analysis herein, because they are based on either completely sequenced genomes or random sequencing projects.

24. D. Grant, P. Cregan, R. C. Showemaker, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4168 (2000).

25. This estimate is the average of two surveys based on analyses of multiple genes in vascular plants (7, 42).

26. L. Skrabanek, K. H. Wolfe, *Curr. Opin. Genet. Dev.* **8**, 694 (1998).

27. A. L. Hughes, *J. Mol. Evol.* **48**, 565 (1999).

28. A. P. Martin, *Am. Nat.* **154**, 111 (1999).

29. P. E. Ahlberg, A. R. Milner, *Nature* **368**, 507 (1994).

30. In large genomic databases like those analyzed here, some sequencing errors may inflate the apparent level of divergence, but this error should be independent of S, and in any event, is unlikely to add more than 0.01 to individual estimates of S. Thus, the impact of such error on our statistical analyses should be negligible.

31. This estimate is the average of the results obtained in three broadly compatible studies, all of which surveyed a large number of genes [S. Easteal, C. Collet, *Mol. Biol. Evol.* **11**, 643 (1994); T. Ohta, *J. Mol. Evol.* **40**, 56 (1995); (7)].

32. A. M. A. Aguinaldo *et al.*, *Nature* **387**, 489 (1997).

33. G. M. Rubin *et al.*, *Science* **287**, 2204 (2000).

34. C. Semple, K. H. Wolfe, *J. Mol. Evol.* **48**, 555 (1999).

35. C. R. Werth, M. D. Windham, *Am. Nat.* **137**, 515 (1991).

36. M. Lynch, A. Force, *Am. Nat.*, in press.

37. H. A. Orr, *Genetics* **144**, 1331 (1996).

38. M. J. D. White, *Modes of Speciation* (Freeman, San Francisco, CA, 1978).

39. G. Fischer, S. A. James, I. N. Roberts, S. G. Oliver, E. J. Louis, *Nature* **405**, 451 (2000).

40. H. R. Parker, D. P. Philipp, G. S. Whitt, *J. Exp. Zool.* **233**, 451 (1985).

41. J. A. Coyne, H. A. Orr, *Evolution* **51**, 295 (1997).

42. M. Lynch, *Mol. Biol. Evol.* **14**, 914 (1997).

43. This research was supported by NIH grant RO1-GM36827. We thank A. Force and A. Wagner for helpful comments.

The Genetic Legacy of Paleolithic *Homo sapiens sapiens* in Extant Europeans: A Y Chromosome Perspective

Ornella Semino,[1,2]*† Giuseppe Passarino,[2,3]† Peter J. Oefner,[4]
Alice A. Lin,[2] Svetlana Arbuzova,[5] Lars E. Beckman,[6]
Giovanna De Benedictis,[3] Paolo Francalacci,[7]
Anastasia Kouvatsi,[8] Svetlana Limborska,[9] Mladen Marcikiæ,[10]
Anna Mika,[11] Barbara Mika,[12] Dragan Primorac,[13]
A. Silvana Santachiara-Benerecetti,[1] L. Luca Cavalli-Sforza,[2]
Peter A. Underhill[2]

A genetic perspective of human history in Europe was derived from 22 binary markers of the nonrecombining Y chromosome (NRY). Ten lineages account for >95% of the 1007 European Y chromosomes studied. Geographic distribution and age estimates of alleles are compatible with two Paleolithic and one Neolithic migratory episode that have contributed to the modern European gene pool. A significant correlation between the NRY haplotype data and principal components based on 95 protein markers was observed, indicating the effectiveness of NRY binary polymorphisms in the characterization of human population composition and history.

Various types of evidence suggest that the present European population arose from the merging of local Paleolithic groups and Neolithic farmers arriving from the Near East after the invention of agriculture in the Fertile Crescent (1–5). However, the origin of Paleolithic European groups and their contribution to the present gene pool have been debated (6, 7). Assuming no selection, local differentiation occurred in isolated and small Paleolithic groups by drift (8, 9). Range expansions and population convergences, which occurred at the end of the Paleolithic, were catalyzed by improved climate and new technologies and spread the present genetic characteristics to surrounding areas (8). The smaller effective population size of the NRY enhances the consequences of drift and founder effect relative to the autosomes, making NRY variation a potentially sensitive index of population composition. Previously, the distribution of two NRY restriction fragment length polymorphism (RFLP) markers suggested Paleolithic and Neolithic contribu-

tions to the European gene pool (10). NRY binary markers (11) representing unique mutational events in human history allow a more comprehensive reconstruction of European genetic history.

Twenty-two relevant binary markers [4 gathered from the literature and 18 detected by denaturing high-performance liquid chromatography (DHPLC) (12)] were genotyped in 1007 Y chromosomes from 25 different European and Middle Eastern geographic regions. More than 95% of the samples studied could be assigned to haplotypes or clades of haplotypes defined by just 10 key mutations (Fig. 1 and Table 1). The frequency distribution of Y chromosome haplotypes revealed here defines the basic structure of the male component of the extant European populations and provides testimony to population history, including the Paleolithic period. Two lineages (those characterized by M173 and M170) appear to have been present in Europe since Paleolithic times. The remaining lineages entered

Europe most likely later during independent migrations from the Middle East and the Urals as they are found at higher frequencies and with more variation of linked microsatellites than in other continents (10–14).

Of the 22 haplotypes that constitute the phylogeny in Fig. 1 (top), Eu18 and Eu19 characterize about 50% of the European Y chromosomes. Although they share M173, the two haplotypes show contrasting geographic distribution. The frequency of Eu18 decreases from west to east, being most frequent in Basques (Fig. 1, bottom, and Table 1). This lineage includes the previously described proto-European lineage that is characterized by the 49a,f haplotype 15 (10). In contrast, haplotype Eu19, which is derived from the M173 lineage and is distinguished by M17, is virtually absent in Western Europe. Its frequency increases eastward and reaches a maximum in Poland, Hungary, and Ukraine, where Eu18 in turn is virtually absent. Both haplotypes Eu18 and Eu19 share the derived M45 allele. The lineage characterized by M3, common in Native Americans

[1]Dipartimento di Genetica e Microbiologia, Università di Pavia, Via Ferrata 1, 27100 Pavia, Italy. [2]Department of Genetics, Stanford University School of Medicine, 300 Pasteur Drive, Stanford, CA 94305–5120, USA. [3]Dipartimento di Biologia Cellulare, Università della Calabria, 87030 Rende, Italy. [4]Stanford Genome Technology Center, 855 California Avenue, Palo Alto, CA 94304, USA. [5]International Medico-Genetic Centre, Hospital Nol, 57 Artem Str, 340000 Donetsk, Ukraine. [6]Department of Oncology, Pathology and Medical Genetics, University of Umeå, S-901 85 Umeå, Sweden. [7]Dipartimento di Zoologia e Antropologia Biologica, Università di Sassari, Via Regina Margherita, 15, 07100 Sassari, Italy. [8]Department of Genetics, Development and Molecular Biology, Aristotle University, 54006 Thessaloniki, Macedonia, Greece. [9]Institute of Molecular Genetics, Russian Academy of Sciences, Kurchatov Square, 2, Moscow 123182, Russia. [10]Clinical Hospital Center Osijek, Department of Pathology Medical School, J Huttlera 4, 31000 Osijek, Croatia. [11]Regionalne Centrum Krwiodawstwa i Krwiolecznictwa w Lublinie–Oddzial w, Zamosciu, ul Legionow 10, 22400 Zamosc, Poland. [12]Samodzielny Publiczny Szpital Wojwodzki im. Papieza Jona Pawla II w, Zamosciu, ul Legionow 10, 22400 Zamosc, Poland. [13]University Hospital Split, Department of Pediatrics, Laboratory for Clinical and Forensic Genetics, Spinèiæeva 1, 21000 Split, Croatia.

*To whom correspondence should be addressed. E-mail: semino@ipvgen.univp.it

†These authors contributed equally to this work.

(12) and a few Siberian populations (15), is also a derivative of M45. This observation suggests that M173 is an ancient Eurasiatic marker that was brought by or arose in the group of *Homo sapiens sapiens* who entered Europe and diffused from east to west about 40,000 to 35,000 years ago (16, 17), spreading the Aurignac culture. This culture also appeared almost simultaneously in Siberia (17), from which some groups eventually migrated to the Americas.

We interpret the differentiation and the distribution of haplotypes Eu18 and Eu19 as signatures of expansions from isolated population nuclei in the Iberian peninsula and the present Ukraine, following the Last Glacial Maximum (LGM). In fact, during this glacial period (20,000 to 13,000 years ago), human groups were forced to vacate Central Europe, with the exception of a refuge in the northern Balkans (16). Similar discrete patterns of the flora and fauna in Europe have been attributed to glaciation-modulated isolation followed by dispersal from climatic sanctuaries (18). This scenario is also supported by the finding that the maximum variation for microsatellites linked to Eu19 is found in Ukraine (19). In turn, the maximum variation for microsatellites linked to 49a,f Ht15 and its derivatives (and then to the Eu18 lineage) is in the Iberian peninsula (19). This is consistent with the diffusion of M173-marked Eu18 from its refuge after the LGM, in agreement with mitochondrial DNA (mtDNA) haplogroup V and some of the H lineages (20). Haplotype Eu19 has been also observed at substantial frequency in northern India and Pakistan (12) as well as in Central Asia (12). Its spread may have been magnified by the expansion of the Yamnaia culture from the "Kurgan culture" area (present-day southern Ukraine) into Europe and eastward, resulting in the spread of the Indo-European language (21). An alternative hypothesis of a Middle Eastern origin of Indo-European languages was proposed on the basis of archaeological data (3).

We estimated the age of M173 by using the variation of three microsatellites, namely DYS19, YCAIIa, and YCAIIb (22). Although an estimate of ~30,000 years for M173 must be interpreted cautiously (23), it is consistent with our hypothesis that M173 marks the Aurignac settlement in Europe or, at least, predates the LGM.

The polymorphism M170 represents another putative Paleolithic mutation whose age has been estimated to be ~22,000 years (22, 23). With the exception of idiosyncratic distributions indicative of recent gene flow, M170 is confined to Europe (Eu7). The mutation is most frequent in central Eastern Europe and also occurs in Basques and Sardinians that have accumulated a subsequent mutation (M26) that distinguishes Eu8. The closest phylogenetic predecessor is the M89 mutation, from which the most important Middle Eastern lineages originated. We propose that M170 originated in Europe in descendants of men that arrived from the Middle East 20,000 to 25,000 years ago, who have been associated with the Gravettian culture (16). This migration may have coincided with that of mtDNA haplogroup H to Europe. It has been suggested that Gravettian and Aurignac groups coexisted for a few thousand years, maintaining their identities despite occasional contacts. During the LGM, Western Europe was isolated from Central Europe, where an Epi-Gravettian culture persisted in the area of present-day Austria, the Czech Republic, and the northern Balkans (16). After climatic improvement, this culture spread north and east (16). This finding is supported by the present Eu7 haplotype distribution. In this scenario, haplotype Eu8 would have originated in the western Paleolithic population during the LGM,
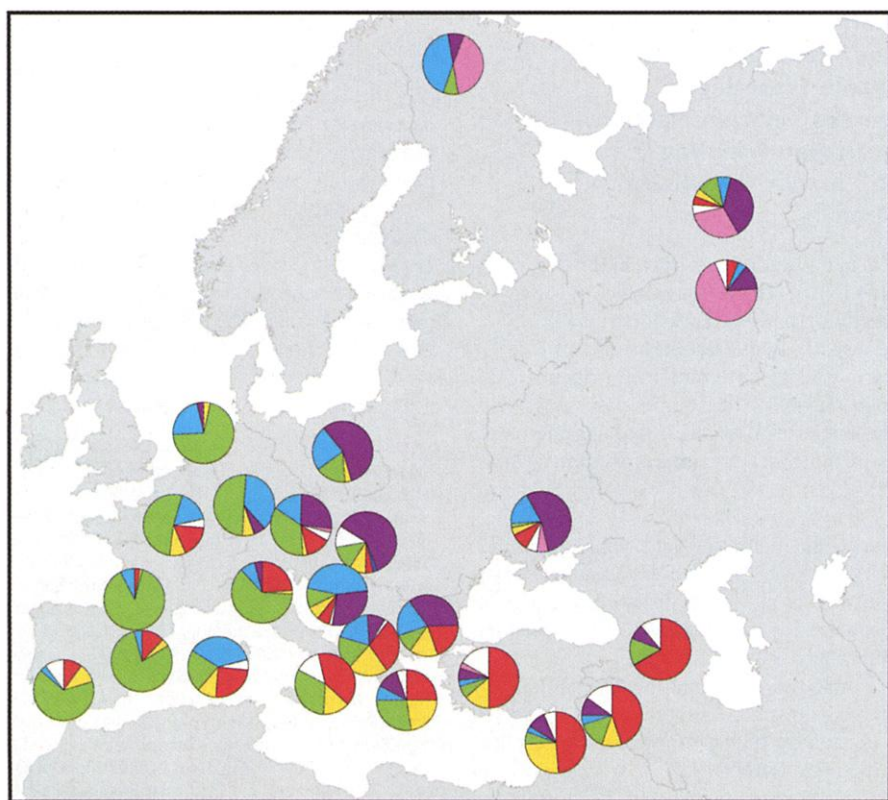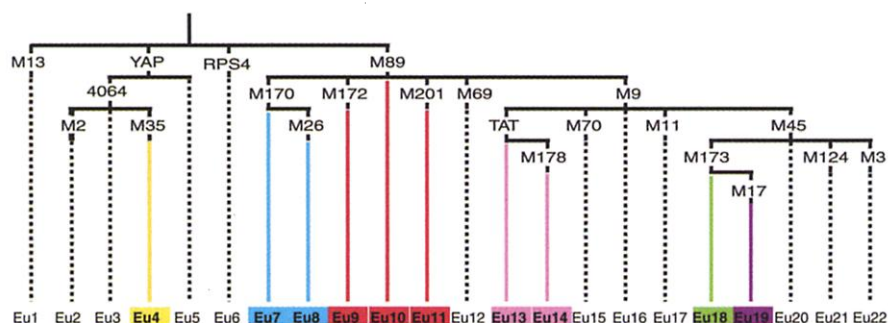
Fig. 1. (Top) Maximum parsimony phylogeny of the NRY markers found in Europe and elsewhere. YAP (32), TAT (14), RPS4 [= RPS4YC711T (33)], and 4064 [= SRY4064 (34)] were previously described. The remaining polymorphisms were identified with DHPLC (11, 12, 27) and are deposited in the National Center for BioTechnology Information (NCBI) dbSNP database (www.ncbi.nlm. nih.gov/SNP). The phylogeny is rooted with the use of great ape sequences. **(Bottom)** The 19 haplotypes observed (Table 1) were pooled into six classes represented by different colors: Yellow indicates haplotype Eu4; blue includes Eu7 and Eu8, which both involve the M170 mutation; red groups three separate haplotypes for reasons explained in the text; pink includes haplotypes Eu13 and Eu14, which both involve the TAT mutation; and green indicates Eu18 and purple indicates Eu19, which despite sharing the M173 mutation are distinguished because they represent a distinct dichotomy in European phylogeography. The other nine observed haplotypes, which catalog the remaining <5% of the total samples, are shown as black dashed lines and are represented in the white sector of relevant pie charts. Three haplotypes, Eu2, Eu5, and Eu21, were not detected. The pie sectors are proportional to the relative frequencies of haplotypes or clades in each population. The two Basque samples have been pooled.

as local differentiation of the M170 lineage. The frequency and the distribution of haplogroup H across Europe support gene flow between Gravettian and Western European Aurignac groups and suggest differential gender migratory phenomena (24).

The cline of frequencies for haplotypes marked by M35 (Eu4), M172 (Eu9), M89 (Eu10), and M201 (Eu11) decreases from the Middle East into Europe. Haplotype Eu4 is phylogenetically distinct from the other three and defines most European YAP$^+$ chromosomes. The Eu4 haplotype appears to correspond to the previously reported Ht-4, defined by the absence of M2 (25). Comparative genotyping with the Y chromosome RFLPs 49a,f and 12f2 [(10) and citations therein] revealed that Eu9 and Eu10 share the 12f2-derived 8Kb allele, whereas Eu11 has the ancestral 12f2-10Kb allele. Haplotypes Eu9, Eu10, and Eu11 share the 49a,f haplotype 8 or its derivatives, which are not observed in any of the other 16 Eu haplotypes (19), suggesting a shared common ancestry. Thus, we have displayed the combined frequencies of haplotypes Eu9, Eu10, and Eu11 in Fig. 1. By correlation between

Ht-4 ≈ Eu4 and 12f2-8Kb ≈ Eu9 and Eu10, the origin of these lineages has been estimated to be about 15,000 to 20,000 years ago (13). A similar date (17,000 years ago) for Eu11 has been estimated (22, 23). The molecular age of a mutation and its corresponding haplotype must predate the demographic migratory event it marks. The age estimates of these haplotypes, especially considering their approximation (22, 23), cannot distinguish whether they came to Europe before or after the LGM. However, the decreasing clinal pattern of haplotypes Eu4, Eu9, Eu10, and Eu11 from the Middle East to Europe would not be compatible with the localization of peoples carrying these Y chromosomes to refuges during the LGM. If these haplotypes were present in Europe before the LGM, we would expect to see a differentiation between the European and Middle Eastern lineages because of temporal and spatial isolation. Unpublished data from a 49a,f system and seven short tandem repeats (STRs) in a large sample of these NRY haplotypes from Europe and the Middle East (19) have revealed that almost all the compound haplotypes observed in Europe were included in the smaller sample of the

Middle East (19). A similar result was observed for mtDNA haplogroup J, which, although considered Paleolithic, is believed to have been introduced to Europe during the Neolithic (6). These observations suggest that the four NRY haplotypes, as well as mtDNA haplogroup J, had sufficient time to differentiate in the Middle East and then migrate toward Europe in sufficiently large numbers to account for most of the existing variation. Therefore, haplotypes Eu4, Eu9, Eu10, and Eu11 represent the male contribution of a demic diffusion of farmers from the Middle East to Europe. The contribution of the Neolithic farmers to the European gene pool seems to be more pronounced along the Mediterranean coast than in Central Europe. This is evident from Fig. 2, in which we have plotted the frequencies of haplotypes Eu4, Eu9, Eu10, and Eu11 against the geographic distances from the Middle East for each population. The regression line accounting for Mediterranean populations has a slope that is significantly different from the other populations, indicating that the diffusion of Neolithic farmers affected Southern more than Central Europe.

Table 1. Frequencies (in percent) of the haplotypes found in the examined European populations.

Population†	n	Eu1	Eu3	Eu4	Eu6	Eu7	Eu8	Eu9	Eu10	Eu11	Eu12	Eu13	Eu14	Eu15	Eu16	Eu17	Eu18	Eu19	Eu20	Eu21
Andalusian Spanish	29			10.3		3.4		6.9	3.4						6.9		3.4	65.5			
Basque French	45			2.2		2.2	4.4		2.2									88.9			
Basque	22						9.1	4.5										86.4			
Catalan	24			4.2		4.2		4.2		8.3								79.2			
French	23			8.7		17.4		13.0	4.3									52.2		4.3	
Dutch	27			3.7		22.2												70.4	3.7		
German	16			6.2		37.5												50.0	6.2		
Czech and Slovakian	45			2.2		15.6		8.9		4.4			2.2	2.2	2.2			35.6	26.7		
Central-northern Italian	50			2.0		8.0		14.0		10.0								62.0	4.0		
Calabrian	37		2.7	13.5				21.6	10.8	8.0					2.7	5.4	32.4		2.7	
Sardinian	77	1.3		10.4	1.3	2.6	35.1	5.2	5.2	14.2				1.3			22.1			1.3
Croatian	58			6.9		44.8		5.2		1.7					1.7		10.3	29.3		
Albanian	51		2.0	21.6		19.6		23.5	4.0	2.0							17.6	9.8		
Greek	76		1.3	22.4	1.3	7.9		21.0	1.3	2.6				1.3		1.3	27.6	11.8		
Macedonian	20			15.0		20.0		15.0	5.0								10.0	35.0		
Polish	55			3.6		23.6											16.4	56.4		
Hungarian	45			8.9		11.1		2.2		2.2				2.0		2.2	13.3	60.0		
Ukrainian	50			4.0		18.0		6.0		4.0	2.0		6.0	2.0			2.0	54.0	2.0	
Georgian	63							33.3	3.2	30.1				1.6	1.6	1.6	14.3	7.9		6.3
Turkish	30		3.3	13.3		3.3		40.0	3.3	6.6			3.3	3.3	3.3	3.3	6.6	6.6	3.3	
Lebanese	31			25.8	3.2	3.2		29.0	16.1	3.2						3.2	6.4	9.7		
Syrian	20		10.0	10.0		5.0		15.0	30.0		5.0						15.0	10.0		
Saami	24					41.7							41.7				8.3	8.3		
Udmurt	43			4.7		7.0			4.7			2.3	27.9		4.6		11.6	37.2		
Mari	46					4.3			6.5			4.3	65.2		6.5			13.0		
Total	1007																			

*The haplotypes are defined by the following markers and the respective derived alleles: Eu1, M13-C; Eu3, YAP$^+$, 4064-A; Eu4, YAP$^+$, 4064-A, M35-C; Eu6, RPS4-T; Eu7, M89-T, M170-C; Eu8, M89-T, M170-C, M26-A; Eu9, M89-T, M172-G; Eu10, M89-T; Eu11, M89-T, M201-T; Eu12, M89-T, M69-C; Eu13, M89-T, M9-G, TAT-C; Eu14, M89-T, M9-G, TAT-C, M178-T; Eu15, M89-T, M9-G, M70-C; Eu16, M89-T, M9-G; Eu17, M89-T, M9-G, M11-G; Eu18, M89-T, M9-G, M45-A, M173-C; Eu19, M89-T, M9-G, M45-A, M173-C, M17(delG); Eu20, M89-T, M9-G, M45-A; Eu21, M89-T, M9-G, M45-A, M124-T. Haplotypes Eu2, Eu5, and Eu22 were not observed. †Several samples were previously described (10, 11, 28). Samples not previously examined included 23 French, 16 Germans, 39 northern Italians, 45 Sardinians, 58 Croats, 20 Macedonians from northern Greece, 55 Poles, 50 Ukrainians, 20 Syrians, 24 Saami, 43 Udmurts, and 46 Mari.

While allelotyping M35 by DHPLC, we found a previously unknown mutation, M178, in 95% of all TAT chromosomes. The latter has been reported to be ~4000 years old and marks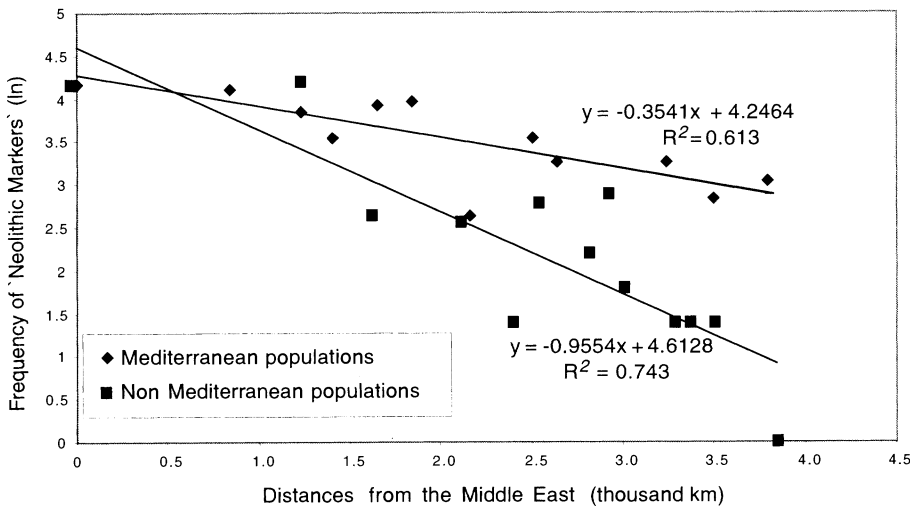 a recent Uralic migration confined to Northern Europe (14). Neither TAT nor M178 was detected in Hungary, where a Uralic language is spoken.

The first two principal components (PC) derived from the data in Table 1 are shown in

Fig. 2. Abscissa: distances in thousands of kilometers of each population from the average of the two Middle Eastern populations (Lebanese and Syrians). Ordinate: logarithm of relative frequencies of Neolithic markers (sum of Eu4, Eu9, Eu10, and Eu11) in the Mediterranean and non-Mediterranean populations. The Middle Eastern point ($X = 0$) was considered for both series of points. The two regression lines are significantly different ($P < 0.01$).

Fig. 3. The Udmurts, Mari, and Saami were excluded because they monopolized the first PC and compressed the rest of the variation because of their high TAT/M178 frequency. In the plot, it is possible to see three clusters of distinct geography and culture. The first comprises Basques and Western Europeans, the second Middle Eastern, and the third Eastern European populations from Croatia, Ukraine, Hungary, and Poland. These three geographic clusters correspond to the major glacial refuges and to the region of origin of the farmers' expansion.
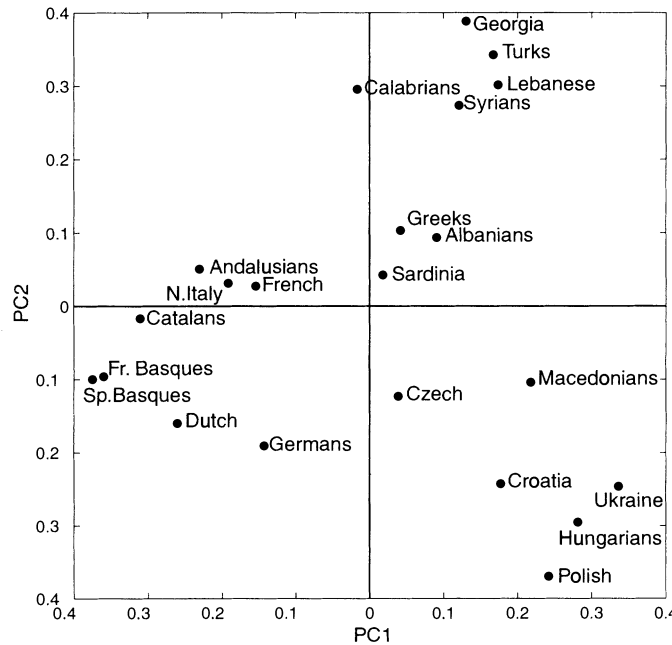
The most comprehensive previous survey of the European gene pool has been the PC analysis of 95 autosomal protein polymorphisms (5, 8). We compared the frequency distribution of the major Eu Y chromosome haplotypes with the first three PCs of Europe (Table 2). Because Sardinians were not included in the original PC analysis because of their pronounced outlier phylogenetic status (5), they were also excluded in our correlation analysis. The first PC, which was proposed to reflect the diffusion of Neolithic farmers (5, 8), correlates with Eu4, Eu9, Eu10, and Eu11. The second PC, whose meaning has never been fully assessed (5, 8), is correlated with the spread of Eu18 from Spain toward Central Europe and, on the opposite pole, with the spread of Uralic TAT/M178 (Eu13 and Eu14). The third PC, the meaning of which has been debated (3, 5, 8), correlates to the M17 mutation (Eu19). The concordance of protein-based PC and NRY data suggests that migration, more than natural selection, has influenced the pattern of NRY variation observed.

Analyses of mtDNA sequence variation in European populations have been conducted (6, 20). These data suggest that the gene pool has ~80% Paleolithic and ~20% Neolithic ancestry. Our data support this observation because haplotypes Eu4, Eu9, Eu10, and Eu11 account for ~22% of European Y chromosomes. Thus, the mtDNA and Y data corroborate the previous observation that the first PC of the 95 classical polymorphisms accounts for ~28% of the overall genetic variation (5, 6). However, some differences exist between the mtDNA and Y data pertaining to the putative Paleolithic components. It has been proposed that mtDNA haplogroup U5 arrived from the Middle East 45,000 years ago (6, 26). We did not detect any corresponding Y haplotypes. Furthermore, most European mtDNA lineages, which account for 60 to 70% of the variation in Europe, have been interpreted as having arrived from the Middle East during the Paleolithic about 25,000 years ago (6). Correspondingly, ~20% of contemporary Y lineages characterized by the M170 mutation derive from deep phylogenetic M89 ancestry, consistent with a Middle Eastern Paleolithic heritage. Moreover, the remaining ~50% of Y lineages associated with the M173 mutation indicate a major influence

Fig. 3. PC analysis of data in Table 1. The first PC accounts for 46.24% of the variance, whereas the second accounts for 34.69%.

Table 2. Correlation between the first three PCs based on autosomal protein markers (5) and the frequency of the major European Y chromosome haplotypes.

	Eu4	Eu9+Eu10+Eu11	Eu7	Eu18	Eu13+Eu14	Eu19
PC1 (28%)†	0.6806**	0.7939**	−0.4450*	−0.3004	−0.3644	−0.1673
PC2 (22%)†	−0.1415	0.0626	0.0594	−0.6967**	0.7570**	0.2074
PC3 (11%)†	−0.3385	−0.1248	0.0924	−0.3836	0.0438	0.7270**

*$P < 0.05$.   **$P < 0.001$.   †Portion of variation accounted for by each PC based on autosomal protein markers (5).

on the extant gene pool from Central Asia ~30,000 years ago. In contrast, Central Asian mtDNA 16223/C haplogroups (I, X, and W) account for only ~7% of the contemporary composition (26). These discrepancies may be due in part to the apparent more recent molecular age of Y chromosomes relative to other loci (27), suggesting more rapid replacement of previous Y chromosomes. Gender-based differential migratory demographic behaviors will also influence the observed patterns of mtDNA and Y variation (24).

The previously categorized Sardinians, Basques, and Saami outliers (5) share basically the same Y binary components of the other Europeans. Their peculiar position with respect to frequency is probably a consequence of genetic drift and isolation. In addition, our analysis highlights the expansion of the Epi-Gravettian population from the northern Balkans.

Almost all of the European Y chromosomes analyzed in the present study belong to 10 lineages characterized by simple biallelic mutations. Furthermore, a substantial portion of the European gene pool appears to be of Upper Paleolithic origin, but it was relocated after the end of the LGM, when most of Europe was repopulated (16).

References and Notes

1. P. Menozzi, A. Piazza, L. L. Cavalli-Sforza, Science 201, 786 (1978).
2. A. J. Ammerman, L. L. Cavalli-Sforza, The Neolithic Transition and the Genetics of Populations in Europe (Princeton Univ. Press, Princeton, NJ, 1984).
3. C. Renfrew, Archaeology and Language; the Puzzle of Indo-European Origin (Jonathan Cape, London, 1987).
4. M. A. Ruhlen, Guide to the World Languages (Stanford Univ. Press, Stanford, CA, 1987).
5. L. L. Cavalli-Sforza, P. Menozzi, A. Piazza, The History and Geography of Human Genes (Princeton Univ. Press, Princeton, NJ, 1994).
6. M. Richards et al., Ann. Hum. Genet. 62, 241 (1998).
7. L. Chikhi et al., Proc. Natl. Acad. Sci. U.S.A. 95, 9053 (1998).
8. L. L. Cavalli-Sforza, P. Menozzi, A. Piazza, Science 259, 639 (1993).
9. M. M. Lahr, R. A. Foley, Am. J. Phys. Anthropol. Suppl. 27, 137 (1998).
10. O. Semino, G. Passarino, A. Brega, M. Fellous, A. S. Santachiara-Benerecetti, Am. J. Hum. Genet. 59, 964 (1996).
11. P. A. Underhill et al., Nature Genet. 26, 358 (2000).
12. P. A. Underhill et al., Genome Res. 7, 996 (1997).
13. M. F. Hammer et al., Proc. Natl. Acad. Sci. U.S.A. 97, 6769 (2000).
14. T. Zerjal et al., Am. J. Hum. Genet. 60, 1174 (1997).
15. J. T. Lell et al., Hum. Genet. 100, 536 (1997).
16. M. Otte, in The World at 18000 BP, O. Soffer, C. Gamble, Eds. (Unwin Hyman, London, 1990), vol. 1, pp. 54–68.
17. R. Klein, Evol. Anthropol. 1, 5 (1992).
18. K. J. Willis, R. J. Whittaker, Science 287, 1406 (2000).
19. A. S. Santachiara-Benerecetti, unpublished data.
20. A. Torroni et al., Am. J. Hum. Genet. 62, 1137 (1998).
21. M. Gimbutas, in Indo-European and Indo-Europeans, G. Cardona, H. M. Hoenigswald, A. M. Senn, Eds. (Univ. of Pennsylvania Press, Philadelphia, PA, 1970), pp. 155–195.
22. The coalescence time for the lineages M173 and M170 was calculated on 209 and 73 Y chromosomes, respectively, by using the variation at YCAIIa, YCAIIb, and DYS19 microsatellite loci (28). The variances obtained from these microsatellite data were computed with equation 2 from Goldstein et al. (29). The initial variance of the CA repeats was considered to be null. A constant population size of 4500 and a generation time of 27 years were assumed as suggested (29). Mutational rates of 5.6×10^{-4} for YCAII loci and 1.1×10^{-3} for DYS19 were used [for a discussion of the mutational rates, see (30)].
23. Many factors confound the estimation of the ages of binary mutations based on Y chromosome microsatellites. First, the mutation rate of microsatellites is uncertain, especially because it is not uniform for all microsatellites (30). Moreover, there is a difference between the mutation rate measured in pedigrees and the mutation rates measured indirectly through phylogenetic analysis (31). To preserve the comparative (if not the absolute) value of the age estimates, the mutation rates we used are those most widely accepted currently (30). In addition, it seems that some demographic and evolutionary mechanism reduces the variation of Y chromosomes in humans, and as a consequence, dating based on such data usually gives an underestimate (27). Last, but not least, it is worth noting that the age of alleles does not correspond to the age of populations, although such estimates can provide insights.
24. M. T. Seielstad, E. Minch, L. L. Cavalli-Sforza, Nature Genet. 20, 278 (1998).
25. M. F. Hammer et al., Genetics 145, 787 (1997).
26. V. Macaulay et al., Am. J. Hum. Genet. 64, 232 (1999).
27. P. Shen et al., Proc. Natl. Acad. Sci. U.S.A. 97, 7354 (2000).
28. L. Quintana-Murci et al., Ann. Hum. Genet. 63, 153 (1999).
29. D. B. Goldstein et al., Mol. Biol. Evol. 13, 1213 (1996). Published erratum appeared in Mol. Biol. Evol. 14, 354 (1997).
30. R. Chakraborty, M. Kimmel, D. N. Stivers, L. J. Davison, R. Deka, Proc. Natl. Acad. Sci. U.S.A. 94, 1041 (1997).
31. F. R. Santos et al., Hum. Mol. Genet. 9, 421 (2000).
32. M. F. Hammer, Mol. Biol. Evol. 11, 749 (1994).
33. A. W. Bergen et al., Ann. Hum. Genet. 63, 63 (1999).
34. L. S. Whitfield, J. E. Sulston, P. N. Goodfellow, Nature 378, 379 (1995).
35. We thank all the men who donated DNA, K. Kyriakou for the Syrian samples, H. Cann for the French samples, A. Piazza for some of the Italian samples, and G. Brumat for helping us in some blood sample collections. We are also grateful to the anonymous reviewers for their constructive criticisms. Supported by NIH grants GM 28428 and GM 55273 to L.L.C.-S. and by funds from the Italian Ministry of the University "Progetti di ricerca ad interesse Nazionale" and PF "Beni culturali" to A.S.S.-B.

5 April 2000; accepted 25 September 2000

Transgene and Transposon Silencing in *Chlamydomonas reinhardtii* by a DEAH-Box RNA Helicase

Dancia Wu-Scharf, Byeong-ryool Jeong, Chaomei Zhang, Heriberto Cerutti*

The molecular mechanism(s) responsible for posttranscriptional gene silencing and RNA interference remain poorly understood. We have cloned a gene (*Mut6*) from the unicellular green alga *Chlamydomonas reinhardtii* that is required for the silencing of a transgene and two transposon families. *Mut6* encodes a protein that is highly homologous to RNA helicases of the DEAH-box family. This protein is necessary for the degradation of certain aberrant RNAs, such as improperly processed transcripts, which are often produced by transposons and some transgenes.

In plants (1–5) and some fungi (6, 7), the overexpression or misexpression of transgenes can induce the silencing of homologous sequences (i.e., transgene and endogenous gene sequences) by a process known as posttranscriptional gene silencing (PTGS). The introduction of double-stranded RNA (dsRNA) triggers a similar phenomenon, called RNA interference (RNAi), in a variety of invertebrate and vertebrate species (7–15). The widespread occurrence of these phenomena in eukaryotes suggests the involvement of one or more ances-

tral mechanisms that may have evolved to limit the expression of parasitic elements, such as transposons and viruses (3–8, 10). Although several models have been proposed to explain PTGS and RNAi (1, 3, 5, 7, 8, 10, 13, 15), and some genes that are essential for these processes have been identified (4–6, 8–11), the molecular machinery responsible for RNA recognition and degradation remains largely unexplored.

To gain insight into the molecular mechanism(s) of PTGS, we have screened for mutants defective in this process in *Chlamydomonas reinhardtii*. In *Chlamydomonas*, the *RbcS2::aadA::RbcS2* transgene is silenced at both transcriptional and posttranscriptional levels (16, 17). This transgene is composed of the coding sequence of the eubacterial *aadA* gene (conferring spectinomycin resis-

School of Biological Sciences and Plant Science Initiative, University of Nebraska–Lincoln, E211 Beadle Center, Post Office Box 880666, Lincoln, NE 68588, USA.

*To whom correspondence should be addressed. E-mail: hcerutti1@unl.edu