view, in favor of a model based on trade and cultural diffusion (Dennell 1983; Barker 1985; Whittle 1996), which would have left the gene pool of prehistoric Europe essentially autochthonous. There is clearly a spectrum of possibilities between these two extremes, including demic diffusion involving a substantial minority of newcomers, perhaps practicing hypergamy (Cavalli-Sforza and Minch 1997), and pioneer colonization involving fewer newcomers and a more substantial contribution from the indigenous Mesolithic population (Zvelebil 1986, 1989; Sherratt 1994).

The study of the geographic distribution and diversity of genetic variation, known as the "phylogeographic approach" (Avise et al. 1987; Templeton et al. 1995), is emerging as a useful tool for the investigation of range expansions, migrations, and other forms of gene flow during prehistory. It is particularly suited to the study of nonrecombining-marker systems such as mtDNA, which is inherited down the female line and evolves rapidly, so that, provided that sufficient characters are assayed, the maternal genealogy can be well resolved. European mtDNAs fall into a number of distinct clusters, or haplogroups (Torroni et al. 1994*a*, 1996; Richards et al. 1998*a;* Macaulay et al. 1999). Most of these clusters are clades defined by particular control-region and/or coding-region motifs, although recurrent mutation, especially in the control region, can sometimes erase diagnostic elements of these motifs. The major clades are H–K, T, U3–U5, and V–X. As has been argued elsewhere (Richards et al. 1998*a;* Macaulay et al. 1999), the RFLP haplogroup U (Torroni et al. 1996) subsumes both haplogroup K and a number of other clusters (U1–U6), including several (U1, U2, and U6) found rarely in Europe but more frequently in the Near East and northern Africa (Macaulay et al. 1999). In addition, lineages are occasionally seen in Europe that belong to clusters more commonly found elsewhere, such as members of haplogroups M from eastern Eurasia (Ballinger et al. 1992; Passarino et al. 1992, 1996; Torroni et al. 1994*b*) or L1 and L2 from Africa (Chen et al. 1995; Watson et al. 1997).

In previous work, it was suggested that much of the extant European mtDNA lineages have their ancestry in Late Glacial expansions within Europe (Richards et al. 1996; Torroni et al. 1998), with only ~10% dating to the earliest Upper Palaeolithic settlement of the continent (Richards et al. 1998*a*) and with ~≤20% dating to fresh immigrations during the early Neolithic. However, these estimates depend on a reliable determination of founder sequence types, since the undetected presence of ancestral heterogeneity in a colonizing population would result in an overestimation of the age. If this were the case, Europe could have been populated far more recently—for example, during the Neolithic—by a much more diverse founder population (Barbujani et al. 1998). A limitation of the initial analysis (Richards et al. 1996) was that it was based on a very small set of published Near Eastern sequences—42 from the Levant and the Arabian peninsula, mainly from the Bedouin (Di Rienzo and Wilson 1991; Richards and Sykes 1998). Although these sequences were difficult to assign, with certainty, to mtDNA clusters, since they encompassed only the first hypervariable segment (HVS-I) of the control region, they appeared to comprise mainly clusters J and T, pre-HV, a few sequences belonging to X, M, and L1/L2, and some probably belonging to cluster U. There were very few or no members of the major European cluster H, which occurs at a frequency of 40%–60% in most European populations, and there were no representatives of either its sister cluster V or of clusters I, W, K, or U5.

More data from the Near East have been published since this initial analysis, suggesting that the Bedouin may be unrepresentative of Near Eastern populations. Both Calafell et al. (1996) and Comas et al. (1996) have presented data from Turkey. These data suggest the presence of substantial frequencies of cluster H (although lower than that in Europe) (Torroni et al. 1998), as well as of I, W, K, and U4, in addition to clusters already identified in the Bedouin. Torroni et al. (1998) also analyzed a sample of Druze from Israel, using high-resolution RFLPs, and concluded that haplogroup H, but not haplogroup V, evolved first in the Near East and subsequently migrated into Europe. Here, we extend the Near Eastern database further, to a total of 1,234 individuals sampled from throughout the region, including ~500 from the vicinity of the Fertile Crescent, where agriculture emerged from the increasingly sedentary Natufian populations at the end of the last Ice Age (Henry 1989).

We have formalized the procedure for founder analysis, investigated the extent of confounding recurrent gene flow between the putative source and derived populations, and developed criteria that take into account the effects of both gene flow and recurrent mutation. This has enabled us to provide an estimate of the contribution, to the present-day mtDNA pool, of immigration events at different times during Europe's past.

Although previous genetic studies, using classical markers, have inferred a demic component to the spread of agriculture into Europe from the Near East (Menozzi et al. 1978; Ammerman and Cavalli-Sforza 1984; Sokal et al. 1991), the present study allows us for the first time to quantify that component realistically—at least for maternal lineages. Furthermore, the founder analysis using mtDNA allows us to trace lineages farther back into prehistory, through the Last Glacial Maximum (LGM), to the first settlement of Europe by anatomically modern humans, almost 50,000 YBP.

Subjects and Methods

Subjects

For the purposes of this analysis, the Near East was taken to include the whole of Turkey, the Fertile Crescent from Israel to western Iran, and the whole of the Arabian peninsula (see Kuhrt 1995, p. 1). The lower Nile (Egypt and northern Sudan) was also included, since this region is often treated historically with the Near East and since the HVS-I sequence data show that a large proportion of typically Near Eastern mtDNAs have penetrated the Nile Valley, where they coexist with sub-Saharan African mtDNAs (Krings et al. 1999). We sampled widely in the Near East, for several reasons. First, we wished to trace the ancestry of European lineages as far back as 50,000 YBP. We therefore needed as wide a source-population database as possible. Second, even though there is a par-

mtDNA phylogeny (Macaulay et al. 1999), being either pre-HV or pre-JT. All samples bore the 11719G (+11718*Hae*III) mutation that is characteristic of HV (Saillard et al., in press), whereas none of them bore the 11251G (–11251*Tsp*509I) mutation that is characteristic of JT (Hofmann et al. 1997; Macaulay et al. 1999). Thus, these mtDNAs were shown to constitute an early branch in the pre-HV cluster.

mtDNA Classification

The mtDNA nomenclature has been described in detail by Richards et al. (1998*a*) and Macaulay et al. (1999). In brief, named clades of the phylogeny typically either refer to early branchings or are distinguished by an interesting geographic distribution. Major clades, by tradition called "haplogroups," are denoted in terms of uppercase roman letters (e.g., H, J, etc.), and nested subclades are denoted by alternating positive integers and lowercase roman letters (e.g., J2, J1a, J1b1, etc.). Superclades, if not denoted by a single letter (e.g., M or N), are denoted by concatenating clade names (e.g., HV)

types, the chance that back-migration or recurrent mutation will be detected is lower—and, for common types, it is higher—so that the criteria might be both too stringent for rare clusters and too weak for common clusters. We therefore introduced an alternative to *f2*, referred to as "*fs*," in which the frequency of the cluster deriving from each candidate founder in Europe was used to scale the number of derivatives required in the Near East in order for the candidate to be counted as a founder type. To this end, we rescaled the (absolute) frequency of founder candidate clusters in Europe by taking logarithms to the base 10, rounding to the nearest integer, and then adding 1, allowing the outcome to be 1–4. This outcome was then used to designate the number of derivatives required in order for the candidate to qualify as a founder. In addition, to investigate the effect of sample size and differential back-migration into the more peripheral Near Eastern populations, we reapplied the *fs* criterion, excluding these populations (a procedure referred to as the "*fs′*" analysis).

Frequency Estimates

We estimated the posterior distribution of the proportion of a group of lineages in the population, given the sample, by using a binomial likelihood and a uniform prior on the population proportion. From this posterior distribution, we calculated a central 95% "credible region" (CR) (Berger 1985).

Dating and Age Classes

Having identified a list of founder types corresponding to each of these criteria, we measured the diversity in the clusters to which they have given rise within Europe, using the statistic ρ, the mean number of transitions from the founder sequence type to the lineages in the cluster (Forster et al. 1996). This is an unbiased estimate of the time to the most common ancestor of the cluster (TMRCA), measured in mutational units. This value was converted to an age estimate, by use of a mutation rate of 1 transition (between nucleotide positions 16090 and 16365) per 20,180 years (Forster et al. 1996), which closely approximates other rates used for HVS-I (Ward et al. 1991; Macaulay et al. 1997). If the underlying genealogy of a cluster is starlike, we can readily calculate the posterior distribution of its TMRCA, given the sequence type of the ancestor, assuming a uniform prior distribution for the TMRCA and a Poisson distribution for the mutational process. From the (gamma-distributed) posterior, we calculated a central 95% CR. We did this for all clusters, regardless of whether their phylogeny was starlike. When the phylogeny is markedly non-starlike, this is highlighted, since this method is expected to underestimate considerably the width of the true CR.

We employed two simple Procrustean models of demographic prehistory to partition the founder clusters, under each criterion, into migration events. The first, or "basic," model assumes four major prehistoric migrations from the Near East to Europe: (i) early Upper Palaeolithic (EUP), 45,000 YBP; (ii) middle Upper Palaeolithic (MUP), 26,000 YBP; (iii) late Upper Palaeolithic (LUP), 14,500 YBP; and (iv) Neolithic, 9,000 YBP. It also employed a fifth class, at 3,000 YBP, in order to distinguish Neolithic from more-recent migration events.

These age classes were chosen by combining archaeological and paleo-climatological information (e.g., see Dansgaard et al. 1993; Strauss 1995) with an eyeballing of the ages of the more common founder clusters (see table 3 and fig. 1). These clusters appeared to fall roughly into at least three age classes, roughly corresponding to the beginning of the EUP, the LUP, and the Neolithic, with some clusters falling broadly between the LUP and the EUP. The MUP date of 26,000 YBP was chosen to allow immigrants arriving during ~30,000–20,000 YBP to register, and it also corresponds to a slight climatic improvement. The EUP and LUP dates also correspond to more-substantial climatic ameliorations, especially the LUP dates, which are based on the rapid onset of the Bølling warm phase (Dansgaard et al. 1993).

With this latter point in mind, we also considered an "extended" model, which included a Mesolithic component during the dramatic rewarming following the Younger Dryas glacial interlude at 11,500 YBP (Dansgaard et al. 1993). This was stimulated by the suggestion, by Adams and Otte (1999), that recovery from this brief cold period, like that from the LGM, may have led to renewed population dispersals in Europe, possibly including some from Near Eastern refugia. Mesolithic events would, of course, be difficult to distinguish from both LUP and Neolithic expansions, but the possibility of a Mesolithic contribution should nevertheless be borne in mind.

Our partition analysis involves making the following assumptions: (i) each cluster can be assigned, in its entirety, to one of the proposed migration phases; (ii) each cluster expanded in Europe, immediately after the migration event, so that, as a result, the genealogy of each founder cluster is starlike, with a time depth closely approximating the time of the migration event; (iii) the mutation-rate estimate is accurate; (iv) the phylogenetic analysis has resolved all mutations; and (v) the founder analysis has correctly identified the sequence types of the founders. We determined the probabilities that each founder cluster took part in each of the migration events, on the basis of the age of the cluster. Then, given the proportion of the modern sample contained in each cluster, we estimated the proportion of the sample (and, by implication, the modern population of Europe) that is derived from each migration event. In detail, the migra-

tion-event times, t_m ($1 \leqslant m \leqslant M$, where $M = 5$ for the basic model and $M = 6$ for the extended model), were first scaled by the mutation rate μ; that is, $\tau_m = \mu t_m$. If we were to know from which event a cluster derived,

Table 1

Estimated Frequencies and Ages of Major Haplogroups and Their Major Subclusters, in the Near East and Europe

HAPLOGROUP

are markedly non-starlike, evidently displaying drift onto rare sequence types, often near the tips of the phylogenies. Although the Caucasian data are therefore difficult to interpret, the presence there of cluster distributions that are similar to those of Europe and the Near East should caution us that both Europe and the Near East could have been populated from a third region, perhaps closer to either the extant Caucasian population or other populations in eastern Europe. More-recent incursions from eastern Europe, particularly during the Bronze Age, are also likely to have taken place.

Most of the major western-Eurasian clades (Macaulay et al. 1999, table 2) occur in the Near East at a frequency of ⩾1%. In addition to these, we here define U7 (HVS-I motif 16318T [Kivisild et al. 1999*a*]), HV1 (HVS-I motif 16067), and a clade in pre-HV (HVS-I motif 16126-16362). We subdivide the haplogroup N defined by 10873T (+10871*Mnl*I) (Quintana-Murci et al. 1999), which encompasses almost all Eurasian mtDNAs (including haplogroups A, B, F, H–J, K, R, and T–Y) that do not fall into haplogroup M. A subcluster N1, characterized by 10238C (+10237*Hph*I), can be identified (Kivisild et al. 1999*b*) that includes haplogroup I and that has distinct subclusters: N1a (tentative HVS-I motif 16147A/G–16172-16223-16248-16355), N1b (probable HVS-I motif 16145-16176G-16223), and N1c (probable HVS-I motif 16223-16265). Another N subcluster with HVS-I motif 16223-16257A-16261 has a predominantly eastern-Eurasian distribution. HV1, the specific clade of pre-HV, N1a–c, and U7 all occur at low frequency in the northern-Caucasian sample. If we enumerate named subclusters of mtDNA clades in the Near East, Europe, and the Caucasus, we also find more in the Near East than in either of the other two regions, again supporting a Near Eastern origin for the main clusters.

The principal exception is cluster V, which seems to have expanded within Europe ~13,000 YBP (Torroni et al. 1998). Cluster U5 is an additional unusual case. Although U5 occurs at ~2% in the Near East, its phylogeography, as we discuss below, suggests that it evolved mainly within Europe during the past ~50,000 years. Haplogroups V and U5 occur in the Near East at ~11% and ~19%, respectively, of their European frequencies, in most cases as occasional haplotypes that are derived from European lineages. These can be regarded as "erratics," in the same way that African and eastern-Eurasian types can be regarded as such in Europe.

Cluster H is the most frequent cluster in the Near East, as it is in Europe; nevertheless, it is present at a frequency of only 25% (95% CR = .222–.270) in the

Table 2

Founder Status of All Founder Candidate Sequence Types, under Five Different Criteria

Founder Status, for Criterion[a]					Haplogroup or Subcluster	HVS-I Sequence Type[b]
f0	*f1*	*f2*	*fs*	*fs′*		
x	x	x	x	x	H	0
x	x		x	x	H	16092
x					H	16092-16311
x					H	16111
x					H	16129
x					H	16172
x					H	16189
x					H	16189-16311
x					H	16212
x						

Table 2 (continued)

Founder Status, for Criterion[a]					Haplogroup or Subcluster	HVS-I Sequence Type[b]
f0	*f1*	*f2*	*fs*	*fs′*		
x	x		x		H	16150-16192
x	x				H	16248
x	x		x	x	H	16356
x	x	x	x	x	H	16266
x	x		x	x	H	16266-16311
x					H	16266-16311-16362
x	x				H	16145
x					H	16188G
x	x	x	x	x	H	16288-16362
x	x	x	x	x	H	16357
x	x	x	x	x	HV	0
x	x		x	x	HV	16129
x					HV	16129-16221
x					HV	16172-16311
x	x				HV	16221
x	x				HV	16311
x	x		x	x	HV	16362
x	x	x	x	x	HV1	16067
x	x		x		HV1	16067-16311
x	x	x	x		HV1	16067-16355
x	x	x	x	x	I	16129-16223
x					I	16129-16172-16223-16311
x	x				I	16129-16223-16311
x	x		x	x	I	16129-16223-16311-16362
x	x		x	x	I	16223-16311
x	x	x	x	x	I	16129-16148-16223
x	x		x	x	I	16129-16223-16311-16319
x	x		x	x	I	16129-16223-16270-16311-16319-16362
x	x	x	x	x	J	16069-16126
x					J	16069-16126-16145
x					J	16069-16126-16148
x					J	16069-16126-16189
x					J	16069-16126-16193-16256-16300-16309
x					J	16069-16126-16241
x	x		x	x	J	16069-16126-16300
x					J	16069-16126-16311
x					J	16069-16126-16319
x					J1	16069-16126-16145-16172-16261
x	x	x	x	x	J1	16069-16126-16145-16261
x	x	x	x	x	J1	16069-16126-16261
x					J1a	16069-16126-16145-16189-16231-16261
x	x	x	x	x	J1a	16069-16126-16145-16231-16261
x	x	x	x	x	J1b	16069-16126-16145-16222-16261
x	x		x	x	J1b	16069-16126-16145-16222-16261-16274
x					J1b1	16069-16126-16145-16172-16222-16261
x	x	x	x	x	J2	16069-16126-16193
x					J2	16069-16126-16193-16278
x					J2	16069-16126-16193-16311
x					K	16093-16224-16234-16311
x			x	x	K	16189-16224-16311
x	x				K	16224-16234-16311
x	x		x	x	K	16224-16291-16311
x	x		x		K	16224-16304-16311
x	x	x	x	x	K	16224-16311
x			x		K	16093-16189-16224-16311
x					K	16093-16224
x					K	16093-16224-16278-16311
x	x	x	x	x	K	16093-16224-16311
				x	N	16223

(*continued*)

Table 2 (continued)

Founder Status, for Criterion[a]					Haplogroup or Subcluster	HVS-I Sequence Type[b]
f0	*f1*	*f2*	*fs*	*fs′*		
x	x				N1a	16147A-16172-16223-16248-16320-16355
x	x	x	x		N1a	16147A-16172-16223-16248-16355
x	x		x		N1b	16126-16145-16176G-16223
x	x		x		N1b	16128G-16145-16176G-16223-16311
x	x	x	x	x	N1b	16145-16176G-16223
x	x		x		N1b	16145-16176G-16223-16256
x	x	x	x	x	N1c	16201-16223-16265
x	x		x	x	pre-HV	16114-16126-16362
x					pre-HV	16126
x	x		x		pre-HV	16126-16264-16362
x	x	x	x	x	pre-HV	16126-16355-16362
x	x	x	x	x	pre-HV	16126-16362
x	x				T	16126-16189-16294-(16296)
x					T	16126-16189-16294-(16296)-16298
x					T	16126-16192-16294-(16296)
x	x	x	x	x	T	16126-16291-16294-(16296)
x	x	x	x	x	T	16126-16294-(16296)
x	x		x		T	16126-16294-(16296)-16311
x	x		x	x	T	16126-16294-(16296)-16362
x	x	x	x	x	T1	16126-16163-16186-16189-16294
x					T2	16093-16126-16294-(16296)-16304
x	x	x	x	x	T2	16126-16294-(16296)-16304
x	x	x	x	x	T3	16126-16146-16292-16294-(16296)
x	x		x	x	T3	16126-16189-16292-16294-(16296)
x	x				T3	16126-16292-16294-(16296)
x	x	x	x		T4	16126-16294-(16296)-16324
x	x	x	x	x	T5	16126-16153-16294-(16296)
x	x	x	x	x	U	0
x	x		x		U	16129-16189-16234
x	x		x		U	16189
		x			U	16189-16234
x	x		x		U	16189-16234-16324
x	x		x	x	U	16189-16362
x	x		x		U	16311
x	x		x	x	U	16362
x					U1a	16092-16189-16249
x	x		x	x	U1a	16129-16189-16249-16288
x					U1a	16129-16189-16249-16288-16362
x	x	x	x	x	U1a	16189-16249
		x			U1a	16189-16249-16288
x					U1a	16189-16249-16362
x	x	x	x	x	U1b	16111-16249-16327
x					U2	16129C-16189-16256
x	x		x	x	U2	16129C-16189-16256-16362
x	x	x	x	x	U2	16129C-16189-16362
x	x		x	x	U3	16093-16343
x	x		x		U3	16189-16343
x					U3	16193-16249-16343
x	x		x	x	U3	16260-16343
x	x		x	x	U3	16301-16343
x					U3	16311-16343
x	x	x	x	x	U3	16343
x	x	x	x	x	U3	16168-16343
x					U4	16179-16356
x					U4	16223-16356
x	x	x	x		U4	16356
x	x				U4	16356-16362

(*continued*)

Table 2 (continued)

Founder Status, for Criterion[a]					Haplogroup	
f0	*f1*	*f2*	*fs*	*fs′*	or Subcluster	HVS-I Sequence Type[b]
x	x				U4	16134-16356
x	x	x			U5	16270
x	x		x		U5	16270-16296
x	x		x		U5a	16187-16192-16270
x					U5a1	16093-16192-16256-16270-16291
x					U5a1	16189-16192-16256-16270
x	x	x	x		U5a1	16192-16256-16270
x	x	x	x		U5a1	16192-16256-16270-16291
x					U5a1	16192-16256-16270-16311
x	x		x		U5a1a	16189-16256-16270
x	x	x	x		U5a1a	16256-16270
x					U5a1a	16256-16270-16295
x	x				U5b	16189-16270
x					U5b	16189-16270-16311
x	x				U5b1	16144-16189-16270
x	x	x	x	x	U7	16318T
x	x	x	x	x	U7	16309-16318C
x	x	x	x	x	U7	16309-16318T
x					V	16216-16261-16298
x					V	16239-16298
x	x		x		V	16274-16298
x	x	x			V	16298
x					V	16298-16311
x					W	16093-16223-16292
x					W	16129-16223-16292
x					W	16172-16223-16231-16292
x	x	x	x		W	16223-16292
x	x	x	x		W	16223-16292-16295
x	x				W	16192-16223-16292-16325
		x	x		W	16223-16292-16325
x					X	16093-16189-16223-16278
x					X	16126-16189-16223-16278
x	x		x		X	16189-16223-16248-16278
x	x		x		X	16189-16223-16265-16278
x	x		x		X	16189-16223-16274-16278
x	x	x	x	x	X	16189-16223-16278
x					X	16189-16223-16278-16293
x	x		x	x	X	16189-16223-16278-16344
x					X	16189-16278

[a] Founder clusters are indicated by bullets (x).

[b] Parentheses denote that, as discussed in the Subjects and Methods section, nucleotide position 16296 in haplogroup T is unstable, making the state of this position in the founder sequence uncertain. Sequence types that are formally founders but whose clusters are empty in the European sample are not included.

Back-Migration from Europe

Recent back-migration can be estimated by an examination of the presence, in the Near East, of clusters that are most likely to have evolved within Europe. Haplogroup U5 is very mry mry

YBP, with recurrent back-migration ever since, a European origin for the U5 cluster seems just as probable. In either case, the U5 cluster itself would have evolved essentially in Europe. U5 lineages, although rare elsewhere in the Near East, are especially concentrated in the Kurds, Armenians, and Azeris. This may be a hint of a partial European ancestry for these populations—not entirely unexpected on historical and linguistic grounds—but may simply reflect their proximity to the Caucasus and the steppes. Of the Near Eastern lineages, 1.8% (95% CR = .012–.027) are members of U5, in contrast to 9.1% (95% CR = .081–.103) in Europe; in the core region of Syria-Palestine through Iraq, the proportion falls to 0.5% (95% CR = .002–.015). Overall, this suggests the presence of as much as 20% of back-migrated mtDNA in the Near East but only ~6% in the core region.

It seems likely that haplogroup V also originated within Europe and subsequently spread eastward (Torroni et al. 1998), although its lower diversity provides less opportunity to differentiate lineages by their ages. A slightly lower figure for back-migration is obtained when V is used: 0.5% (95% CR = .002–.011) of samples in the Near East (in Turkey, Azerbaijan, and Syria) versus 4.6% (95% CR =

Table 3

Ages of the Major Founder Clusters Identified under Four Different Criteria

Haplogroup	A

Table 4

Percentage, of Extant European mtDNA Pool, Derived, in Each Migration Event, from Near Eastern Founder Lineages

Migration Event	Mean	Root-Mean-Square Error, of Contribution, for
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

Table 5

Percentage, of Extant European mtDNA Pool, in Each Migration Event, by Region

	MEAN 5 ROOT-MEAN-SQUARE ERROR, FOR									
	Southeastern Europe (n = 166)	Eastern Mediterranean (n = 233)	Central Mediterranean (n = 302)	Alps (n = 218)	North-Central Europe (n = 332)	Western Mediterranean (n = 217)	Basque Country (n = 156)	Northwestern Europe (n = 456)	Scandinavia (n = 316)	Northeastern Europe (n = 407)
Migration Event:										
Bronze Age/recent	8.2 5 3.3	19.5 5 3.7	4.6 5 1.6	6.9 5 2.7	8.9 5 3.5	6.3 5 2.5	5.4 5 2.6	4.6 5 1.5	7.4 5 3.3	5.5 5 3.0
Neolithic	19.7 5 6.0	10.7 5 5.0	9.2 5 5.1	15.1 5 5.4	17.1 5 5.6	12.0 5 4.2	6.7 5 4.7	21.7 5 4.5	11.7 5 4.5	18.0 5 4.5
LUP	44.1 5 6.5	43.0 5 4.9	49.1 5 5.8	54.4 5 5.8	52.2 5 12.1	56.3 5 5.0	58.8 5 4.7	52.7 5 4.5	52.8 5 4.7	43.3 5 4.4
MUP	14.6 5 4.9	13.2 5 4.7	21.1 5 4.4	14.7 5 4.8	11.8 5 11.7	14.4 5 4.8	13.0 5 3.4	12.2 5		

America also has begun (Ruiz Linares et al. 1996; Karafet et al. 1999). Both situations are thought to be likely to be amenable to such an analysis, in that they have relatively well-defined source regions, only one major dispersal event, and probably minimal postsettlement gene exchange with those regions (although they undoubtedly are more complex than usually is supposed; see Terrell 1986).

Europe clearly presents a more difficult case. The time depth is such that it is unclear whether the Near East represents a suitable source population stretching back prior to the LGM. Settlement seems likely to have occurred in multiple waves from the east and to have been subsequently obscured by millennia of recurrent gene flow. There may well have been significant levels of gene flow throughout Eurasia, from the Upper Palaeolithic to the present, particularly during the Holocene (the "Holocene filter"), which would obscure the signals of earlier dispersals. The problem is particularly acute for the Near East, since the latter forms the junction between three continents. Therefore, it is important to take into account recurrent gene flow when a founder analysis of Europe is performed.

Sample size may also be an issue. Despite a source-population sample ($n = 1{,}234$) much larger than has been used in all previous studies, there are reasons to be cautious. Both the higher diversity and degree of substructure in the Near East, in comparison with Europe, and the greater number of potential founder lineages raise the possibility that some founders may be missed in the sampling.

Our aim was to identify the principal founder lineages that have entered Europe and to date the times of their entry, in order to quantify the contribution that the main episodes of new settlement during European prehistory have made to the modern mtDNA pool. As regularly has been pointed out (e.g., see Barbujani et al. 1998), the divergence time estimated on the basis of the genetic diversity *of the population as a whole* will not, in general, indicate the time of settlement. This is because some of the preexisting diversity of the source population is expected to be carried into the derived population, so that some of the earlier branches in the genealogy will

3. A further assumption for the founder methodology is the infinite-alleles model—that is, that recur-

[redacted]

[redacted]

forwarding the Nile Valley sequences. This research received support from The Wellcome Trust and from Italian Consiglio Nazionale delle Ricerche grant 99.02620.CT04 (to A.T.), by the Faculty of Biological Sciences, University of Urbino (support to A.T.), by the Ministero dell'Università e Ricerca Scientifica e Tecnologica: Grandi Progetti Ateneo (support to R.S.), by the Progetti Ricerca Interesse Nazionale 1999 (support to R.S. and S.S.-B.), and by the Progetto Finalizzato C.N.R. "Beni Culturali" (Cultural Heritage, Italy). N.A.-Z. was supported by an International Center for Genetic Engineering and Biotechnology fellowship. The Romanian contribution to this work was supported by the EC Commission, Directorate General XII (50%), and by the Romanian Ministry of Research and Technology (50%), within the framework of the Cooperation in Science and Technology with Central and Eastern European Countries (supplementary agreement ERBCIPDCT 940038 to the contract ERBCHRXCT 920032 coordinated by Prof. Alberto Piazza, University of Turin, Turin). M.R. was supported by a Wellcome Trust Bioarchaeology fellowship, V.M. was supported by a Wellcome Trust Research Career Development fellowship, and H.-J.B. was supported by a Deutscher Akademischer Austauschdienst travel grant.

Electronic-Database Information

The URL for data in this article is as follows:

Authors' data available online, http://www.stats.ox.ac.uk/~macaulay/founder2000/index.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) The Neolithic transition and the genetics of populations in Europe. Princeton University Press, Princeton, NJ

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465

Avise J, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC (1987) Intraspecific phylogeography: the molecular bridge between population genetics and systematics. Annu Rev Ecol Syst 18:489–522

Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, Hassan K, Chen K-H, Wallace DC (1992) Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. Genetics 130:139–152

Bandelt H-J, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. Genetics 141:743–753

Barbujani G, Bertorelle G, Chikhi L (1998) Evidence for Paleolithic and Neolithic gene flow in Europe. Am J Hum Genet 62:488–491

Barker G (1985) Prehistoric farming in Europe. Cambridge University Press, Cambridge

Bar-Yosef O (1998) On the nature of transitions: the Middle to Upper Palaeolithic and the Neolithic Revolution. Camb Archeol J 8:141–163

AJH,ou AJH,salun-33(ou-433(R7(SWntr7(oeva33(R,)D433(R,)K0.ydjieva3339(

Diamond JM (1997) The language steamrollers. Nature 389: 544–546

Di Rienzo A, Wilson AC (1991) Branching pattern in the evolutionary tree for human mitochondrial DNA. Proc Natl Acad Sci USA 88:1597–1601

Dolukhanov P (1994) Environment and ethnicity in the ancient Middle East. Avebury, Aldershot

Finnilä, S, Hassinen IE, Ala-Kokko L, Majamaa K (2000) Phylogenetic network of the mtDNA haplogroup U in northern Finland based on sequence analysis of the complete coding region by conformative-sensitive gel electrophoresis. Am J Hum Genet 66:1017–1026

Forster P, Harding R, Torroni A, Bandelt H-J (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. Am J Hum Genet 59:935–945

Francalacci P, Bertranpetit J, Calafell F, Underhill PA (1996) Sequence diversity of the control region of mitochondrial DNA in Tuscany and its implications for the peopling of Europe. Am J Phys Anthropol 100:443–460

Gamble C (1986) The Palaeolithic settlement of Europe. Cambridge University Press, Cambridge

——— (1999) The Palaeolithic societies of Europe. Cambridge University Press, Cambridge

Gilead I (1991) The Upper Palaeolithic Period in the Levant. J World Prehist 5:105–154

Harris DR (1996) The origins and spread of agriculture and pastoralism in Eurasia. University College London Press, London

Hasegawa M, Di Rienzo A, Kocher TD, Wilson AC (1993) Toward a more accurate time scale for the human mitochondrial DNA tree. J Mol Evol 37:347–354

Helgason A, Sigurðardóttir S, Gulcher JR, Ward R, Stefánsson K (2000) mtDNA and the origin of the Icelanders: deciphering signals of recent population history. Am J Hum Genet 66:999–1016

Henry DO (1989) From foraging to agriculture. University of Pennsylvania Press, Philadelphia

Hofmann S, Jaksch M, Bezold R, Mertens S, Aholt S, Paprotta A, Gerbitz KD (1997) Population genetics and disease susceptibility: characterization of central European haplogroups by mtDNA gene mutations, correlations with D loop variants and association with disease. Hum Mol Genet 6: 1835–1846

Housley RA, Gamble CS, Street M, Pettitt P (1997) Radiocarbon evidence for the Lateglacial human recolonisaton of northern Europe. Proc Prehist Soc 63:25–54

Jochim M (1987) Late Pleistocene refugia in Europe. In: Soffer O (ed) The Pleistocene Old World. Plenum Press, New York, pp 317–331

Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz K, Harihara S, de Knijff 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

Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, Reidla M, Laos S, Parik J, Watkins WS, Dixon ME, Papiha SS, Mastana SS, Mir MR, Ferak V, Villems R (1999*a*) Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. Curr Biol 9:1331–1334

Kivisild T, Kaldma K, Metspalu M, Parik J, Papiha S, Villems R (1999*b*) The place of the Indian mitochondrial DNA variants in the global network of maternal lineages and the peopling of the Old World. In: Papiha S, Deka R, Chakraborty R (eds) Genomic diversity: applications in human population genetics. Plenum Press, New York, pp 135–152

Krings M, Salem AH, Bauer K, Geisert H, Malek A, Chaix L, Simon C, Welsby D, Di Rienzo A, Utermann G, Sajantila A, Pääbo S, Stoneking M (1999) mtDNA analysis of Nile River Valley populations: a genetic corridor or a barrier to migration? Am J Hum Genet 64:1166–1176

Kuhrt A (1995) The ancient Near East. Vol 1: 3000–330 B.C.

zerland reveal striking homogeneity of European populations. Biol Chem Hoppe Seyler 375:837–840

Quintana-Murci L, Semino O, Bandelt H-J, Passarino G, McElreavey K, Santachiara-Benerecetti AS (1999) Genetic evidence for an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. Nat Genet 23:437–441

Rando JC, Pinto F, González AM, Hernández M, Larruga JM, Cabrera VM, Bandelt H-J (1998) Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-Eastern, and sub-Saharan populations. Ann Hum Genet 62:531–550

Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro ASM, Stoneking M (1995) Evolutionary history of the COII/tRNALys intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. Mol Biol Evol 12:604–615

Redgate AE (1998) The Armenians. Blackwell, Oxford

Renfrew C (1998) The origins of world linguistic diversity: an archaeological perspective. In: Jablonski NG, Aiello LC (eds) The origins and diversification of language. California Academy of Sciences, San Francisco, pp 171–192

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 MB, Macaulay VA, Bandelt H-J, Sykes BC (1998*a*) Phylogeography of mitochondrial DNA in western Europe. Ann Hum Genet 62:241–260

Richards M, Oppenheimer S, Sykes B (1998*b*) mtDNA suggests Polynesian origins in eastern Indonesia. Am J Hum Genet 63:1234–1236

Richards M, Sykes B (1998) Reply to Barbujani et al. Am J Hum Genet 62:491–492

Ruiz Linares A, Nayar K, Goldstein DB, Hebert JM, Seielstad MT, Underhill PA, Lin AA, Feldman MW, Cavalli-Sforza LL (1996) Geographic clustering of human Y-chromosome haplotypes. Ann Hum Genet 60:401–408

Saillard J, Magalhães PJ, Schwartz M, Rosenberg T, Nørby S. Mitochondrial DNA variant 11719G is a marker for the mtDNA haplogroup cluster HV. Hum Biol (in press)

Sajantila A, Lahermo P, Anttinen T, Lukka M, Sistonen P, Savontaus ML, Aula P, Beckman L, Tranebjaerg L, Geddedahl T, Isseltarver L, Dirienzo A, Paabo S (1995) Genes and languages in Europe—an analysis of mitochondrial lineages. Genome Res 5:42–52

Sajantila A, Salem AH, Savolainen P, Bauer K, Gierig C, Pääbo S (1996) Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. Proc Natl Acad Sci USA 93:12035–12039

Salas A, Comas D, Lareu MV, Bertranpetit J, Carracedo A (1998) mtDNA analysis of the Galician population: a genetic edge of European variation. Eur J Hum Genet 6:365–375

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

Sherratt A (1994) The transformation of early agrarian Europe: the later Neolithic and Copper Ages. In: Cunliffe B (ed) The Oxford illustrated prehistory of Europe. Oxford University Press, Oxford

Soffer O (1987) Upper Paleolithic connubia, refugia, and the archaeological record from eastern Europe. In: Soffer O (ed) The Pleistocene Old World. Plenum Press, New York, pp 333–348

Weiss KM, Wallace DC (1992) Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. Genetics 130:153–162

Torroni A, Sukernik RI, Schurr TG, Starikovskaya YB, Cabell MF, Crawford MH, Comuzzie AG, Wallace DC (1993*b*) mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. Am J Hum Genet 53: 591–608

Tubb JN (1998) Canaanites. British Museum Press, London /F14108