The Eurasian Heartland: A continental perspective on Y-chromosome diversity

R. Spencer Wells[a,b], Nadira Yuldasheva[a,c], Ruslan Ruzibakiev[c], Peter A. Underhill[d], Irina Evseeva[e], Jason Blue-Smith[d], Li Jin[f], Bing Su[f], Ramasamy Pitchappan[g], Sadagopal Shanmugalakshmi[g], Karuppiah Balakrishnan[g], Mark Read[h], Nathaniel M. Pearson[i], Tatiana Zerjal[j], Matthew T. Webster[k], Irakli Zholoshvili[l], Elena Jamarjashvili[l], Spartak Gambarov[m], Behrouz Nikbin[n], Ashur Dostiev[o], Ogonazar Aknazarov[p], Pierre Zalloua[q], Igor Tsoy[r], Mikhail Kitaev[s], Mirsaid Mirrakhimov[s], Ashir Chariev[t], and Walter F. Bodmer[a,u]

[a]Imperial Cancer Research Fund Cancer and Immunogenetics Laboratory and Wellcome Trust Centre for Human Genetics, University of Oxford, Headington OX3 9DS, United Kingdom; [c]Institute of Immunology, Academy of Sciences, Tashkent, Uzbekistan; [d]Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305; [e]Department of Medical Genetics, Archangelsk State Medical Academy, Archangelsk, Russia; [f]Department of Human Genetics, School of Public Health, University of Texas, Houston, TX 77030; [g]Centre for Advanced Studies in Functional Genomics, School of Biological Sciences, Madurai Kamaraj University, Madurai, India; [h]Mark Read Photography, 2 Beatty Road, London N16 8EB, United Kingdom; [i]Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637; [j]Department of Biochemistry, University of Oxford, Oxford OX3 1QU, United Kingdom; [k]Institute of Molecular Medicine, University of Oxford, Headington OX3 9DS, United Kingdom; [l]Department of Pediatrics, State Medical Institute, Tbilisi, Georgia; [m]Immunology Laboratory, Mikaelin Surgical Institute, Yerevan, Armenia; [n]Department of Immunology, University of Medical Sciences, Tehran 14496, Iran; [o]Tajik State Medical Institute, Dushanbe, Tajikistan; [p]Pamir Biological Laboratory, Khorog, Tajikistan; [q]Genetics Research Laboratory, Chronic Care Center, Hazmieh, Lebanon; [r]Institute of Nutrition, Almaty, Kazakstan; [s]Institute of Cardiology, Bishkek, Kyrgyzstan; and [t]State Medical Institute, Ashgabad, Turkmenistan

Contributed by Walter F. Bodmer, June 15, 2001

The nonrecombining portion of the human Y chromosome has proven to be a valuable tool for the study of population history. The maintenance of extended haplotypes characteristic of particular geographic regions, despite extensive admixture, allows complex demographic events to be deconstructed. In this study we report the frequencies of 23 Y-chromosome biallelic polymorphism haplotypes in 1,935 men from 49 Eurasian populations, with a particular focus on Central Asia. These haplotypes reveal traces of historical migrations, and provide an insight into the earliest patterns of settlement of anatomically modern humans on the Eurasian continent. Central Asia is revealed to be an important reservoir of genetic diversity, and the source of at least three major waves of migration leading into Europe, the Americas, and India. The genetic results are interpreted in the context of Eurasian linguistic patterns.

Human population genetics uses the distribution of genetic markers in extant human populations to gain an insight into demographic and migrational history. Population studies of protein (“classical”) polymorphisms, HLA variation, and mitochondrial DNA have all contributed greatly to our understanding of human origins and early dispersals (1). New molecular biological techniques, including surveys of autosomal microsatellite diversity (2) and single nucleotide polymorphism (SNP) haplotypes (3, 4) have revealed similar patterns. The genetic evidence has consistently pointed to a recent common origin in Africa, with subsequent dispersal(s) of modern humans throughout the rest of the world.

Recently, diversity on the nonrecombining portion of the Y chromosome (NRY) has been applied to the study of human history (5, 6). Two features of the NRY make it particularly well suited for these investigations. First, like mtDNA, it is passed from parent to offspring (in this case, father to son) without the “shuffling” effect of recombination; this allows the evolution and retention of a wide variety of stable haplotypes with varying ages, related through a clear, stepwise mutational process. The second feature is the highly specific geographic pattern of NRY variation: NRY diversity within populations is lower than that seen for other markers, and variation among populations is higher (7). Thus, the Y chromosome provides anthropologists and geneticists with an extremely powerful tool for historical and demographic studies.

The ease with which recently described biallelic NRY markers (8) can be typed and assigned to haplotypes promises to reveal many details of human history that were unclear from the analysis of other genetic markers. In particular, the Y data may shed some light on the path(s) followed by anatomically modern humans after they left Africa and settled other continents. In this paper we present the analysis of 23 NRY biallelic polymorphisms in a sample of 1,935 men from 49 Eurasian populations. These results are compared with data from other populations in an effort to reconstruct the history of early human migrations in Eurasia, as well as more recent events in the region of Central Asia.

Methods

Samples. Samples were collected from healthy adult men, primarily during three expeditions in 1996, 1998, and 2000 (see http://popgen.well.ox.ac.uk/eurasia for details). Informed consent was obtained from all donors. We made an effort to sample a variety of linguistic groups, including those speaking languages belonging to the Afro-Asiatic, Indo-European, Dravidian, South Caucasian, North Caucasian, Altaic, Uralic, and Sino-Tibetan language families. The Sourashtran, Yadhava, and Kallar samples were collected as described elsewhere (S.S., R.S.W., K.B., and R.P., unpublished work). The Nenets, Northern Russian, Pomor, and Saami samples were collected in 1996 and 1998, the Tuvinian in 1996, and the Orkney samples in 1994. For some samples, cell lines were prepared by Epstein–Barr virus transformation of peripheral blood leukocytes. For most, however, blood was collected by venipuncture, and the red cells were lysed by two rounds of osmotic shock and centrifugation in 115 mM NH_4Cl. The resulting white cell pellet was resuspended in a white cell lysis buffer (100 mM Tris·CL, pH 7.6/50 mM NaCl/40 mM EDTA, pH 8.0/0.2% SDS/0.5% sodium azide), rendering the DNA stable for several weeks under field conditions. After returning to the laboratory, the DNA was purified by using a standard salting-out procedure (9).

Genotyping. The polymorphisms typed in this study are shown in Table 1 and in the supplemental data (Tables 2 and 3), which are

Abbreviation: NRY, nonrecombining portion of the Y chromosome.

[b]To whom correspondence should be addressed. E-mail: rswells@well.ox.ac.uk.

[u]To whom reprint requests should be addressed at: Imperial Cancer Research Fund Cancer and Immunogenetics Laboratory, Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DS, United Kingdom. E-mail: Walter.Bodmer@hertford.ox.ac.uk.

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

Table 1. Y-chromosome haplotype frequencies in 49 Eurasian populations, listed according to geographic region

			Haplotype																						
Population	Language Family	N	YAP(+	M96	M174	M130	M48	M89	M170	M172	M52	M9	M175	M122	M119	M95	M20	M46	M45	M120	M124	M3	M173	M17	M87	Diversity
WESTERN EUROPE																										
British	Indo-European	25		0.04				0.08	0.16														0.72			0.467
Orkney	Indo-European	26							0.08														0.65	0.27		0.514
RUSSIA																										
Pomor	Indo-European	28						0.04	0.11	0.04		0.04						0.43						0.36		0.698
Russian/North	Indo-European	49		0.02				0.02	0.27	0.04		0.02						0.20						0.43		0.716
Russian/Tashkent	Indo-European	89		0.03				0.03	0.12	0.08		0.06						0.13					0.07	0.47		0.736
Kazan Tatar	Altaic	38		0.03				0.13	0.18	0.11		0.08					0.03	0.13	0.05				0.03	0.24		0.876
Saami	Uralic	23		0.09					0.17	0.04								0.39					0.09	0.22		0.787
Nenets	Uralic	54							0.06			0.50						0.30					0.04	0.11		0.658
MIDDLE EAST																										
Lebanon																										
Lebanese	Afro-Asiatic	50		0.32				0.30		0.30							0.02						0.06			0.728
Iran																										
Tehran	Indo-European	24		0.21				0.46		0.21							0.04						0.04	0.04		0.728
Shiraz	Indo-European	12		0.08				0.42		0.25		0.08					0.08		0.08							0.803
Esfahan	Indo-European	16		0.25				0.31		0.25				0.06					0.06		0.06					0.817
CAUCASUS																										
Georgia																										
Svanetian	South Caucasian	25						0.92																0.08		0.153
Kazbegi	South Caucasian	25						0.12	0.04	0.72													0.08	0.04		0.477
Ossetian	Indo-European	17		0.18				0.41		0.24													0.12	0.06		0.772
Azerbaijan																										
Lezgi	North Caucasian	12		0.17				0.58															0.17	0.08		0.652
Azeri	Altaic	21		0.05				0.10	0.05	0.48		0.10											0.14	0.10		0.757
Armenia																										
Armenian	Indo-European	47		0.04				0.13	0.04	0.21		0.06					0.04	0.02					0.36	0.09		0.808
CENTRAL ASIA																										
Turkmenistan																										
Turkmen	Altaic	30						0.13		0.17		0.13							0.10		0.03		0.37	0.07		0.814
Kurd	Indo-European	17						0.35		0.18	0.06												0.29	0.12		0.787
Uzbekistan																										
Sinte Romani	Indo-European	15						0.07		0.20	0.13			0.07							0.53					0.695
Iranian/Samarkand	Indo-European	53		0.09		0.02		0.19	0.02	0.30	0.04	0.06					0.02				0.08		0.08	0.11		0.851
Tajik/Samarkand	Indo-European	40		0.10		0.05	0.03	0.03		0.20			0.03	0.05			0.10		0.05		0.03		0.10	0.25		0.879
Arab/Bukhara	Afro-Asiatic	42		0.02			0.05	0.19	0.10	0.17							0.10		0.02				0.10	0.19	0.07	0.885
Crimean Tatar	Altaic	22		0.05		0.09		0.18	0.05	0.14			0.05	0.05									0.09	0.32		0.861
Karakalpak	Altaic	44				0.20	0.02	0.09		0.09		0.07		0.11			0.05	0.02			0.07		0.09	0.18		0.895
Uzbek/Kashkadarya	Altaic	19				0.16		0.21		0.11				0.05					0.11				0.21	0.16		0.883
Uzbek/Bukhara	Altaic	58		0.02		0.09		0.16		0.16	0.02	0.05		0.12			0.02		0.02		0.02		0.07	0.28		0.860
Uzbek/Surkhandarya	Altaic	68			0.06	0.12		0.03		0.16		0.12	0.01	0.03			0.01	0.04	0.04		0.01		0.06	0.29		0.859
Uzbek/Khorezm	Altaic	70		0.07		0.06	0.04	0.04	0.01	0.11	0.03	0.09	0.01	0.01			0.01	0.01	0.09		0.01		0.09	0.30		0.873
Uzbek/Tashkent	Altaic	43			0.05	0.07		0.09	0.07	0.14		0.05					0.05	0.02	0.05		0.02		0.12	0.28		0.882
Uzbek/Fergana Valley	Altaic	63		0.03	0.03	0.13		0.05	0.05	0.10	0.02	0.06		0.05		0.02	0.03		0.05		0.05		0.13	0.22		0.905
Uzbek/Samarkand	Altaic	45		0.02		0.16	0.02	0.09	0.02	0.16	0.04	0.04		0.02			0.09		0.07		0.02		0.11	0.13		0.915
Tajikistan																										
Ishkashimi	Indo-European	25						0.08									0.12				0.08		0.04	0.68		0.530
Bartangi	Indo-European	30						0.23		0.03									0.13		0.17		0.03	0.40		0.763
Shugnan	Indo-European	44		0.11		0.02		0.16		0.11							0.16		0.14				0.07	0.23		0.868
Yagnobi	Indo-European	31				0.03				0.32		0.03					0.10		0.03				0.32	0.16		0.778
Tajik/Khojant	Indo-European	22				0.05	0.05	0.05		0.09							0.05				0.09			0.64		0.597
Tajik/Dushanbe	Indo-European	16		0.06				0.06		0.31	0.13		0.06				0.13				0.06			0.19		0.875
Kyrgyzstan																										
Kyrgyz	Altaic	52				0.08	0.08	0.02	0.02	0.02		0.02		0.02	0.06			0.02	0.02				0.02	0.63		0.590
Dungan	Sino-Tibetan	40				0.03		0.05	0.03	0.13		0.03		0.40	0.05			0.03		0.08	0.05		0.05	0.10		0.817
Kazakstan																										
Kazak	Altaic	54			0.02	0.09	0.57	0.02					0.02	0.09				0.02	0.06		0.02		0.06	0.04		0.656
Uighur	Altaic	41				0.15		0.10	0.02	0.20	0.02		0.05	0.12			0.02	0.02	0.07					0.22		0.879
SOUTH INDIA																										
Sourashtran	Indo-European	46				0.07		0.04		0.02	0.15					0.02	0.26				0.04			0.39		0.763
Kallar	Dravidian	84				0.07		0.18		0.01	0.15					0.01	0.48		0.01		0.05			0.04		0.717
Yadhava	Dravidian	129				0.03		0.12		0.20	0.19			0.01		0.01	0.19		0.03		0.09			0.13		0.852
SIBERIA/EAST ASIA																										
Tuvinian	Altaic	42				0.10	0.07					0.26		0.02				0.21	0.17				0.02	0.14		0.842
Mongolian	Altaic	24			0.04	0.13	0.46	0.08				0.13		0.08	0.04									0.04		0.772
Korean	Korean	45			0.07	0.16		0.02				0.02	0.31	0.36	0.04					0.02						0.762

Haplotypes are defined in Table 2.

published on the PNAS web site, www.pnas.org; the nomenclature is that of Underhill *et al.* (8), and marker details are available on dbSNP (http://www.ncbi.nlm.nih.gov/SNP). The polymorphism M3, not found in our samples, was typed by using dHPLC with the analysis conditions described (10). YAP was typed by using the method of Hammer and Horai (11). All other polymorphisms were typed by using allele-specific PCR, with primers designed to amplify only in the presence of the particular polymorphism; the primer sets are given in Table 3. PCR reactions were carried out in a 10 μl reaction volume, containing 10 pmol of each primer, 30 ng of genomic DNA, 0.2 units AmpliTaq Gold (Perkin–Elmer), 175 μM of each dNTP, 1X AmpliTaq Gold PCR buffer, and 2.3 mM $MgCl_2$. Primers were optimized to allow the use of the following cycling conditions for all markers: 95° for 10 min, followed by 29 cycles of 95° for 30s, 62° for 30s, and 72° for 30s, followed by a 10 min extension at 72°. Products were resolved by gel electrophoresis or alkaline-mediated differential interaction (AMDI) (12).

An evolutionarily redundant typing strategy was used, based on the relationships presented by Underhill *et al.* (8); "ancestral" and "derived" states were determined from comparisons to non-human primate sequences. Every sample was analyzed in a "first round" typing plate containing all primer sets for the following markers: M20A (primers specific for the ancestral state), M9G, M17del, M45A, M89T, RPS4YT, M173C, and M175del (all specific for the derived state). Thus, an M20G derived sample would show no band for the M20A ancestral primer set, but would show bands for the M9G and M89T derived sets. All subsequent typing for further resolution ("subtyping") was carried out with the most evolutionarily derived marker from the first round, as well as additional primer sets specific for the new markers. For example, individuals typed as M89T in the first round would be further subtyped for M52C, M170C, and M172G, in addition to M89T (see Table 2). Positive and negative control samples were included on each subtyping plate. This typing strategy was possible because of the absolute conservation of evolutionary relationships among the haplotypes, and the method was confirmed by testing 170 samples

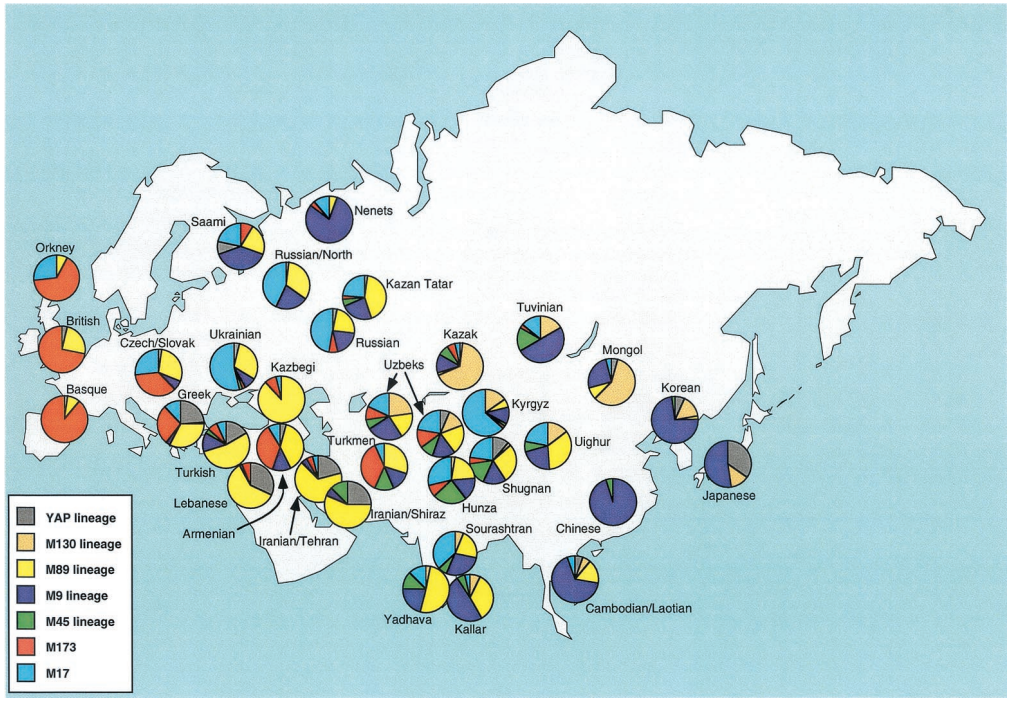

Fig. 1. Geographic distribution of Y-chromosome haplotypes in selected Eurasian populations. Evolutionarily related haplotypes were combined to clarify their display. Colors are those shown in Table 1.

previously typed by dHPLC and sequencing (8), including at least one example of each haplotype presented here. No samples in this study had haplotypes other than those shown in Table 1.

For dating the biallelic lineages, the following microsatellites were analyzed in a subset of the samples containing the lineages of interest: DYS19, DYS388, DYS389I, DYS390, DYS391, DYS392, and DYS395. These were typed on an Applied Biosystems 373 or 377 as described (13).

Analysis. Haplotype assignments were made based on the typing results and the evolutionary relationship of the markers. As described above, haplotypes were named for their most derived marker, because markers were added during the course of the study and will continue to be added in the future as additional polymorphisms are discovered. Haplotype frequencies and diversities [heterozygosity multiplied by $n/(n-1)$; ref. 14] were calculated in Microsoft EXCEL, and the tree analysis was carried out with the GENDIST, SEQBOOT, NEIGHBOR, and CONSENSE programs of PHYLIP (15).

For estimating the age of a haplotype-defining biallelic polymorphism, the equation $t = -N_e\ln(1 - V/N_e\mu)$ was used (16). This equation was derived from the single step mutation model for a haploid population assuming constant population size, where N_e is the effective population size, V is the variance of repeat numbers in the population, and μ is the mutation rate. If the population goes through a strong bottleneck event followed by a rapid population expansion, it can be shown that this formula is still approximately valid (16). Following Semino *et al.* (17), we used $N_e = 4{,}500$ and $\mu = 0.0011$. For each haplotype, the average V over all loci was used for the age estimate. The results for DYS388 were excluded from the analysis because it appears to have undergone at least one multistep mutation (on the lineage ancestral to M172, such that all M172 haplotypes contain higher numbers of DYS388 repeats).

Results

Y chromosome haplotype frequencies are shown in Table 1. Central Asian populations show the highest haplotype diversity for the markers studied here, particularly the Uzbeks, Uighurs, and Karakalpaks, which have very evenly distributed haplotype frequencies.

The haplotype frequencies in selected populations are shown graphically in Fig. 1, with related haplotypes combined into common colors. Data for Basque, Japanese, Middle Eastern, Macedonian, Greek, Cambodian/Laotian, Chinese, Taiwanese, Czech/Slovak, Turkish, Ukrainian, and Hunza populations were taken from the literature (8, 17). There are several frequency variations across Eurasia. The first is that of the M89 lineages (including M89, M170, and M172), which distinguishes between the western and eastern extremes of the continent. It is noteworthy that M172 is a major subset of the 12f2 8kb allele, which has been attributed to the spread of farming from the Near East (6, 17). The M130 haplotypes (M130 + M48) show another pattern, from a maximum in Northeast Asia (e.g., Mongolia), to lower in the south and west. The M45 haplotypes (M45 + M124) show a maximum in Central Asia, with much lower frequencies in Europe, the Middle East, and East Asia. The haplotype defined by M173, at highest frequency in Western Europe (particularly in Basque and British populations), is found at much lower frequency throughout Central Asia and the Middle East. Finally, the M17 haplotypes (M17 + M87) are found at high frequency in Russia, Ukraine, the Czech/Slovak Republics, and throughout Central Asia, but are rare in East Asia and Western Europe.

The neighbor-joining tree (Fig. 2) shows several population clusters defined by branches from a central point. Cluster I is composed primarily of European populations, all characterized by high frequencies of the M173 haplotype. Cluster II contains Middle Eastern populations (Turks, Lebanese, and Iranians), as well as populations from the Caucasus, with some suggestion of a separate branch for the Iranian/Samarkand, Turkish, Kazbegi, and Azeri populations. Cluster III contains the Tuvinian and Nenets populations. Cluster IV is an East Asian cluster, distinguished by high frequencies of M122, M119, M174, and M130. Cluster V contains populations united by high frequencies of

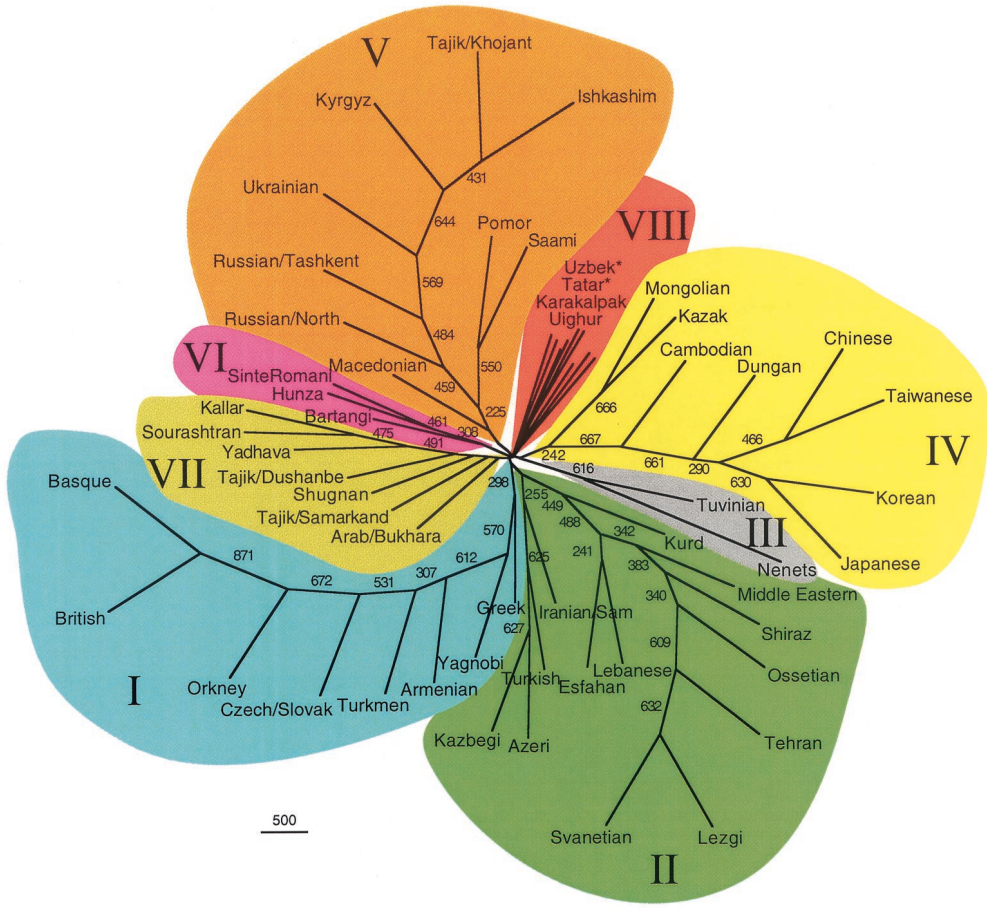

Fig. 2. Neighbor-joining tree of 61 Eurasian populations, based on Y-chromosome biallelic haplotype frequencies. Nei genetic distances were used in a neighbor-joining analysis. Internal numbers are bootstrap values (1000 replicates); values less than 200 are not shown. Roman numerals denote population clusters described in the text. Data for the following populations were taken from the literature: Basque, Cambodian/Laotian, Chinese, Hunza, Japanese, Middle Eastern, and Taiwanese from ref. 8; Czech/Slovak, Greek, Macedonian, Turkish, and Ukrainian from ref. 17. Uzbek* and Tatar* include all Uzbek and Tatar populations shown in Table 1.

M17, and includes Eastern European (Macedonia, Russian, and Ukrainian) and Central Asian (Kyrgyz, Tajik/Khojant, and Ishkashim) populations. Cluster VI includes the Sinte Romani (Gypsies), Hunza (a northern Pakistani population speaking a divergent language, Burushaski), and Bartangi (Pamir region) populations. Cluster VII includes the Indian populations (Kallar, Sourashtran, and Yadhava), as well as those Central Asian populations closest to them geographically (Tajik/Dushanbe, Shugnan, Tajik/Samarkand, and Arab/Bukhara). Finally, cluster VIII includes the settled populations of Central Asia (Uzbeks, Uighurs, Tatars, and Karakalpaks). The populations in the eighth cluster also show the highest haplotype diversity; their close proximity to the center of the tree is consistent with this, because they share haplotypes in common with all other populations.

The analysis of microsatellite diversity on haplotypes defined by biallelic polymorphisms has been widely applied as a method to assess the age of the haplotypes (5, 18). The ages of haplotypes M122 and M172 have been given elsewhere (16, 17), and we have therefore focused on the ages of three other important haplotypes in Central Asia: M45, M173, and M17. M45 is an ancient lineage, estimated to date from 40,000 years ago; surprisingly, for such an old marker, it is almost entirely confined to Central Asia. M173, a descendant lineage of M45, has an estimated age of 30,000 years, a value consistent with the age obtained by Semino *et al.* (17). Finally, M17, a descendant of M173, is apparently much younger, with an inferred age of ≈15,000 years. It must be noted that these age estimates are dependent on many, possibly invalid, assumptions about mutational processes and population structure. Nonetheless, they constitute a basis for comparison to other data.

Discussion

The high haplotype diversity values in the settled Central Asian populations may reflect a bias in the selection of markers that were informative in these populations. Although this bias cannot be ruled out, most of the markers examined here are located relatively deep within the NRY evolutionary tree (8), and thus there is no *a priori* reason to expect a bias toward Central Asia. In addition, Y chromosome microsatellites indicate that Central Asian (Pakistani) populations are the most diverse in Eurasia (19). The consistency of the biallelic and microsatellite results suggests that Central Asian populations are among the oldest on the continent. This pattern of high diversity is consistent with an early settlement of Central Asia by anatomically modern humans, perhaps 40,000–50,000 years ago (see below), followed by subsequent migrations into Europe, America, and India, dispersing M45-, M173-, and M17-derived lineages.

The evolutionary relationship among NRY haplotypes can be used to infer details of past migrations (6). One intriguing feature of the Eurasian haplotype distribution is that M45, the ancestor of haplotypes M173 [the major European haplotype

(17)] and M3 [the major Native American haplotype (10, 20)], is found at polymorphic frequencies mainly in Central Asia. This finding, as well as geography, strongly suggests that the source of both of these migrations was an ancient Central Asian population, consistent with the results of previous studies (21, 22). It is notable that the inferred age of M45 (≈40,000 years) coincides reasonably with the first appearance of anatomically modern humans and their toolkits in southern Siberia (23, 24), during a period that saw the desertification of southern Central Asia and the disappearance of human remains from the southern Central Asian lowlands (25, 26). There may have been a general movement from southern to northern Central Asia during this period, with human populations following migrating herds of large ungulates into the steppe zone. This initial migration may have facilitated subsequent movements to the East and West. The M173 haplotype is thought to delineate the earliest expansion into Europe, during the Upper Paleolithic ≈30,000 years ago (17). It is likely that M173 arose initially in Central Asia, and that M173-carrying subpopulations migrated westward into Europe soon thereafter. The extremely high frequency of this haplotype in Western Europe is probably the result of drift, consistent with an inferred population bottleneck during the Last Glacial Maximum (4, 17).

The American descendant of M45, defined by the marker M3, may be as little as 2,000 years old (10); this finding, as well as the fact that it is not found in Central Asia or Siberia, suggests that the expansion of this haplotype occurred entirely within the Americas. An assessment of the upper limit to the date of entry of humans into the Americas therefore awaits the identification of further markers on the M45 lineage that are ancestral to M3 and are found in both Central Asia and America. What seems clear, though, is that an ancient M45-containing population living in Central Asia was the source of much modern European and Native American Y-chromosome diversity.

The current distribution of the M17 haplotype is likely to represent traces of an ancient population migration originating in southern Russia/Ukraine, where M17 is found at high frequency (>50%). It is possible that the domestication of the horse in this region around 3,000 B.C. may have driven the migration (27). The distribution and age of M17 in Europe (17) and Central/Southern Asia is consistent with the inferred movements of these people, who left a clear pattern of archaeological remains known as the Kurgan culture, and are thought to have spoken an early Indo-European language (27, 28, 29). The decrease in frequency eastward across Siberia to the Altai-Sayan mountains (represented by the Tuvinian population) and Mongolia, and southward into India, overlaps exactly with the inferred migrations of the Indo-Iranians during the period 3,000 to 1,000 B.C. (27). It is worth noting that the Indo-European-speaking Sourashtrans, a population from Tamil Nadu in southern India, have a much higher frequency of M17 than their Dravidian-speaking neighbors, the Yadhavas and Kallars (39% vs. 13% and 4%, respectively), adding to the evidence that M17 is a diagnostic Indo-Iranian marker. The exceptionally high frequencies of this marker in the Kyrgyz, Tajik/Khojant, and Ishkashim populations are likely to be due to drift, as these populations are less diverse, and are characterized by relatively small numbers of individuals living in isolated mountain valleys.

Intriguingly, the population of present-day Iran, speaking a major Indo-European language (Farsi), appears to have had little genetic influence from the M17-carrying Indo-Iranians. It is possible that the pre-Indo-European population of Iran—effectively an eastern extension of the great civilizations of Mesopotamia—may have reached sufficient population densities to have swamped any genetic contribution from a small number of immigrating Indo-Iranians. If so, this may have been a case of language replacement through the "elite-dominance" model (29). Alternatively, an Indo-Iranian language may have been the *lingua franca* of the steppe nomads and the surrounding settled populations, facilitating communication between the two. Over time, this language could have become the predominant language in Persia, reinforced and standardized by rulers such as Cyrus the Great and Darius in the mid-first millennium B.C. Whichever model is correct, the Iranians sampled here (from the western part of the country) appear to be more similar genetically to Afro-Asiatic-speaking Middle Eastern populations than they are to Central Asians or Indians. This finding contrasts with a recent analysis of Eastern Iranian populations, which have high frequencies of Y-chromosome haplogroup 3, defined by the M17 analogue SRY-1532A (30). It is likely that the Dasht-e Kavir and Dasht-e Lut deserts in the center of the country have acted as significant barriers to gene flow.

The Turkish and Azeri populations are atypical among Altaic speakers (Table 1) in having low frequencies of M130, M48, M45, and M17 haplotypes. Rather, these two Turkic-speaking groups seem to be closer to populations from the Middle East and Caucasus, characterized by high frequencies of M96- and/or M89-related haplotypes. This finding is consistent with a model in which the Turkic languages, originating in the Altai-Sayan region of Central Asia and northwestern Mongolia (31), were imposed on the Caucasian and Anatolian peoples with relatively little genetic admixture—another possible example of elite dominance-driven linguistic replacement.

The Sinte Romani, or Gypsies, group with the Hunza and Bartangi (from the Pamir region of Central Asia) populations in our tree. This finding is primarily due to the M124 haplotype, which is present at high frequency in all three populations. M124 is not found in Eastern Europe (17), where the Sinte Romani lived before being resettled in Central Asia in the 1940s. It is, however, common in Central and Southern Asia. Thus, the Y-chromosome results provide clear genetic evidence of a link between the Gypsies and their Asian kin.

Bodmer (32) suggested that the Celtic populations of Britain trace their origins to an early settlement of the British Isles by Paleolithic Europeans, rather than by a later migration associated with the spread of the Celtic culture from central Europe in the first millennium B.C. (27). The results given here support this view, with the British sample grouping closely with the Basque—a presumed remnant of the pre-Neolithic European population (33); both populations show similarly high frequencies of the M173 haplotype. The Orkney population is also characterized by a high frequency of M17, which was not found in our British sample. It is likely that the presence of M17 in the Orcadians is due to admixture with Viking invaders in the ninth and tenth centuries, as noted recently by Wilson *et al.* (33).

Altheide and Hammer (34) have suggested that haplotypes defined by the presence of the YAP insertion originated in Asia and spread back to Africa. One prediction of this model is that the ancestral state of this lineage, which would be YAP(+) but ancestral for both the eastern (M174C) and western (M96C) sublineages (8), should be found in the Asian population(s) where the insertion originally occurred. We do not find any such ancestral chromosomes in our study. Although we cannot rule out the possibility that an ancestral YAP(+) chromosome will be found as more samples are analyzed, the current survey of ≈2,000 men does not support an Asian origin for the YAP(+) lineage, consistent with the results of Underhill *et al.* (6).

The sketch of Eurasian population movements outlined here is, admittedly, based entirely on Y-chromosome evidence. The actual history of these populations presumably has included the migration of women, and thus the "genetic history" of Eurasia awaits further study of mitochondrial DNA and autosomal markers. In addition, the smaller population size of the Y chromosome (one-fourth that of the autosomes), as well as the possible confounding effects of sexual and/or natural selection, makes it imperative that additional markers are examined to

provide a complement to the Y data. Despite these caveats, it is clear from the pattern of Y-chromosome diversity that Central Asia has played a critical role in human history.

Note Added in Proof. A recently published analysis of craniofacial data from Old and New World populations supports a common origin for Upper Paleolithic Europeans and the earliest Native Americans (35). The consistency of these craniometric results and the genetic results described here may provide an explanation for the "unusual" morphological characteristics of many early Native American remains, such as those from Kennewick, USA.

This project would have been impossible without the cooperation of the 1,935 men who donated blood samples—our sincere thanks for their trust and interest in this study. This project was supported by a grant to R.S.W. from the Alfred P. Sloan Foundation and to W.F.B. from the Imperial Cancer Research Fund. Sample collections were supported in part by sponsorship from Land Rover, Brockbank, Virgin Atlantic Airways, Olympus, Kodak, and Garmont. N.Y. was supported by a Royal Society/North Atlantic Treaty Organization (NATO) Postdoctoral Research Fellowship, and I.E. by a Royal Society Visiting Fellowship. P.A.U. was supported by National Institutes of Health Grant GMS28428.

1. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. (1994) *The History and Geography of Human Genes* (Princeton Univ. Press, Princeton).
2. Bowcock, A. M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J. R. & Cavalli-Sforza, L. L. (1994) *Nature (London)* **368,** 455–457.
3. 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.
4. Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J., Lavery, T., Kouyoumjian, R., Farhadian, S. F., Ward, R. & Lander, E. S. (2001) *Nature (London)* **411,** 199–204.
5. Jobling, M. A. & Tyler-Smith, C. (1995) *Trends Genet.* **11,** 449–456.
6. Underhill, P. A., Passarino, G., Lin, A. A., Shen, P., Foley, R. A., Mirazon-Lahr, M., Oefner, P. J. & Cavalli-Sforza, L. L. (2001) *Ann. Hum. Genet.* **65,** 43–62.
7. Seielstad, M. T., Minch, E. & Cavalli-Sforza, L. L. (1998) *Nat. Genet.* **20,** 278–280.
8. Underhill, P. A., Shen, P., Lin, A. A., Jin, L., Passarino, G., Yang, W. H., Kauffman, E., Bonné-Tamir, B., Bertranpetit, J., Francalacci, P., *et al.* (2000) *Nat. Genet.* **26,** 358–361.
9. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (2001) *Current Protocols in Molecular Biology* (Wiley, New York).
10. Underhill, P. A., Jin, L., Zemans, R., Oefner, P. J. & Cavalli-Sforza, L. L. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 196–200.
11. Hammer, M. F. & Horai, S. (1995) *Am. J. Hum. Genet.* **56,** 951–962.
12. Bartlett, S., Straub, J., Tonks, S., Wells, R. S., Bodmer, J. G. & Bodmer, W. F. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 2694–2697. (First Published February 20, 2001; 10.1073/pnas.041619998)
13. Perez-Lezaun, A., Calafell, F., Seielstad, M., Mateu, E., Comas, D., Bosch, E. & Bertranpetit, J. (1997) *J. Mol. Evol.* **45,** 265–270.
14. Nei, M. & Tajima, F. (1981) *Genetics* **97,** 145–163.
15. Felsenstein, J. (1989) *Cladistics* **5,** 164–166.
16. 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.
17. Semino, O., Passarino, G., Oefner, P. J., Lin, A. A., Arbuzova, S., Beckman, L. E., de Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., *et al.* (2000) *Science* **290,** 1155–1159.
18. Zerjal, T., Dashnyam, B., Pandya, A., Kayser, M., Roewer, L., Santos, F. R., Schiefenhovel, W., Fretwell, N., Jobling, M. A., Harihara, S., *et al.* (1997) *Am. J. Hum. Genet.* **60,** 1174–1183.
19. Seielstad, M. T., Bekele, E., Ibrahim, M., Touré, A. & Traoré, M. (1999) *Genome Res.* **9,** 558–567.
20. 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.
21. Karafet, T. M., Zegura, S. L., Posukh, O., Osipova, L., Bergen, A., Long, J., Goldman, D., Klitz, W., Hrihara, S., de Knijff, P., *et al.* (1999) *Am. J. Hum. Genet.* **64,** 817–831.
22. Santos, F. R., Pandya, A., Tyler-Smith, C., Pena, S. D. J., Schanfield, M., Leonard, W. R., Osipova, L., Crawford, M. H. & Mitchell, R. J. (1999) *Am. J. Hum. Genet.* **64,** 619–628.
23. Okladnikov, A. P. (1964) in *The Peoples of Siberia*, eds. Levin, M.G. & Potapov, L. P. (Univ. of Chicago Press, Chicago), pp. 13–98.
24. Klein, R. G. (1999) *The Human Career* (Univ. of Chicago Press, Chicago), 2nd Ed.
25. Kohl, P. L. (1984) *L'Asie Central des Origines a l'Age de Fer* (Editions sur les Civilisations, Paris).
26. Madeyska, T. (1992) in *Atlas of Paleoclimates and Paleoenvironments of the Northern Hemisphere: Late Pleistocene-Holocene*, eds. Frenzel, B., Pécsi, M. & Velichko, A. A. (Hungarian Academy of Sciences, Budapest), pp. 61–63.
27. Mallory, J. P. (1989) *In Search of the Indo-Europeans* (Thames and Hudson, London).
28. Gimbutas, M. (1970) in *Indo-European and Indo-Europeans*, eds. Cardona, G., Koenigswald, H. M. & Senn, A. (Univ. of Pennsylvania Press, Philadelphia), pp. 155–198.
29. Renfrew, C. (1987) *Archaeology and Language* (Cambridge Univ. Press, Cambridge, U.K.).
30. Quintana-Murci, L., Krausz, C., Zerjal, T., Sayar, S. H., Hammer, M. F., Mahdi, S. Q., Ayub, Q., Qamar, R., Mohyuddin, A., Radhakrishna, U., *et al.* (2001) *Am. J. Hum. Genet.* **68,** 537–542.
31. Comrie, B. (1981) *Languages of the Soviet Union* (Cambridge Univ. Press, Cambridge, U.K.).
32. Bodmer, W. F. (1992) *Proc. Br. Acad.* **82,** 37–57.
33. Wilson, J. F., Weiss, D. A., Richards, M., Thomas, M. G., Bradman, N. & Goldstein, D. B. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 5078–5083. (First Published April 3, 2001; 10.1073/pnas.071036898)
34. Altheide, T. K. & Hammer, M. F. (1997) *Am. J. Hum. Genet.* **61,** 462–466.
35. Brace, C. L., Nelson, A. R., Seguchi, N., Oe, H., Sering, L., Qifeng, P., Yongyi, L. & Tumen, D. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 10017–10022. (First Published July 31, 2001; 10.1073/pnas.171305898)