Population genetic implications from sequence variation in four Y chromosome genes

Peidong Shen*†, Frank Wang*, Peter A. Underhill‡, Claudia Franco*, Wei-Hsien Yang*, Adriane Roxas*, Raphael Sung*, Alice A. Lin‡, Richard W. Hyman*, Douglas Vollrath‡, Ronald W. Davis*, L. Luca Cavalli-Sforza‡, and Peter J. Oefner*

*Stanford DNA Sequencing and Technology Center, 855 California Avenue, Palo Alto, CA 94304; and ‡Department of Genetics, Stanford University, Stanford, CA 94305

Contributed by L. Luca Cavalli-Sforza, April 4, 2000

Some insight into human evolution has been gained from the sequencing of four Y chromosome genes. Primary genomic sequencing determined gene *SMCY* to be composed of 27 exons that comprise 4,620 bp of coding sequence. The unfinished sequencing of the 5′ portion of gene *UTY1* was completed by primer walking, and a total of 20 exons were found. By using denaturing HPLC, these two genes, as well as *DBY* and *DFFRY*, were screened for polymorphic sites in 53–72 representatives of the five continents. A total of 98 variants were found, yielding nucleotide diversity estimates of 2.45×10^{-5}, 5.07×10^{-5}, and 8.54×10^{-5} for the coding regions of *SMCY*, *DFFRY*, and *UTY1*, respectively, with no variant having been observed in *DBY*. In agreement with most autosomal genes, diversity estimates for the noncoding regions were about 2- to 3-fold higher and ranged from 9.16×10^{-5} to 14.2×10^{-5} for the four genes. Analysis of the frequencies of derived alleles for all four genes showed that they more closely fit the expectation of a Luria–Delbrück distribution than a distribution expected under a constant population size model, providing evidence for exponential population growth. Pairwise nucleotide mismatch distributions date the occurrence of population expansion to ≈28,000 years ago. This estimate is in accord with the spread of Aurignacian technology and the disappearance of the Neanderthals.

The human Y chromosome consists of a nonrecombining region (NRY), which makes up 95% of its length, flanked by pseudoautosomal regions (1). Aside from the absence of recombination, NRY differs from all other nuclear human chromosomes by its presence in males only, its common ancestry and persistent meiotic relationship with the X chromosome, and the tendency of its genes to degenerate during evolution. Other than for its role in male sex determination, mediated by the *SRY* gene (2), the Y chromosome has been perceived as functionally desolate. Recently, 20 genes or gene families have been identified in the NRY (3). They fall into two classes. One class of nine genes is expressed in many organs and has X homologs that escape X inactivation. Among them are the single-copy genes *DBY*, *DFFRY*, and *UTY*, all of which have been mapped into the deletion interval 5C, and *SMCY*, located in the interval 5O (3). Most of the other NRY genes are highly duplicated, including the 11 NRY genes that are expressed specifically in testes. By deletion mapping, putative roles in sex determination, germ-cell tumorigenesis, determination of stature, and spermatogenesis have been established for most of the NRY genes (3). In particular, deletion of the *AZFa* region that contains *DBY*, *DFFRY*, and *UTY* has been shown to disrupt spermatogenesis, causing infertility in otherwise healthy men (4). *SMCY*, on the other hand, has been found to contain the human H-Y epitope, H-Y/HLA-B7, which alone or in part accounts for the male-specific transplantation antigen (5). *SMCY*'s 65% homology to the retinoblastoma 2 gene suggests that the gene may code for a transcription factor (5).

The NRY has been also considered devoid of sequence variation (6). However, recently, biallelic variation has been revealed, including three variants in an 18.3-kilobase (kb) stretch of the *SRY* sequence (7), two variants within 949 bp of the *RPS4Y* gene (8), and five variants in the *DFFRY* gene (9). In other studies that used sequenced-tagged sites dispersed throughout the NRY (10) as templates for PCR, direct sequencing alone or in combination with DHPLC led to the discovery of more than 30 biallelic sites in a wide range of ethnicities (11, 12).

The present study reports a total of 98 simple sequence polymorphisms that were discovered by DHPLC in the coding and intronic regions of the *DBY*, *DFFRY*, *SMCY*, and *UTY1* genes screened in a set of 53–72 males representing as many as 46 diverse populations from the five continents. We also provide evidence for exponential population growth and its onset with the spread of Aurignacian technology.

Materials and Methods

DNA Samples. The continental and ethnic affiliations of the 53–72 males used for screening *SMCY*, *DBY*, *DFFRY*, and *UTY1* are given in Tables 1 and 3. All samples were collected according to approved human subject protocols.

Reference DNA Sequences. A contiguous 75,850-bp reference sequence containing the *SMCY* gene was obtained by joining two cosmid sequences (GenBank accession nos. AC003031 and AC003094), which were shotgun sequenced at the Stanford DNA Sequencing and Technology Center. The REPEATMASKER2 program (http://ftp.genome.washington.edu) was used to identify human repeat DNA sequences. The respective reference genomic and cDNA sequences of *DFFRY*, AC002531, and AF000986 and of *DBY*, AC004474, and AF000985 were downloaded from GenBank, as was the genomic sequence of the first 16 exons of *UTY1* (AC006376). A 9-kb sequence gap containing exons 17–20 was closed by long-range PCR (PCR SuperMix High Fidelity kit, Life Technologies, Rockville, MD) by using exon 16 and the 3′ untranslated region (AF000996) as priming sites.

PCR. Primers designed for *SMCY* covered all unique sequences and repeat elements other than long interspersed elements (LINEs), yielding amplicons 300–500 bp in length (http://www.ncbi.nlm.nih.gov/SNP). For *DBY*, *DFFRY*, and *UTY1*, only the coding and flanking noncoding sequences were amplified. The PCR protocol comprised an initial denaturation at 95°C for 10 min to activate AmpliTaq Gold, 14 cycles of denaturation at 94°C for 20 s, primer annealing at 63–56°C with 0.5°C decrements, and extension at 72°C for 1 min, followed by 20 cycles at 94°C for 20 s, 56°C for 1 min, and 72°C for 1 min, and a final 5-min extension at 72°C. Each 50-μl PCR contained 1 unit of AmpliTaq Gold polymerase, 10 mM

Abbreviation: NRY, nonrecombining region of the Y chromosome; DHPLC, denaturing HPLC; kb, kilobase.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AC003031, AC003094, AF265575, and AF273841)

See commentary on page 6927.

†To whom reprint requests should be addressed. E-mail: shen@genome.stanford.edu.

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

Table 1. Nucleotide position, defined with respect to the GenBank entry U52195, and geographic distribution of the 47 SMCY polymorphisms detected in 53 representatives from Africa, Asia, Europe, America, and Oceania

Population	Haplotype	1-31760	1-31474	1-29454	1-26251	1-24954	1-23603	1-23580	1-19429	1-18817	1-17646	1-10620	1-842	1-401	425+162	425+354	504-280	627-9	932+565	933-726
Orangutan		-	G	-	A	-	A	5bp	-	-	G	G	9A	A	-	-	A	T	-	-
Gorilla		-	G	G	A	T	A	5bp	C	A	G	G	7A	A	A	A	A	T	-	4A
Chimpanzee		G	G	G	A	T	A	5bp	C	A	G	G	9A	A	A	A	A	T	T	4A+C+4A
Human, ancestral		G	G	G	A	T	A	5bp	C	A	G	G	9A	A	A	A	A	T	T	9A
AF09 Khoisan	1	.	.	.	.	.	.	.	.	.	.	.	8A	.	.	.	.	.	.	.
AF11 Surma	2	.	.	.	.	.	.	.	.	G	.	.	8A	.	.	.	.	.	.	.
AF01, AF08, OC02†	3	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AF02, AF03, AF05‡	4	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AF07 Lisongo	5	.	T	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AF04 Mbuti Zaire Pygmy	6	.	T	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.
AF06 Lisongo	7	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AF12 Bozo	8	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AF13 Tuareg	9	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AS18 Sindhi	10	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	8A
AS21 Kazak	11	.	T	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.
AS06 Japanese	12	.	T	.	.	.	G	.	.	.	.	.	.	.	.	.	.	C	.	.
AS07 Japanese	13	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AF10 and others§	14	.	T	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.
AS10 and others¶	15	.	T	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.
AS19 Arab, Central Asia	16	.	T	.	.	.	.	.	.	.	.	A	.	.	G	.	.	C	.	.
AS20 Arab, Central Asia	17	.	T	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.
AS22 Uzbek	18	.	T	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.
EU04 Italian	19	.	T	.	.	.	.	O	.	.	.	A	.	.	.	.	.	.	.	.
AS02, AS11, AS17‖	20	.	T	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.
OC03, 04 Melanessian	21	.	T	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.
OC01 New Guinean	22	.	T	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.
EU05, 06 Basque	23	.	T	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.
AS01 Cambodian	24	A	T	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.
AS08 Han Chinese	25	A	T	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	C	.
AS15 Brahul	26	.	T	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.
EU01 Sardinian	27	.	T	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.
AS14 Baloochi	28	.	T	.	.	.	.	.	.	.	.	A	.	.	.	.	G	.	.	.
EU02 Italian	29	.	T	.	.	.	.	.	G	.	.	A	.	.	.	.	.	.	.	.
EU03 German	30	.	T	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.
AS03 Atayal	31	.	T	A	.	.	.	.	.	.	.	A	.	G	.	.	.	.	.	.

Population	Haplotype	1208+977	1208+1022	1367+23	1367+63	1368-124	1368-7	1487+51	1551*	1646+1086	1646+3481	1646+3891	1646+4783	1991+792	1991+1185	1992-207	2306+93	2762-930	2762-528	2762-380	3356+123	4043	4298	4344+8	4989UTR	5467+468	5467+817	5467+884	5467+1487
Orangutan		-	-	G	T	-	A	T	G	-	C	C	-	-	T	T	O	-	T	A	C	C	A	A	C	T	G	T	A
Gorilla		-	-	G	T	AT	A	T	G	G	T	C	AC	7T	T	T	O	G	T	G	C	C	A	A	C	T	-	-	C
Chimpanzee		C	G	G	T	AT	A	T	G	G	C	C	AC	TA+7T	T	T	O	G	T	A	T	C	A	A	C	T	G	T	C
Human, ancestral		C	G	G	T	AT	A	T	G	G	C	C	AC	8T	T	T	O	G	T	A	C	C	A	A	C	T	G	T	C
AF09 Khoisan	1	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.
AF11 Surma	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.
AF01, AF08, OC02†	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
AF02, AF03, AF05‡	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	A	.	.	.
AF07 Lisongo	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	A	.	.	.
AF04 Mbuti Zaire Pygmy	6	.	T	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	T	.	.	.	A	.	.	.
AF06 Lisongo	7	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	A	.	.	.
AF12 Bozo	8	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
AF13 Tuareg	9	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
AS18 Sindhi	10	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
AS21 Kazak	11	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
AS06 Japanese	12	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	.	.
AS07 Japanese	13	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	.	.
AF10 and others§	14	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
AS10 and others¶	15	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	T	A	.	.	.
AS19 Arab, Central Asia	16	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	T	A	.	.	.
AS20 Arab, Central Asia	17	.	.	.	.	.	.	.	.	.	.	T	O	.	.	.	.	.	.	.	.	.	.	.	T	A	.	.	.
AS22 Uzbek	18	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	A	.	.	.	.	.	T	A	.	.	.
EU04 Italian	19	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	T	A	.	.	.
AS02, AS11, AS17‖	20	.	.	.	.	O	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
OC03, 04 Melanessian	21	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	C	.
OC01 New Guinean	22	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	A	.	C	.
EU05, 06 Basque	23	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	T
AS01 Cambodian	24	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
AS08 Han Chinese	25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
AS15 Brahul	26	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
EU01 Sardinian	27	.	.	.	.	.	.	.	.	.	.	.	.	9T	.	A	.	.	.	.	.	.	.	.	.	A	.	.	.
AS14 Baloochi	28	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	A	.	.	.
EU02 Italian	29	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
EU03 German	30	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
AS03 Atayal	31	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	A	.	.	.

AF, Africa; AS, Asia; EU, Europe; AM, America; OC, Oceania; -, sequence not available.
*Nonsynonymous substitution.
†AF01, Biaka Pygmy; AF08, Khoisan; OC02, New Guinean.
‡AF02, 03, Biaka Pygmy; AF05, Mbuti Zaire Pygmy.
§AF10, Berta; AS04, Ami; AS05, Japanese; AS09, Han Chinese; AS12, 13, Pathan; AS16, Sindhi; AS23, Druze; OC06, Australian Aborigine.
¶AS10, Brushaski; EU07, Basque; OC05, Australian Aborigine; AM01, Karitiana; AM02, Surui; AM03, Colombian Indian; AM04, Mayan Indian.
‖AS02, Cambodian; AS11, Brushaski; AS17, Sindhi.

Tris·HCl (pH 8.3), 50 mM KCl, 2.5 mM $MgCl_2$, 0.1 mM each of the four deoxyribonucleotide triphosphates, 0.2 μM each of forward/reverse primers, and 50 ng of genomic DNA. PCR yields were determined semiquantitatively on ethidium bromide-stained agarose gels.

DHPLC Analysis. Unpurified PCR products were mixed at an equimolar ratio with a reference Y chromosome and subjected to a 3-min 95°C denaturing step followed by gradual reannealing from 95°C to 65°C over 30 min. Each mixture (10 μl) was loaded onto a DNASep column (Transgenomic, San Jose, CA), and the amplicons were eluted in 0.1 M triethylammonium acetate (pH 7) with a linear acetonitrile gradient at a flow rate of 0.9 ml/min (11). Under appropriate temperature conditions, which were optimized by computer simulation (available at http://insertion.stanford.edu/melt.html), mismatches were recognized by the appearance of two or more peaks in the elution profiles.

DNA Sequencing. Polymorphic and reference PCR samples were purified with Qiagen (Valencia, CA) QIAquick spin columns and sequenced with the PE Biosystems (Foster City, CA) Dye Terminator Cycle Sequencing Kit and a model 373A DNA sequencer. Chimpanzee, gorilla, and orangutan samples were sequenced for the entire 4.62-kb coding region of *SMCY* and, whenever possible, for those NRY segments that were polymorphic in humans.

EVOLUTION

Statistical Analysis. The program DNAML in PHYLIP (version 3.57c) was used to construct phylogenetic networks. Insertions and deletions were treated as single-nucleotide substitutions. The estimated nucleotide diversity, π, in a sample of n chromosomes was calculated from the equation

$$\pi = \sum_{i<j} \pi_{ij}/n_c\,, \qquad [1]$$

where π_{ij} is the number of nucleotide differences between the ith and jth DNA sequences and $n_c = n(n-1)/2$ (13).

Tajima's D statistic (14) is defined by the equation

$$D = d/\sqrt{\hat{V}(d)}, \quad \text{where } d = \pi - S/\sum_{i=1}^{n-1} 1/i\,, \qquad [2]$$

S is the number of segregating nucleotides, and $\hat{V}(d)$ an estimator of the variance of d. The value of D is 0 for selectively neutral mutations in a constant population infinite sites model. A negative value of D indicates either purifying selection or population expansion (15).

The expected Luria–Delbrück/Lea–Coulson (16, 17) distribution of the number of mutants for each gene was fitted by maximum likelihood, treating each nucleotide of the screened sequence as analogous to a parallel, independent bacterial culture. The distributions under the expectation of constant population size were calculated according to the method of Watterson (18).

For the mismatch distribution, the mean number of differences (m) between pairs of chromosomes and variance (ν) of this number were computed for each gene. The corresponding parameters were θ ($\theta = 2\mu N_0$, where N_0 is the effective population size before a single instantaneous expansion), μ (the mutation rate for each gene), and τ (time in mutational units since the expansion of the population, assuming a generation length of 25 years). For estimation of θ and τ, we used the following equations (see ref. 19)

$$\hat{\theta} = (\nu - m)^{1/2} \qquad [3]$$

and

$$\hat{\tau} = m - \hat{\theta}. \qquad [4]$$

The standard errors for $\hat{\tau}$ were estimated by simulation (20), averaging the mean number of differences in 1,000 coalescent trees, each of which was constructed randomly with a sample size and an estimate of θ matched to the observed distribution. Errors from estimating θ itself and any uncertainties from the estimation of mutation rates are not included. The actual time (T) since the expansion was estimated by $\hat{\tau}/2\,\mu l$, where μ is the mutation rate (based on the number of sequence differences observed between human and chimpanzee and assuming a divergence time of 4.9 million years) and l is the number of base pairs tested. The mutation rate per nucleotide for *SMCY* was 1.53×10^{-9}, calculated on the basis of 597 observed differences between human and chimpanzee over 39,931 bp, whereas that of *DBY*, *DFFRY*, and *UTY1* was 1.25×10^{-9} (470 differences over 38,468 bp).

Results and Discussion

SMCY. The genomic structure of *SMCY* is shown in Fig. 1. The gene consists of 27 exons. The shadowed areas (35.9 kb) represent LINE elements that have been excluded from DHPLC analysis. A total of 39,931 bp (4,620 bp coding) was screened in 53 human male samples from the five continents, and 47 polymorphic sites (Table 1) were detected: 30 (64%) transitions, 10 (21%) transversions, and 7 (15%) deletions or insertions including two runs of As or Ts, respectively. Three variants in the coding region (exon 11, G1,551A, Val426Met; exon 24, C4,043T, Pro1,256Pro; exon 25, A4,298G, Gly1,341Gly) were observed only once in a Sindhi, Zaire Pygmy, and Khoisan, respectively. One site was recurrent, namely T627-9C in intron 4. Of the 47 biallelic sites, 15, 19, 5, and 2 were observed exclusively in Africa, Asia, Europe, and Oceania, respectively. Three polymorphisms were found in all five continents. Two additional sites were detected everywhere except in Africa, and one was present in both a Sardinian and a Baloochi.

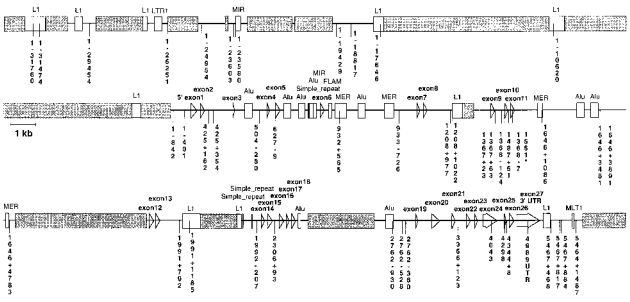

Fig. 1. Genomic structure of *SMCY* and the nucleotide positions of 47 sequence variants detected in 53 individuals. *SMCY* consists of 27 exons spaced over 39.5 kb, 11.6% of which are coding. The AUG start codon is located in exon 2. Shadowed areas (35.9 kb) represent LINE elements that were excluded from the variation search. The numbers indicate the nucleotide positions of the polymorphic sites with respect to the published cDNA sequence (GenBank accession no. U52195).

The 47 polymorphisms were used to infer a phylogenetic network comprising 31 distinct haplotypes (Table 1 and Fig. 2*A*). Asia and Africa have the most haplotypes with 15 and 10, respectively, followed by Europe, Oceania, and America with 6, 5, and 1, respectively. The root of the network, based on primate sequence data, is located in Africa, with the most ancestral patrilineages represented today by a Khoisan and an East African Surma. As suggested by the three clades centered at haplotypes 3, 14 and 15, respectively, anatomically modern humans migrated in, at least, two waves from Africa into Asia and New Guinea, followed by a later dispersal probably out of

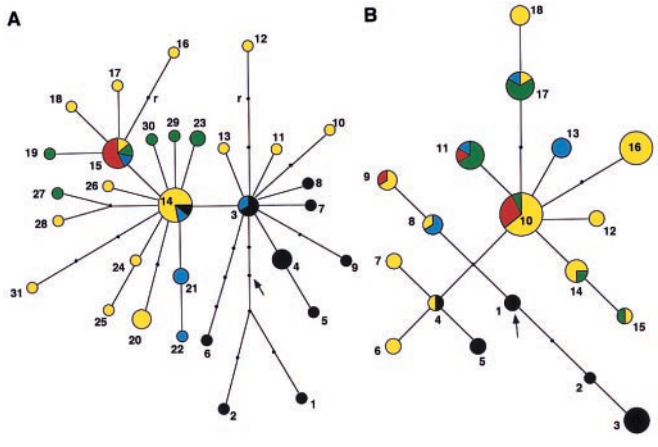

Fig. 2. Maximum likelihood networks of the *SMCY* haplotypes (*A*) and the major haplotypes constructed from the combined variant sites in *DFFRY*, *DBY*, and *UTY1* (*B*). Haplotype numbers are from Tables 1 and 3. The representations of color are black, Africa; yellow, Asia; green, Europe; red, America; and blue, Oceania. The areas of the circles represent the number of individuals carrying each haplotype. The arrows indicate the location of the most likely root in the phylogenies. The occurrence of a single recurrent mutation (r) did not generate any ambiguity in the parsimonious mutational pathway when considered in the context of other polymorphisms.

Table 2. Number of sequence differences between humans and apes in coding and flanking noncoding sequences of SMCY

Species	Coding (4,620 bp)*	Noncoding (4,758 bp)
Human/chimpanzee	29 (51.7)	79
Human/gorilla	41 (70.7)	109
Human/orangutan	89 (69.7)	258
Chimpanzee/gorilla	40 (55.0)	130
Chimpanzee/orangutan	94 (64.9)	275
Gorilla/orangutan	95 (67.4)	282

*The percentage of silent mutations is given in parentheses.

Western Asia into Central Asia, Europe, and the Americas. Independent corroborating evidence for the occurrence of these three major prehistoric migratory episodes has been derived from the geographic distribution of haplotypes from a chromosome 21 region (21). A detailed coalescence analysis and dating of the key mutations in the *SMCY* phylogeny is provided in the companion paper by Thomson *et al.* (22).

The three primates (chimpanzee, gorilla, and orangutan) were sequenced over the entire coding region of *SMCY* (4,620 bp) and 4,758 bp of flanking noncoding sequences. In the coding region, humans differ in a total of 125 positions (available on request) from apes, 80% of which are transitions. Nucleotide position 3,917 is triallelic (C in human and chimpanzee, A in gorilla, and G in orangutan). There are four nonsynonymous differences separating the human from all great apes, only one of which resulted in a nonconservative amino acid change (Val1,186Ala). The two transitions and two transversions were found in exon 3 (A480G, Val69Ile), exon 7 (A963C, His230Asn), exon 10 (C1,423G, Ser383Thr), and exon 24 (C3,832T, Val1,186Ala). Generally, nonsynonymous differences are more frequent in the C-terminal part of the gene. No nonsynonymous change was detected in the H-Y epitope region (nucleotides 3,160–3,200). Table 2 summarizes the number of sequence differences found among the four species. Assuming 4.9 million years for the divergence time of human and chimpanzee, the divergence time of human and gorilla was estimated from the log proportion of identical nucleotides as 7.0 and 7.4 million years for coding and noncoding segments, respectively. The corresponding divergence times for human and orangutan were 16.2 and 16.9 million years. These numbers are in fairly good agreement with published divergence times of 6.6 and 13.0 million years, respectively, for gorilla and orangutan based on mitochondrial DNA substitution rates (23).

***DFFRY*, *DBY*, and *UTY1*.** Table 3 lists the polymorphic sites identified in *DFFRY*, *DBY*, and *UTY1* in 70 (72 for *UTY1*) Y chromosomes representing 41 ethnic groups. For *DFFRY*, 17 polymorphisms were discovered over 16,699 bp (7,668 bp coding). Seven of the variant sites are located in exons 6, 11, 18, 22, 30, 33, and 37. Three are nonsynonymous, resulting in two nonconservative (exon 6, C2,295T, Arg211Cys; exon 22, G4,842A, Ala1,060Thr) and one conservative amino acid change (exon 33, G6,777T, Ala1,705Ser). G4,842A had been already observed three times in 576 infertile men and once in 96 fertile men (9). In the present study, G4,842A was detected only once in an individual from Pakistan. The other three polymorphisms reported by Sun *et al.* (9) were not observed in our populations.

For *DBY*, a total of 14 polymorphic sites were found over 9,000 bp (1,983 bp coding), and they are all noncoding. This observation is in concordance with the study of Sun *et al.* (9), who also failed to detect a single coding variant. For *UTY1*, 20 polymorphic sites were discovered over 15,317 bp (3,240 bp coding). Two sites, both synonymous, are located in exons 1 and 14. By combining the polymorphisms observed in *DFFRY*, *DBY*, and *UTY1* and excluding those observed only once, 18 major haplotypes could be inferred, most of them found in Asia ($n = 10$) and Africa ($n = 5$). The similarity of the phylogeny of the three genes (Fig. 2*B*) that are located within the same deletion interval on the NRY (5C) to that of *SMCY* (deletion interval 5O) attests to the coherence of the NRY, which, therefore, can be viewed evolutionarily as if it were a single locus. For instance, among the individuals present in both ascertainment sets, the same Asian individuals (As06, As07, and As18) and the New Guinean (OC02) cluster with the Africans, supporting the hypothesis of at least two distinct migratory waves out of Africa. A more detailed analysis of *DBY* and *DFFRY* including coalescence time estimates is provided in the companion paper by Thomson *et al.* (22).

Nucleotide Diversity and Tajima's *D* Values. The following nucleotide diversity estimates ($\pm$ SEM) were obtained for the coding and noncoding segments, respectively, of the four NRY genes investigated: *DBY*, 0 and $10.5 \pm 8.75 \times 10^{-5}$; *DFFRY*, $5.07 \pm 5.34 \times 10^{-5}$ and $9.85 \pm 7.72 \times 10^{-5}$; *SMCY*, $2.45 \pm 4.44 \times 10^{-5}$ and $9.16 \pm 5.33 \times 10^{-5}$; and *UTY1*, $8.54 \pm 10.3 \times 10^{-5}$ and $14.2 \pm 9.29 \times 10^{-5}$. Both the higher than expected percentage of synonymous mutations (67%) and the approximately 2-fold lower nucleotide diversity in coding versus noncoding segments imply selection against deleterious alleles. However, the ratios of nucleotide diversity estimates for coding and noncoding segments on the NRY are very similar to those observed for 16 autosomal genes, with an arithmetic mean of 2.6 and a range of 0.63 to 6.67. Further, the average nucleotide diversity for the coding regions of the four NRY genes (0.51×10^{-4}) is about five times lower than the corresponding estimate of $2.62 \pm 0.51 \times 10^{-4}$ (based on 223 polymorphic sites over 58,093 bp) for the 16 autosomal genes *AMPD1*, *ATM*, *BRCA1*, *BRCA2*, *COL*, *COX2*, *CYP1A1*, *EA2*, *FBN1*, *IL2*, *IL4*, *MTE*, *STR*, *Tubulin M40*, *Werner*, and *XRCC1*. A similar ratio was obtained comparing the noncoding regions of the same 16 autosomal genes ($5.07 \pm 0.55 \times 10^{-4}$, based on 479 polymorphic sites over 79, 445 bp) to the 4 NRY genes (0.99×10^{-4}). The ratio is somewhat lower than expected given that the effective population size of autosomes relative to the Y chromosome is only 4-fold greater. To test for reduced variability on the NRY, a pairwise comparison of loci linked to noncoding segments of the NRY and the autosomal genes was performed with the Hudson–Kreitman–Aguadé test (24) and chimpanzee sequences for homologous genes (data not shown). The NRY clearly shows evidence of reduced variability to the autosomes ($\chi^2_{[1]} = 4.0$; $P = 0.045$), confirming similar comparisons of smaller numbers of polymorphisms on previously reported NRY sequences with X-linked human genes (25, 26). Possible explanations include not only selection on NRY but also differences between male and female effective population sizes and/or in the dispersal of males and females.

The Tajima's *D* values for the combined coding and noncoding segments of *DBY*, *DFFRY*, *SMCY*, and *UTY1* were -2.15 ($P < 0.05$), -1.89, -2.31 ($P < 0.05$), and -1.57, respectively. At first glance, these values suggest purifying selection on the human Y chromosome. However, unpublished Tajima's *D* values obtained for mtDNA (-2.11; $P < 0.05$), *AMPD1* (-2.07; $P < 0.05$), *ATM* (-2.25; $P < 0.05$), *BRCA2* (-2.10; $P < 0.05$), *COL* (-1.19), *COX2* (-2.16; $P < 0.05$), *CYP1A1* (-2.20, $P < 0.05$), *EA2* (-1.72), *FBN1* (-1.80), *IL4* (-1.19), *MTE* (-1.71), *STR* (-1.09), *Tubulin M40* (-1.21), *XRCC1* (-1.45), *DIAPH2* (-1.36), and *MCM* (-0.96) are also all negative. Other than being characteristic of selection, negative *D* values are also characteristic of growing populations where initially rare alleles have a stronger effect on S than on π (15). The relative homogeneity of values obtained for 20 genes of four different genetic systems is therefore more in agreement with population growth, which will affect all genes equally. Signatures of population expansion were also reported for microsatellite repeat data (27).

EVOLUTION

Table 3. Variable sites in 72 individuals for *DFFRY*, *DBY*, and *UTY1*, defining 18 major haplotypes, not including polymorphic sites observed once

		DFFRY†																
			1		2	2		3			5	5	6				8	9
	H	1	5	2	4	9	3	0	4	4	4	9	0	6	6	7	8	1
	a	2	4	2	3	8	0	8	2	8	7	0	5	1	7	9	8	9
	p	7	4	9	7	1	7	7	9	4	8	3	0	5	7	9	6	4
	l	7	+	5	+	+	7	-	2	2	-	-	+	5	7	1	-	+
	o	U	7	*	1	9		4		*	1	1	9		*		3	4
	t	T			3	8		6			0	0					1	8
	y	R			1						6	9						
	p																	
Population	e																	
Orangutan		-	-	C	-	-	T	-	A	G	T	-	-	-	G	T	-	T
Gorilla		-	-	C	-	-	-	-	-	G	T	-	-	-	-	T	-	-
Chimpanzee		T	C	C	-	T	T	C	A	G	T	-	-	C	G	T	A	T
Human, ancestral		T	C	C	A	T	T	C	A	G	C	G	O	C	G	T	A	T
AF09 Khoisan	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AF15 Falasha, Ethiopia	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
AF14 Nuba	2	.	.	.	.	.	C	T	.	.	.	.	.	.	.	.	.	.
AF02 Biaka Pygmy	3	.	.	.	.	.	C	T	.	.	.	.	.	.	.	.	.	.
AF03 Biaka Pygmy	3	.	.	.	.	.	C	T	.	.	.	.	.	.	.	.	.	.
AF05 Mbuti Zaire Pygmy	3	C	.	.	.	.	C	T	.	.	.	.	.	.	.	.	.	.
AF07 Lissongo	3	.	.	.	.	.	C	T	.	.	.	.	.	.	.	.	.	.
AF04 Mbuti Zaire Pygmy	4	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AS37 Bedouine	4	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AF01 Biaka Pygmy	5	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.
AF06 Lissongo	5	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.
AF16 Nuba‡	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AS06 Japanese	6	.	T	T	.	.	.	.	.	.	.	.	.	.	.	C	.	.
AS33 Hmong, Laos	6	.	T	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.
AS18 Sindhi	7	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AS38 Bedouine	7	.	T	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.
AS07 Japanese	8	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.
OC02 New Guinean	8	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
OC07 Australian Aborigine	8	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AS28, AS30, AM06§	9	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AS02 and others¶	10	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AM02 Surul	10	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AM04 Mayan	10	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AS11 Brushaski	10	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AS12 Pathan	10	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AS15 Brahui	10	.	T	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.
EU03 and others‖	11	.	T	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.
AS16, AS25**	12	.	T	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.
OC01 New Guinean	13	.	T	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.
OC03, OC04 Melanessian	13	.	T	.	.	.	.	.	.	.	.	O	.	.	T	.	.	.
AS14, AS29, AS35, EU08††	14	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
AS40 Yemenite	15	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
EU02 Italian	15	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
AS01 and others‡‡	16	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AS09 Han Chinese	16	.	T	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.
EU04 and others§§	17	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EU13 Italian‡	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AS10, AS24¶¶	18	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AS39 Arab, Central Asia	18	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

		DBY†													
									1	1	2	2	2	2	2
	H					1	6	7	0	9	3	3	3	3	3
	a					7	0	4	9	7	1	1	1	1	1
	p	1	1	1	1	5	9	5	1	4	2	2	2	2	2
	l	-	-	-	-	+	-	-	-	+	+	+	+	+	+
	o	2	2	2	2	1	3	5	1	3	3	4	8	1	1
	t	5	2	2	4	2	2	2	3	8	9	6	1	0	1
	y	1	7	2	0	0					0	0	9	9	3
	p	5	2	7										9	0
Population	e														
Orangutan		C	O	T	G	T	T	-	G	-	-	-	-	A	G
Gorilla		-	-	-	-	-	-	-	-	T	-	-	-		-
Chimpanzee		C	O	T	A	T	T	A	G	T	C	C	O	A	G
Human, ancestral		C	O	T	A	T	T	A	G	T	C	C	O	A	G
AF09 Khoisan	1	.	.	.	.	.	.	.	.	.	G	.	.	.	.
AF15 Falasha, Ethiopia	1	.	.	.	.	.	.	.	.	G	.	.	.	.	.
AF14 Nuba	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AF02 Biaka Pygmy	3	T	.	.	.	.	.	.	.	.	.	.	.	.	.
AF03 Biaka Pygmy	3	T	.	.	.	.	.	.	.	.	.	.	.	.	.
AF05 Mbuti Zaire Pygmy	3	T	.	.	.	.	.	.	.	.	.	.	.	.	.
AF07 Lissongo	3	T	.	.	.	.	.	.	.	.	.	.	.	.	.
AF04 Mbuti Zaire Pygmy	4	.	.	.	.	.	.	.	.	.	.	.	.	.	A
AS37 Bedouine	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AF01 Biaka Pygmy	5	.	.	.	.	G	.	.	.	.	.	.	.	.	.
AF06 Lissongo	5	.	.	.	G	.	.	.	.	.	.	.	.	.	.
AF16 Nuba‡	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AS06 Japanese	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AS33 Hmong, Laos	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AS18 Sindhi	7	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AS38 Bedouine	7	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AS07 Japanese	8	.	.	.	.	.	.	.	.	.	.	.	.	.	.
OC02 New Guinean	8	.	.	.	.	.	.	.	.	.	.	.	.	.	.
OC07 Australian Aborigine	8	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AS28, AS30, AM06§	9	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AS02 and others¶	10	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AM02 Surul	10	.	.	.	.	.	.	.	.	.	.	.	G	.	.
AM04 Mayan	10	.	.	C	.	.	.	.	.	.	.	.	.	.	.
AS11 Brushaski	10	.	.	.	.	.	C	.	.	.	.	.	.	.	.
AS12 Pathan	10	.	.	.	.	.	.	.	T	.	.	.	.	G	.
AS15 Brahui	10	.	4bp	.	.	.	.	.	.	.	.	.	.	.	.
EU03 and others‖	11	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AS16, AS25**	12	.	.	.	.	.	.	.	.	.	.	.	.	.	.
OC01 New Guinean	13	.	.	.	.	.	.	.	.	.	.	.	.	.	.
OC03, OC04 Melanessian	13	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AS14, AS29, AS35, EU08††	14	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AS40 Yemenite	15	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EU02 Italian	15	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AS01 and others‡‡	16	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AS09 Han Chinese	16	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EU04 and others§§	17	.	.	.	.	.	.	C	.	.	.	.	.	.	.
EU13 Italian‡	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AS10, AS24¶¶	18	.	.	.	.	.	.	C	.	.	.	T	.	.	.
AS39 Arab, Central Asia	18	.	.	.	.	.	.	C	.	.	.	T	.	.	.

		UTY1†																			
			1	1	1	1	1	1	1	1	1	1	1	1		3	3	3	3	3	3
	H	1	1	2	2	3	3	3	3	3	3	3	6	9	2	6	6	6	6	6	7
	a	0	5	2	2	3	3	3	3	8	8	8	1	7	3	7	7	7	7	7	8
	p	1	8	1	1	0	0	0	0	0	1	1	5	1	5	8	8	9	9	9	4
	l	4	-	+	+	+	+	+	+	+	-	-	+	-	8	+	+	-	-	-	+
	o		4	3	3	1	5	6	6	1	2	7	8	6		5	7	2	2	3	1
	t			6	6	8	7	2	2	6	2	8	8	0		3	6	8	9	2	6
	y			2	7		9	1	2	2						7	8	1	4	9	5
	p			4	1		8	1	1	8											
Population	e									3											
Orangutan		C	A	-	-	T	-	-	-	-	-	-	5bp	-	T	G	T	G	-	-	-
Gorilla		C	-	-	-	T	-	-	-	-	G	-	-	-	-	-	T	-	-	-	C
Chimpanzee		C	A	A	A	T	G	T	T	G	G	A	5bp	A	T	G	T	G	T	A	C
Human, ancestral		C	A	A	A	T	G	T	T	G	G	A	5bp	A	T	G	T	G	T	A	C
AF09 Khoisan	1	.	.	.	C	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.
AF15 Falasha, Ethiopia	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	G	.
AF14 Nuba	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
AF02 Biaka Pygmy	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AF03 Biaka Pygmy	3	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
AF05 Mbuti Zaire Pygmy	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AF07 Lissongo	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AF04 Mbuti Zaire Pygmy	4	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AS37 Bedouine	4	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AF01 Biaka Pygmy	5	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AF06 Lissongo	5	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AF16 Nuba‡	-	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AS06 Japanese	6	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AS33 Hmong, Laos	6	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AS18 Sindhi	7	G	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
AS38 Bedouine	7	G	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
AS07 Japanese	8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
OC02 New Guinean	8	.	.	.	.	.	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.
OC07 Australian Aborigine	8	.	.	.	.	.	.	.	A	.	.	.	.	.	.	A	.	.	.	.	.
AS28, AS30, AM06§	9	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	G	.	.	.	.
AS02 and others¶	10	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.
AM02 Surul	10	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.
AM04 Mayan	10	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.
AS11 Brushaski	10	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.
AS12 Pathan	10	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.
AS15 Brahui	10	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.
EU03 and others‖	11	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.
AS16, AS25**	12	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.
OC01 New Guinean	13	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.
OC03, OC04 Melanessian	13	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.
AS14, AS29, AS35, EU08††	14	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.
AS40 Yemenite	15	.	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	T
EU02 Italian	15	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	T
AS01 and others‡‡	16	.	.	.	.	.	.	.	.	.	.	G	O	G	.	.	.	.	.	.	.
AS09 Han Chinese	16	.	.	.	.	.	.	C	.	.	.	G	O	G	.	.	.	.	.	.	.
EU04 and others§§	17	.	.	.	.	C	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.
EU13 Italian‡	-	.	.	.	.	C	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.
AS10, AS24¶¶	18	.	.	.	.	C	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.
AS39 Arab, Central Asia	18	.	C	.	.	C	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.

AF, Africa; AS, Asia; EU, Europe; AM, America; OC, Oceania; -, sequence not available.
*Nonsynonymous substitution.
†Variant nucleotide positions are defined with respect to the cDNA sequences in GenBank for *DFFRY* (AF000986), *DBY* (AF000985), and *UTY1* (AF000996).
‡Screened for UTY1 only.
§AS28, Japanese; AS30, Hazara; AM06, Navajo Indian.
¶AS02, Cambodian; AS17, Sindhi; AS23, Druze; AS32, Tamil; AS34, 36, Bedouine; EU12, Ashkenazi; AM01, Karitiana; AM03, Colombian Indian.
‖EU03, German; EU05, 06, Basque; EU09, Finn; OC06, Australian Aborigine; AM05, Muskogee.
**AS16, Sindhi; AS25, Makrani.
††AS14, Baloochi; AS29, Hazara; AS35; Sepharadim; EU08, Ashkenazi.
‡‡AS01, Cambodian; AS03, Atayal; AS04, Ami; AS05, 27; Japanese; AS08, 26, Han Chinese; AS13, Pathan.
§§EU04, Italian; EU07, Basque; EU10, Finn; EU11, Georgia; AS31, Haraza; OC05, Australian Aborigine.
¶¶AS10, Brushaski; AS24, Makrani.

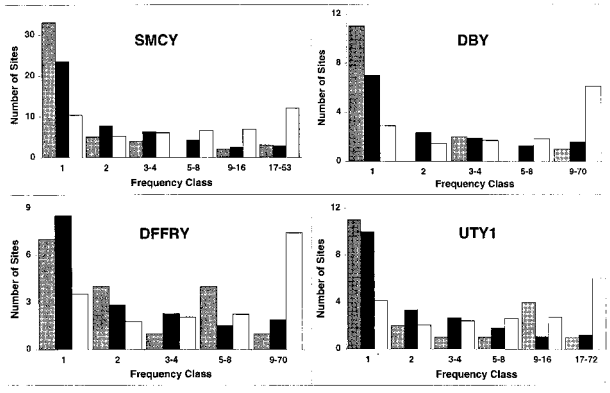

Fig. 3. Allele frequency spectra of the four Y chromosome genes. Shaded columns show the observed distribution of allele frequencies. Black columns depict the expected numbers of derived alleles under the expectation of the Luria–Delbrück/Lea–Coulson theory (16, 17), assuming that each nucleotide in the screened regions is analogous to a parallel, independent bacterial culture. The concordance of the observed and expected distributions is consistent with a significant population expansion. In contrast, there is strong incongruity with the expectation of constant population size (blank columns) with the distribution of alleles estimated by Watterson (18).

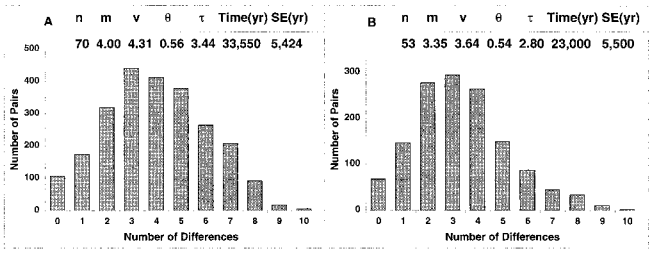

Fig. 4. Histograms display the numbers of sequence differences among all possible pairs of sequences in the set of 70 individuals for 51 segregating sites in *DFFRY*, *DBY*, and *UTY1*, over 41,016 bp (*A*) and 53 individuals for 47 segregating sites in *SMCY* over 39,931 bp (*B*). The mean values and standard errors (SE) for human expansion times were estimated as described in the text.

An independent observation points unequivocally to population growth. We analyzed the distribution of the number of mutants at independent sites in an exponentially growing population by fitting the Luria–Delbrück distribution (16, 17) to the observed frequency spectra in the four NRY genes (Fig. 3). Although far from perfect, the marginal fits obtained for the former (*SMCY*, $\chi^2_{[4]} = 9.99$, $P = 0.041$; *DBY*, $\chi^2_{[3]} = 6.08$, $P = 0.108$; *DFFRY*, $\chi^2_{[3]} = 5.96$, $P = 0.114$; and *UTY1*, $\chi^2_{[4]} = 10.38$, $P = 0.034$) are strikingly better than those observed with distributions generated under the hypothesis of constant population size (18). The respective χ^2 values for the latter distributions were 66.96 (df = 4; $P \approx 10^{-13}$), 20.51 (df = 3; $P \approx 0.0001$), 13.73 (df = 3; $P \approx 0.003$), and 18.03 (df = 4; $P \approx 0.001$) for *SMCY*, *DBY*, *DFFRY*, and *UTY1*, respectively. Deviations from the Luria–Delbrück distribution may result from population growth rates not exactly constant in time or space, variable mutation rates, random genetic drift, and the difficulty of obtaining a truly random sample of the world population.

Mismatch Distributions. To investigate further the hypothesis of population expansion and the time of its occurrence, we computed the distribution of pairwise differences from the segregating sites of both the 70 individuals studied for all of DBY, *DFFRY*, and *UTY1*, and the 53 *SMCY* sequences (Fig. 4). In the two cases, nearly unimodal and smooth mismatch distributions were observed with a raggedness (r) of 0.0133 and 0.0231, respectively (20). A similar distribution has been reported for segregating sites in the hypervariable segments of human mitochondrial DNAs, with an estimated time of population expansion of 40,000 years ago. Although based on a simple model of a sudden expansion to a very large population size (19), mismatch distributions provide a means of comparing different genetic systems. In this regard, it is interesting that our NRY data suggest a more recent expansion time of ≈28,000 years ago. This time falls within the range of 7,000–41,000 years, with a mean of 18,000 years, that has been obtained with a rejection algorithm for eight NRY microsatellites surveyed in 445 individuals (28). The discrepancy in expansion time estimates between the NRY genes on the one hand and mtDNA on the other hand is most likely due to the reduced variability on the former. Nevertheless, the NRY expansion time estimates seem to be correlated with the spread of Aurignacian technology (29) and the disappearance of Neanderthals, who lived in Europe as recently as 28,000 years ago (30).

In conclusion, polymorphic genes on the NRY are evolutionarily neutral or nearly so. Largely immune to recurrent mutation, they provide a valuable system for the reconstruction of the history of our species.

We thank the DNA donors and the investigators who provided the samples: M. E. Ibrahim, T. Jenkins, J. Kidd, S. Q. Mehdi, P. Parham, M. T. Seielstad, and R. S. Wells. Marc Feldman and Steve Sherry kindly provided constructive criticism. This work was supported by National Institutes of Health Grants GM55273 and HG01707.

1. Cooke, H. J., Brown, W. R. & Rappold, G. A. (1985) *Nature (London)* **317,** 687–692.
2. Sinclair, A. H., Berta, P., Palmer, M. S., Hawkins, J. R., Griffiths, B. L., Smith, M. J., Foster, J. W., Frischauf, A. M., Lovell-Badge, R. & Goodfellow, P. N. (1990) *Nature (London)* **346,** 240–244.
3. Lahn, B. T. & Page, D. C. (1997) *Science* **278,** 675–680.
4. Vogt, P. H., Edelmann, A., Kirsch, S., Henegariu, O., Hirschmann, P., Kiesewetter, F., Kohn, F. M., Schill, W. B., Farah, S., Ramos, C., *et al.* (1996) *Hum. Mol. Genet.* **5,** 933–943.
5. Kent-First, M. G., Maffitt, M., Muallem, A., Brisco, P., Shultz, J., Ekenberg, S., Agulnik, A. I., Agulnik, I., Shramm, D., Bavister, B., *et al.* (1996) *Nat. Genet.* **14,** 128–129.
6. Dorit, R. L., Akashi, H. & Gilbert, W. (1995) *Science* **268,** 1183–1185.
7. Whitfield, L. S., Sulston, J. E. & Goodfellow, P. N. (1995) *Nature (London)* **368,** 379–380.
8. Bergen, A. W., Wang, C. Y., Tsai, J., Jefferson, K., Dey, C., Smith, K. D., Pear, S. C., Tsai, S. J. & Goldman, D. (1999) *Ann. Hum. Genet.* **63,** 62–80.
9. Sun, C., Skaletsky, H., Birren, B., Devon, K., Tang, Z., Silber, S., Oates, R. & Page, D. (1999) *Nat. Genet.* **23,** 429–432.
10. Vollrath, D., Foote, S., Hilton, A., Brown, L. G., Beer-Romero, P., Bogan, J. S. & Page, D. C. (1992) *Science* **258,** 52–59.
11. Underhill, P. A., Jin, L., Lin, A. A., Mehdi, S. Q., Jenkins, T., Vollrath, D., Davis, R. W., Cavalli-Sforza, L. L. & Oefner, P. J. (1997) *Genome Res.* **7,** 996–1005.
12. Su, B., Xiao, J., Underhill, P., Deka, R., Zhang, W., Akey, J., Huang, W., Shen, D., Lu, D., Luo, J., *et al.* (1999) *Am. J. Hum. Genet.* **65**, 1718–1724.
13. Nei, M. (1987) *Molecular Evolutionary Genetics*, (Columbia Univ. Press, New York), pp. 256–258.
14. Tajima, F. (1989) *Genetics* **123,** 585–595.
15. Tajima, F. (1989) *Genetics* **123,** 597–601.
16. Luria, S. E. & Delbrück, M. (1943) *Genetics* **28,** 491–511.
17. Lea, D. E. & Coulson, A. C. (1949) *Genetics* **49,** 264–285.
18. Watterson, G. A. (1975) *Theor. Popul. Biol.* **7,** 256–276.
19. Sherry, S. T., Rogers, A. R., Harpending, H., Soodyall, H., Jenkins, T. & Stoneking, M. (1994) *Hum. Biol.* **66,** 761–775.
20. Harpending, H. C., Sherry, S. T., Rogers, A. R. & Stoneking, M. (1993) *Curr. Anthropol.* **34**, 483–496.
21. Jin, L., Underhill, P. A., Doctor, V., Davis, R. W., Shen, P., Cavalli-Sforza, L. L. & Oefner, P. J. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 3796–3800.
22. Thomson, R., Pritchard, J. K., Shen, P., Oefner, P. J. & Feldman, M. W. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 7360–7365.
23. Horai, S., Hayasaka, K., Kondo, R., Tsugane, K. & Takahata, N. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 532–536.
24. Hudson, R. R., Kreitman, M. & Aguadé, M. (1987) *Genetics* **116,** 153–159.
25. Nachman, M. W. (1998) *Mol. Biol. Evol.* **15,** 1744–1750.
26. Jaruzelska, J., Zietkiewicz, E. & Labuda, D. (1999) *Mol. Biol. Evol.* **16,** 1633–1640.
27. Kimmel, M., Chakraborty, R., King, J. P., Bamshad, M., Watkins, W. S. & Jorde, L. B. (1998) *Genetics* **148,** 1921–1930.
28. Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A. & Feldman, M. W. (1999) *Mol. Biol. Evol.* **16,** 1791–1798.
29. Klein, R. G. (1999) *The Human Career: Human Biological and Cultural Origins* (Univ. Chicago Press, Chicago), 2nd Ed.
30. Smith, F. H., Trinkaus, E., Pettitt, P. B., Karavanic, I. & Paunovic, M. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 12281–12286.