Y chromosome sequence variation and the history of human populations

Peter A. Underhill[1], Peidong Shen[2], Alice A. Lin[1], Li Jin[3], Giuseppe Passarino[1], Wei H. Yang[2], Erin Kauffman[2], Batsheva Bonné-Tamir[4], Jaume Bertranpetit[5], Paolo Francalacci[6], Muntaser Ibrahim[7], Trefor Jenkins[8], Judith R. Kidd[9], S. Qasim Mehdi[10], Mark T. Seielstad[11], R. Spencer Wells[12], Alberto Piazza[13], Ronald W. Davis[2], Marcus W. Feldman[14], L. Luca Cavalli-Sforza[1] & Peter. J. Oefner[2]

Binary polymorphisms associated with the non-recombining region of the human Y chromosome (NRY) preserve the paternal genetic legacy of our species that has persisted to the present, permitting inference of human evolution, population affinity and demographic history[1]. We used denaturing high-performance liquid chromatography (DHPLC; ref. 2) to identify 160 of the 166 bi-allelic and 1 tri-allelic site that formed a parsimonious genealogy of 116 haplotypes, several of which display distinct population affinities based on the analysis of 1062 globally representative individuals. A minority of contemporary East Africans and Khoisan represent the descendants of the most ancestral patrilineages of anatomically modern humans that left Africa between 35,000 and 89,000 years ago.

We deduced a phylogenetic tree from 167 NRY polymorphisms on the principle of maximum parsimony (Fig. 1). Of the 167 polymorphisms, 7 had been detected by means other than DHPLC and were taken from the literature. Of the 160 polymorphisms detected by DHPLC, 73 had been reported previously[3,4]. Of the remaining 87 unreported polymorphisms, 53 were discovered in a set of 53 individuals of diverse geographic origin during the screening of the unique sequences and repeat elements, other than long interspersed elements, contained in 3 overlapping cosmid sequences (GenBank accession numbers AC003032, AC003095, AC003097) and a few small fragments scattered throughout the NRY. Finally, we detected 34 during genotyping. In total, the marker panel is composed of 91 transitions, 53 transversions, 22 small insertions or deletions, and 1 *Alu* insertion. All polymorphisms are bi-allelic, except a double transversion (M116) that has three alleles, A, C or T, defining different haplotypes. Two non-CpG associated transitions (M64 and M108) show evidence of recurrence, but generate no ambiguities when considered in the context of other markers. We placed the root of the phylogeny using sequence information generated from the three great ape species. The sequential succession of mutational events is unequivocal, except for those appearing in the same tree segment (for example, M42, M94, M139). The phylogeny is composed of 116 haplotypes and their frequencies in 21 general populations are given (Table 1). Forty-two haplotypes (36.2%) are represented by just one individual. Several haplotypes, however, have higher frequencies and/or geographic associations that disclose patterns of population affinities apparent from a maximum likelihood analysis (Fig. 2) performed on the haplotype frequencies (Table 1). To facilitate presentation, we grouped the 116 haplotypes into 10 haplogroups as defined by either the presence or the absence of mutations occupying strategic internal positions in the phylogeny. Haplogroups VI, VIII and X, although polyphyletic, are distinguished by criteria (Table 2).

Three mutually reinforcing mutations, M42, M94 and M139 (two transversions and a 1-bp deletion), distinguish haplogroup I, which is represented today by a minority of Africans—mainly Sudanese, Ethiopians and Khoisans (Table 1). All non-Africans, except a single Sardinian, and most African males sampled carry only the derived alleles at the three sites. This implies that modern extant human Y chromosomes trace ancestry to Africa and that the descendants of the derived lineage left Africa and eventually replaced archaic human Y chromosomes in Eurasia[5].

An important property of a phylogeny is the randomness of number of mutations per segment of the tree. Of the 166 segments, 41 carry no mutation, whereas 98, 16, 8, 2 and 1 segment have 1, 2, 3, 4 and 8 mutations, respectively. The mean number of mutations per segment is 1.024 with a variance of 0.945. Applying the G-test for goodness of fit and Williams' correction to the observed G, the data do not fit a Poisson distribution (G_{adj}=34.98, d.f.=3, P~10^{-7}). This is due to an excess of segments with one mutation, as expected in an exponentially growing population. Similar results were obtained recently for the separate analysis of four Y chromosome genes[4]. Further support that the human population has undergone a major expansion comes from the consistently negative values of Tajima's D (ref. 6) for not only the Y chromosome, but also for mitochondrial DNA, X-chromosomal and autosomal genes[4]. Notably, NRY shows evidence of significantly reduced variability to the other genetic systems[4], confirming a similar comparison of a smaller number of polymorphisms on previously reported NRY sequences with 8 X-linked[7,8] and 16 autosomal human genes[4]. Possible explanations include positive selection on NRY (ref. 9) and a difference between male and female effective population sizes[10].

Assuming expansion, the age of the most recent common ancestor (T_{MRCA}) was previously estimated at 59,000 years, with a 95% probability interval of 40,000–140,000 years[11]. This value is similar

[1]*Department of Genetics, Stanford University, Stanford, California, USA.* [2]*Stanford DNA Sequencing and Technology Center, Palo Alto, California, USA.* [3]*University of Texas-Houston, Human Genetics Center, Houston, Texas, USA.* [4]*Sackler Faculty of Medicine, Human Genetics, Tel-Aviv University, Tel-Aviv, Israel.* [5]*Unitat de Biologia Evolutiva, Facultat de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra, Barcelona, Catalonia, Spain.* [6]*Dipartimento di Zoologia e Antropologia Biologica, Università di Sassari, Sassari, Italy.* [7]*Institute of Endemic Diseases, University of Khartoum, Sudan.* [8]*Department of Human Genetics, School of Pathology, South African Institute for Medical Research and the University of Witwatersrand, Johannesburg, South Africa.* [9]*Department of Genetics, Yale University School of Medicine, New Haven, Connecticut, USA.* [10]*Dr. A. Q. Khan Research Laboratories, Biomedical & Genetic Engineering Laboratories, Islamabad, Pakistan.* [11]*Harvard School of Public Health, Program for Population Genetics, Boston, Massachusetts, USA.* [12]*Wellcome Trust Centre for Human Genetics, University of Oxford, Headington, UK.* [13]*Department of Genetics, Biology and Biochemistry, Department of Genetics, University of Torino, Torino, Italy.* [14]*Department of Biological Sciences, Herrin Laboratories, Stanford University, California, USA. Correspondence should be addressed to P.A.U. (e-mail: under@stanford.edu).*

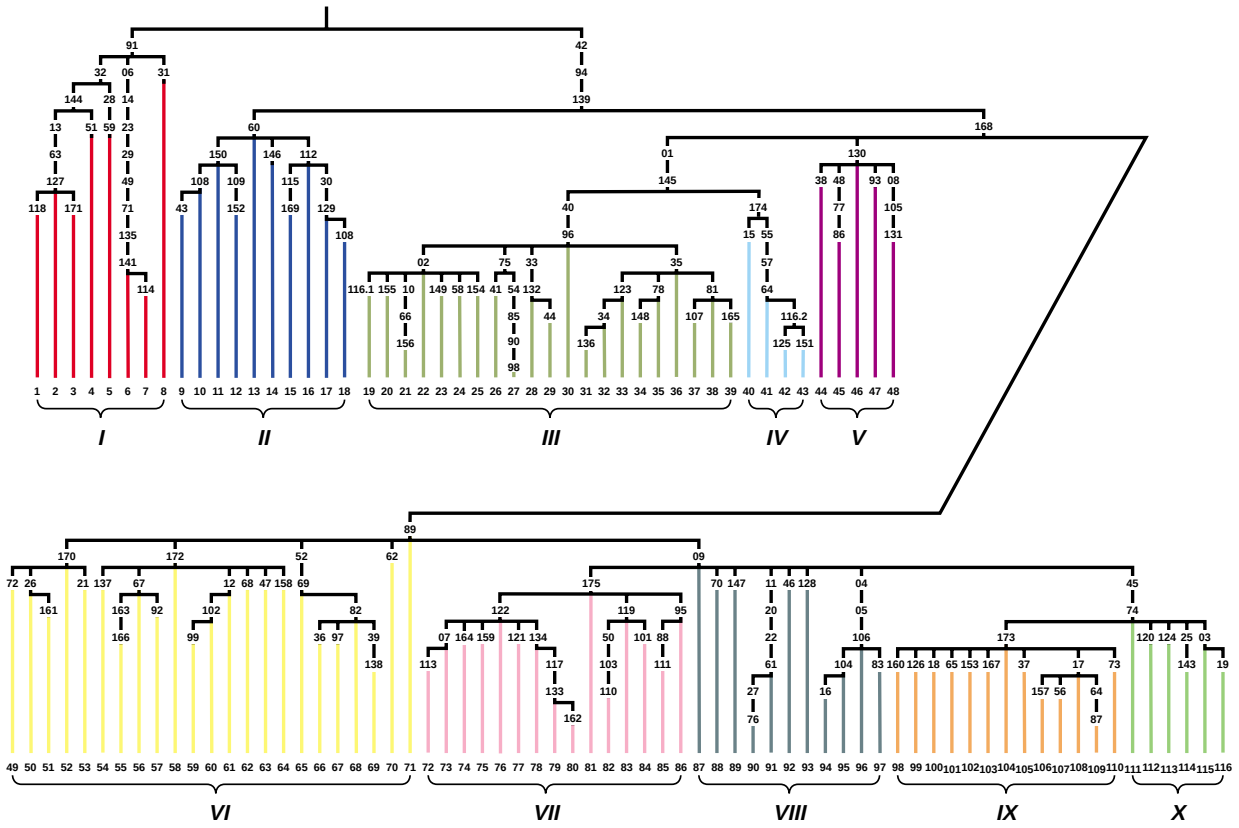

Fig. 1 Maximum parsimony phylogeny of human NRY chromosome bi-allelic variation. The tree is rooted with respect to non-human primate sequences. The 116 numbered compound haplotypes were constructed from 167 mutations, of which 160 were discovered by DHPLC. The remaining seven were taken from the literature and included YAP (M1)[17], DYS271 (M2)[18], PN3 (M29)[19], SRY 4064 (M40)[5], TAT (M46)[20], RPS4YC711T (M130)[21] and SRY 2627 (M167)[22]. Marker numbers indicated on the segments are discontinuous because of the removal of all but one polymorphism associated with tandem repeats and homopolymer tracts whose ancestral state is uncertain. Haplotypes are assorted into 10 haplogroups (I–X) using criteria given in Table 2. Haplogroup I members, ancestral for M42, M94 and M139, also share the only homopolymer-associated marker M91. All haplogroup I individuals have an 8-T length variant, whereas 1,009 men in haplogroups II–X have 9 and in 2 cases 10-T length variants (not shown). Only one inconsistent haplogroup X individual had an 8-T length variant (not shown). Haplogroups I and II, both of which are almost exclusively represented in Africa, share the ancestral allele of M168. haplogroup III is generally the most frequent one in Africa. Its frequency decreases with increasing distance from Africa, from 27% in the Mid-East to a few per cent in Northern Europe and South and Central Asia. Haplogroup IV, related to the former through M1 and M145, is found mainly in Japan. Haplogroups V and VIII are prevalent in New Guinea and Australia, but they are also found at varying though smaller frequencies throughout Asia. Haplogroup VIII represents the relevant source of Haplogroups VII, IX and X. Haplogroups VI and IX are found mostly in Europe and the Indus Valley. They are not observed in East Asia, where haplogroup VII dominates, suggesting that this part of the world where agriculture developed independently resisted effectively subsequent gene flow[23]. The distinction between Eurasians and East Asians was also observed with mtDNA (ref. 24) and autosomal genes[25]. Haplogroup X is common in the Americas, although its origin may have been in Central Asia where traces of it persist (Table 1).

to an estimate of 46,000–91,000 years based on 8 Y chromosome microsatellites[12] and, therefore, is considerably less than estimates of greater than 100,000 years obtained previously[5]. Of course, this assumes that selection or population structure has not had a major effect on NRY diversity, an assumption that may be wrong in light of our findings of significantly reduced variability on NRY. As the average number of mutations of all segments departing from the root is 8.60 (Table 2), and with a T_{MRCA} value of 59,000 years, the average time for adding a new mutation to the tree is approximately 6,900 years. This puts the age of M168, which marks the expansion of anatomically modern humans out of Africa, at approximately 44,000 years, in agreement with a previous estimate of 47,000 years with 95% probability intervals of 35,000–89,000 years using the program GENETREE (ref. 11). This concurs with recent archeological[13] and mtDNA data[14], and is also consistent, though at a compressed time scale, with the weak Garden-of-Eden hypothesis[15]. Under this hypothesis, a small subgroup of behaviourally modern humans[13] left Africa and separated into several fairly isolated groups represented today by the major haplogroups III–X. Those groups remained small throughout the last glaciation before they underwent roughly simultaneous expansions in size as suggested by a star-like genealogy (Fig. 1).

The new levels of bi-allelic variation revealed here indicate a recent ancestry of the paternal lineages of our species from Africa and testify to the informativeness of the Y chromosome in deciphering the evolution of humankind.

Methods

DNA samples. The ascertainment set consisted of the following 53 samples with their subsequently determined haplogroup designations: Africa: 3 Central African Republic Biaka II, III (1); 2 Zaire Mbuti II, III; 2 Lissongo II, III; 2 Khoisan I, III; 1 Berta VI; 1 Surma I; 1 Mali Tuareg III; 1 Mali Bozo III; Europe: 1 Sardinian VI; 2 Italian VI IX; 1 German VI; 3 Basque VI, IX (2); Asia: 3 Japanese IV, V, VII; 2 Han Chinese VII, 1 Taiwan Atayal VII, 1 Taiwan Ami, VII, 2 Cambodian VI, VII; Pakistan: 2 Hunza VI, IX; 2 Pathan VI, VII; 1 Brahui VIII; 1 Baloochi VI; 3 Sindhi III, VI, VIII; Central Asia: 2 Arab IX; 1 Uzbek IX; 1 Kazak V; MidEast: 1 Druze VI; Pacific: 2 New Guinean V, VIII; 2 Bougainville Islanders VIII; 2 Australian VI, X: America: 1 Brazil Surui, 1 Brazil Karatina, 1 Columbian, 1 Mayan all X. We genotyped an additional 1,009 chromosomes, representing 21 geographic regions, by DHPLC for all markers other than those on the terminal branches of the phylogeny. We genotyped the latter only in individuals from the haplogroup to which those markers belonged. This hierarchic genotyping protocol was necessitated by the limited amounts of genomic DNA available for most samples.

PCR. We used the RepeatMasker2 program (http://ftp.genome.washington.edu) to identify human repeat DNA sequences. We designed primers to amplify unique sequences and repeat elements other than LINE as confirmed by a negative female control, yielding amplicons 300–500 bp in length. The description of the 167 Y markers are given in Table A (http://genetics.nature.com/supplementary_info/). All primers had a uniform annealing temperature, which allowed a single PCR protocol to be used. It 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 using 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, 72 °C for 1 min and a final 5-min extension at 72 °C. Each 50-µl PCR reaction contained 1 U AmpliTaq Gold polymerase, 10 mM 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 genomic DNA. PCR yields were determined semi-quantitatively on ethidium bromide stained agarose gels.

Denaturing high-performance liquid chromatography analysis. We mixed unpurified PCR products at an equimolar ratio with a reference Y chromosome and then subjected the mixture to a 3 min, 95 °C denaturing step followed by gradual reannealing from 95–65 °C over 30 min. We loaded 10 µl of each mixture onto a DNASep column (Transgenomic), 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[2]. We recognized heteroduplex mismatches by the appearance of two or more peaks in the elution profiles under appropriate temperature conditions, which were optimized by computer simulation (available at http://insertion.stanford.edu/melt.html).

DNA sequencing. We purified polymorphic and reference PCR samples with QIAquick spin columns (Qiagen). We sequenced both strands to determine the location and chemical nature of any polymorphic sites, using the amplimers as sequencing primers and ABI Dye-terminator cycle sequencing reagents (PE Biosystems). Each cycle sequencing reaction contained 6 µl purified PCR product, 4 µl dye

Table 1 • Distribution of Y-chromosome haplotypes by geographic population group

Haplotype #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Sudan		17	1									5	1													2			1						7			2		
Ethiopia	6	5			1					3	1	4	1									3				15				16		2			20	6				
Mali								1	3		1			1						1	1	7						13	2								1	12			
Morocco																						2													1			1			
C. Africa										1	1				1	7	1	1			1	20				3														
Khoisan				11		5	1									11						7													4					
S. Africa				3								7										28	1	3	2	8			1						1					
Europe																															1									
Sardinia		1																													1			4						
Basque																																					1			
Mid-east																						2						1			1			2				1		
C. Asia + Siberia																															2	1		1						
Pakistan + India												2																		2			1							
Hunza																																								
Japan																																								1
China																																								
Taiwan																																								
Cambo + Laos																																								1
New Guinea																																								
Australia																																								
America																						2													1					
Total	6	23	1	14	1	5	1	1	3	3	3	19	2	1	1	18	1	1	1	1	1	71	1	3	2	17	12	14	2	17	2	7	1	1	36	11	1	16	1	2

Haplotype #	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81
Sudan																															4										
Ethiopia																															4										
Mali																																									
Morocco										1								3													14										
C. Africa																																									
Khoisan																																									
S. Africa																																									
Europe									1	1		8		1				2	1												9										
Sardinia										11						1															2										
Basque										2		1																			1										
Mid-east															1	2		2													8										
C. Asia + Siberia					10	16										2	1	12		4		1	1		2			1		1	17					6	9		2	
Pakistan + India						1											4	3		3	2			1	1		1	4			7					1	1				
Hunza						1												1		3						1		2			1						1				
Japan	1	5	1			1	1	1																											2		2	1		6
China																																1		1	2	1	2	4	1	2
Taiwan																																	4		18					
Cambo + Laos						1																	2						1				1			1		1		
New Guinea				4																																				
Australia						3							1																												
America						1					1	1																			1										
Total	1	5	1	4	10	24	1	1	1	15	1	10	1	1	1	5	5	23	1	10	2	1	1	3	3	1	1	7	1	1	68	1	4	1	1	22	2	12	16	1	10

Haplotype#	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	98	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	Total
Sudan																																				40
Ethiopia																						1														88
Mali																						1														44
Morocco							1															5														28
C. Africa																																				37
Khoisan																																				39
S. Africa																																				53
Europe																	3	1				29					3									60
Sardinia																			2																	22
Basque																				2	7	5	26													45
Mid-east									2													2														24
C. Asia + Siberia		2			1	5			2	2	12	1										10			1	30	3	6	3		12	6			184
Pakistan + India							1	2	8	2												6		1		28			2		4				88
Hunza						2				3												3				11			2		7				38
Japan												1																								23
China		3	1	1																										1						20
Taiwan	5	46				1																														74
Cambo + Laos		1		1	6	1																				1										18
New Guinea						7							2	5	4	1																				23
Australia																						1	2													7
America						1	1															5							5				83	4	106
Total	5	52	1	2	7	17	3	2	12	7	12	2	2	5	4	1	3	1	2	2	7	5	89	2	1	1	73	3	6	12	1	23	6	83	4	1062

Table 2 • Defining features of haplogroups

Haplogroup	Most recent defining mutation	Avg. no. of mutations from root to individual haplotypes*	Total no. of individuals	No. mutations per haplogroup minus defining mutation(s)	No. haplotypes per haplogroup
I	M91	6.1±0.95	52	20	8
II	M60	6.1±0.41	52	12	10
III	M96	10.4±0.24	218	27	21
IV	M174	10.5±0.96	9	7	4
V	M130	6.6±0.60	40	8	5
VI	M89 & absence of M9	7.4±0.25	163	25	23
VII	M175	9.3±0.35	137	18	15
VIII	M9 & absence of M175 and M45	8.9±0.68	67	16	11
IX	M173	10.2±0.20	195	13	13
X	M74 & absence of M173	9.2±0.1	129	6	6
Totals	–	8.59±0.20	1062	152	116

*Mean and standard error.

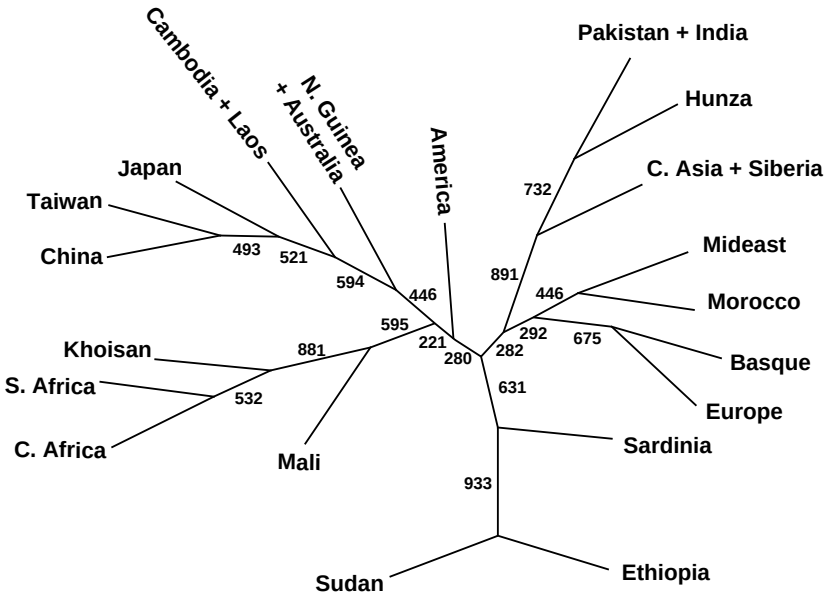

Fig. 2 Maximum likelihood network inferred from the haplotype frequencies reported in Table 1. The gene frequencies of New Guineans and Australian aborigines were grouped together because of the small sample size of the latter. Values at nodes indicate number of 1,000 bootstrap trees presenting cluster distal of node. Sudanese and Ethiopians are distinct from the other Africans and appear to be more associated with samples from the Mediterranean basin. This may reflect either repeated genetic contact between Arabia and East Africa during the last 5,000–6,000 years or a Middle Eastern origin with subsequent acquisition of African alleles on the way southwest with agricultural expansion[26]. The Moroccan samples are under-represented with respect to Group III (J.B., unpublished data). Native Americans are located between Eurasians and East Asian indicating common ancestry with both. This network is consistent with the first two principal components capturing 18% of the variation present in the 116 haplotypes.

terminator reaction mix and 0.8 µl primer (5 µM). Cycle sequencing was started at 94 °C for 1 min, followed by 25 cycles of 96 °C for 10 s, 50 °C for 2 s and 60 °C for 4 min. We purified the cycle sequencing reactions using Centrifex gel filtration cartridges (Edge Biosystems), which were then analysed on a PE Biosystems 373A sequencer.

Statistical analysis. We used the program CONTML in PHYLIP, version 3.57c, to construct a frequency based maximum likelihood network.

Accession numbers. Most of the NRY sequence surveyed was derived from 5 cosmid sequences retrievable from GenBank using the accession numbers AC003031, AC003032, AC003094, AC003095, and AC003097. Six polymorphisms were affiliated with genomic regions for DFFRY (AC002531), one each for DBY (AC004474) and UTY1 (AC006376), 3 for SRY (NM003140), and 15 for random genomic STSs reported by Vollrath and collaborators[16].

Acknowledgements

We thank the 1,062 men who donated DNA; R.G. Klein, J. Mountain and M. Ruhlen for helpful discussions; D. Vollrath, R. Hyman and F.S. Dietrich for Y-specific cosmid sequences; and J. Block, D. Soergel, K. Prince, C. Edmonds and A. Rojas for technical help. A.W. Bergen made the RPS4YC711T marker (M130) information available to us before its publication. This work was supported in part by the NIH, NIHGR and L.S.B. Leakey Foundation.

Received 21 April; accepted 9 September 2000.

1. Hammer, M.F. & Zegura, S.L. The role of the Y chromosome in human evolutionary studies. *Evol. Anthropol.* **5**, 116–134 (1996).
2. Oefner, P.J. & Underhill, P.A. DNA mutation detection using denaturing high-performance liquid chromatography. *Current Protocols in Human Genetics*. Suppl **19**, 7.10.1–7.10.12 (Wiley & Sons, New York, 1998).
3. Underhill, P.A. *et al.* Detection of numerous Y chromosome biallelic polymorphisms by denaturing high performance liquid chromatography (DHPLC). *Genome Res.* **7**, 996–1005 (1997).
4. Shen, P. *et al.* Population genetic implications from sequence variation in four Y chromosome genes. *Proc. Natl Acad. Sci. USA* **97**, 7354–7359 (2000).
5. Hammer, M.F. *et al.* Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Mol. Biol. Evol.* **15**, 427–441 (1998).
6. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism *Genetics* **123**, 585–595 (1989).
7. Nachman, M.W. Y chromosome variation of mice and men. *Mol. Biol. Evol.* **15**, 1744–1750 (1998).
8. Jaruzelska, J., Zietkiewicz, E. & Labuda, D. Is selection responsible for the low level of variation in the last intron of the *ZFY* locus? *Mol. Biol. Evol.* **16**, 1633–1640 (1999).
9. Wyckoff, G.J., Wang, W. & Wu, C.I. Rapid evolution of male reproductive genes in the descent of man. *Nature* **403**, 304–309 (2000).
10. Jorde, L.B. *et al.* The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am. J. Hum. Genet.* **66**, 979–988 (2000).
11. Thomson, R. *et al.* Recent common ancestry of human Y chromosomes: Evidence from DNA sequence data. *Proc. Natl Acad. Sci. USA* **97**, 7360–7365 (2000).
12. Pritchard, J.K., Seielstad, M.T., Perez-Lezaun, A. & Feldman, M.W. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol. Biol. Evol.* **16**, 1791–1798 (1999).
13. Klein, R.G. *The Human Career: Human Biological and Cultural Origins* (University of Chicago Press, Illinois, 1999).
14. Quintana-Murci, L. *et al.* Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. *Nature Genet.* **23**, 437–441 (1999).
15. Rogers, A.R. Genetic evidence for a Pleistocene population explosion. *Evolution* **49**, 608–615 (1995).
16. Vollrath, D. *et al.* The human Y chromosome: a 43-interval map based on naturally occurring deletions. *Science* **258**, 52–59 (1992).
17. Hammer, M.F. & Horai, S. Y chromosomal DNA variation and the peopling of Japan. *Am. J. Hum. Genet.* **56**, 951–962 (1995).
18. Seielstad, M.T. *et al.* Construction of human Y-chromosome haplotypes using a new polymorphic A to G transition. *Hum. Mol. Genet.* **3**, 2159–2161 (1994).
19. Hammer, M.F. *et al.* The geographic distribution of human Y chromosome variation. *Genetics* **145**, 787–805 (1997).
20. Zerjal, T. *et al.* Genetic relationships of Asians and northern Europeans, revealed by Y-chromosomal DNA analysis. *Am. J. Hum. Genet.* **60**, 1174–1183 (1997).
21. Bergen, A.W. *et al.* An Asian-native American paternal lineage identified by RPS4Y resequencing and by microsatellite haplotyping. *Ann. Hum Genet.* **63**, 63–80 (1999).
22. Bianchi, N.O. *et al.* Origin of Amerindian Y-chromosomes as inferred by the analysis of six polymorphic markers. *Am. J. Phys. Anthropol.* **102**, 79–89 (1997).
23. Diamond, J. *Guns, Germs, and Steel* (Norton, New York, 1999).
24. Macaulay, V. *et al.* The emerging tree of west Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am. J. Hum. Genet.* **64**, 232–249 (1999).
25. Jin, L. *et al.* Distribution of haplotypes from a chromosome 21 region distinguishes multiple prehistoric human migrations. *Proc. Natl Acad. Sci. USA* **96**, 3796–3800 (1999).
26. Cavalli-Sforza, L.L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes* (Princeton University Press, Princeton, New Jersey, 1994).