Am. J. Hum. Genet. 67:1526–1543, 2000

Y-Chromosomal Diversity in Europe Is Clinal and Influenced Primarily by Geography, Rather than by Language

Zoë H. Rosser,[1] Tatiana Zerjal,[2] Matthew E. Hurles,[1,*] Maarja Adojaan,[5] Dragan Alavantic,[6] António Amorim,[7] William Amos,[8] Manuel Armenteros,[9] Eduardo Arroyo,[10] Guido Barbujani,[11] Gunhild Beckman,[12] Lars Beckman,[12] Jaume Bertranpetit,[13] Elena Bosch,[13,†] Daniel G. Bradley,[14] Gaute Brede,[15] Gillian Cooper,[8] Helena B. S. M. Côrte-Real,[16] Peter de Knijff,[17] Ronny Decorte,[18] Yuri E. Dubrova,[1] Oleg Evgrafov,[19] Anja Gilissen,[18] Sanja Glisic,[6] Mukaddes Gölge,[20] Emmeline W. Hill,[14] Anna Jeziorowska,[21] Luba Kalaydjieva,[22] Manfred Kayser,[23,‡] Toomas Kivisild,[3] Sergey A. Kravchenko,[24] Astrida Krumina,[25] Vaidutis Kučinskas,[26] João Lavinha,[16] Ludmila A. Livshits,[24] Patrizia Malaspina,[27] Syrrou Maria,[28] Ken McElreavey,[29] Thomas A. Meitinger,[30] Aavo-Valdur Mikelsaar,[4] R. John Mitchell,[31] Khedoudja Nafa,[32] Jayne Nicholson,[3] Søren Nørby,[33] Arpita Pandya,[2] Jüri Parik,[5] Philippos C. Patsalis,[28] Luísa Pereira,[7] Borut Peterlin,[34] Gerli Pielberg,[5] Maria João Prata,[7] Carlo Previderé,[35] Lutz Roewer,[23] Siiri Rootsi,[5] D. C. Rubinsztein,[36] Juliette Saillard,[33] Fabrício R. Santos,[2,§] Gheorghe Stefanescu,[37] Bryan C. Sykes,[32] Aslihan Tolun,[38] Richard Villems,[5] Chris Tyler-Smith,[2] and Mark A. Jobling[1]

Clinal patterns of autosomal genetic diversity within Europe have been interpreted in previous studies in terms of a Neolithic demic diffusion model for the spread of agriculture; in contrast, studies using mtDNA have traced many founding lineages to the Paleolithic and have not shown strongly clinal variation. We have used 11 human Y-chromosomal biallelic polymorphisms, defining 10 haplogroups, to analyze a sample of 3,616 Y chromosomes belonging to 47 European and circum-European populations. Patterns of geographic differentiation are highly nonrandom, and, when they are assessed using spatial autocorrelation analysis, they show significant clines for five of six haplogroups analyzed. Clines for two haplogroups, representing 45% of the chromosomes, are continentwide and consistent with the demic diffusion hypothesis. Clines for three other haplogroups each have different foci and are more regionally restricted and are likely to reflect distinct population movements, including one from north of the Black Sea. Principal-components analysis suggests that populations are related primarily on the basis of geography, rather than on the basis of linguistic affinity. This is confirmed in Mantel tests, which show a strong and highly significant partial correlation between genetics and geography but a low, nonsignificant partial correlation between genetics and language. Genetic-barrier analysis also indicates the primacy of geography in the shaping of patterns of variation. These patterns retain a strong signal of expansion from the Near East but also suggest that the demographic history of Europe has been complex and influenced by other major population movements, as well as by linguistic and geographic heterogeneities and the effects of drift.

Introduction

The earliest accepted date for the occupation of Europe by anatomically modern humans is ~40,000 years before the present (YBP) (Boyd and Silk 1997). Population size during the Paleolithic was probably stable and small, limited by the resources available from a hunting-gathering economy (Landers 1992). The development of agriculture (the Neolithic transition) was important, because the abundance of food supplies allowed populations to expand (Hassan 1973).

The origins of agriculture have become the focus of

Received July 10, 2000; accepted for publication September 25, 2000; electronically published November 9, 2000.

Address for correspondence and reprints: Dr. Mark A. Jobling, Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, United Kingdom. Email: maj4@leicester.ac.uk

0002-9297/2000/6706-0018$02.00

[1]Department of Genetics, University of Leicester, Leicester; [2]CRC Chromosome Molecular Biology Group, Department of Biochemistry, and [3]Institute of Molecular Medicine, University of Oxford, Oxford; [4]Institute of General and Molecular Pathology, University of Tartu and [5]Estonian Biocentre, Tartu, Estonia; [6]Laboratory for Radiobiology and Molecular Genetics, Institute of Nuclear Sciences "Vinca," Belgrade; [7]5 IPATIMUP and Faculdade de Ciências, Universidade do Porto, Porto, Portugal; [8]Department of Zoology, University of Cambridge, Cambridge; [9]Centro de Investigacion y Criminalistica, Laboratorio de ADN, Policia Judicial, Guardia Civil, and [10]Laboratorio de Biología Forense, Departamento de Toxicología y Legislación Sanitaria, Univ-

attempts to interpret the genetic landscape of modern Europe. The fact that agriculture arose in the Near East ~10,000 YBP (evinced by the dating of archaeological sites) is not disputed; the argument has arisen over the mechanism of its subsequent dispersal. In the demic diffusion model (Ammerman and Cavalli-Sforza 1984), the spread is thought to be due to a movement of people and would therefore have substantially changed the genetic composition of European populations; the contrasting, cultural diffusion model (Dennell 1983; Zvelebil and Zvelebil 1988) holds that the ideas and technologies were transferred without substantial population movement and thus suggests that current patterns of genetic diversity should have their roots in the Paleolithic.

These opposing hypotheses are undoubtedly overly simplistic but have been widely adopted as models in genetic studies (Sokal et al. 1991; Cavalli-Sforza et al. 1993; Barbujani et al. 1994; Piazza et al. 1995; Semino et al. 1996; Chikhi et al. 1998*a,* 1998*b;* Richards and Sykes 1998; Simoni et al. 2000*a*), since they predict patterns of diversity that should be easily recognizable—in particular, demic diffusion is expected to result in clines with foci in the Near East. Principal components (PC) analysis of classical gene-frequency data reveals clines within Europe, and the first principal component, which indeed has a Near Eastern focus, has been taken to support the demic diffusion hypothesis (Menozzi et al. 1978; Cavalli-Sforza et al. 1993). A similar pattern has been observed in spatial autocorrelation analysis of DNA-based polymorphisms, including microsatellites, which have identified geographic patterns compatible with a substantial directional demographic expansion affecting much of the continent (Chikhi et al. 1998*a*). However, although these patterns in the genetic data are impressive and suggest major east-west population movements, their time depths are not known, and associating them with particular demographic events is usually speculative. They could be just as well due to the original peopling of Europe during the Upper Paleolithic as to the Neolithic transition. In this regard, some support for the latter does come from the finding of significant partial correlations between classical marker frequencies and the relative dates for the origin of agriculture in different locations (Sokal et al. 1991).

By contrast, analysis of diversity in European mtDNA reveals a relatively homogeneous landscape (Comas et al. 1997), with clines detectable only in the south (Simoni et al. 2000*a*). However, this is a contentious area, and conclusions may depend on the depth of analysis—for example, which sublineages are studied. An east-west gradient of pairwise differences has been discerned and claimed to be compatible with expansion from the Middle East (Comas et al. 1997). However, attempts to identify and date founding lineages (Richards et al. 1996) have suggested that Paleolithic lineages may persist in Europe to a degree that is inconsistent with the demic diffusion hypothesis, although an ancient origin of certain alleles or haplogroups (HGs) is certainly compatible with a later spread of those alleles within a geographic region (Langaney et al. 1992; Templeton 1993).

Language can provide additional evidence about past demography (Renfrew 1989), although direct information about past languages on the basis of writing is limited to the past 5,000 years, and inferences before that time are controversial (Renfrew 2000). Europe is remarkable for its linguistic homogeneity, languages of the Indo-European (IE) family being spoken by most populations from India to Ireland (Renfrew 1989). In one persuasive view, demic diffusion from the Near East provides a common explanation for the spread of both

ersidad Complutense, Madrid; [11]Dipartimento di Biologia, Universita di Ferrara, Ferrara, Italy; [12]Umeå University, Department of Medical Genetics, Umeå, Sweden; [13]Unitat de Biologia Evolutiva, Facultat de Ciecies de la Salut I de la Vida, Universitat Pompeu Fabra, Barcelona; [14]Department of Genetics, Trinity College, Dublin; [15]University of Oslo, Centre for Biotechnology, Oslo; [16]Instituto Nacional de Saúde Dr. Ricardo Jorge, Lisbon; [17]Forensic Laboratory for DNA Research, MGC-Department of Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands; [18]Laboratory for Forensic Genetics and Molecular Archaeology, Center for Human Genetics, K.U. Leuven, Leuven, Belgium; [19]Research Centre for Medical Genetics, Russian Academy of Medical Sciences, UFA Science Centre, Department of Biochemistry and Cytochemistry, Moscow; [20]Department of Physiology, University of Kiel, Kiel; [21]Department of Medical Genetics, Institute of Endocrinology, Medical University of Lódz, Lódz, Poland; [22]Department of Human Biology, Edith Cowan University, Joondalup Campus, and Western Australian Institute for Medical Research, Royal Perth Hospital, Perth; [23]Genetic Research Laboratory, Institute of Legal Medicine, Medical Faculty (Charité), Humboldt-University Berlin, Berlin; [24]Institute of Molecular Biology and Genetics, National Academy of Science of Ukraine, Kiev; [25]Department of Medical Biology and Genetics, Medical Academy of Latvia, Riga; [26]Center of Human Genetics, University of Vilnius, Vilnius, Lithuania; [27]Department of Biology, University of Rome "Tor Vergata," Rome; [28]The Cyprus Institute of Neurology and Genetics, Nicosia; [29]Unité d'Immunogénétique Humaine, Institut Pasteur, Paris; [30]Department of Medical Genetics, Kinderpoliklinik, Munich; [31]La Trobe University, School of Genetics and Human Variation, Bundoora, Australia [32]Department of Human Genetics, Memorial Sloan-Kettering Cancer Center, New York; [33]Laboratory of Biological Anthropology, Institute of Forensic Medicine, University of Copenhagen, Copenhagen; [34]Division of Medical Genetics, Department of Obstetrics and Gynaecology, Ljubljana, Slovenia; [35]Dipartimento di Medicina Legale e Sanita Pubblica, Pavia, Italy; [36]I.C. Biologice, Iasi, Romania; and [37]Bogazici University, Department of Molecular Biology and Genetics, Istanbul

[*] Present affiliation: McDonald Institute for Archaeological Research, University of Cambridge, Cambridge.

[†] Present affiliation: Department of Genetics, University of Leicester, Leicester, United Kingdom.

[‡] Present affiliation: Max Planck Institute for Evolutionary Anthropology, Department of Evolutionary Genetics, Leipzig.

[§] Present affiliation: Departamento de Biologia Geral, Instituto Ciências Biológicas/Universidade Federal de Minas Gerais, Minas Gerais, Brazil.

agriculture and IE languages (Renfrew 1987). Other ideas have been put forward, however; one, which has been adopted by some geneticists because of its apparent compatibility with the pattern seen in the third principal component of variation of classical gene frequencies (Cavalli-Sforza et al. 1994), is that the IE language was spread by the movement, from north of the Caspian Sea, of the Kurgan people, pastoral nomads who domesticated the horse (Gimbutas 1970). An alternative view has it that the spread of IE language preceded the origins of agriculture and was due to the reexpansion of hunter-gatherers after the end of the Last Glacial Maximum (Adams and Otte 2000).

Despite the hegemony of IE languages, there is diversity within them, and some members of other language families also exist; one example, Basque, clearly represents a survival from an earlier era. Various methods for the detection of genetic barriers in autosomal gene frequencies within Europe (Barbujani 1991) show that most of these barriers correlate with linguistic boundaries, and it may be that language and geographic proximity are equally good predictors of genetic affinity (Barbujani 1997). However, some examples of non-IE languages reflect not persistence but recent acquisition through "elite dominance": for example, the Hungarians acquired their Uralic language from the invading Magyars only ~1,100 YBP (Cavalli-Sforza et al. 1994), and the Altaic language of the Turks was acquired as a result of the Turkic invasions during the 11th–15th centuries (Renfrew 1989). This process of language acquisition by elite dominance is not expected to be accompanied by a high degree of genetic admixture, and, if this is so, populations such as the Hungarians and Turks are unlikely to be separated from surrounding populations by genetic barriers.

Use of the Y chromosome to investigate human population histories (Jobling and Tyler-Smith 1995) is increasing as convenient polymorphic markers become available. However, the effective population size of this chromosome is one-quarter that of any autosome, and this means that it is particularly influenced by drift. Effective population size may be further reduced through the variance in the number of sons that a father has and perhaps by selective sweeps (Jobling and Tyler-Smith 2000). Conclusions about populations on the basis of this single locus must therefore be made with caution. One useful property of the Y chromosome is its high degree of geographic differentiation, compared with other parts of the genome, which has been explained by drift and a greater effective migration of women than of men, through the phenomenon of patrilocality (Seielstad et al. 1998), in which women are more likely to move from their birthplace after marriage than are men. The Y chromosome may therefore be a sensitive system for detecting the population movements that have shaped European genetic diversity; there again, it may be so susceptible to drift that ancient patterns have been obscured.

Published data on European Y-chromosome diversity are not extensive; markers have been of limited informativeness, and the distribution of population samples has often been unsatisfactory. By use of two "classical" Y-chromosome markers—the complex and highly polymorphic 49f/*Taq*I system (Ngo et al. 1986; Lucotte and Loirat 1999) and the biallelic marker 12f2 (Casanova et al. 1985)—patterns of diversity have been demonstrated that have been claimed to be clinal and to support the demic diffusion model (Semino et al. 1996). Subsequent analysis using Y-chromosome–specific microsatellites (Quintana-Murci et al. 1999) and a combination of microsatellites and two biallelic markers (Malaspina et al. 1998) showed similar east-west gradients. 49f has been exploited more fully to analyze the correlation between Y-chromosome diversity, mtDNA diversity, and language in a global sample, and it has been suggested that the Y chromosome shows the stronger correlation with language (Poloni et al. 1997).

Recent progress in the development of Y-chromosome polymorphic markers that can be assayed by use of PCR now allows us to explore these issues in greater detail. In this study, we use 11 such markers to assay the diversity of Y-chromosomal lineages in a large sample of men from 47 populations distributed over most of Europe.

Subjects and Methods

Subjects

Y chromosomes from 3,616 men from 47 populations (table 1) were included in this study; the majority were classified by birthplace of the paternal grandfather. DNA samples were from collections of the authors, and informed consent was obtained. A total of 311 samples from the Baltic region are from the study by T. Zerjal, L. Beckman, G. Beckman, A.-V. Mikelsaar, A. Krumina, V. Kučinskas, M. E. Hurles, and C. Tyler Smith (unpublished data). The 257 Irish Y chromosomes included 221 chromosomes studied elsewhere (Hill et al. 2000), which were typed here with three additional markers. The 129 North African samples were those studied elsewhere by Bosch et al. (1999); chromosomes with the M9 G allele and 92R7 C allele were additionally typed with LLY22g (see below). The 172 East Anglian samples were studied elsewhere by Cooper et al. (1996).

Biallelic Markers

A total of 11 biallelic markers were used in this study (fig. 1). These were chosen on the basis of previous work by us and by others (Santos and Tyler-Smith 1996; Sem-

Table 1

HG Frequency Data in 47 Populations

Population (no.)	Location	Language Family (Subfamily)	No. (%) of Individuals with HG											
			1	2	3	4	7	8	9	12	16	21	22	26
Icelandic (28)	64°1′N, 21°6′W	IE (Germanic)	13 (46)	9 (32)	6 (21)	0	0	0	0	0	0	0	0	0
Saami (48)	68°N, 22°E	Uralic (Finno-Ugric)	3 (6)	15 (31)	10 (21)	0	0	0	0	0	20 (42)	0	0	0
Northern Swedish (48)	63°7′N, 20°3′E	IE (Germanic)	11 (23)	22 (48)	9 (19)	0	0	0	1 (2)	0	4 (8)	1 (2)	0	0
Gotlander (64)	57°5′N, 18°5′E	IE (Germanic)	11 (17)	38 (59)	10 (16)	0	0	0	0	0	4 (6)	0	0	1 (2)
Norwegian (52)	59°9′N, 10°8′E	IE (Germanic)	15 (29)	17 (33)	16 (31)	0	0	0	1 (2)	0	2 (4)	1 (2)	0	0
Danish (56)	55°7′N, 12°6′E	IE (Germanic)	28 (50)	18 (32)	4 (7)	0	0	0	4 (7)	0	1 (2)	1 (2)	0	0
Finnish (57)	60°1′N, 25°E	Uralic (Finno-Ugric)	1 (2)	13 (23)	6 (10)	0	0	0	0	1 (2)	35 (61)	1 (2)	0	0
Estonian (207)	59°4′N, 24°7′E	Uralic (Finno-Ugric)	18 (9)	30 (14)	56 (27)	0	0	0	2 (1)	8 (4)	76 (37)	6 (3)	0	11 (5)
Latvian (34)	56°9′N, 24°1′E	IE (Balto-Slavic)	5 (15)	4 (12)	14 (41)	0	0	0	0	0	11 (32)	0	0	0
Lithuanian (38)	54°7′N, 25°3′E	IE (Balto-Slavic)	2 (5)	5 (13)	13 (34)	0	0	0	0	0	18 (47)	0	0	0
Russian (122)	55°8′N, 37°7′E	IE (Balto-Slavic)	8 (7)	21 (17)	57 (47)	0	0	0	5 (4)	5 (4)	17 (14)	8 (7)	0	1 (1)
Belarusian (41)	53°9′N, 27°5′E	IE (Balto-Slavic)	4 (10)	14 (34)	16 (39)	0	0	0	1 (2)	0	1 (2)	4 (10)	0	1 (2)
Ukrainian (27)	50°4′N, 30°5′E	IE (Balto-Slavic)	1 (4)	13 (48)	8 (30)	0	0	0	0	0	4 (11)	1 (4)	0	0
Mari (48)	56°5′N, 48°E	Uralic (Finno-Ugric)	5 (10)	2 (4)	14 (29)	0	0	0	3 (6)	8 (17)	16 (33)	0	0	0
Chuvash (17)	55°5′N, 47°E	Altaic (Turkic)	2 (12)	4 (24)	3 (18)	0	0	0	1 (6)	0	3 (18)	1 (6)	0	3 (18)
Georgian (64)	41°5′N, 44°5′E	Caucasian (Southern Caucasian)	12 (19)	31 (48)	4 (6)	0	0	0	15 (23)	0	0	1 (2)	0	1 (2)
Ossetian (47)	43°1′N, 44°5′E	IE (Indo-Iranian)	20 (43)	5 (11)	1 (2)	0	0	0	16 (34)	0	0	3 (6)	0	2 (4)
Armenian (89)	40°2′N, 44°5′E	IE (Armenian)	22 (25)	28 (31)	5 (6)	0	0	0	26 (29)	0	3 (3)	3 (3)	0	2 (2)
Turkish (167)	41°N, 29°E	Altaic (Turkic)	34 (20)	41 (25)	8 (5)	0	0	0	55 (33)	2 (1)	2 (1)	17 (10)	0	8 (5)
Cypriot (45)	35°3′N, 33°4′E	IE (Greek)	4 (9)	10 (22)	1 (2)	0	0	0	15 (33)	1 (2)	0	12 (27)	0	2 (4)
Greek (36)	38°N, 23°7′E	IE (Greek)	4 (11)	8 (22)	3 (8)	0	0	0	10 (28)	0	0	10 (28)	0	1 (3)
Bulgarian (24)	42°7′N, 23°3′E	IE (Balto-Slavic)	4 (17)	10 (42)	3 (12)	0	0	0	3 (12)	0	0	4 (17)	0	0
Czech (53)	50°2′N, 14°5′E	IE (Balto-Slavic)	10 (19)	10 (19)	20 (38)	0	0	0	6 (11)	3 (6)	0	4 (8)	0	0
Slovakian (70)	48°1′N, 17°1′E	IE (Balto-Slavic)	12 (17)	12 (17)	33 (47)	0	0	0	2 (3)	1 (1)	2 (3)	7 (10)	0	1 (1)
Romanian (45)	44°4′N, 26°1′E	IE (Italic)	8 (18)	12 (27)	9 (20)	0	0	0	11 (24)	0	0	3 (7)	1 (2)	1 (2)
Yugoslavian (100)	44°8′N, 20°5′E	IE (Balto-Slavic)	11 (11)	49 (49)	16 (16)	0	0	0	8 (8)	2 (2)	0	13 (13)	0	1 (1)
Slovenian (70)	46°1′N, 14°5′E	IE (Balto-Slavic)	15 (21)	19 (27)	26 (37)	0	0	0	4 (6)	0	0	5 (7)	1 (1)	0
Hungarian (36)	47°5′N, 19°1′E	Uralic (Finno-Ugric)	11 (30)	10 (28)	8 (22)	0	0	0	1 (3)	0	0	6 (17)	0	0
Polish (112)	51°7′N, 19°5′E	IE (Balto-Slavic)	20 (18)	19 (17)	61 (54)	0	0	0	4 (4)	1 (1)	5 (4)	2 (2)	0	0
Italian (99)	41°9′N, 12°5′E	IE (Italic)	44 (44)	14 (14)	2 (2)	0	0	0	20 (20)	0	0	13 (13)	0	6 (6)
Sardinian (10)	39°2′N, 9°1′E	IE (Italic)	3 (30)	4 (40)	0	0	0	1 (10)	0	0	0	2 (20)	0	0
Bavarian (80)	48°1′N, 11°6′E	IE (Germanic)	38 (48)	18 (23)	12 (15)	0	0	0	4 (5)	0	0	6 (8)	2 (3)	0
German (30)	52°5′N, 13°4′E	IE (Germanic)	12 (40)	6 (20)	9 (30)	0	0	0	1 (3)	0	1 (3)	0	0	1 (3)
Dutch (84)	52°3′N, 4°9′E	IE (Germanic)	36 (43)	27 (32)	11 (13)	0	0	0	6 (7)	0	0	3 (8)	1 (1)	0
French (40)	48°9′N, 2°3′E	IE (Italic)	20 (50)	10 (25)	2 (5)	0	0	1 (3)	2 (5)	0	0	3 (8)	2 (5)	0
Belgian (92)	50°8′N, 4°3′E	IE (Germanic)	58 (63)	21 (23)	4 (4)	0	0	0	5 (5)	0	0	2 (2)	1 (1)	1 (1)
Western Scottish (120)	57°2′N, 6°2′W	IE (Celtic)	87 (72)	23 (19)	8 (7)	0	0	0	0	0	0	2 (2)	0	0
Scottish (43)	56°N, 3°2′W	IE (Celtic)	34 (79)	5 (12)	3 (7)	0	0	0	0	0	0	0	0	1 (2)
Cornish (51)	50°3′N, 4°4′W	IE (Celtic)	42 (82)	9 (18)	0	0	0	0	0	0	0	0	0	0
East Anglian (172)	52°6′N, 1°3′E	IE (Germanic)	97 (56)	52 (30)	15 (9)	0	0	0	1 (1)	0	0	5 (3)	1 (1)	1 (1)
Irish (257)	53°3′N, 6°3′W	IE (Celtic)	207 (81)	39 (15)	2 (1)	0	0	0	2 (1)	0	1 (.5)	6 (2)	0	0
Basque (26)	43°3′N, 2°9′W	Basque (Basque)	19 (73)	2 (8)	0	0	0	0	0	0	0	0	5 (19)	0
Spanish (126)	40°4′N, 3°7′W	IE (Italic)	86 (68)	17 (13)	3 (2)	0	0	0	4 (3)	0	0	12 (10)	3 (2)	1 (1)
Southern Portuguese (57)	38°7′N, 9°1′W	IE (Italic)	32 (56)	8 (14)	1 (2)	0	0	0	5 (9)	0	0	10 (17)	0	1 (2)
Northern Portuguese (328)	41°2′N, 8°6′W	IE (Italic)	203 (62)	54 (16)	0	0	0	0	21 (6)	0	0	35 (11)	6 (2)	9 (3)
Algerian (27)	36°5′N, 3°E	Afro-Asiatic (Semitic)	0	1 (4)	0	0	0	1 (4)	11 (41)	0	0	14 (52)	0	0
Northern African (129)	35°5′N, 5°7′W	Afro-Asiatic (Berber and Semitic)	5 (4)	4 (3)	0	0	0	6 (5)	15 (12)	0	0	99 (77)	0	0
Total (3,616)			1,337 (37)	803 (22)	512 (14)	0	0	9 (0.3)	291 (8)	32 (1)	226 (6)	326 (9)	23 (0.7)	57 (2)

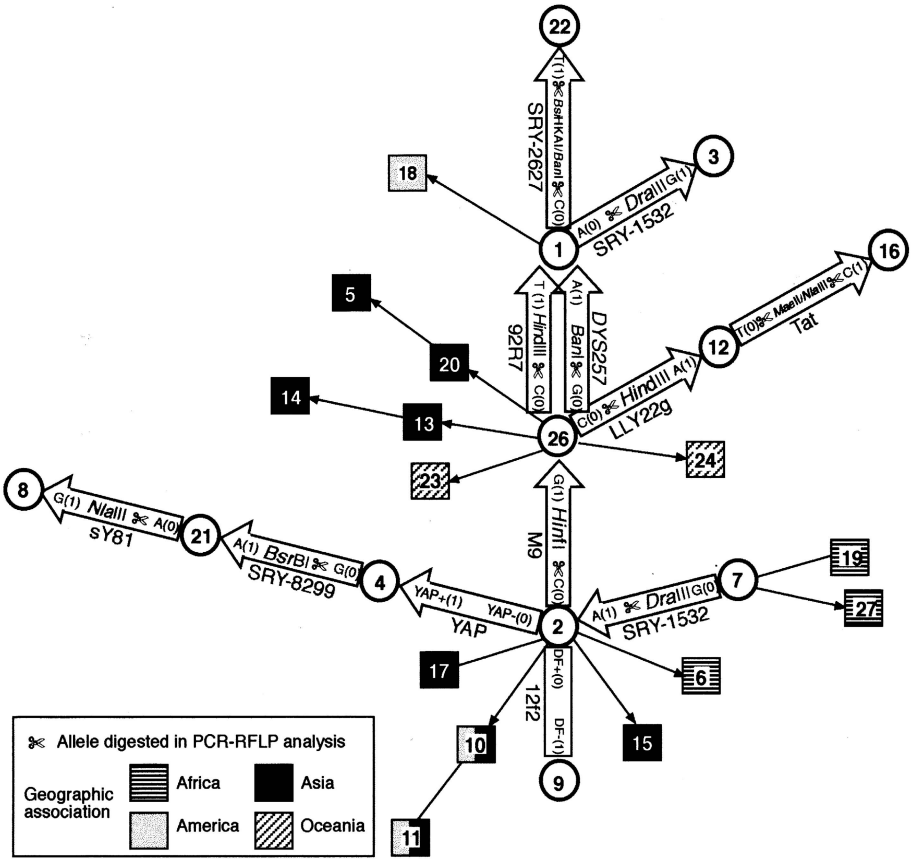

Figure 1 Maximum-parsimony network of Y-chromosomal biallelic HGs. Circles and squares represent compound haplotypes, or HGs; numbers within them are their arbitrarily assigned names; and arrows or lines between them represent the defining biallelic mutations. The order of occurrence of the 92R7 and *DYS257* mutations is not known, because the intermediate HG has not been found; arrows for these polymorphisms are shown adjacent to each other. Where ancestral state is known, arrows point to the derived state. HGs analyzed in this study are indicated by circles; arrows or boxes between them give the nature of the mutation (0, ancestral; 1, derived), and, where appropriate, the restriction enzyme used and the allele cleaved in PCR-RFLP analysis. For HGs not analyzed (*squares*), information on geographic association is provided by shading. The correspondence of some of these HGs with the haplotype nomenclature of Karafet et al. (1999) and Hammer et al. (2000), whose work is referred to in the text, is as follows: HGs 1 + 22, haplotype 1C; HG 3, haplotype 1D; HG 4, haplotype 3G; HG 7, haplotypes 1A + 2; HG 8, haplotype 5; HGs 12 + 26, haplotype 1U; HG 16, haplotype 1I; HG 21, haplotypes 3A + 4; and HG 9, haplotype "Med."

ino et al. 1996; Underhill et al. 1997; Zerjal et al. 1997; Hammer et al. 1998; Hurles et al. 1999), indicating that the HGs that they define are likely to be found within European populations. There are several nomenclature systems currently in use for Y-chromosomal lineages, and, since we refer to the data of Karafet et al. (1999) and Hammer et al. (2000) in the text, we give some correspondences in the legend to figure 1. HG 7 is specific to sub-Saharan African populations (Karafet et al. 1999) but is typed here by default, since it is defined by the ancestral state of the recurrent *SRY*-1532 polymorphism (fig. 1). Maximum-parsimony analysis of haplotypes defined by these markers generates a unique tree (figs. 1 and 2) in which *DYS257* (Hammer et al. 1998) and 92R7 (Mathias et al. 1994) are phylogenetically equivalent (Jobling et al. 1998; Z. H. Rosser, M. E. Hurles, and M. A. Jobling, unpublished data). For this part of the phylogeny, 92R7 was typed routinely, and *DYS257* was typed when necessary to confirm results. Nine of the markers have been described elsewhere: YAP (Hammer 1994) was typed according to the method of Hammer and Horai (1995), *SRY*-1532 (Whitfield et al. 1995) according to Kwok et al. (1996), *SRY*-2627 according to Veitia et al. (1997), 92R7 (Mathias et al. 1994) according to Hurles et al. (1999), *DYS257* according to Hammer et al. (1998), M9 (Underhill et al. 1997) according to Hurles et al. (1998), sY81 according to Seielstad et al. (1994), Tat according to Zerjal et al. (1997), and *SRY*-8299 (Whitfield et al. 1995) according to Santos et al. (1999).

12f2 (Casanova et al. 1985) was typed using a newly developed PCR assay. This polymorphism was originally suggested to be an ~2-kb insertion/deletion, but our analysis suggests that its molecular basis is more com-

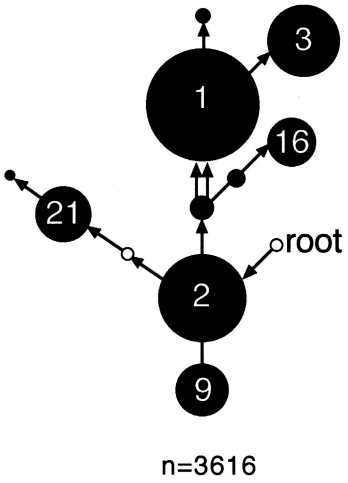

Figure 2 HG profile of the entire sample set. HG diversity within the complete sample set of 3,616 Y chromosomes, summarized on a simplified version of the network shown in figure 1. The area of each black circle is proportional to the frequency of the HG. Small unblackened circles indicate unobserved HGs (4 and 7). The position of the HG closest to the root (HG 7) is indicated.

plex than this. The PCR assay generates a 500-bp product from chromosomes carrying the *Taq*I/10-kb allele, but this product is absent from *Taq*I/8-kb–allele chromosomes (HG 9). An 820-bp amplicon from the *SRY* region, present in all chromosomes, is amplified as a control. Analysis of the 12f2 region gives no information about ancestral state, but we assume that presence of the 500-bp amplicon is ancestral. Primer sequences for the 12f2 amplicon are 12f2D (5′-CTG ACT GAT CAA AAT GCT TAC AGA TC-3′) and 12f2F (5′-TCT TCT AGA ATT TCT TCA CAG AAT TG-3′), and those for the *SRY* control amplicon are 3′*SRY*15 (5′-CTT GAT TTT CTG CTA GAA CAA G-3′) and 3′*SRY*16 (5′-TGT CGT TAC ATA AAT GGG CAC-3′). PCR conditions were 33–35 cycles of 94°C for 30 s, 59°C for 30 s, and 72°C for 45 s. An alternative assay, generating shorter amplicons, was used with degraded DNAs. The primers 12f2D (see above) and 12f2G (5′-GGA TCC CTT CCT TAC ACC TTA TAC-3′) produce an 88-bp product from *Taq*I/10-kb–allele chromosomes (and no product from *Taq*I/8-kb–allele chromosomes), which is coamplified with the Tat 112-bp amplicon (Zerjal et al. 1997) as a control, under the following conditions: 33–36 cycles of 94°C for 30 s, 59°C for 30 s, and 72°C for 30 s. All chromosomes known, from previous hybridization analysis, to carry *Taq*I/8-kb alleles lacked the 12f2 test amplicons in both of these assays. However, some YAP+ chromosomes belonging to HG 4 also lack the 12f2 amplicons, suggesting that the polymorphism may be recurrent (Blanco et al. 2000).

The LLY22g *Hin*dIII polymorphism was typed by a PCR-RFLP assay that will be described elsewhere (E. Righetti and C. Tyler-Smith, unpublished data). The deep-rooting markers *SRY*-1532, M9, YAP, and 92R7 were typed on all samples. For many samples, all other markers were also typed. However, in some cases, remaining markers were typed hierarchically—for instance, *SRY*-8299 and sY81 were, in some cases, typed only on chromosomes classified as YAP+.

Experimental Procedures

Haplotyping was carried out in Leicester; Oxford (both laboratories); Barcelona; Belgrade; Dublin; Leuven, Belgium; Lisbon; Porto, Portugal; Rome; and Tartu, Estonia. Procedures were based on those described by Hurles et al. (1998). To verify typing methodologies, a set of 12 quality-control DNAs was satisfactorily typed blindly by all participating laboratories.

Statistical Analysis

Spatial autocorrelation analysis was done by AIDA (Bertorelle and Barbujani 1995), for the entire data set, and SAAP (Sokal and Oden 1978), for individual HGs. PC analysis of covariances was carried out according to the method of Harpending and Jenkins (1973).

Mantel (1967) tests, done by ARLEQUIN version 2.0 (Schneider et al. 2000), were used to determine whether language or geography has the stronger impact on genetic differentiation. Genetic distances (as a pairwise F_{ST} matrix) were computed within ARLEQUIN, and geographic distances were calculated from latitude and longitude by use of great-circle distances, in a program written in Interactive Data Language 5.1 (Research Systems Inc.) by M. E. Hurles. Within IE languages, linguistic distances were adapted from Dyen et al. (1992), who used the lexicostatistical method of Swadesh (1952) on comparisons of 200-word lists: percentage similarities were first converted to dissimilarities, and these numbers then assigned as nonpercentage distances between languages (ranging from 9 [Czech to Slovak] to 88 [Armenian to Irish]). All IE languages within the data set were represented, with the exception of Scottish, which was assigned a distance of 10 from Irish; we also tested the effects of other values, in the range 5–20. The Belgian sample was divided into its two linguistic groups—those speaking French (56 individuals) and those speaking Dutch (36). An arbitrary and conservative, larger value, 200, was then assigned as a distance between language families. As was done by Poloni et al. (1997), Mantel tests were also performed using different inter–language-family distances, of 400 and 1,000. Two of the non–IE language families, Altaic and Uralic, are represented by more than one language within our data set. On the basis of a consideration of the classification by Ruhlen (1991) and of the inter–IE-language distances of Dyen et al. (1992), plausible distances were assigned within these families, and the effect of altering these values over

a range was tested. Within Uralic, values were as follows: Finnish to Estonian, 25 (altered value range 10–30); Finnish-Estonian to Saami, 30 (20–40); Finnish-Estonian-Saami to Mari, 40 (30–70); and Hungarian to all other Uralic languages, 80 (40–90). Values for Chuvash and Turkish (Altaic) were 40 (20–60).

To locate zones of abrupt genetic change, or genetic boundaries, and to assess their significance, we used the program ORINOCO, written in Interactive Data Language 5.1 (Research Systems) by M.E. Hurles (Hurles 1999), which adapts a method known as "wombling" (Barbujani et al. 1989), initially developed for the analysis of allele frequencies. First, an inverse-distance–squared weighted algorithm was used to interpolate the frequencies for each of the eight observed HGs at each grid point within a 100 × 100 array (with account taken of the curvature of the earth and with correspondence to a grid point every 0.36° latitude and 0.72° longitude). The derivatives of these eight interpolated surfaces were then calculated at every node of the grid, and the magnitudes of the derivatives were summed, thus giving a measurement of the slope of the combined surfaces—that is, the overall rate of Y-chromosomal genetic change in 10,000 rectangles covering Europe. The significance of these gradients was considered in two ways, both of which take into account isolation by distance within the landscape (Barbujani et al. 1989). First, a simple significance threshold was applied, with only the top 5% of values. Second, a Monte-Carlo algorithm was used to permute the HG data 1,000 times, and summed derivatives were calculated for each permutation. This algorithm maintains the observed sample sizes and positions and therefore controls for the conflated effects, in the generation of false positives, of sampling and heterogeneity in distances between sample sites. Grid points obtained with the original HG data were then retained only if the values of their summed derivatives were >95% of the values obtained from the permuted data. Grid points could then be plotted on a map, color coded to indicate the strength of the barrier, to show the positions of significant barriers, and were also displayed on Delaunay triangulation connections (Brassel and Reif 1979) between sample sites. The Algerian and northern-African samples were excluded from the barrier analysis, since their high degree of difference from all other samples (as shown in PC analysis) represents a strong genetic barrier that would bias the detection of barriers elsewhere.

Results

Y chromosomes from 3,487 males belonging to 47 populations (fig. 3*A*) were haplotyped using biallelic markers and were classified into HGs (table 1); data on 129 northern-African Y chromosomes (Bosch et al. 1999) were also included (see the Subjects and Methods section), giving a total of 3,616. The resulting frequency data for the entire sample are summarized in figure 2. Two HGs, 7 and 4, are absent, which is consistent with published information: HG 7 has been discussed above (see the Subjects and Methods section), and HG 4 is restricted to eastern and central Asia (Karafet et al. 1999).

No single population has a frequency distribution resembling that of the overall sample (fig. 2), emphasizing the strong geographic differentiation of Y-chromosomal variation in Europe. This is evident in the HG frequency data in figure 3: distributions of HGs are highly nonrandom, with, for example, a concentration of HG 1 chromosomes in the west, HG 9 chromosomes in the southeast, HG 16 chromosomes in the northeast, and HG 3 chromosomes in central and eastern Europe.

Clinal Distribution of Y-Chromosomal Lineages

To examine the geographic differentiation of these HGs more quantitatively, we used spatial autocorrelation analyses (Sokal and Oden 1978). These methods give a measure of the average level of genetic similarity, between populations within particular geographic distance classes, that can be represented as correlograms (fig. 4), and they allow clinal variation, reflecting population movement or natural selection, to be distinguished both from isolation by distance, reflecting short-range dispersals and drift, and from nonsignificance. We first used AIDA (Bertorelle and Barbujani 1995), which takes into account molecular distances between HGs and provides autocorrelation indices (Moran's II) for the entire data set, including the rare HGs. The pattern (fig. 4*A*) is strongly clinal, recognized as a change from positive to negative autocorrelation indices with increasing distance class. The SAAP analysis (fig. 4*B*–*G*), omitting low-frequency HGs (HGs 8, 12, 22, and 26), confirms this clinal pattern and reveals information about individual lineages. The distributions of all of the HGs examined, with the exception of HG 2, are strongly clinal (fig. 4), confirming the visual impression given by figure 3. In two cases (HGs 3 and 16), values become positive or zero in the longest-distance class (a "depression"), indicating the regional—rather than continentwide—influence of these clines.

HG 2 is the most ancestral lineage that we find within Europe, and it lies at a starlike node within the tree; chromosomes within this HG are essentially undefined and are likely to consist of a set of discrete sublineages that themselves probably have greater geographic coherence. Consistent with this, HG 2 chromosomes are widely distributed across the whole landscape and constitute the only high-frequency lineage that does not show clinal variation (figs. 3*B* and 4*C*). Because of this

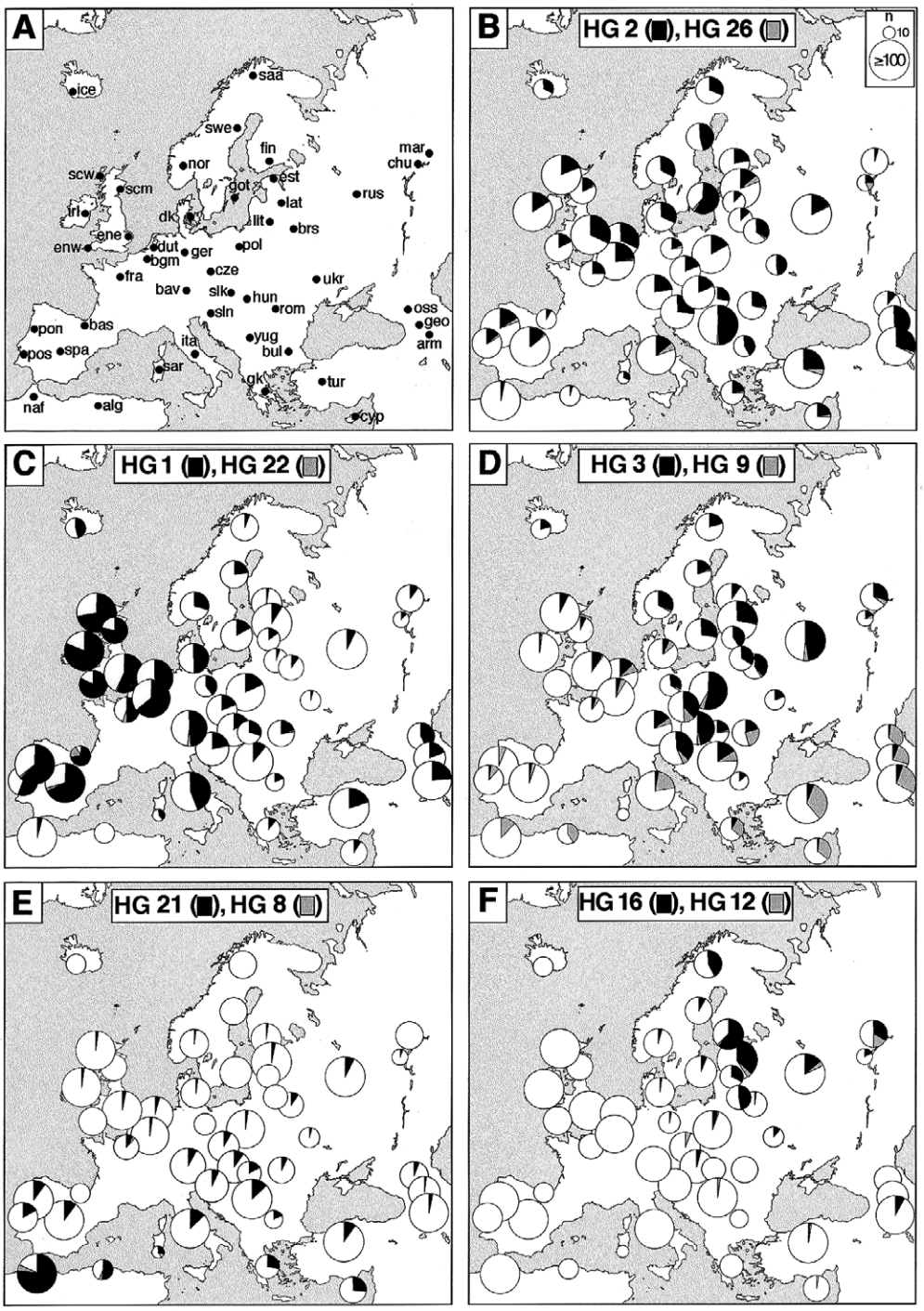

Figure 3 Distribution of populations sampled and geographic distribution of Y-chromosomal HG diversity. *A*, Abbreviated population names. alg = Algerian; arm = Armenian; bas = Basque; bav = Bavarian; bgm = Belgian; brs = Belarusian; bul = Bulgarian; chu = Chuvash; cyp = Cypriot; cze = Czech; dk = Danish; dut = Dutch; ene = East Anglian; enw = Cornish; est = Estonian; fin = Finnish; fra = French; geo = Georgian; ger = German; gk = Greek; got = Gotlander; hun = Hungarian; ice = Icelandic; irl = Irish; ita = Italian; lat = Latvian; lit = Lithuanian; mar = Mari; naf = northern African; nor = Norwegian, oss = Ossetian; pol = Polish; pon = northern Portuguese; pos = southern Portuguese; rom = Romanian; rus = Russian; saa = Saami; sar = Sardinian; scm = Scottish; scw = western Scottish; slk = Slovakian; sln = Slovenian; spa = Spanish; swe = northern Swedish; tur = Turkish; ukr = Ukrainian; yug = Yugoslavian. For a list of linguistic affiliations, see table 1. *B–F*, HG diversity within each of 47 populations, summarized on a map of Europe. The area of each pie chart is proportional to the sample size, up to a number of ⩾100; sizes are indicated schematically within *B*. The area of each black or gray sector is proportional to the frequency of the corresponding HG.

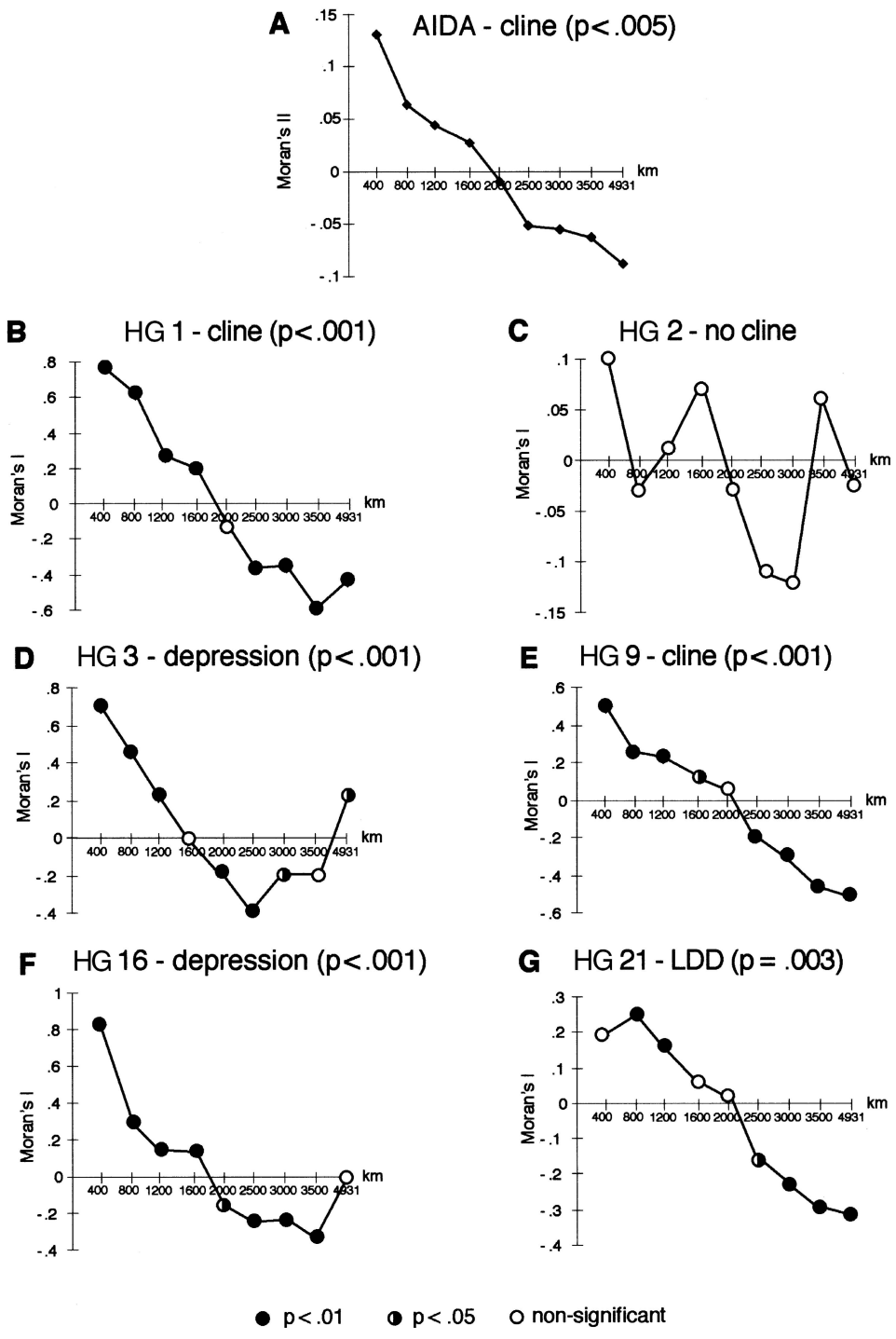

Figure 4 Spatial autocorrelation analyses. *A,* Correlogram, calculated using AIDA, for the entire data set. Overall significance is given. *B–G,* Correlograms, calculated using SAAP, for the six most frequent HGs. The significance of each point is indicated by its symbol, and the overall significance of each correlogram is also given. LDD = long-distance differentiation. In all correlograms, the *X*-axes show distance classes (km).

uninformativeness, HG 2 will not be further considered here. HG 26 occurs at low frequency (fig. 3*B*); like HG 2, it lies at a deep internal node within the tree and probably contains unidentified coherent sublineages.

We find two other HGs at low frequency—HGs 8 and 22. HG 8 is common in sub-Saharan Africa (Karafet et al. 1999) and is present in our northern-African samples at ~5% (fig. 3*E*). Only two European examples exist,

in Sardinia and France, which may represent recent admixture.

HG 22 chromosomes (fig. 3*C*) reach appreciable frequencies only in the French (5%) and Basques (19%). This HG has been analyzed in detail in a study elsewhere (Hurles et al. 1999), which suggested that it has a recent Iberian origin and that non-Iberian examples represent migrants. The distribution here is consistent with this analysis.

A Major Cline Consistent with the Demic Diffusion Model

HGs 1 and 9 show complementary clines on the continental scale, from the southeast of Europe to the northwest (figs. 3*C* and *D* and 4*B* and *E*): indeed, when the Irish sample is further subdivided on the basis of geographic information contained within surnames (Hill et al. 2000), HG 1 reaches near-fixation (98.5%) in the west of Ireland. HG 9 reaches its highest frequencies (~33%) in the Caucasus and in Anatolia (fig. 3*D*), where it is thought that agriculture originated (Cavalli-Sforza et al. 1994). The strong clinal pattern of these two HGs, which together account for almost half (45%) of the chromosomes in our study, resembles the first principal component of genetic variation of classical loci and is consistent with the demic diffusion hypothesis. However, distributions of the remaining HGs are very different from these and cannot be interpreted as a simple reflection of population movement from the Near East.

A Northeast/Southwest Cline Signaling an Expansion from North of the Black Sea

The distribution of HG 3 chromosomes is also strongly clinal (fig. 4*D*), but with a very different axis (fig. 3*D*) and more on a regional scale, and is likely to reflect population-historical events distinct from those responsible for the distributions of HGs 1 and 9. It reaches its highest frequencies in central-eastern Europe, comprising approximately half of the chromosomes in the Russian, Polish, and Slovakian samples; frequencies in the southeast and southwest are low. This distribution resembles the third principal component of variation of classical gene frequencies, which has been interpreted by some geneticists (Cavalli-Sforza et al. 1994) as marking the movement, from north of the Caspian Sea, of the Kurgan people, dated to ~7,000 YBP.

A North-South Cline: A Northern-African Influence?

Within Europe, HG 21 chromosomes are concentrated in the south (fig. 3*E*). Their frequency in the two northern-African samples is very high (52% and 77%), and their frequencies in the Greek and Cypriot samples are also high (~27%), which might reflect a barrier to gene flow between Africa and Europe, as is also shown by the analysis of autosomal protein markers (Simoni et al. 1999) and microsatellites (Bosch et al. 2000). In other southern-European populations, such as those in Spain, Portugal, Sardinia, Italy, Turkey, and Yugoslavia, frequencies are in the range of 10%–20%. The decline in frequencies to the north is rather uniform. This regional cline (fig. 4*G*) has similiarities to that detected in the second principal component of classical gene frequencies (Cavalli-Sforza et al. 1994), which has been interpreted on a climatic basis.

A Lineage Concentrated in the Northeast: A Contribution of Uralic Speakers?

HG 16 is at high frequency in the north, east of the Baltic Sea (fig. 3*F*), a distribution consistent with that noticed previously in a global survey (Zerjal et al. 1997). Its pattern is again clinal but regional (fig. 4*F*). HG 12, ancestral to HG 16, is at low frequency in the sample overall. However, its distribution overlaps that of HG 16, with no examples in the western half of the continent, and is concentrated more in the south (fig. 3*F*). It is most frequent (17%) in the Mari, who may be the population of origin of the Tat mutation, which defines HG 16 (T. Zerjal and C. Tyler-Smith, unpublished data).

With the exception of the Hungarians, who acquired their Uralic language through elite dominance by the Magyars during recent times (Cavalli-Sforza et al. 1994), all Uralic-speaking populations tested (Finnish, Estonians, Saami, and Mari) show a high frequency of HG 16. However, two nearby populations, the Lithuanians and Latvians, also show HG 16 at high frequency but speak languages of the IE family—for this lineage at least, the association appears to be geographic rather than linguistic. In the following section, we use methods that summarize variation among all lineages, to examine this issue in more detail.

Geography and Language as Causes of Genetic Differentiation

Population comparisons through PC analysis.—PC analysis is a method that allows the graphic display, in a few dimensions, of the maximum amount of variance within a multivariate data set, with minimum loss of information. Figure 5 shows the results of a PC analysis of the Y-chromosome HG data, in which populations are labeled according to linguistic affiliation. PC1–PC3 summarize 71.4% of the variance.

The major division is between the two populations from northern Africa and the others. This is unsurprising, given their high frequencies of HGs 21 and 9 and their near absence of HG 1, and indicates that the Mediterranean, even at its narrowest point, has represented a barrier to gene flow, as has been suggested previously by autosomal DNA analysis. The Mediterranean pop-

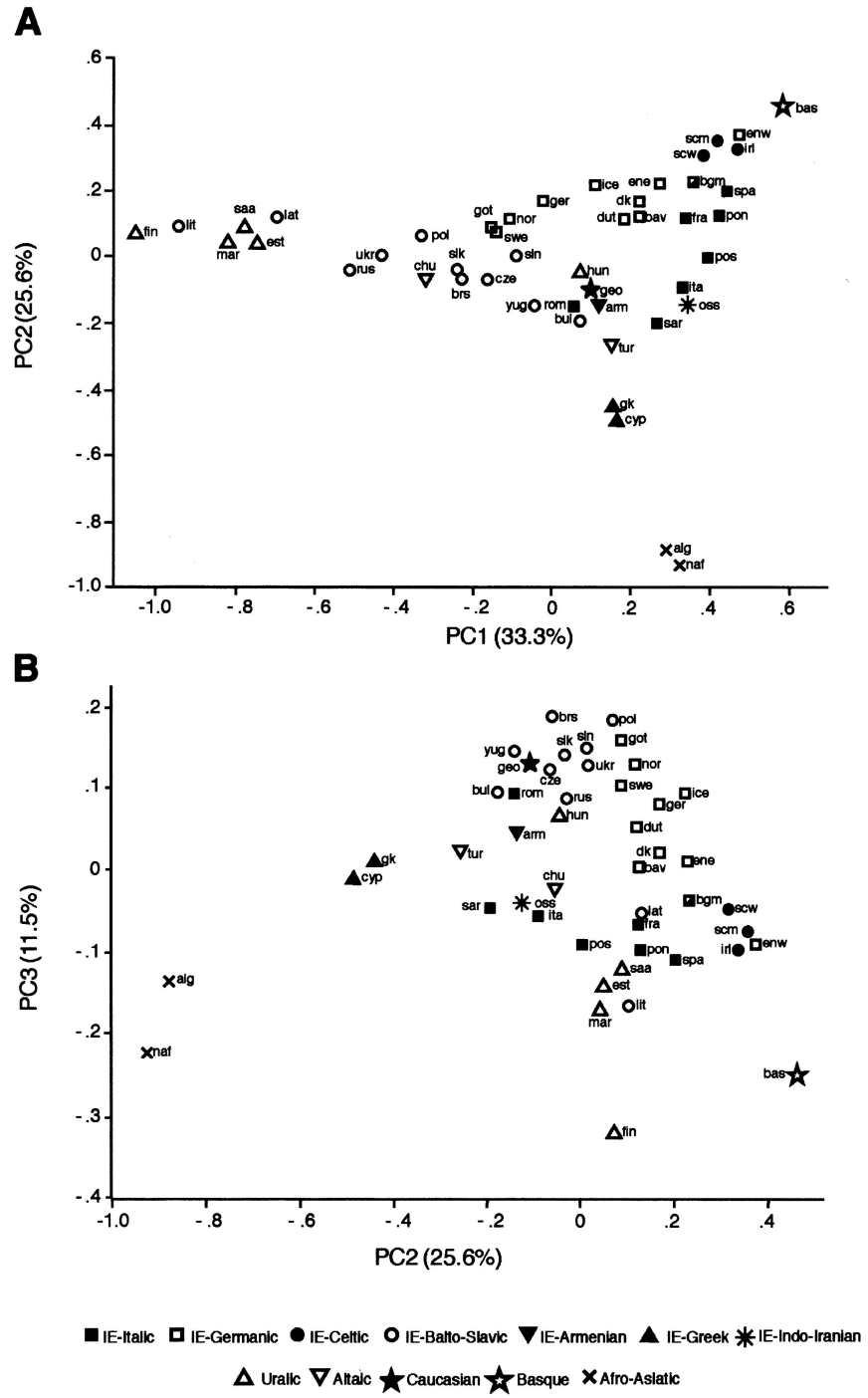

Figure 5 PC analysis of Y-chromosomal HG diversity. *A*, PC2 plotted against PC1. *B*, PC3 plotted against PC2. The percentage of variance explained by each component is given on the axes. Linguistic affiliation for each population is indicated symbolically; the Belgian sample is part Dutch-/part French-speaking and has a hybrid symbol. Abbreviations are as in figure 3.

ulations of Greece and Cyprus occupy an intermediate position between the northern Africans and the rest.

Basques speak a non-IE language unrelated to any other language (Ruhlen 1991) and thus represent the most striking example of a linguistic isolate in Europe. This isolation seems to be reflected in the PC analysis, in which they are separated from other populations (fig. 5*A*); however, this may be due to high frequency of a young lineage (HG 22; Hurles et al. 1999), rare elsewhere, rather than to persistence of ancient ones. Their

closest neighbors in the PC analysis are not the geographically close populations of Iberia but those of the Atlantic fringe, most of which speak Celtic-IE languages. In this context, the Cornish sample ("enw" in Figs. 3 and 5) is grouped not with the eastern English sample (ene) but with the Scottish and Irish—a reflection of geography or of the original Celtic language of this region (Ruhlen 1991) or both.

Among Uralic-speaking populations, this analysis confirms the impression given by figure 3*F:* with the exception of the Hungarians, who lie close to IE language speakers, these populations are grouped together with the Finns separated from the rest in PC3 (fig. 5*B*). Also within this group are the Lithuanians and Latvians, supporting the idea that this is primarily a geographic association.

The overall impression from figure 5 is that geographic proximity may be a better predictor of Y-chromosomal genetic affinity than is language: as well as the examples discussed above, the Italic-IE language–speaking Romanians are distant from other Italic language speakers, and the Turks lie between the geographically neighoring but linguistically distant Armenians and Greeks.

Correlating Geography, Language, and Genetics through Mantel Testing

Mantel (1967) tests provide an objective way of assessing the relative importance of different factors in the shaping of genetic diversity. In this method, correlation coefficients between pairs of factors (from genetics, geography, and language) can be calculated, together with significance values; partial correlation coefficients are then calculated between genetics and geography and between genetics and language, with the third factor kept constant to control for the strength of the correlation between geography and language. The populations from northern Africa are linguistically remote and geographically peripheral, and the PC analysis has shown their genetic differentiation. We therefore excluded them from the Mantel analysis, to examine effects within Europe itself. Genetics and geography (table 2) are strongly and significantly correlated ($P < .001$), and the correlation between genetics and language is less strong but still significant ($P = .014$). The partial correlation of genetics and geography, with language kept constant, is again strong and significant ($P < .001$); in contrast, the partial correlation of genetics and language is low and nonsignificant ($P = .095$). We examined the effect of changing the values that we had assigned to distances within Uralic and within Altaic and between Irish and Scottish (see the Subjects and Methods section), and this had a negligible influence on our results. Increasing the distance assigned between language families had the effect of reducing still further the partial correlation between genetics and language, as well as its significance. This analysis confirms the primacy of geography, rather than language, in the shaping of Y-chromosomal genetic diversity within Europe.

Table 2

Correlation and Partial Correlation Coefficients between Genetic, Geographic, and Linguistic Distance

Distance Considered	Correlation Coefficient	P[a]
Genetics and geography	.387	<.001
Genetics and language	.198	<.01
Genetics and geography, language held constant	.349	<.001
Genetics and language, geography held constant	.088	NS

[a] NS = not significant.

Location of Y-Chromosomal Genetic Barriers within Europe

Although the analysis above indicates a lack of large-scale correlation between language and genetics, it does not address local genetic differentiation, which may reflect local effects of language. Genetic-barrier analysis, which locates the zones of sharpest genetic change within a landscape, provides a way to do this.

Figure 6 shows the results of a genetic-barrier analysis of the Y-chromosome HG data for 45 populations, for the top 5% of barriers and a 95% significance filter (see the Subjects and Methods section). Within western Europe, minor barriers separate the Basques from some neighboring populations, the western from the eastern English, and the Dutch from the Belgians. In the east, there are two major barriers, one between the Uralic-speaking Mari and Altaic-speaking Chuvash and one between the Georgians and the Ossetians, who speak languages belonging to different families and who are also separated by the Caucasus Mountains. Most of the major barriers lie in the middle of the European landscape, running from Italy in the south to the Baltic Sea in the north, including one barrier around the island population of Gotland.

To what extent are linguistic differences contributing to Y-chromosomal barriers within Europe? Since 37 different languages are spoken among our 45 sample sites, we expect most genetic barriers to fall between populations speaking different languages. However, if language differences do constitute barriers to gene flow, then we might expect that the degree of linguistic difference between a pair of populations should correlate with the chance of a genetic barrier occurring—that is, the greatest proportion of genetic barriers should fall between populations speaking languages from different families, a lesser proportion between those speaking languages from different subfamilies, and the least between those speaking languages within a subfamily. There are

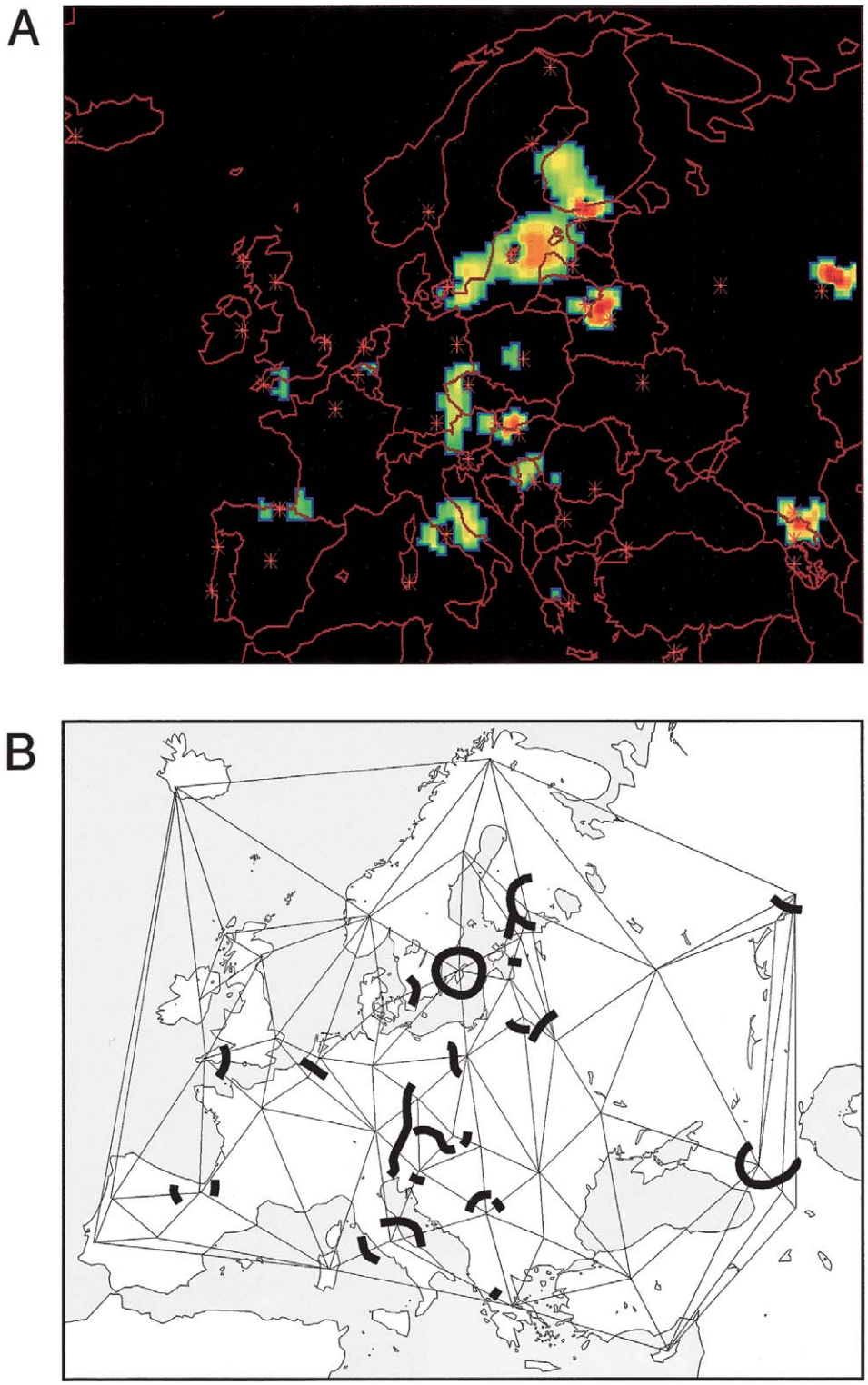

Figure 6 Significant Y-chromosomal genetic barriers within Europe. *A*, Output from the ORINOCO program. Positions of genetic barriers showing 95% significance after permutation (see the Subjects and Methods section) are indicated by blue through red areas on the black background, with sample sites indicated by stars. A three-dimensional animation of the actual output from the program can be viewed at the Molecular Genetics Laboratory of the McDonald Institute for Archaeological Research Web site. *B*, Schematic version of the output shown in *A*, with the positions of barriers indicated as thick lines on Delaunay connections (*thin lines*) between sample sites.

122 Delaunay connections in figure 6*B*, 48 of which are crossed by a genetic barrier. We count the proportion of connections that are crossed by a genetic barrier in each of the three classes, between language families, between subfamilies, and within subfamilies; these values are 46.2% (18/39), 40.5% (15/37), and 32.6% (15/46), respectively. Although the ranking of these three values is that expected under the hypothesis, differences between them are not significant ($P > .1$, three-way χ^2 test). This suggests that language may not be the primary force contributing to genetic barriers here. However, this analysis does not take into account the fact that two non-IE languages, Hungarian and Turkish, have been acquired recently: the PC analysis and the relative absence of Y-chromosomal genetic barriers around these populations supports the idea that elite dominance was not accompanied by extensive genetic admixture. If we remove these two populations and repeat the above analysis, differences between the proportions increase (to 50.0% [13/26], 43.2% [19/44], and 31.9% [15/47], respectively) but remain not significant ($P > .1$).

Discussion

We have described the most detailed survey to date of human Y-chromosomal diversity within Europe. Samples were distributed over most of the continent, including its western and eastern fringes; inclusion of these regions, omitted from some other studies, has allowed both the detection of influences from the east and clines extending to the extreme west, for example. However, some regions remain poorly sampled, and, if the possible effects of local differentiation are to be studied, more-extensive sampling is needed. At the eastern edge of Europe lie the steppes, which stretch uninterrupted to China. Analogous studies of Asian Y chromosomes are under way and will place the European data within a broader context (W. Bao, S. Zhu, M. E. Hurles, T. Zerjal, M. A. Jobling, J. Xu, Q. Shu, R. Du, H. Yang, and C. Tyler-Smith, unpublished data).

We used 11 biallelic markers in this study, but there is still a need for more. For instance, HG 2, constituting 22% of the total sample and as much as 49% in the sample from Yugoslavia, is poorly defined and therefore constitutes a potential source of error in our analyses, since equal weight is given both to this and to well-defined HGs. The pace of new marker discovery is increasing (Underhill et al. 1997; Shen et al. 2000), and soon the resources will be available to adequately define all major European lineages.

Consistent with global surveys (Underhill et al. 1997; Karafet et al. 1999), this continental study confirms the high degree of geographic differentiation of Y-chromosomal lineages. This differentiation makes the Y chromosome a sensitive indicator of either admixture, as demonstrated in studies of Polynesia (Hurles et al. 1998), South America (Bianchi et al. 1997), and Uruguay (Bravi et al. 1997), for example, or an absence of admixture, as has been shown in Jewish populations in Europe and northern Africa (Hammer et al. 2000). Knowledge about admixture is of particular importance in the choice of populations for studies that use linkage-disequilibrium analysis (McKeigue 1997) in both simple and complex disorders.

Clines of Y-Chromosomal HGs

The effects of drift on human Y-chromosome diversity are likely to be great. It is striking, therefore, to observe clear clinal variation in five of the six major lineages within Europe—this suggests that drift has not erased the patterns of variation established by past population movement. Natural selection on Y chromosomes (Jobling and Tyler-Smith 2000) provides an alternative explanation for such clines; possible effects of geographically variable factors (such as temperature) on fertility within specific lineages have yet to be investigated, but, in the absence of evidence to the contrary, we assume that the variation that we are assaying is selectively neutral and can therefore be interpreted in terms of population history.

The contrast between the clinal variation of Y-chromosomal lineages and the lack of clines in mtDNA data (Simoni et al. 2000*a*) is marked, although the latter is still a matter of debate (Simoni et al. 2000*b;* Torroni et al. 2000). It seems consistent with studies of global genetic diversity (Seielstad et al. 1998), which have ascribed such differences to patrilocality. However, direct evidence about mating practices in European prehistory is lacking—indeed, populations in some regions, such as northern Iberia, may have practiced matrilocality (Collins 1986).

Clines for HGs 1 and 9, encompassing 45% of the chromosomes—and doing so on a continental scale—show a pattern similar to that seen both in the first principal component of classical gene-frequency data and in the autocorrelation analysis of six Y-chromosomal microsatellites (Casalotti et al. 1999). A simplistic interpretation is that HG 9 chromosomes were carried in a major demographic expansion of agricultural migrants from the Near East and that HG 1 chromosomes were a preexisting predominant European lineage. Estimates of the ages of these lineages, from coalescent analysis, are not inconsistent with this scenario: the mutation defining HG 1 has been dated at ~23,000 YBP (Karafet et al. 1999), and that defining HG 9 has been dated at 14,800 ± 9,700 YBP (Hammer et al. 2000).

Demic diffusion—and, indeed, any major directional gene-flow process—is generally expected to generate clines for only a fraction of the alleles at one locus (Sokal

et al. 1989, 1997). Although two HGs show clines compatible with expansion from the Near East, three further lineages show different clinal patterns, indicating distinct population movements: southward and westward from north of the Black Sea (HG 3), from eastern Europe or northern Asia westward to the Baltic Sea (HG 16), and from south to north (HG 21). These clines are more regionally localized than those for HGs 1 and 9, pointing to phenomena affecting only part of the continent. It is tempting to assign known or surmised population-historical movements to these genetic gradients, but this should be done with caution.

The distribution of HG 3 chromosomes resembles the third principal component of variation of classical gene frequencies. There are several possible interpretations of this pattern. One explanation (Cavalli-Sforza et al. 1994) is that it marks the Kurgan expansion from north of the Caspian Sea, dated to ~7,000 YBP. However, alternative explanations—such as the spread of pastoralism, or east-to-west movements of people such as the Scythians, Mongols, and Huns—seem equally likely (Renfrew 2000). Globally, HG 3 chromosomes are absent from Africa and the Americas, but their distribution is wide within Asia as well as in Europe (Zerjal et al. 1999), consistent with their association with a recent and major expansion within Eurasia. Microsatellite diversity analysis (Zerjal et al. 1999) used the mutation-rate estimates of Heyer et al. (1997) to date the most recent common ancestor of a set of European and Asian HG 3 chromosomes to 3,800 YBP (95% confidence interval [CI] 1,600–13,000 YBP); the use of more-recent mutation-rate estimates (Kayser et al. 2000) would yield a date of 2,550 YBP (95% CI 1,650–4,260 YBP). Coalescent analysis has dated the *SRY*-1532 mutation defining HG 3 to ~7,500 YBP (Karafet et al. 1999). If these dates are to be relied on, they seem to suggest that the expansion of HG 3 chromosomes was due to population movements later than those of the Kurgan people.

Currently, dates cannot be attached to the clines, and the modern distributions of lineages are the outcome of many millennia of population movement. Assigning plausible dates to demographic movements is important, and here the Y chromosome can potentially contribute. Finer-scale definitions of monophyletic lineages within Europe, by use of new markers, and the analysis of these, by use of microsatellites, offers the possibility that timescales for the major demographic events can be inferred.

Language, Geography, and Y-Chromosomal Diversity

The Mantel tests demonstrate that patterns of Y-chromosomal genetic variation do not correlate as well with language as with geography. However, it should be borne in mind that geography and language together explain only 16.8% of the genetic variance (data not shown); therefore, other forces, such as founder effects and genetic drift, have also been important in determining the current patterns of spatial variation. Our findings seem at odds with those of Poloni et al. (1997), who showed that most of the population differentiation of Y-chromosome haplotypes was due to language. However, there are important differences between the two studies. The samples of Poloni et al. (1997) were global, rather than from a single continent, and showed a correspondingly greater linguistic and genetic diversity. The populations that we have studied are located within a single continent, and most speak languages belonging to one language family, IE; indeed, much of the genetic patterning that we now see may have its roots in the spread of that language family (Renfrew 1987). The effect of increasing genetic, geographic, and linguistic diversity in the input to the Mantel tests can be seen by including the northern-African samples (data not shown), which are both geographically and linguistically distant from most other populations. This increases the partial correlations between genetics and geography and between genetics and language and also increases the significance of the latter to $P = .024$, which, however, is still lower than the significance of the genetics-geography partial correlation ($P < .001$).

The results of genetic-barrier analysis (fig. 6) need to be interpreted with caution when, as in this case, sample distribution is uneven; the method is likely to be sensitive to the introduction of new populations, especially between existing sample sites that are far apart. However, the analysis has suggested that there is little correlation between genetic barriers and levels of linguistic separation, even when elite dominance is taken into account by removing the Hungarians and Turks from the analysis. Although cultural factors other than language (such as politics and religion) might also be associated with genetic barriers, we have examined language because it has the greatest time depth. However, this is still likely to be less than the age of geographic barriers, the relative importance of which cannot easily be analyzed. Twenty-five of 48 Delaunay connections crossed by genetic barriers also coincide with geographic barriers (under a conservative definition that considers only large stretches of water and the two major mountain ranges, the Alps and the Caucasus), which seems to emphasize the greater importance of geographic factors in subdividing populations, resulting in large differences in Y-chromosomal HG frequencies.

In synthesis, it seems that many kinds of barriers are probably recent, on an evolutionary timescale (see Renfrew 1987); after they have been established, fluctuations of allele frequencies have become partly or largely independent in the populations separated by those barriers. Therefore, it is perhaps not surprising to find little

correlation between the degree of language differentiation at a language boundary and the amount of genetic change observed across that boundary. As has been shown in the analysis of protein polymorphisms (Sokal et al. 1990), linguistic differences tend to cause some degree of population subdivision, regardless of whether such differences are between language families, between languages of the same family, or even between dialects of the same language.

Although we have dichotomized the forces of geography and language, in reality they work together; spatially coincident weak geographic and linguistic barriers may together form strong barriers to gene flow. Some of the strongest genetic barriers observed, in central Europe, coincide with neither strong linguistic nor strong geographic barriers. Linguistic and geographic heterogeneities and the effects of drift, on a background retaining a strong signal of expansion from the Near East and of other migrations, have combined to shape the genetic landscape of Europe.

Acknowledgments

We thank the DNA donors for making this study possible, and we thank Laurent Excoffier for assistance. Z.H.R. was supported by a BBSRC Studentship, T.Z. by a Wellcome Trust Bioarchaeology Studentship, M.E.H. by an MRC Studentship, F.R.S. by the Leverhulme Trust, and L.P. by Ph.D. grant PRAXIS XXI/BD/13632/97 from Fundação para a Ciência e a Tecnologia. D.C.R. is a Glaxo Wellcome Research Fellow. C.T.-S. is supported by the CRC, and M.A.J. is a Wellcome Trust Senior Fellow in Basic Biomedical Science, supported by grant 057559. Iberian sample collection was partially funded by multidisciplinary project grant PR182/96 6745 from Complutense University.

Electronic-Database Information

The URL for data in this article is as follows:

Molecular Genetics Laboratory of the McDonald Institute for Archaeological Research, http://www-mcdonald.arch.cam.ac.uk/Genetics/home.html

References

Adams J, Otte M (1999) Did Indo-European languages spread before farming? Curr Anthropol 40:73–77

Ammerman AJ, Cavalli-Sforza LL (1984) Neolithic transition and the genetics of populations in Europe. Princeton University Press, Princeton, NJ

Barbujani G (1991) What do languages tell us about human microevolution? Trends Ecol Evol 6:151–156

——— (1997) DNA variation and language affinities. Am J Hum Genet 61:1011–1014

Barbujani G, Oden NL, Sokal RR (1989) Detecting regions of abrupt change in maps of biological variables. Syst Zool 38: 376–389

Barbujani G, Pilastro A, de Domenico S, Renfrew C (1994) Genetic variation in North Africa and Eurasia: neolithic demic diffusion vs. paleolithic colonisation. Am J Phys Anthropol 95:137–154

Bertorelle G, Barbujani G (1995) Analysis of DNA diversity by spatial autocorrelation. Genetics 140:811–819

Bianchi NO, Bailliet G, Bravi CM, Carnese RF, Rothhammer F, Martínez-Marignac VL, Pena SDJ (1997) Origin of Amerindian Y-chromosomes as inferred by the analysis of six polymorphic markers. Am J Phys Anthropol 102:79–89

Blanco P, Shlumukova M, Sargent CA, Jobling MA, Affara N, Hurles ME (2000) Divergent outcomes of intra-chromosomal recombination on the human Y chromosome: male infertility and recurrent polymorphism. J Med Genet 37:752–758

Bosch E, Calafell F, Pérez-Lezaun A, Clarimón J, Comas D, Mateu E, Martínez-Arias R, Morera B, Brakez Z, Akhayat O, Sefiani A, Hariti G, Cambon-Thomsen A, Bertranpetit J (2000) Genetic structure of north-west Africa revealed by STR analysis. Eur J Hum Genet 8:360–366

Bosch E, Calafell F, Santos FR, Pérez-Lezaun A, Comas D, Benchemsi N, Tyler-Smith C, Bertranpetit J (1999) Variation in short tandem repeats is deeply structured by genetic background on the human Y chromosome. Am J Hum Genet 65:1623–1638

Boyd R, Silk JB (1997) How humans evolved. WW Norton, New York

Brassel KE, Reif D (1979) A procedure to generate Thiessen polygons. Geogr Anal 11:289–303

Bravi CM, Sans M, Bailliet G, Martinez-Marignac VL, Portas M, Barreto I, Bonilla C, Bianchi NO (1997) Characterization of mitochondrial DNA and Y-chromosome haplotypes in a Uruguayan population of African ancestry. Hum Biol 69: 641–652

Casalotti R, Simoni L, Belledi M, Barbujani G (1999) Y-chromosome polymorphisms and the origins of the European gene pool. Proc R Soc Lond B Biol Sci 266:1959–1965

Casanova M, Leroy P, Boucekkine C, Weissenbach J, Bishop C, Fellous M, Purrello M, Fiori G, Siniscalco M (1985) A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. Science 230: 1403–1406

Cavalli-Sforza LL, Menozzi P, Piazza A (1993) Demic expansions and human evolution. Science 259:639–646

——— (1994) The history and geography of human genes. Princeton University Press, Princeton, NJ

Chikhi L, Destro-Bisol G, Bertorelle G, Pascali V, Barbujani G (1998*a*) Clines of nuclear DNA markers suggest a largely neolithic ancestry of the European gene pool. Proc Natl Acad Sci USA 95:9053–9058

Chikhi L, Destro-Bisol G, Pascali V, Baravelli V, Dobosz M, Barbujani G (1998*b*) Clinal variation in the nuclear DNA of Europeans. Hum Biol 70:643–657

Collins R (1986) The Basques. Blackwell, Oxford

Comas D, Calafell F, Mateu E, Pérez-Lezaun A, Bosch E, Bertranpetit J (1997) Mitochondrial DNA variation and the origin of the Europeans. Hum Genet 99:443–449

Cooper G, Amos W, Hoffman D, Rubinsztein DC (1996) Network analysis of human Y microsatellite haplotypes. Hum Mol Genet 5:1759–1766

Dennell R (1983) European economic prehistory: a new approach. Academic Press, London

Dyen I, Kruskal JB, Black P (1992) An Indoeuropean classification: a lexicostatistical experiment. Trans Am Philos Soc 82:1–132

Gimbutas M (1970) Proto-Indo-European culture: the Kurgan culture during the fifth, fourth and third millennia B.C. In: Cardona G, Hoenigswald HM, Senn A (eds) Indo-European and Indo-Europeans. University of Pennsylvania Press, Philadelphia, pp 155–195

Hammer MF (1994) A recent insertion of an Alu element on the Y chromosome is a useful marker for human population studies. Mol Biol Evol 11:749–761

Hammer MF, Horai S (1995) Y-chromosomal DNA variation and the peopling of Japan. Am J Hum Genet 56:951–962

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

Hammer MF, Redd AJ, Wood ET, Bonner MR, Jarjanazi H, Karafet T, Santachiara-Benerecetti S, Oppenheim A, Jobling MA, Jenkins T, Ostrer H, Bonné-Tamir B (2000) Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes. Proc Natl Acad Sci USA 97:6769–6774

Harpending H, Jenkins T (1973) Genetic distance among Southern African populations. In: Crawford MH, Workman PL (eds) Methods and theories of anthropological genetics. University of New Mexico Press, Albuquerque, pp 177–199

Hassan FA (1973) On mechanisms of population growth during the neolithic. Curr Anthropol 14:535–542

Heyer E, Puymirat J, Dieltjes P, Bakker E, de Knijff P (1997) Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. Hum Mol Genet 6:799–803

Hill EW, Jobling MA, Bradley DG (2000) Y chromosomes and Irish origins. Nature 404:351–352

Hurles ME (1999) Mutation and variability of the human Y chromosome genetics. University of Leicester, Leicester

Hurles ME, Irven C, Nicholson J, Taylor PG, Santos FR, Loughlin J, Jobling MA, Sykes BC (1998) European Y-chromosomal lineages in Polynesia: a contrast to the population structure revealed by mitochondrial DNA. Am J Hum Genet 63:1793–1806

Hurles ME, Veitia R, Arroyo E, Armenteros M, Bertranpetit J, Pérez-Lezaun A, Bosch E, Shlumukova M, Cambon-Thomsen A, McElreavey K, López de Munain A, Röhl A, Wilson IJ, Singh L, Pandya A, Santos FR, Tyler-Smith C, Jobling MA (1999) Recent male-mediated gene flow over a linguistic barrier in Iberia, suggested by analysis of a Y-chromosomal DNA polymorphism. Am J Hum Genet 65: 1437–1448

Jobling MA, Tyler-Smith C (1995) Fathers and sons: the Y chromosome and human evolution. Trends Genet 11:449–456

——— (2000) New uses for new haplotypes: the human Y chromosome, disease, and selection. Trends Genet 16:356–362

Jobling MA, Williams G, Schiebel K, Pandya A, McElreavey K, Salas L, Rappold GA, Affara NA, Tyler-Smith C (1998) A selective difference between human Y-chromosomal DNA haplotypes. Curr Biol 8:1391-1394

Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, deKnijff P, Wiebe V, Griffiths RC, Templeton AR, Hammer MF (1999) Ancestral Asian source(s) of New World Y-chromosome founder haplotypes. Am J Hum Genet 64:817–831

Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, Krüger C, Krawczak M, Nagy M, Dobosz T, Szibor R, de Knijff P, Stoneking M, Sajantila A (2000) Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. Am J Hum Genet 66:1580–1588

Kwok C, Tyler-Smith C, Medonca BB, Hughes I, Berkovitz GD, Goodfellow PN, Hawkins JR (1996) Mutation analysis of 2kb 5′ to SRY in XY females and XX intersex subjects. J Med Genet 33:465–468

Landers J (1992) Reconstructing ancient populations. In: Jones S, Martin R, Pilbeam D (eds) The Cambridge encyclopedia of human evolution. Cambridge University Press, Cambridge, pp 402–405

Langaney A, Roessli D, van Blyenburgh NH, Dard P (1992) Do most human populations descend from phylogenetic trees? Hum Evol 7:47–61

Lucotte G, Loirat F (1999) Y-chromosome DNA haplotype 15 in Europe. Hum Biol 71:431–437

Malaspina P, Cruciani F, Ciminelli BM, Terrenato L, Santolamazza P, Alonso A, Banyko J, Brdicka R, Garcia O, Gaudiano C, Guanti G, Kidd KK, Lavinha J, Avila M, Mandich P, Moral P, Qamar R, Mehdi SQ, Ragusa A, Sefanescu G, Caraghin M, Tyler-Smith C, Scozzari R, Novelletto A (1998) Network analyses of Y-chromosomal types in Europe, northern Africa, and western Asia reveal specific patterns of geographic distribution. Am J Hum Genet 63:847–860

Mantel NA (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220

Mathias N, Bayés M, Tyler-Smith C (1994) Highly informative compound haplotypes for the human Y chromosome. Hum Mol Genet 3:115–123

McKeigue PM (1997) Mapping genes underlying ethnic differences in disease risk by linkage disequilibrium in recently admixed populations. Am J Hum Genet 60:188–196

Menozzi P, Piazza A, Cavalli-Sforza LL (1978) Synthetic maps of human gene frequencies in Europeans. Science 201: 786–792

Ngo KY, Vergnaud G, Johnsson C, Lucotte G, Weissenbach J (1986) A DNA probe detecting multiple haplotypes of the human Y chromosome. Am J Hum Genet 38:407–418

Piazza A, Rendine S, Minch E, Menozzi P, Mountain J, Cavalli-Sforza LL (1995) Genetics and the origin of European languages. Proc Natl Acad Sci USA 92:5836–5840

Poloni ES, Semino O, Passarino G, Santachiara-Benerecetti AS, Dupanloup L, Langaney A, Excoffier L (1997) Human genetic affinities for Y-chromosome P49a,f/*Taq*I haplotypes show strong correspondence with linguistics. Am J Hum Genet 61:1015–1035

Quintana-Murci L, Semino O, Minch E, Passarino G, Brega A, Santachiara-Benerecetti AS (1999) Further characteristics of proto-European Y chromosomes. Eur J Hum Genet 7:603–608

Renfrew C (1987) Archaeology and language: the puzzle of Indo-European origins. Jonathan Cape, London

——— (1989) The origins of Indo-European languages. Sci Am 261:106–114

——— (2000) At the edge of knowability: towards a prehistory of languages. Camb Archaeol J 10:7–34

Richards M, Côrte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A, Papiha S, Hedges R, Bandelt H-J, Sykes B (1996) Paleolithic and neolithic lineages in the European mitochondrial gene pool. Am J Hum Genet 59: 185–203

Richards M, Sykes B (1998) Evidence for Paleolithic and Neolithic gene flow in Europe. Am J Hum Genet 62:491–492

Ruhlen M (1991) A guide to the world's languages. Edward Arnold, London

Santos FR, Carvalho-Silva DR, Pena SDJ (1999) PCR-based DNA profiling of human Y chromosomes. In: Epplen JT, Lubjuhn T (eds) Methods and tools in biosciences and medicine. Birkhaüser Verlag, Basel, pp 133–152

Santos FR, Tyler-Smith C (1996) Reading the human Y chromosome: the emerging DNA markers and human genetic history. Braz J Genet 19:665–670

Schneider S, Roessli D, Excoffier L (2000) ARLEQUIN ver 2.0: a software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva

Seielstad MT, Hebert JM, Lin AA, Underhill PA, Ibrahim M, Vollrath D, Cavalli-Sforza LL (1994) Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition. Hum Mol Genet 3:2159–2161

Seielstad MT, Minch E, Cavalli-Sforza LL (1998) Genetic evidence for a higher female migration rate in humans. Nat Genet 20:278–280

Semino O, Passarino G, Brega A, Fellous M, Santachiara-Benerecetti AS (1996) A view of the Neolithic demic diffusion in Europe through two Y chromosome-specific markers. Am J Hum Genet 59:964–968

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

Simoni L, Calafell F, Pettener D, Bertranpetit J, Barbujani G (2000*a*) Geographic patterns of mtDNA diversity in Europe. Am J Hum Genet 66:262–278

——— (2000*b*) Reconstruction of prehistory on the basis of genetic data. Am J Hum Genet 66:1177–1179

Simoni L, Gueresi P, Pettener D, Barbujani G (1999) Patterns of gene flow inferred from genetic distances in the Mediterranean region. Hum Biol 71:399–415

Sokal RR, Harding RM, Oden NL (1989) Spatial patterns of human gene frequencies in Europe. Am J Phys Anthropol 80:267–294

Sokal RR, Oden NL (1978) Spatial autocorrelation in biology. Biol J Linn Soc 10:199–249

Sokal RR, Oden NL, Legendre P, Fortin MJ, Kim J, Thomson BA, Vaudor A, Harding RM, Barbujani G (1990) Genetics and language in European populations. Am Nat 135: 157–175

Sokal RR, Oden NL, Thomson BA (1997) A simulation study of microevolutionary inferences by spatial autocorrelation analysis. Biol J Linn Soc 60:73–93

Sokal RR, Oden NL, Wilson C (1991) Genetic evidence for the spread of agriculture in Europe by demic diffusion. Nature 351:143–145

Swadesh M (1952) Lexico-statistic dating of prehistoric ethnic contacts: with special reference to North American Indians and Eskimos. Proc Am Philos Soc 96:452–463

Templeton AR (1993) The "Eve" hypothesis: a genetic critique and reanalysis. Am Anthropol 95:51–72

Torroni A, Richards M, Macaulay V, Forster P, Villems R, Nørby S, Savontaus M-L, Huoponen K, Scozzari R, Bandelt H-J (2000) mtDNA haplogroups and frequency patterns in Europe. Am J Hum Genet 66:1173–1177

Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Res 7:996–1005

Veitia R, Ion A, Barbaux S, Jobling MA, Souleyreau N, Ennis K, Ostrer H, Tosi M, Meo T, Chibani J, Fellous M, McElreavey K (1997) Mutations and sequence variants in the testis-determining region of the Y chromosome in individuals with a 46,XY female phenotype. Hum Genet 99:648–652

Whitfield LS, Sulston JE, Goodfellow PN (1995) Sequence variation of the human Y chromosome. Nature 378:379–380

Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhövel W, Fretwell N, Jobling MA, Harihara S, Shimizu K, Semjidmaa D, Sajantila A, Salo P, Crawford MH, Ginter EK, Evgrafov OV, Tyler-Smith C (1997) Genetic relationships of Asians and northern Europeans, revealed by Y-chromosomal DNA analysis. Am J Hum Genet 60: 1174–1183

Zerjal T, Pandya A, Santos FR, Adhikari R, Tarazona E, Kayser M, Evgrafov O, Singh L, Thangaraj K, Destro-Bisol G, Thomas MG, Qamar R, Mehdi Q, Rosser ZH, Hurles ME, Jobling MA, Tyler-Smith C (1999) The use of Y-chromosomal DNA variation to investigate population history: recent male spread in Asia and Europe. In: Papiha SS, Deka R, Chakraborty R (eds) Genomic diversity: applications in human population genetics. Plenum Press, New York, pp 91–102

Zvelebil M, Zvelebil KV (1988) Agricultural transition and Indo-European dispersal. Antiquity 62:574–583