels of genetic variation, which could explain the relatively low levels we observe in the *CYP1A2 5′* region. These low levels of variation are also consistent with the reduced recombination observed in the region (Aquadro et al. 2001; Nachman 2001). Finally, the presence of four high-frequency derived SNPs at frequencies >0.9 matches patterns observed by Fay and Wu (2000) in simulations of positive natural selection. Such selection would not need to be acting directly on the *CYP1A2 5′* region. Ding et al. (2002), for example, found positive selection in markers linked to *DRD4.* The variation patterns that we see could be caused by the action of natural selection on *CYP1A2* exons or even on other loci (Braverman et al. 1995; Fay and Wu 2000).

The hypotheses of population growth and natural selection are not mutually exclusive. It is possible for both to occur at the same time in the same populations. Whether both occur in the *CYP1A2 5′* region and relative importance of each are still unclear. Further analyses of diversity in *CYP1A2,* especially in coding regions, may provide more information about the relative importance of demographic and selective effects by allowing comparisons between synonymous and nonsynonymous nucleotide substitutions (Fay et al. 2001).

Implications

Patterns of population genetic variation in the *CYP1A2 5′* region shed some light on questions, about the relationship between *CYP1A2* diversity, ethnicity, and geography, raised by earlier association studies. There are appreciable regional differences in allele frequency. Haplotype 9, for example, was found in ~20% of Asian and African samples but was rare in the European and Indian samples. Furthermore, although haplotype cluster B was found with a frequency of ~10% in the Africans, Indians, and Asians, Europeans lacked haplotypes from cluster B altogether.

Our sequences do not extend far enough upstream or downstream from the *CYP1A2* promoter to allow comparisons with most polymorphisms identified in intron 1 (Sachse et al. 1999) and in 5′ regions (Aitchison et al. 2000). However, one of the polymorphic sites identified in our sample, position −2963, was studied by Nakajima et al. (1999), who found the A nucleotide at a frequency of 0.23 in a Japanese study population. We find the same variant at a similar frequency (0.22) in the Asian population that we studied. On the basis of both analyses of rates of caffeine metabolism in Japanese subjects and gel-retardation assays that detected differential protein binding to the A and G variants at position −2963, Nakajima et al. (1999) concluded that variation at that position accounts for some of the phenotypic variation in *CYP1A2* expression. Our results do not contradict those of Nakajima et al. (1999), but they do suggest that other sites may be important as well. The sites that distinguish haplotype clusters A and B may be of particular interest.

Diversity patterns in the *CYP1A2* 5′ region also suggest a number of hypotheses about phenotypic variation in *CYP1A2* expression and activity. To date, most association studies on *CYP1A2* have compared the mean values of phenotypes, rather than their variances. However, the geographic distribution of genetic diversity that we observe suggests that phenotypic variances should be analyzed as well. One hypothesis is that, if haplotypes in the *CYP1A2* 5′ region are associated with phenotypes, then phenotypic variance should be greatest in Africans, followed by Asians and then Europeans.

A simple test of the hypothesis that *CYP1A2*-related phenotypes are more diverse in Africans than in Europeans can be performed with the data of Basile et al. (2000), who used the abnormal-involuntary-movement scale (AIMS) to measure phenotypes in African Americans and Europeans who were treated with antipsychotic drugs thought to interact with *CYP1A2* to cause tardive dyskinesia. Basile et al. found significant differences between individuals with genotypes defined by an A/C SNP (Basile et al. 2000, but see Schulze et al. 2002). Within genotypes, Basile et al. (2000) found no significant difference between the mean AIMS scores of African Americans and Europeans. We found, however, that the phenotypic variance measured using the AIMS is significantly greater in African Americans than in Europeans, in two of the three reported genotypes. An analysis of variance yields $P < .01$ for Basile et al.'s AA and AC types (see fig. 2 in Basile et al. 2000). The greater phenotypic variance found in Africans could be associated with the greater diversity in African *CYP1A2* haplotypes.

Whether clusters A and B contain functionally divergent haplotypes cannot be determined directly from our analyses. More important is the fact that diversity patterns in the *CYP1A2* 5′ region reflect patterns of population genetic variation, including broad-scale similarities and differences among geographical regions, that can be used to improve the design and implementation of future studies. The presence of these clusters is likely to be of some significance in further studies of the association between *CYP1A2* genotypes and phenotypes.

DNA sequences from chimpanzee and gorilla were a key source of information in our analyses. We anticipated that inclusion of these species would resolve the ambiguity of ancestral and derived character states in only a small number of variable positions. However, these outgroup comparisons were crucial in the identification of high-frequency variants. The usual assumption in the absence of outgroup data is that the minor allele is in the derived state. With such an assumption, we would have described all four high-frequency variants as low-frequency variants, and the outcomes of statistical tests would have been flawed. The importance of outgroup comparisons is well recognized, and chimpanzees are becoming widely adopted as a standard in human genetic analyses (Hacia et al. 1999). Our results reemphasize the importance of the inclusion of outgroups in analyses of human genetic variation.

The PDR is being extensively used to detect new human SNPs, so the comparison between results based on this sample and those based on the Old World sample is of some significance. It is expected that high-frequency alleles should be ancient and thus should be widely shared among populations. This expectation was confirmed in the comparison undertaken here: the more common polymorphisms were indeed found in both samples, indicating that the PDR sample should be effective for the detection of most such polymorphisms. One polymorphism, with a frequency of 11% in the African sample, was not, however, detected in the PDR sample. This likely reflects a broader sampling of African diversity in the Old World sample than in the PDR sample.

The wider representation of diversity in the Old World sample also resulted in higher average nucleotide diversity than in the PDR sample and in a larger number of detected low-frequency polymorphisms (although the latter is also partially due to a larger number of chromosomes in the Old World sample [226] than in the PDR sample [180]). Although low-frequency polymorphisms are less informative for linkage and linkage-disequilibrium analyses (Ott and Rabinowitz 1997), they can sometimes be useful for association-based mapping. For example, if a disease-causing mutation occurs on a chromosome that contains a rare marker, the disease-marker association will be stronger than it would be if the mutation were to occur on a chromosome that contains a polymorphism that is common in the population (Jorde 2000).

Patterns of genetic diversity observed in the *CYP1A2* 5′ region bring attention to some basic limitations of the PDR as a tool for the study of patterns of human genetic variation. The PDR is intended primarily to "facilitate the discovery of DNA sequence variants" (Collins et al. 1998, p. 1229), not for population genetic analysis. We found that, although patterns of genetic diversity in the Old World and PDR samples are similar in some respects, a clear advantage of the Old World sample is that genetic diversity can be measured and compared in different subpopulations. Between-population comparisons can be an important step in the characterization of genetic diversity, as was demonstrated by our comparisons of haplotype-cluster frequencies in different geographical regions. In the case of the *CYP1A2* 5′ region, the effects of population substructuring and of natural selection would probably be indistinguishable without information about geographic provenance. These results underscore the continued importance of the sampling of diverse populations across broad geographical regions.

Acknowledgments

We would like to thank Ed Meenen, for technical assistance, and Chris Ricker, Henry Harpending, Alan Rogers, Jon Seger, Fred Adler, Bradley Demarest, and Thomas Hills, for helpful discussions on this manuscript. Justin Fay provided assistance with analyses. This research was supported by National Institutes of Health grants GM-59290, RR-00064, and ES-10058 and by NSF grants SBR-9514733 and SBR-9818215.

Electronic-Database Information

The accession number and URLs for data presented herein are as follows:

dbSNP Home Page, http://www.ncbi.nlm.nih.gov/SNP/ (submitter handle is uugc)

GenBank, http://www.ncbi.nlm.nih.gov/Genbank/ (for human *CYP1A2* genomic sequence [accession number U02993])

GeneSNPs Public Internet Resource, http://www.genome.utah.edu/genesnps/

NCBI Map Viewer, http://www.ncbi.nih.gov/sitemap/#MapViewer

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for cytochrome P450 gene family [MIM 124060])

References

Aitchison KJ, Gonzalez FJ, Quattrochi LC, Sapone A, Zhao JH, Zaher H, Elizondo G, Bryant C, Munro J, Collier DA, Makoff AJ, Kerwin RW (2000) Identification of novel polymorphisms in the 5′ flanking region of *CYP1A2*, characterization of interethnic variability, and investigation of their functional significance. Pharmacogenetics 10:695–704

Alonso S, Armour JAL (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

Aquadro CF, Dumont VB, Reed FA (2001) Genome-wide variation in the human and fruitfly: a comparison. Curr Opin Genet Dev 11:627–634

Ardlie KG, Kruglyak L, Sielstad M (2002) Patterns of linkage disequilibrium in the human genome. Nat Rev Genet 3:299–309

Bamshad M, Mummidi S, Gonzalez E, Ahuja SS, Dunn DM, Stone AC, Jorde LB, Ahuja SK, Weiss RB. Evidence of balancing selection in the 5′ *cis*-regulatory region of CCR5. Proc Natl Acad Sci USA (in press)

Basile VS, Osdemir V, Masellis M, Walker ML, Meltzer HY, Lieberman JA, Potkin SG, Alva G, Kalow W, Macciardi FM, Kennedy JL (2000) A functional polymorphism of the cytochrome P450 1A2 (*CYP1A2*) gene: association with tardive dyskinesia in schizophrenia. Mol Psychiatry 5:410–417

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

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

Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W (1995) The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics 140:783–796

Calafell F, Shuster A, Speed WC, Kidd JR, Kidd KK (1997) Short tandem repeat polymorphism evolution in humans. Eur J Hum Genet 6:38–49

Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nat Genet 22:231–238

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

Chen FC, Li WH (2001) Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. Am J Hum Genet 68:444–456

Collins FS, Brooks LD, Chakravarti A (1998) A DNA polymorphism discovery resource for research on human genetic variation. Genome Res 8:1229–1231

Collins FS, Lonjou C, Morton NE (1999) Genetic epidemiology of single-nucleotide polymorphisms. Proc Natl Acad Sci USA 96:15173–15177

Deinard A, Kidd K (1999) Evolution of a *HOXB6* intergenic region within the great apes and humans. J Hum Evol 36: 687–703

Deka R, Shriver MD, Yu LM, Ferrell RE, Chakraborty R (1995) Intra- and inter-population diversity at short tandem repeat loci in diverse populations of the world. Electrophoresis 16:1659–1664

Deka R, Shriver MD, Yu LM, Heidreich EM, Jin L, Zhong Y, McGarvey ST, et al (1999) Genetic variation at twenty-three microsatellite loci in sixteen populations. J Genet 78:99–121

Ding Y, Chi H, Grady D, Morishima A, Kidd JR, Kidd KK, Flodman P, Spence MA, Schuck S, Swanson JM, Zhang Y, Moyzis RK (2002) Evidence of positive selection acting at the human dopamine receptor D4 locus. Proc Natl Acad Sci USA 99:309–314

Di Rienzo A, Donnelly P, Toomajian C, Sisk B, Hill A, Petzl-Erler ML, Haines GK, Barch DH (1998) Heterogeneity of microsatellite mutations within and between loci, and implications for human demographic histories. Genetics 148: 1269–1284

Di Rienzo A, Wilson AC (1991) Branching pattern in the evolutionary tree for human mitochondrial DNA. Proc Natl Acad Sci USA 88:1597–1601

Dorit RL, Akashi H, Gilbert W (1995) Absence of polymorphism at the *ZFY* locus on the human Y chromosome. Science 268:1183–1185

Evett IW, Weir BS (1998) Interpreting DNA evidence: statistical genetics for forensic scientists. Sinauer Associates, Sunderland, MA

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using *phred*. I. Accuracy assessment. Genome Res 8:175–185

Excoffier L, Slatkin M (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Mol Biol Evol 12:921–927

Eyre-Walker A, Keightley PD (1999) High genomic deleterious mutation rates in hominids. Nature 397:344–347

Fay J (2002) *H*-test, version 1.0. Lawrence Berkeley Laboratories, Berkeley

Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. Genetics 155:1405–1413

Fay JC, Wyckoff GJ, Wu CI (2001) Positive and negative selection on the human genome. Genetics 158:1227–1234

Felsenstein J (1993) PHYLIP: phylogenetic inference package, version 3.5d. Department of Genetics, University of Washington, Seattle

Forster P, Rohl A, Lunnemann P, Brinkmann C, Zerjal T, Tyler-Smith C, Brinkmann B (2000) A short tandem repeat-based phylogeny for the human Y chromosome. Am J Hum Genet 67:182–196

Fu YX (1996) New statistical tests of neutrality for DNA samples from a population. Genetics 143:557–570

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709

Fullerton SM, Harding RM, Boyce AJ, Clegg JB (1994) Molecular and population genetic analysis of allelic sequence diversity at the human β-globin locus. Proc Natl Acad Sci USA 91:1805–1809

Gabriel S, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. Science 296:2225–2229

Goddard KA, Hopkins PJ, Hall JM, Witte JS (2000) Linkage disequilibrium and allele-frequency distributions for 114 single-nucleotide polymorphisms in five populations. Am J Hum Genet 66:216–234

Goldstein DB, Zhivotovsky LA, Nayar K, Linares AR, Cavalli-Sforza LL, Feldman MW (1996) Statistical properties of the variation at linked microsatellite loci: implications for the history of human Y chromosomes. Mol Biol Evol 13:1213–1218

Gonser R, Donnelly P, Nicholson G, Di Rienzo A (2000) Microsatellite mutations and inferences about human demography. Genetics 154:1793–1807

Hacia JG, Fan JB, Ryder O, Jin L, Edgemon K, Ghandour G, Mayer RA, Sun B, Hsieh L, Robbins CM, Brody LC, Wang D, Lander ES, Lipshutz R, Fodor SPA, Collins FS (1999) Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. Nat Genet 22:164–167

Halushka MK, Fan JB, Bentley K, Hsie L, Shen N, Weder A, Cooper R, Lipshutz R, Chakravarti A (1999) Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nat Genet 22:239–247

Hammer MF (1995) A recent common ancestry for human Y chromosomes. Nature 378:376–378

Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, Griffiths RC, Templeton AR, Zegura SL (1998) Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. Mol Biol Evol 15: 427–441

Hammer MF, Spurdle AB, Karafet T, Bonner MR, Woodi ET, Novelletto A, Malaspina P, Mitchell RJ, Horai S, Jenkins T, Zegura SL (1997) The geographic distribution of Y chromosome variation. Genetics 145:787–805

Harding RM, Fullerton SM, Griffiths RC, Bond J, Cox MJ, Schneider JA, Moulin DS, Clegg JB (1997) Archaic African and Asian lineages in the genetic ancestry of modern humans. Am J Hum Genet 60:722–789

Harpending HC, Batzer MA, Gurven M, Jorde LB, Rogers AR (1998) Genetic traces of ancient demography. Proc Natl Acad Sci USA 95:1961–1967

Harpending HC, Rogers AR (2000) Genetic perspectives on human origins and differentiation. Annu Rev Genomics Hum Genet 1:361–385

Hey J (2001) HKA. Department of Genetics, Rutgers University, Piscataway, NJ

Hudson RR, Kreitman M, Aguadé M (1987) A test of neutral molecular evolution based on nucleotide data. Genetics 116: 153–159

Ikeya K, Jaiswal AK, Owens RA, Jones JE, Nebert DW, Kimura S (1989) Human *CYP1A2*: sequence, gene structure, comparison with the mouse and rat orthologous gene, and differences in liver 1A2 mRNA expression. Mol Endocrinol 3: 1399–1408

Ingman M, Kaessmann H, Pääbo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. Nature 408:708–713

International SNP Map Working Group (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409:928–933

Jin L, Baskett ML, Cavalli-Sforza LL, Zhivotovsky LA, Feldman MW, Rosenberg NA (2000) Microsatellite evolution in modern humans: a comparison of two data sets from the same populations. Ann Hum Genet 64:117–134

Jorde LB (2000) Linkage disequilibrium and the search for complex disease genes. Genome Res 10:1435–1444

Jorde LB, Bamshad MJ, Rogers AR (1998) Using mitochondrial and nuclear DNA markers to reconstruct human evolution. BioEssays 20:126–136

Jorde LB, Rogers AR, Bamshad M, Watkins WS, Krakowiak P, Sung S, Kere J, Harpending HC (1997) Microsatellite diversity and the demographic history of modern humans. Proc Natl Acad Sci USA 94:3100–3103

Jorde LB, Watkins WS, Bamshad MJ (2001) Population genomics: a bridge from evolutionary history to genetic medicine. Hum Mol Genet 10:2199–2207

Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Sielstad MT, Batzer MA (2000) The distribution of human genetic diversity: a comparison of mitochondrial, autosomal and Y-chromosome data. Am J Hum Genet 66:979–988

Jorde LB, Watkins WS, Carlson M, Groden J, Albertsen H, Thliveris A, Leppert M (1994) Linkage disequilibrium predicts physical distance in the adenomatous polyposis coli region. Am J Hum Genet 54:884–898

Kaessmann H, Heißig F, von Haeseler A, Pääbo S (1999) DNA sequence variation in a non-coding region of low recombination on the human X chromosome. Nat Genet 22:78–81

Kidd JR, Pakstis AJ, Zhao H, Lu RB, Okonofua FE, Odunsi A, Grigorenko E, Tamir BB, Friedlander J, Schulz LO, Parnas J, Kidd KK (2000) Haplotypes and linkage disequilibrium at the phenylalanine hydroxylase locus, *PAH*, in a global representation of populations. Am J Hum Genet 66: 1882–1899

Kidd KK, Morar B, Castigione CM, Zhao H, Pakstis AJ, Speed WC, Bonné-Tamir B, Lu RB, Goldman D, Lee C, Nam YS, Grandy DK, Jenkins T, Kidd JR (1998) A global survey of haplotype frequencies and linkage disequilibrium at the *DRD2* locus. Hum Genet 103:211–227

Kimmel M, Chakraborty R, King JP, Bamshad M, Watkins WS, Jorde LB (1998) Signatures of population expansion in microsatellite repeat data. Genetics 148:1921–1930

Labuda D, Zietkiewicz E, Yotova V (2000) Archaic lineages in the history of modern humans. Genetics 156:799–808

Lewontin RC (1964) The interaction of selection and linkage. I. General considerations: heterotic models. Genetics 49:49–67

Lewontin RC, Hubby JL (1966) A molecular approach to the study of genetic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. Genetics 54:595–609

Li WH, Sadler LA (1991) Low nucleotide diversity in man. Genetics 129:513–523

Mountain JL (1998) Molecular evolution and modern human origins. Evol Anthropol 7:21–37

Nachman MW (2001) Single nucleotide polymorphisms and recombination rate in humans. Trends Genet 117:481–485

Nachman MW, Bauer VL, Crowell SL, Aquadro CF (1998) DNA variability and recombination rates at X-linked loci in humans. Genetics 150:1133–1141

Nachman MW, Crowell SL (2000*a*) Contrasting evolutionary histories of two introns of the Duchenne muscular dystrophy gene, *DMD*, in humans. Genetics 155:1855–1964

——— (2000*b*) Estimate of the mutation rate per nucleotide in humans. Genetics 156:297–304

Nakajima M, Yokoi T, Mizutani M, Kinoshita M, Funayama M, Tamataki T (1999) Genetic polymorphism in the 5′-flanking region of human *CYP1A2* gene: effect on the *CYP1A2* inducibility in humans. J Biochem 125:803–808

Nei M, Graur D (1984) Extent of protein polymorphism and the neutral mutation theory. Evol Biol 17:73–118

Nei M, Livshits G, Ota T (1993) Genetic variation and evolution of human populations. In: Sing CF, Hanis CL (eds) Genetics of cellular, individual, family, and population variability. Oxford University Press, New York, pp 239–252

Nickerson DA, Taylor SL, Fullerton SM, Weiss KM, Clark AG, Stengård JH, Salomaa V, Boerwinkle E, Sing CF (2000) Sequence diversity and large-scale typing of SNPs in the human apolipoprotein E gene. Genome Res 10:1532–1545

Nickerson DA, Taylor SL, Weiss KM, Clark AG, Hutchinson RG, Stengård J, Salomaa V, Vartiainen E, Boerwinkle E, Sing CF (1998) DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. Nat Genet 19:233–240

Nickerson DA, Tobe VO, Taylor SL (1997) PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. Nucleic Acids Res 25:2475–2451

Nordberg M, Tavaré S (2002) Linkage disequilibrium: what history has to tell us. Trends Genet 18:83–90

Ott J, Rabinowitz D (1997) The effect of marker heterozygosity on the power to detect linkage disequilibrium. Genetics 147:927–930

Owens K, King MC (1999) Genomic views of human history. Science 286:451–453

Pääbo S (1999) Human evolution. Trends Cell Biol 9:M13–M16

Perez-Lezaun A, Calafell F, Mateu E, Comas D, Ruiz-Pacheco R, Bertranpetit J (1997) Microsatellite variation and the differentiation of modern humans. Hum Genet 99:1–7

Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans: models and data. Am J Hum Genet 69:1–14

Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol Biol Evol 16: 1791–1798

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES (2001) Linkage disequilibrium in the human genome. Nature 411:199–204

Reich DE, Goldstein DB (1998) Genetic evidence for a Paleolithic human population expansion in Africa. Proc Natl Acad Sci USA 95:8119–8123

Relethford JH, Jorde LB (1999) Genetic evidence for larger African population size during recent human evolution. Am J Phys Anthropol 108:251–260

Rieder MJ, Taylor SL, Clark AG, Nickerson DA (1999) Sequence variation in the human angiotensin converting enzyme. Nat Genet 22:59–62

Rogers AR, Harpending HC (1992) Population growth makes waves in the distribution of pairwise genetic differences. Mol Biol Evol 9:552–569

Rogers AR, Jorde LB (1995) Genetic evidence on modern human origins. Hum Biol 67:1–36

Sachse C, Brockmöller J, Bauyer S, Roots I (1999) Functional significance of a C→A polymorphism in intron 1 of the cy-

tochrome P450 *CYP1A2* gene tested with caffeine. Br J Clin Pharmacol 44:445–449

Schneider S, Roessli D, Excoffier L (2000) Arlequin, version 2.000: a software for population genetics data analysis. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Geneva

Schulze TG, Schumacher J, Müller DJ, Krauss H, Alfter D, Maroldt A, Ahle G, Maroldt AO, Novo y Fernández A, Weber T, Held T, Propping P, Maier W, Nöthen MM, Rietschel M (2002) Lack of association between a functional polymorphism of the cytochrome P450 1A2 (CYP1A2) gene and tardive dyskinesia in schizophrenia. Am J Med Genet 105:498–501

Schweikl H, Taylor JA, Kitareewan S, Linko P, Nagorney D, Goldstein JA (1993) Expression of *CYP1A1* and *CYP1A2* genes in human liver. Pharmacogenetics 3:239–249

Seielstad M, Bekele D, Ibrahim M, Touré A, Traoré M (1999) A view of modern human origins from Y chromosome microsatellite variation. Genome Res 9:558–567

Shen P, Wang F, Underhill PA, Franco C, Yang WH, Roxas A, Sung R, Lin AA, Hyman RW, Vollrath D, Davis R, Cavalli-Sforza LL, Oefner PJ (2000) Population genetic implications from sequence variation in four Y chromosome genes. Proc Natl Acad Sci USA 97:7354–7359

Sherry ST, Harpending HC, Batzer MA, Stoneking M (1997) Alu evolution in human populations: using the coalescent to estimate effective population size. Genetics 147:1977–1982

Shriver MD, Jin L, Ferrell RE, Deka R (1997) Microsatellite data support an early population expansion in Africa. Genome Res 7:586–591

Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics 129:555–562

Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE, Jiang R, et al (2001) Haplotype variation and linkage disequilibrium in 313 human genes. Science 293: 489–493

Stoneking M (1997) Recent African origin of human mitochondrial DNA: review of the evidence and current status of the hypothesis. In: Donnelly PJ, Tavaré S (eds) Progress in population genetics and evolution. Springer-Verlag, New York

Tajima F (1983) Evolutionary relationships of DNA sequences in finite populations. Genetics 105:437–460

——— (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585–595

Takahata N (1993) Allelic genealogy and human evolution. Mol Biol Evol 10:2–22

Takahata N, Satta Y, Klein J (1995) Divergence time and population size in the lineage leading to modern humans. Theor Popul Biol 48:198–221

Terwilliger JD, Göring HH (2000) Gene mapping in the 20th and 21st centuries: statistical methods, data analysis and experimental design. Hum Biol 72:63–132

Thompson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW (2000) Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. Proc Natl Acad Sci USA 97:7360–7365

Tishkoff SA, Dietzsch E, Speed W, Pakstis AJ, Kidd JR, Cheung K, Bonné-Tamir B, Santachiara-Benerecetti AS, Moral P, Krings M (1996) Global patterns of linkage disequilibrium at the *CD4* locus and modern human origins. Science 271: 1380–1387

Tishkoff SA, Goldman A, Calafell F, Speed WC, Deinard AS, Bonné-Tamir B, Kidd JR, Pakstis AJ, Jenkins T, Kidd KK (1998) A global haplotype analysis of the myotonic dystrophy locus: implications for the evolution of modern humans and for the origin of myotonic dystrophy mutations. Am J Hum Genet 62:1389–1402

Tishkoff SA, Pakstis AJ, Ruano G, Kidd KK (2000) The accuracy of statistical methods for estimation of haplotype frequencies: an example from the *CD4* locus. Am J Hum Genet 67:518–522

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonné-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ (2000) Y chromosome sequence variation and the history of human populations. Nat Genet 26:358–361

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. Science 253:1503–1507

Wall JD, Przeworski M (2000) When did the human population size start increasing? Genetics 155:1865–1874

Watkins WS, Ricker CE, Bamshad MJ, Carroll ML, Nguyen V, Batzer MA, Harpending HC, Rogers AR, Jorde LB (2001) Patterns of ancestral human diversity: an analysis of *Alu*-insertion and restriction-site polymorphisms. Am J Hum Genet 68:738–752

Watson E, Forster P, Richards M, Bandelt HJ (1997) Mitochondrial footprints of human expansions in Africa. Am J Hum Genet 61:691–704

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:256–276

Wise CA, Sraml M, Easteal S (1998) Departure from neutrality at the mitochondrial NADH dehydrogenase subunit 2 gene in humans, but not in chimpanzees. Genetics 148:409–421

Yu N, Chen FC, Ota S, Jorde LB, Pamilo P, Patthy L, Ramsay M, Jenkins T, Shyue SK, Li WH (2002) Larger genetic differences within Africans than between Africans and Eurasians. Genetics 161:269–274

Yu N, Zhao Z, Fu YX, Sambuughin N, Ramsay M, Jenkins T, Leskinen E, Patthy L, Jorde L, Kuromori T, Li WH (2001) Global patterns of human DNA sequence variation in a 10-kb region on chromosome 1. Mol Biol Evol 18:214–222

Zhao Z, Jin L, Fu YX, Ramsay M, Jenkins T, Leskinen E, Pamilo P, Trexler M, Patthy L, Jorde LB, Ramos-Onsins S, Yu N, Li WH (2000) Worldwide DNA sequence variation in a 10-kilobase noncoding region on human chromosome 22. Proc Natl Acad Sci USA 97:11354–11358

Zhivotovsky LA, Bennett L, Bowcock AM, Feldman MW (2000) Human population expansion and microsatellite variation. Mol Biol Evol 17:757–767

Zietkiewicz E, Votova V, Jarnik M, Koran-Laskowska M, Kidd KK, Modiano D, Scozzari R, Stoneking M, Tishkoff S, Batzer M, Labuda D (1998*a*) Genetic structure of the ancestral population of modern humans. J Mol Evol 47: 146–155