

# Estimating the Impact of Prehistoric Admixture on the Genome of Europeans

Isabelle Dupanloup,<sup>\*1</sup> Giorgio Bertorelle,<sup>\*</sup> Lounès Chikhi,<sup>†</sup> and Guido Barbujani<sup>\*</sup>

<sup>\*</sup>Dipartimento di Biologia, Università di Ferrara, Ferrara, Italy; and <sup>†</sup>UMR Evolution et Diversité Biologique, Université Paul Sabatier, Toulouse, France

We inferred past admixture processes in the European population from genetic diversity at eight loci, including autosomal, mitochondrial and Y-linked polymorphisms. Admixture coefficients were estimated from multilocus data, assuming that most current populations can be regarded as the result of a hybridization process among four or less potential parental populations. Two main components are apparent in the Europeans' genome, presumably corresponding to the contributions of the first, Paleolithic Europeans, and of the early, Neolithic farmers dispersing from the Near East. In addition, only a small fraction of the European alleles seems to come from North Africa, and a fourth component reflecting gene flow from Northern Asia is largely restricted to the northeast of the continent. The estimated Near Eastern contribution decreases as one moves from east to west, in agreement with the predictions of a model in which (Neolithic) immigrants from the Near East contributed a large share of the alleles in the genome of current Europeans. Several tests suggest that probable departures from the admixture models, due to factors such as choice of the putative parental populations and more complex demographic scenarios, may have affected our main estimates only to a limited extent.

## Introduction

Recent analyses of human diversity have reached different conclusions on the origin of the European gene pool. There is general agreement that genes coming from anatomically archaic people, the Neandertal people, represent either an extremely small fraction of the contemporary genome or none at all (Krings et al. 2000, Relethford 2001; Caramelli et al. 2003; Stringer 2003). Archaeological data show that the first anatomically modern Europeans entered from the Near East in Paleolithic times, 45,000 years ago or less (Otte 2000), and lived in different parts of the continent until the end of the last glaciation. They also show that another large-scale expansion from the Near East accompanied the spread of the technologies for food production in Neolithic times, between 10,000 and 5,000 years ago (Zvelebil 1986; Pinhasi, Foley, and Mirazòn Lahr 2000). However, the relative contribution of Paleolithic hunting-gathering and Neolithic farming ancestors to the genome of current Europeans cannot be easily inferred or quantified from archaeological data.

Most studies based on nuclear protein (Menozzi, Piazza, and Cavalli-Sforza 1978; Sokal, Oden, and Wilson 1991; Cavalli-Sforza, Menozzi, and Piazza 1993; Barbujani et al. 1994) and DNA (Chikhi et al. 1998; Barbujani and Bertorelle 2001) polymorphisms, including the Y chromosome (Rosser et al. 2000; Chikhi et al. 2002), suggest that the Neolithic spread of farming entailed a large-scale population replacement, also termed demic diffusion (Ammerman and Cavalli-Sforza 1984). The main genetic evidence for Neolithic dispersal from the Near East is represented by the broad genetic gradients affecting many loci over much of Europe (Menozzi, Piazza, and Cavalli-Sforza 1978; Cavalli-Sforza, Menozzi, and Piazza 1993).

On the other hand, 75% or so of the current mitochondrial and Y-chromosome European lineages can be

traced back to ancestral lineages that originated in Paleolithic times. Some interpreted this finding as evidence that 75% of the ancestors of current Europeans were already in Europe in Paleolithic times, before the Neolithic transition (Richards et al. 1996, 2002; Macaulay et al. 1999; Semino et al. 2000; Torroni et al. 2001). A single percentage value for all of Europe is not very informative, because populations are likely to differ in their history and genetic composition, as allele-frequency gradients (Menozzi, Piazza, and Cavalli-Sforza 1978; Sokal, Oden, and Wilson 1991; Cavalli-Sforza, Menozzi, and Piazza 1993; Barbujani et al. 1994) clearly suggest. However, the question remains whether the ancestors of the current Europeans were mostly local Paleolithic hunters and gatherers (hereafter: Paleolithic model, or PM), or mostly Neolithic farmers who dwelt out of Europe until comparatively recent times (hereafter: Neolithic model, or NM).

Aside from its anthropological relevance, this question has implications of obvious medical and epidemiological relevance. Genes that influence multifactorial diseases are expected to be easier to identify in isolated communities, which were only marginally affected by admixture, if at all (Goldstein and Chikhi 2002). In addition, depending on the model of European prehistory that one assumes, the patterns shown by various pathological or disease-resistance alleles may have different explanations. For example, the relative frequencies of the  $\Delta F508$  cystic fibrosis allele decline from northwest to southeast Europe (Estivill, Bancells, and Ramos 1997). Under an NM it is reasonable to regard that gradient as a result of Near Eastern gene flow into a Paleolithic population with higher  $\Delta F508$  allele frequencies. However, if just a few people entered Europe in Neolithic times (as stated by the PM) a continent-wide gradient must reflect processes other than Neolithic gene flow, possibly some form of geographically variable heterozygote advantage (Wiuf 2001). Similar problems exist for many other mutations of clinical relevance.

A common ground for supporters of either model is the recognition that, as a rule, current European populations are hybrids, containing variable proportions of alleles derived from both Paleolithic settlers and Neolithic immigrants. This means that quantifying admixture is

<sup>1</sup> Present address: Center of Integrative Genomics, University of Lausanne, Lausanne, Switzerland.

Key words: human populations, admixture, Europe, Paleolithic, Neolithic.

E-mail: g.barbujani@unife.it.

*Mol. Biol. Evol.* 21(7):1361–1372. 2004

doi:10.1093/molbev/msh135

Advance Access publication March 24, 2004

**Table 1**  
**List of Samples Used to Estimate the Contributions of 4 Parental Populations to European Populations**

Population	<i>mtDNA</i>		<i>YRosser</i>		<i>YSemino</i>	
	N1	N2	N1	N2	N1	N2
Basques	1	106	1	26	2	67
Near East	5	261	1	167	3	81
North Africa	2	179	2	156	4	176
North-Eastern Europe	2	117	2	65	2	89
Balkans	2	78	6	320	4	205
British Isles	3	261	5	643		
CEE	3	389	4	250	2	43
CEW	3	216	2	132	1	23
EE	1	117	8	411	4	195
Finland	2	107	2	264		
IberiaP	1	54	2	385		
IberiaS	3	181	1	126	2	53
ItalyN	1	49	1	72	1	50
ItalyS	2	100	1	99	1	37
Sardinia	1	72	1	10	1	77
Scandinavia	2	62	3	164		
Total	34	2349	42	3290	27	1096

NOTE.—N1: number of samples, N2: number of individuals. Population label definitions and abbreviations: Near East (Iran, Iraq, Israel, Jordan, Lebanon, Saudi Arabia, Syria, Turkey), North Africa (Algeria, Egypt, Libya, Morocco, Tunisia), North-Eastern Europe (Russia), Balkans (Albania, Bosnia, Bulgaria, Croatia, Greece, Macedonia, Slovenia, Rumania, Yugoslavia), British Isles (England, Ireland), CEE: Central Europe East (Denmark, Germany, The Netherlands), CEW: Central Europe West (Belgium, France, Switzerland), EE: Eastern Europe (Austria, Belarus, Czech Republic, Hungary, Poland, Slovakia, Ukraine), Finland (Estonia, Finland), IberiaP: Portugal, IberiaS: Spain, ItalyN: Northern Italy, ItalyS: Southern Italy (except Sardinia), Scandinavia (Norway, Sweden).

a way to see which model, whether the PM or the NM, better describes the composition of the European gene pool. In the only comparable study available to date, Chikhi et al. (2002) estimated that between approximately 70% (in the Balkans) and 30% (in Iberia) of the current Y-chromosome haplotypes should be attributed to (presumably Neolithic) immigration from the Near East. Here we extend the analysis to several genome regions, inherited both uni- and biparentally. Indeed, variation at a single locus reflects a combination of population-specific demographic factors (including drift, gene flow, and admixture) and locus-specific, or even allele-specific (Holtkemper et al. 2001), mutational differences and selective pressures. Unless selection can be ruled out, which is not the case for widely used markers such as mitochondrial DNA (Mishmar et al. 2003) and the Y chromosome (Jobling and Tyler-Smith 2000), the effects of selection and demographic history are difficult to disentangle at the single-locus level (see, e.g., Pääbo 1999, Dupanloup et al. 2003).

Because admixture affects the genome as a whole, in this study we estimated admixture rates in twelve European regions, each representing an aggregate of populations, on the basis of the largest available set of loci that proved suitable for that purpose. The method we used (Dupanloup and Bertorelle 2001) infers admixture coefficients considering several potential parental populations, which also gave us the opportunity to quantify the potential contributions of immigrants from North Africa and Northern Asia.

**Table 2**  
**List of Samples Used to Estimate the Contributions of 2 Parental Populations to European Populations**

	<i>DQalpha</i>		<i>FES/FPS</i>		<i>FXIII</i>		<i>HUMTH01</i>		<i>VWA31A</i>	
	N1 <sup>a</sup>	N2 <sup>b</sup>	N1	N2	N1	N2	N1	N2	N1	N2
Basques	3	1138	2	1054	1	416	2	1018	3	1130
Near East	2	488	2	670	1	254	2	752	3	958
Balkans	4	798	1	470	2	522	2	698	2	880
British Isles	1	402	2	1324	2	1330	3	1734	3	1534
CEE	8	3662	7	5209	4	2486	7	5276	7	4724
CEW	5	1900	3	1518	6	3482	4	1392	3	1940
EE	2	514	9	5110	4	1900	8	3228	9	4688
Finland	1	224	1	508			2	8006	3	748
IberiaP	5	1560	3	1550	4	1846	5	2184	4	2062
IberiaS	7	2224	3	1248	2	684	9	2824	5	1484
ItalyN	13	2442	14	5199			12	2504	17	3956
ItalyS	4	538	4	1116	2	826	3	570	6	1414
Sardinia	1	172	1	200	2	160	1	206	1	172
Scandinavia	3	152	2	1290	3	1980	4	1202	2	1290
Total	59	16214	54	26466	33	15886	64	31594	68	26980

NOTE.—N1: number of samples, N2: number of individuals. Population labels as defined in table 1.

## Materials and Methods

### Data Sets

We searched the available literature for sets of DNA markers that had been typed on a sufficiently large number of populations, covering in sufficient detail the map of Europe. In this way, we selected eight data sets, for which information exists about at least one (but generally, many more) population dwelling in each of 12 arbitrarily defined European regions. Therefore, each such region is an aggregate of geographically close populations and is treated as a hybrid between two, or four, parental populations (see tables 1 and 2 and figs. 1 and 2).

Each of the eight data sets corresponds to a different locus, mitochondrial or nuclear, except for two sets of Y-chromosome polymorphisms. Because the individuals in these two data sets are different, in agreement with Sokal, Oden, and Wilson (1991) we shall use the term *system* to refer to an independent data set. In this way, we shall be analyzing eight genetic systems, representing seven different loci, namely:

- (1) 2,349 sequences of the mitochondrial hypervariable region I (HVR-I) from 34 samples, collected by Simoni et al. (2000), and repeatedly updated (see Vernesi et al. 2002; Caramelli et al. 2003);
- (2) eleven binary markers from the non-recombining region of the Y chromosome (hereafter NRY) in 42 populations for a total of 3,290 individuals, from Rosser et al. (2000);
- (3) 22 binary markers of NRY in 27 populations for a total of 1,096 individuals, from Semino et al. (2000) and (for North Africa) Underhill et al. (2000);
- (4–8) five nuclear DNA loci from Chikhi et al. (1998), updated by a Medline search of the recent literature. Four of the Chikhi et al. (1998) loci are tetranucleotide microsatellites (FES/FPS, FXIII, HUMTH01, VWA31A), whereas DQ $\alpha$  is a highly polymorphic gene coding for the  $\alpha$ -chain of the HLA-DQ molecule. For each of these five markers, the

number of population samples ranged between 33 and 68 with a mean of 55 samples and a total of 278. Overall, 117,140 chromosomes (or 58,570 individuals) were studied, for an average of 427 chromosomes per population. The number of chromosomes analyzed at each locus varied between 15,886 and 31,594.

For each system, we initially tested whether it is legitimate to clump the data from different samples by performing an AMOVA analysis (Excoffier, Smouse, and Quattro 1992). The AMOVA technique allowed us to partition genetic variance in three components, corresponding respectively to differences (1) among individuals within population, (2) among populations within a region, and (3) among regions. The percentage of the global European variation corresponding to population differentiation within regions was always below 5% (range: 0.16%–4.85%), which does not indicate a substantial genetic heterogeneity among the population samples that we clumped.

#### Choice of the Parental Populations

Admixture coefficients estimate the likely components of the contemporary European gene pool contributed by two or more parental populations whose members hybridized at a certain moment in the past. For all the loci of this study we considered possible admixture between two parental populations, namely (1) Neolithic people from the Levant and (2) Paleolithic inhabitants of Europe. Whenever sufficient data were available (i.e., for the mitochondrial and Y-chromosome data sets), we also considered as potential parental groups populations from (3) North Africa and (4) North-Eastern Europe.

A preliminary step of the analysis was to select the modern populations that better represent the genetic characteristics of these parental populations. Archaeological, linguistic, and genetic evidences suggest choices that are largely shared by all previous studies on European genetic diversity (see, e.g., Ammerman and Cavalli-Sforza 1984; Richards et al. 2002).

As a proxy for Neolithic farmers, all studies we are aware of chose populations from the Near East and Anatolia (Menozi, Piazza, and Cavalli-Sforza 1978; Semino et al. 1996, 2000; Richards et al. 2000; Wilson et al. 2001; Chikhi et al. 2002), which is where the first archaeological evidence of farming was found (Renfrew 1987). As for the Paleolithic component of the genome, in principle any population could be used under the PM model, because this model considers current populations as derived, with very few changes, from local Paleolithic ancestors. However, there is a general consensus that the Basques represent the most direct descendants of the hunter-gatherers who dwelt in Europe before the spread of agriculture, based on both linguistic and genetic evidence (Menozi, Piazza, and Cavalli-Sforza 1978; Bertranpetit and Cavalli-Sforza 1991; Cavalli-Sforza and Piazza 1993; Bertranpetit et al. 1995; Semino et al. 2000; Wilson et al. 2001). When sufficient data were available to test for more complex scenarios (i.e., for mitochondrial and Y-chromosome data sets), we also considered present-day North

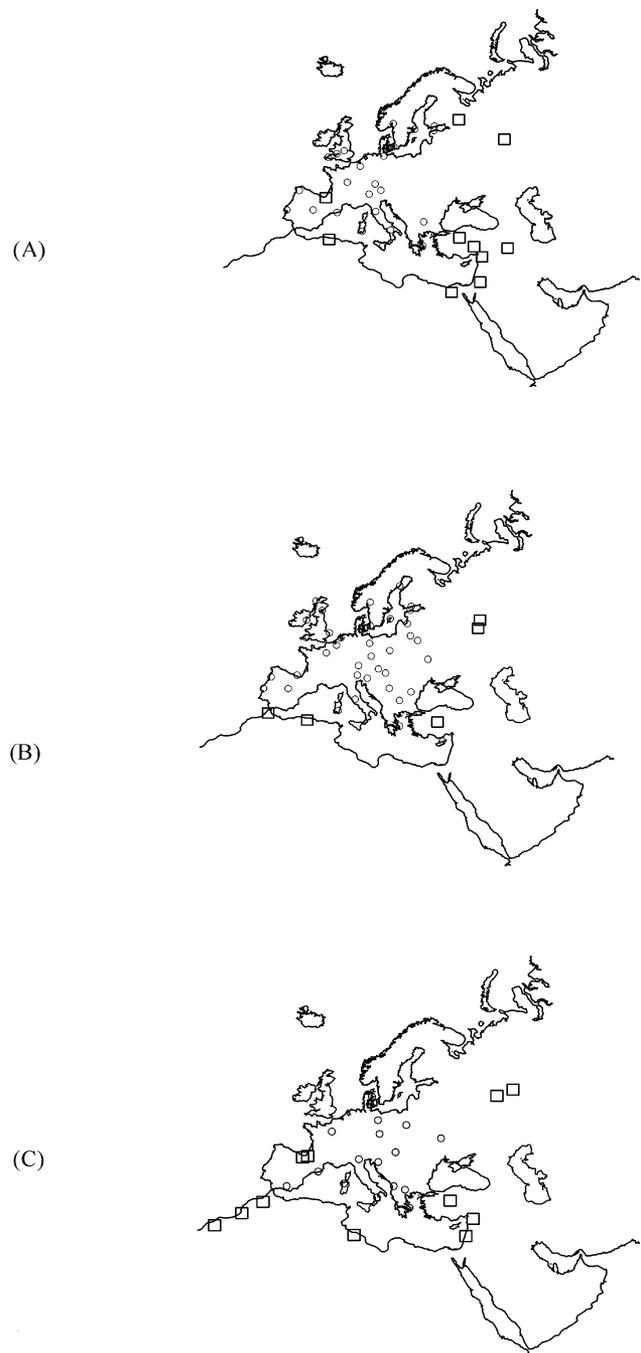


FIG. 1.—(A) Distribution of the 34 populations tested for mtDNA HVRI sequences polymorphisms (Simoni et al. 2000). (B) 42 populations tested for 11 NRY binary markers (Rosser et al. 2000). (C) 27 populations tested for 22 NRY binary markers (Semino et al. 2000). Squares show the location of the parental population samples, and circles show the location of the European samples.

Africans and North-Eastern Europeans as parental populations, thus modeling admixture as a process potentially involving up to four groups. In this way, we looked for the possible genetic consequences of gene flow through the Mediterranean Sea (see, e.g., Rando et al. 1998; Bosch et al. 2001), and from Northern Asia, as suggested, among others, by Rosser et al. (2000).



FIG. 2.—(A) Distribution of the 59 populations tested for DQa and (B) the samples analyzed for one to four tetranucleotide microsatellites polymorphisms.

#### The Admixture Model and Estimation Methodology

Our method allows the estimation of the relative contribution of  $d$  parental populations into a hybrid group, using either allele-frequency differences or both such differences and the degree of molecular divergence between alleles (Bertorelle and Excoffier 1998; Dupanloup and Bertorelle 2001). We consider an ancestral population splitting into  $d$  parental populations (PPs) that evolve independently for  $\tau$  generations. At that point, a hybrid population (HP) is instantaneously created by combining  $d$  fractions, each indicated by  $\mu_i$ , of genes taken at random

from each PP. From that moment on, for  $t_A$  generations, the HP and PPs evolve independently, under random genetic drift.

Under this model, the mean coalescence time between a gene drawn from the HP and a gene drawn from the  $i^{\text{th}}$  PP,  $\bar{t}_{h,i}$ , is simply given by

$$\bar{t}_{h,i} = \mu_i(\bar{t}_{i,i} + t_A) + \sum_{j \neq i} \mu_j \bar{t}_{i,j} \quad (1)$$

where  $\bar{t}_{i,i}$  is the mean coalescence time between two genes sampled in the same PP  $i$ ,  $\bar{t}_{i,j}$  is the mean coalescence time

**Table 3**  
**Weighted Average Across Loci, and Standard Deviations (SD), of the Estimated Contributions of 4 Parental Populations to European Populations**

Population	Basques			Near East			North Africa			North-Eastern Europe		
	Av.	SD	$\chi^2$	Av.	SD	$\chi^2$	Av.	SD	$\chi^2$	Av.	SD	$\chi^2$
Balkans	19.78%	4.42%	2.809	62.84%	7.57%	3.631	9.03%	3.57%	2.900	8.49%	3.23%	9.368*
British Isles <sup>a</sup>	8.08%	3.72%	4.817	69.21%	7.92%	0.889	5.75%	2.94%	0.146	17.70%	4.36%	1.563
	69.74%	7.81%	3.533	42.16%	13.61%	0.020	-8.05%	4.81%	0.013	-11.01%	5.65%	3.229
CEE	70.46%	7.89%	7.628*	21.02%	10.14%	0.002	-1.12%	2.15%	0.359	3.74%	6.90%	6.863*
	53.71%	5.50%	4.977	38.99%	8.39%	1.628	-4.16%	3.35%	0.367	2.17%	3.15%	6.469*
CEW	46.63%	5.41%	10.566*	36.20%	7.19%	1.425	0.65%	1.98%	3.020	13.11%	3.63%	5.499
	61.12%	7.27%	0.363	48.18%	11.31%	0.999	-5.22%	4.32%	0.028	-6.92%	3.69%	1.976
EE	47.66%	6.63%	10.620**	43.27%	8.72%	0.379	1.10%	2.10%	0.908	1.71%	3.78%	7.070*
	40.42%	4.65%	7.697*	-0.27%	9.73%	10.572*	8.41%	3.81%	7.775*	49.85%	6.28%	1.700
Finland <sup>d</sup>	10.00%	3.48%	4.958	28.89%	8.80%	3.451	-0.41%	2.35%	2.612	57.64%	6.80%	7.894*
	-3.60%	9.34%	3.732	-3.62%	13.15%	0.001	2.21%	3.89%	0.271	99.64%	13.23%	3.078
IberiaP <sup>a</sup>	-0.19%	7.17%	1.551	1.78%	10.98%	1.965	1.45%	2.38%	0.009	90.02%	9.47%	8.329**
	63.86%	6.52%	0.166	43.13%	14.11%	1.228	5.20%	5.19%	0.312	-14.17%	5.22%	0.642
IberiaS	71.37%	8.34%	0.087	22.68%	10.52%	1.146	8.20%	3.24%	3.425	-5.83%	6.28%	5.613*
	72.75%	5.98%	2.783	26.20%	8.83%	0.179	2.51%	3.94%	0.176	-3.22%	2.90%	3.425
ItalyN	73.95%	6.18%	5.963	17.72%	7.89%	0.251	3.42%	2.37%	2.843	-0.97%	2.16%	5.663
	61.74%	7.05%	2.633	42.34%	11.06%	0.797	-0.80%	4.83%	6.369*	-5.67%	3.16%	4.137
ItalyS	68.38%	7.01%	2.465	28.83%	9.66%	2.475	1.61%	3.22%	13.111**	-3.66%	3.36%	4.709
	34.03%	6.74%	1.269	67.16%	10.67%	1.828	3.92%	5.12%	0.594	-7.85%	3.91%	0.699
Sardinia	39.38%	6.41%	1.322	56.07%	9.51%	2.618	4.03%	3.00%	1.612	-6.27%	3.72%	2.430
	25.00%	8.63%	3.459	37.29%	12.60%	2.851	16.57%	6.25%	0.112	5.94%	6.28%	20.743**
Scandinavia <sup>a</sup>	25.66%	6.96%	2.965	59.74%	14.23%	4.309	-0.08%	4.02%	1.107	2.41%	5.87%	2.343
	27.47%	7.71%	0.670	71.61%	13.16%	0.903	-10.65%	4.58%	0.002	8.85%	7.20%	0.209
	13.05%	8.28%	0.344	54.49%	14.13%	0.746	-4.12%	3.32%	5.537*	39.05%	9.35%	0.007

NOTE.—Molecular (first row) and frequency (second row) estimates. Significant heterogeneity between estimates for the different loci at the 5% level (\*) and at the 0.5% level (\*\*); Y-chromosome and mitochondrial data sets only. Population labels as defined in table 1.

<sup>a</sup> For these regions, no data was available for the 22 NRY binary markers from Semino et al. (2000).

between two genes sampled in two different PPs,  $i$  and  $j$  (a quantity equal to  $\bar{t}_{i,j}$ ), and  $\mu_i$  (or  $\mu_j$ ) is the relative contribution of the  $i^{\text{th}}$  (or  $j^{\text{th}}$ ) PP into the HP.

Least-squares estimators of  $\mu_i$ ,  $mY_i$ , were derived minimizing the sum of the squared differences between the left and the right-hand sides of equation (1) computed for each parental population. The  $mY_i$  estimators can be applied to any type of molecular data (such as DNA sequences, Restriction-Fragment Length Polymorphisms (RFLPs), or microsatellite data) for which the extent of molecular diversity is related to coalescence times. For DNA sequences, assuming that each new mutation occurs at a previously monomorphic site (the infinite-site model), coalescence times are estimated from the number of pairwise differences. For microsatellites, assuming a stepwise mutation model, coalescence times are estimated from the average squared difference in allele size.

When the number of substitutions (or, for microsatellites, the length differences) between alleles are disregarded, the estimated  $\mu_i$  fractions become equivalent to conventional admixture rates, estimated from haplotype or allele frequencies (Chakraborty 1986).

For each system, we estimated twice the contributions of the putative parental populations to the 12 European regions, either considering or not considering the molecular differences between alleles (hereafter, we shall refer to these estimates respectively as *molecular* and *frequency* admixture rates). Standard errors  $\sigma_{mY_i}$  were computed by a bootstrap procedure (Efron, 1982) that consists in drawing, with replacement, the alleles from the original samples, as described in Bertorelle and Excoffier (1998). A weighted average across systems was then computed (Cavalli-Sforza

and Bodmer 1971), and the heterogeneity among the contributions estimated at  $k$  different systems was tested by means of a  $\chi^2$  as suggested by Cavalli-Sforza and Bodmer (1971):

$$\chi^2_{k-1} = \sum_{i=1}^k \frac{(mY_i - \overline{mY})^2}{\sigma_{mY_i}^2} \quad (2)$$

This  $\chi^2$  test does not account for the stochasticity of the coalescent process. Consequently, nominally significant tests may reflect either real heterogeneity among systems, or random differences among realizations of the same stochastic process at different loci, or both. Therefore, the number of nominally significant results is expected to be higher than the real number of populations whose estimated admixture rates are heterogeneous across systems. Finally, to identify geographical trends in the admixture proportions, we summarized by linear regression the relationships of the admixture estimates in each hybrid population with the distances from the geographical barycenters of the Basque, Near Eastern, North African, and North-Eastern Europe samples, respectively.

## Results

### Admixture Proportions (4 Parental Populations): Y Chromosome and mtDNA

With few exceptions, the mean admixture proportions estimated from mitochondrial and NRY data (table 3) fall in the range [0%–100%]. Values exceeding this range would indicate that a population considered a hybrid has more extreme characteristics than one of the parental populations. That may occasionally happen if recent

genetic drift was strong, but a large number of values greater than 100% or smaller than 0 would suggest errors either in the model used or in the parental populations chosen. However, only slightly negative values, not exceeding  $-15\%$ , are occasionally observed for the North African and North-Eastern Europe contribution to European groups of samples. Standard deviations are in most cases lower than 10% but can reach 15% in some regions.

Even after Bonferroni's correction for multiple tests (Sokal and Rohlf 1995), significant heterogeneity between loci is observed for several groups of samples (see table 3). This finding is in agreement with previous results indicating that the Y chromosome and mtDNA have different distributions in Europe (Dupanloup et al. 2003) and indeed worldwide (Seielstad, Minch, and Cavalli-Sforza 1998; Harris and Hey 1999).

The estimated North African contribution to the European gene pools is low, less than 2% on average (range:  $-10.7\%$  in Scandinavia,  $16.6\%$  in Sardinia for molecular estimates;  $-4.1\%$  in Scandinavia,  $8.2\%$  in Portugal, for frequency estimates). In more than one-third of the samples, especially in Northern Europe, the estimated North African admixture does not differ significantly from zero, suggesting that genes from North Africa essentially do not occur in the gene pools of these regions. In general, the estimated contributions from North-Eastern Europe are higher than the African contributions, but they still represent a small component of genetic diversity, accounting for between  $10.5\%$  (molecular estimates) and  $17.4\%$  (frequency estimates) of the total. Variation among regions is high, and most groups show little or no North-Eastern Europe admixture. The exceptions are Finland and Eastern Europe, where roughly 95% and 50% of the gene pools, respectively, seem to come from North-Eastern European ancestors.

The main components in the European genomes appear to derive from ancestors whose features were similar to those of modern Basques and Near Easterners, with average values greater than 35% for both these parental populations, regardless of whether or not molecular information is taken into account. The lowest degree of both Basque and Near Eastern admixture is found in Finland, whereas the highest values are, respectively, 70% in Spain and more than 60% in the Balkans.

#### Admixture Proportions (2 Parental Populations): All Loci

With the increase of the number of systems considered (6 to 8 mitochondrial and nuclear systems, depending on the number of autosomal loci available in each population), the statistical errors of the admixture coefficients decrease substantially (all below 8%; table 4). The Near Eastern contribution is generally high, with a mean of 49.4% across Europe (range: 20.8% in England, 79.0% in the Balkans) when considering molecular information and 54.5% (22.1% in England, 95.6% in Finland) when considering only the frequency of haplotypes. However, there is reason to mistrust the estimates obtained for Finland. Indeed, more than 90% of the alleles observed there seem to have come from North-Eastern Europe (table 3), so its population can by no means be

regarded as a hybrid between Basques and Near Easterners (table 4). The extent to which an incorrect choice of parental populations leads to wrong results is investigated by simulation in a successive section of this paper. At any rate, when Finland is excluded from calculations, the average Near Eastern contributions become 48.3% (molecular estimates) and 50.7% (frequency estimates).

Heterogeneity among the estimates computed for the different systems is nominally significant in central Eastern Europe, Eastern Europe, Finland and Scandinavia, and remains significant in the Balkans even after Bonferroni's correction for multiple tests. Note that, with the test we used, the probability to reject the null hypothesis (homogeneity across loci) when true was higher than the nominal 5%. However, this result confirms that analyses of single markers are likely to yield inaccurate estimates of demographic parameters.

#### Regression Analysis

In figure 3, no significant correlation is apparent between North African admixture and geography. Genetic exchanges across the Mediterranean Sea, and especially in its western-most part where the geographic distance between continents is smallest, seem to have been limited or very limited (Simoni et al. 1999; Bosch et al. 2001). By contrast, when a Bonferroni correction for multiple tests is applied, admixture from North-Eastern Europe and from the Basque area are significantly associated with the distance from the populations of interest (see fig. 3), with a decrease, respectively, of 30% and 35% every 1,000 km, in a range of 2,000 to 2,500 km from the barycenter of the parental samples.

The Near Eastern contribution to European samples is significantly correlated with geography when frequency data are used but not when molecular information is taken into account. In this case, the distribution of points reveals a reduction of admixture rates with distance, but these estimates have a large variance within each distance class. The mean proportion of Near Eastern genes in European samples does decrease with distance from the Near East, but, even after 3,000 km, it is still high and different from 0.

To quantify more precisely the relative importance of what seem to be the two main components of the European genome, we re-estimated admixture using all systems but only two parental populations, the Basques and the Near Easterners. The relationship between admixture rates and geographic distances becomes stronger ( $r_{NE} = -0.709$ ,  $p = 0.010$ , molecular estimates;  $r_{NE} = -0.638$ ,  $p = 0.026$ , frequency estimates), and a rather clear geographical pattern is evident (fig. 4).

#### Testing the Choice of Hybrid and Parental Populations

In all previous analyses, we assumed that the parental and the hybrid populations were unambiguously defined. As shown by the estimates obtained for Finland in table 4, violations of that assumption may lead to erroneous conclusions. However, there is a way to validate it. As suggested by Bertorelle and Excoffier (1998), a misidentification of the parental populations in the analysis results in many coefficients outside the range [0%–100%], and/or

**Table 4**  
**Weighted Average Across Six to Eight Loci, and Standard Deviations (SD), of the Estimated Near Eastern Contribution to European Gene Pools**

Population	NumLoci	Av.	SD	$\chi^2$
Balkans	8	78.97%	4.87%	25.681**
British Isles <sup>a</sup>	7	82.80%	3.69%	54.378**
CEE	8	22.13%	5.67%	10.119
CEW	8	47.64%	4.62%	36.985**
EE	8	31.44%	5.23%	4.412
		47.25%	4.99%	13.037
EE	8	50.66%	4.23%	19.936*
		77.05%	3.76%	65.390**
Finland <sup>a</sup>	6	63.06%	6.46%	15.929*
		95.62%	7.59%	20.982**
IberiaP <sup>a</sup>	7	40.34%	6.46%	3.250
		32.73%	5.41%	3.594
IberiaS	8	34.59%	4.97%	6.526
		29.87%	5.07%	13.651
ItalyN	7	44.68%	6.09%	7.985
		36.01%	5.82%	8.447
ItalyS	8	62.76%	6.25%	5.345
		60.44%	5.29%	12.789
Sardinia	8	75.77%	7.91%	12.344
		69.25%	6.18%	21.976**
Scandinavia <sup>a</sup>	7	52.13%	5.40%	17.081*
		52.74%	5.54%	34.245**

NOTE.—Molecular (first row) and frequency (second row) estimates. Significant heterogeneity between estimates for the different loci at the 5% level (\*) and at the 0.5% level (\*\*). Population labels as defined in table 1.

in high errors associated with the estimates. We ran five simulation experiments in which we selected four random populations as parentals from the set of populations here considered. We then re-estimated the hypothetical contribution of these four populations into the 12 (hypothetically admixed) populations left using one Y-chromosome data set (Rosser et al. 2000).

As is evident in table 5, the range, and especially the standard errors, of the admixture estimates become extremely large in random demographic scenarios alternative to the admixture model considered throughout this study. A greater number of randomization tests would be necessary to prove that the populations we used as parental represent the best possible choice, and that would be overly time-consuming. However, we can at least conclude that implausible results are evident when clearly implausible parental populations are used to estimate admixture. Because when we used what we consider plausible parental populations the results were clearly different, it seems reasonable to conclude that the evolutionary scenario tested in this study is, by and large, at least realistic. That does not come as a surprise, because that scenario is supported by, and was originally designed using, up-to-date archaeological information.

## Discussion

The questions asked in this and in comparable studies are of the type: When did a certain group of people come to occupy a certain area? How extensive was the admixture between them and other groups? These are questions about

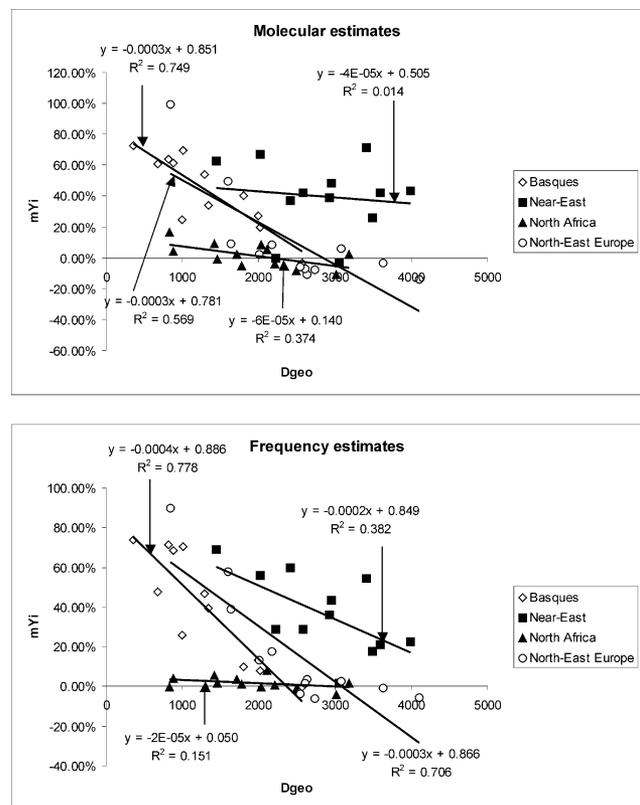


FIG. 3.—Linear regression of the contributions of the four parental populations to the 12 European groups of samples on the geographic distances between them (Y-chromosome and mitochondrial data sets only). Regression line equations and Pearson correlation coefficients for the frequency estimates of admixture proportions:  $m_{BA} = -3 \times 10^{-4} D_{GEO} + 0.851$  ( $R^2 = 0.749$ ,  $r = -0.865$ ,  $p < 0.001$ ),  $m_{NE} = -4 \times 10^{-5} D_{GEO} + 0.505$  ( $R^2 = 0.014$ ,  $r = -0.119$ ,  $p = 0.712$ ),  $m_{NA} = -6 \times 10^{-5} D_{GEO} + 0.140$  ( $R^2 = 0.374$ ,  $r = -0.612$ ,  $p = 0.054$ ),  $m_{NEE} = -3 \times 10^{-4} D_{GEO} + 0.781$  ( $R^2 = 0.569$ ,  $r = -0.754$ ,  $p = 0.005$ ). Regression line equations and Pearson correlation coefficients for the molecular estimates of admixture proportions:  $m_{BA} = -4 \times 10^{-4} D_{GEO} + 0.886$  ( $R^2 = 0.778$ ,  $r = -0.882$ ,  $p < 0.001$ ),  $m_{NE} = -2 \times 10^{-4} D_{GEO} + 0.849$  ( $R^2 = 0.382$ ,  $r = -0.618$ ,  $p = 0.032$ ),  $m_{NA} = -2 \times 10^{-5} D_{GEO} + 0.050$  ( $R^2 = 0.151$ ,  $r = -0.389$ ,  $p = 0.212$ ),  $m_{NEE} = -3 \times 10^{-4} D_{GEO} + 0.866$  ( $R^2 = 0.706$ ,  $r = -0.840$ ,  $p = 0.001$ ).

population history, and they need be addressed considering simultaneously as many independent alleles as possible. Analyses of single or physically linked alleles or haplotypes, no matter how informative they appear to be, are unlikely to contain all the information needed to infer and quantify population processes, and may also, if selected a posteriori, produce biased inferences.

With one exception, previous estimates of the Paleolithic and Neolithic contributions to the European gene pools did not consider the entire genetic diversity in the populations of interest. Rather, admixture rates were equated with the frequencies of haplotypes whose distribution was supposed to be a result of Neolithic admixture (Semino et al. 2000; Richards et al. 2002). In the only study so far that explicitly models the admixture process at the population level, Chikhi et al. (2002) described Y-chromosome patterns supporting a significantly greater genetic contribution of Neolithic farmers than did previous studies based on the same data (Semino et al. 2000) and an

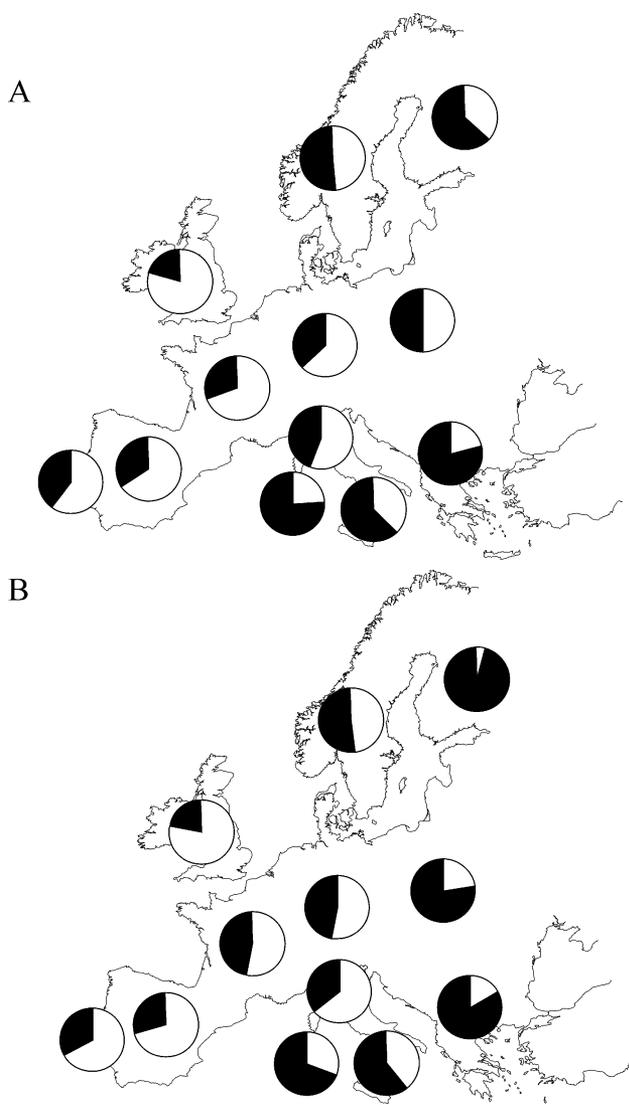


FIG. 4.—Pie diagrams showing the distribution of Basque (white) and Near East (black) contributions to the 12 European groups of samples in Europe: (A) molecular and (B) frequency admixture rates. The corresponding admixture estimates are given in table 4.

east-west gradient of Neolithic admixture across Europe. In this study, we found similar patterns across the genome, which implies that we are unlikely to have been misled by the effects of selection (Luikart et al. 2003).

The Y chromosome, and mtDNA, can be regarded as single, if very large and polymorphic, loci. Because gene flow processes, including admixture, affect the entire genome, the greater the number of systems considered, the more robust the inferences about admixture (e.g., Bertorelle and Excoffier 1998). Eight systems are not many, but this is the first admixture study of Europe based on multiple loci. Its results suggest that the main components in the genomes of Europeans may be referred to admixing populations whose genes resembled, respectively, the modern Basque and Near Eastern populations. Only a small fraction of the European alleles seems to come from North Africa, whereas a fourth component of Northern European

**Table 5**  
Proportion of Admixture Estimates Outside the Range [0%–100%] and Mean Associated Standard Errors in Different Admixture Models Using the Y-Chromosome Data Set (Rosser et al. 2000)

Admixture Model	Out of Range [0%–100%]	Mean SD
Real	31.2% [–29.33%, 114.00%]	9.98%
Simulation 1	62.5% [–1017.21%, 720.10%]	202.64%
Simulation 2	45.8% [–363.86%, 450.49%]	209.21%
Simulation 3	77.1% [–585.00%, 788.17%]	861.25%
Simulation 4	56.3% [–631.74%, 725.52%]	176.76%
Simulation 5	56.3% [–302.89%, 281.40%]	220.70%

NOTE.—The first row of the table shows the results obtained using the admixture model considered throughout this study. The other results correspond to alternative admixture model in which the parental populations (PPs) were chosen at random from the set of populations tested for NRY diversity (simulation 1: PPs: Basques, Balkans, Portugal, Northern Italy; simulation 2: PPs: Balkans, Northern Italy, Sardinia, Scandinavia; simulation 3: PPs: British Isles, Central Europe West, Portugal, Sardinia; simulation 4: PPs: Near East, Eastern Europe, British Isles, Finland; simulation 5: PPs: North Africa, Central Europe West, Eastern Europe, Southern Italy).

(and ultimately, perhaps, Northern Asian) origin is nonzero, but it is largely restricted to the northeast of the continent. Near Eastern admixture is less than 30% only in the British Isles and exceeds 50% over much of the continent, with a decrease of this contribution as the geographic distance from the Near East increases (figs. 3 and 5).

In agreement with essentially all published literature, we took the genes in current Basque and Near Eastern populations as the best available approximation to the genes of the people inhabiting, respectively, Europe and the Near East before the Neolithic dispersals. To the extent that this assumption is realistic, the results indicate that a large fraction of alleles in the European genomes can be traced to a Neolithic origin, certainly much higher than the 15–20% proposed by Richards et al. (2000, 2002) and Semino et al. (2000). The spatial distribution of these fractions is the one expected under a NM model, in which the genes of Neolithic farmers got diluted as they moved away from the Near East.

Any analysis of admixture relies on the validity of the underlying model, and every model is a simplification of a set of evolutionary phenomena that would otherwise be difficult or impossible to address quantitatively. Here we assumed that up to four parental populations determined the current gene pool of all European populations and that other gene flow processes were negligible. In addition, we assumed that after admixture genetic drift and new mutations could be neglected. There is no doubt that genetic exchanges in historic and prehistoric Europe have been multiple and complex (e.g., Sokal et al. 1997), and that five to ten thousand years of genetic drift and mutation must have left a mark in the populations we considered. The question is whether or not, by disregarding these additional phenomena, one ends up with unreliable admixture estimates.

As for the effects of additional gene flow, negative admixture estimates accompanied by large standard deviations are commonly observed when more complex exchanges occurred than a simple admixture event, or when the parental populations are improperly chosen

(Bertorelle and Excoffier 1998). We showed that replacing our four parental populations with other, implausible parental populations leads to evidently implausible results (table 5). On the contrary, most values estimated using what we consider plausible parental populations were in the range 0%–100%, and standard errors were always below 15%. Therefore, by and large these results do not suggest that the admixture model we chose grossly misrepresents the population processes leading to the current European genetic diversity.

The question of how important drift was after the admixture event is a complicated issue that we could only partly address. First of all, in general, low levels of genetic differentiation are observed among present-day European populations at the genomic level (Romualdi et al. 2002; Rosenberg et al. 2002), which does not support the idea that drift was the main evolutionary force affecting them. Mitochondrial data suggest that the European populations expanded in the last ten millennia (Excoffier and Schneider 1999), and genetic drift is known to be less effective in expanding populations (e.g., Terwilliger et al. 1998). Chikhi et al. (2002) are the only ones so far who inferred the impact of drift after the Neolithic transition. Their results, based on Y-chromosome diversity, suggest a limited effect in the Near East and increasing, but never large, effects for populations that acquired farming at later times. In addition, we note that these loci are uniparentally transmitted, and hence their effective population size is one-fourth that of the autosomal loci we considered. As a consequence, we expect a lesser effect of drift on most genes considered in our study. Finally, the stochastic nature of drift should tend to produce an increase in the errors associated with our estimates, but averaging several independent loci should make a systematic bias unlikely.

Although we could not model the effects of drift, we could get further insight on the effect of mutation after admixture by re-analyzing one Y-chromosome data set (Rosser et al. 2000), after introducing a mutational parameter,  $t_A$ , in equation (1). Assuming that admixture occurred 10,000 or 5,000 years ago, and a mutation rate of  $2.5 \times 10^{-8}$ /site/year (Hammer 1995; Jobling, Pandya, and Tyler-Smith 1997), or even a rate ten times higher, no estimated coefficients changed by more than a few percent, and in no population was the change significant when evaluated by a Mann-Whitney test (results not shown). To observe substantial changes of admixture rates, the mutation rate had to be at least 1,000-fold as high.

Molecular and frequency estimates of admixture are not identical, and they should not be expected to be so. Indeed, DNA sequences evolve mainly by the accumulation of mutations, occurring over millennia, whereas the frequencies of allelic variants, no matter whether estimated at the protein or DNA level, diverge more rapidly because of drift (Sajantila et al. 1995; Barbujani 1997). Thus, there is little doubt that in the 10,000 years elapsed from the origin of agriculture the European genetic diversity was affected more by drift and migration than by mutation. As a consequence, estimates based on molecular distances may incorporate, to an undefined extent, the effect of mutations that predate the admixture events we planned to describe, so figure 4B probably represents a more reliable

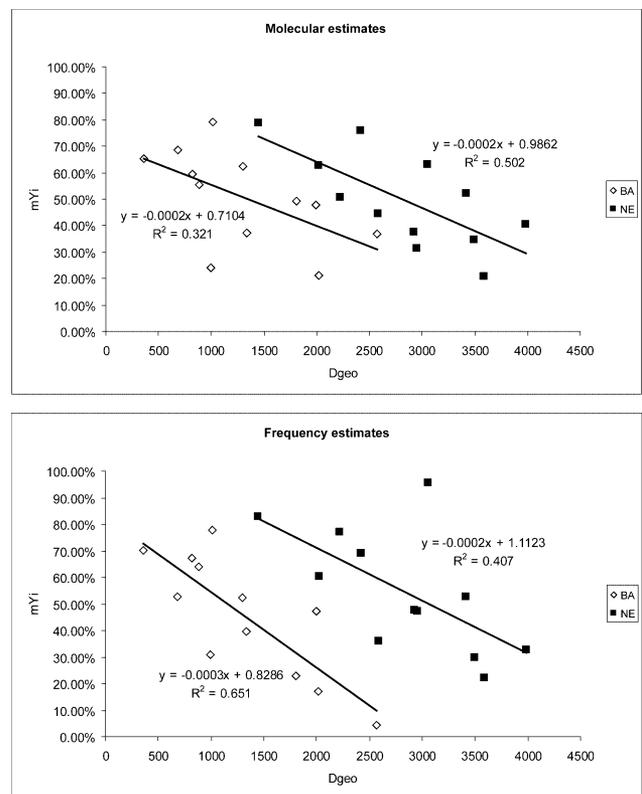


Fig. 5.—Linear regression of the contributions of the Basque and Near East populations to the 12 European groups of samples on the geographic distances between them (Y-chromosome, mitochondrial and nuclear data sets). Regression line equations are shown in the charts.

summary of European admixture than figure 4A. Nevertheless, the similarity between the two parts of figure 4 indicates that, for the main question addressed here, it is not terribly important whether or not molecular differences among alleles are considered. Indeed, the differences between estimates relative to the same population are usually below 10% (see tables 3 and 4) and show roughly parallel trends across Europe.

In brief, our results corroborate Chikhi et al.'s (2002) conclusion that the Neolithic shift to agriculture entailed major population dispersal from the Near East, by increasing significantly the amount of data considered. There is a single important difference between the results of this study and theirs. Chikhi et al. (2002) found a very limited (if any) amount of Y-chromosome introgression from the Near East into Sardinia. This led them to suggest that Sardinians might be, like the Basques, descendants of hunter-gatherers whose genomes were only mildly affected by incoming farmers. The method used by Chikhi et al. (2002) accounted for the effect of genetic drift after admixture, whereas the method used here only accounts for drift through the use of independent loci. As Chikhi et al. (2002) noted, the uncertainty on a particular population for a given locus is often large. One possible explanation of the difference observed for the Sardinians, then, is that a higher level of introgression may have occurred at nuclear loci, which is possible in principle, but is difficult to prove using the available data. If so, the results

presented here could be different without being contradictory. Other explanations can be envisaged. At this stage it is clear that, unless a method is able to account for the stochasticity of genetic drift, the study of single loci should be avoided, in favor of a multi-locus approach.

But we can now ask what these admixture calculations actually mean. In particular, what is a Neolithic or a Paleolithic ancestry, and does the high Neolithic admixture in, say, Scandinavia mean that the mutations generating the alleles we currently observe in Scandinavia occurred in Neolithic times? Or does it mean that gene flow was high from the Near East into Sweden in Neolithic times? The answer is no in both cases. First, the age of a mutation is not the moment at which that mutation entered a population, because the depth of the gene genealogy associated with a mutation is greater than that of the evolutionary process that gave rise to its present-day distribution (Barbujani, Bertorelle, and Chikhi 1998; Edwards and Beerli 2000; Nichols 2001). Despite some past disagreement, most authors have now come to acknowledge that there is no necessary correlation between the timing of migrations and the age of mitochondrial or Y-chromosome clades (Stumpf and Goldstein 2001; Richards et al. 2002). Second, admixture rates measure the fraction of alleles that can be traced back to the people who, respectively, were already in Europe, or entered it, with the Neolithic expansion, but they tell us nothing about where exactly these people dwelt at that time. Therefore, a high Paleolithic or Neolithic component in a gene pool does not mean that a region was colonized in Paleolithic or Neolithic times, respectively. Under the assumptions of our model, a 52% Neolithic component in Scandinavia means that roughly half of the Scandinavians' alleles are probably descended from ancestors who entered Europe (not Scandinavia) during the Neolithic dispersal and reached Scandinavia at an unspecified, later time.

In the future, it will be important to incorporate detailed archeological information into the population models, so that the assumptions will become both more complicated and more realistic. In addition, we need more sophisticated genetic methods to discriminate between the effects of isolation by distance and historical migration. Indeed, both phenomena may have contributed to the generation of the European gradients, although simulation studies have rejected the hypothesis that isolation by distance *by itself* might have caused such a strong patterning of genetic diversity in Europe (Barbujani, Sokal, and Oden 1995). In the next years, a greater number of polymorphic markers will also become available. Considering greater numbers of loci will progressively reduce the importance of loci whose peculiar evolutionary history, possibly including selection, renders them statistical outliers (Luikart et al. 2003). Therefore, in the not-so-distant future, there are good chances to achieve more robust admixture estimates for Europe and to define with greater confidence the timing of the admixture. At present, this study, the largest so far, shows that a component of the Europeans' genome of Near Eastern origin is large, and it decreases as one moves west. Neither finding is in agreement with the predictions of a model in which Neolithic immigrants from the Near East contributed a

small share of the alleles in the genome of current Europeans.

## Acknowledgments

This study was supported by grants of the Italian National Research Council (CNR), within the European Science Foundation (ESF) Eurocores programme *The origins of man, language and languages*, (project JA03-B02); by the Swiss National Science Foundation (FNRS); and by funds of the University of Ferrara. We thank two anonymous referees for their comments and suggestions.

## Literature Cited

- Ammerman, A. J., and L. L. Cavalli-Sforza. 1984. *The Neolithic transition and the genetics of populations in Europe*. Princeton University Press, Princeton.
- Barbujani, G. 1997. DNA variation and language affinities. *Am. J. Hum. Genet.* **61**:1011–1014.
- Barbujani, G., and G. Bertorelle. 2001. Genetics and the population history of Europe. *Proc. Natl. Acad. Sci. USA* **98**:22–25.
- Barbujani, G., G. Bertorelle, and L. Chikhi. 1998. Evidence for Paleolithic and Neolithic gene flow in Europe. *Am. J. Hum. Genet.* **62**:488–492.
- Barbujani, G., A. Pilastro, S. De Domenico, and C. Renfrew. 1994. Genetic variation in North Africa and Eurasia: Neolithic demic diffusion vs. Paleolithic colonisation. *Am. J. Phys. Anthropol.* **95**:137–154.
- Barbujani, G., R. R. Sokal, and N. L. Oden. 1995. Indo-European origins: a computer-simulation test of five hypotheses. *Am. J. Phys. Anthropol.* **96**:109–132.
- Bertorelle, G., and L. Excoffier. 1998. Inferring admixture proportions from molecular data. *Mol. Biol. Evol.* **15**:1298–1311.
- Bertranpetit, J., and L. L. Cavalli-Sforza. 1991. A genetic reconstruction of the history of the population of the Iberian Peninsula. *Ann. Hum. Genet.* **55**:51–56.
- Bertranpetit, J., J. Sala, F. Calafell, P. A. Underhill, P. Moral, and D. Comas. 1995. Human mitochondrial DNA variation and the origin of Basques. *Ann. Hum. Genet.* **59**:63–81.
- Bosch, E., F. Calafell, D. Comas, P. J. Oefner, P. A. Underhill, and J. Bertranpetit. 2001. High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula. *Am. J. Hum. Genet.* **68**:1019–1029.
- Caramelli, D., C. Lalueza-Fox, C. Vernesi, M. Lari, A. Casoli, F. Mallegni, B. Chiarelli, I. Dupanloup, J. Bertranpetit, G. Barbujani, and G. Bertorelle. 2003. Evidence for a genetic discontinuity between Neandertals and 24,000-year-old anatomically modern Europeans. *Proc. Natl. Acad. Sci. USA* **100**:6593–6597.
- Cavalli-Sforza, L. L., and W. F. Bodmer. 1971. *The genetics of human populations*. W.H. Freeman and Company, San Francisco, Calif.
- Cavalli-Sforza, L. L., P. Menozzi, and A. Piazza. 1993. Demic expansions and human evolution. *Science* **259**:639–646.
- Cavalli-Sforza, L. L., and A. Piazza. 1993. Human genomic diversity in Europe: a summary of recent research and prospects for the future. *Eur. J. Hum. Genet.* **1**:3–18.
- Chakraborty, R. 1986. Gene admixture in human populations: models and predictions. *Yearb. Phys. Anthropol.* **29**:1–43.
- Chikhi, L., G. Destro-Bisol, G. Bertorelle, V. Pascali, and G. Barbujani. 1998. 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., R. A. Nichols, G. Barbujani, and M. A. Beaumont. 2002. Y genetic data support the Neolithic demic diffusion model. *Proc. Natl. Acad. Sci. USA* **99**:10008–10013.
- Dupanloup, I., and G. Bertorelle. 2001. Inferring admixture proportions from molecular data: extension to any number of parental populations. *Mol. Biol. Evol.* **18**:672–675.
- Dupanloup, I., L. Pereira, G. Bertorelle, F. Calafell, M. J. Prata, A. Amorim, and G. Barbujani. 2003. A recent shift from polygyny to monogamy in humans is suggested by the analysis of worldwide Y-chromosome diversity. *J. Mol. Evol.* **57**:85–97.
- Edwards, S. V., and P. Beerli. 2000. Gene divergence, population divergence, and the variance in coalescence times in phylogeographic studies. *Evolution* **54**:1839–1854.
- Efron, B. 1982. The jackknife, the bootstrap and other resampling plans. Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, Philadelphia.
- Estivill, X., C. Bancells, and C. Ramos. 1997. Geographic distribution and regional origin of 272 cystic fibrosis mutations in European populations. The Biomed CF Mutation Analysis Consortium. *Hum. Mutat.* **10**:135–154.
- Excoffier, L., and S. Schneider. 1999. Why hunter-gatherer populations do not show signs of Pleistocene demographic expansions. *Proc. Natl. Acad. Sci. USA* **96**:10597–10602.
- Excoffier, L., P. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**:479–491.
- Goldstein, D. B., and L. Chikhi. 2002. Human migration and population structure: what we know and why it matters. *Annu. Rev. Genom. Hum. Genet.* **3**:129–152.
- Hammer, M. F. 1995. A recent common ancestry for human Y chromosomes. *Nature* **378**:376–378.
- Harris, E. E., and J. Hey. 1999. Human demography in the Pleistocene: do mitochondrial and nuclear genes tell the same story? *Evol. Anthropol.* **8**:81–86.
- Holtkemper, U., B. Rolf, C. Hohoff, P. Forster, and B. Brinkmann. 2001. Mutation rates at two human Y-chromosomal microsatellite loci using small pool PCR techniques. *Hum. Mol. Genet.* **10**:629–633.
- Jobling, M. A., A. Pandya, and C. Tyler-Smith. 1997. The Y chromosome in forensic analysis and paternity testing. *Int. J. Legal Med.* **110**:118–124.
- Jobling, M. A., and C. Tyler-Smith. 2000. New uses for new haplotypes. The human Y chromosome, disease and selection. *Trends. Genet.* **16**:356–362.
- Krings, M., C. Capelli, F. Tschentscher, H. Geisert, S. Meyer, A. von Haeseler, K. Grossschmidt, G. Possnert, M. Paunovic, and S. Pääbo. 2000. A view of Neandertal genetic diversity. *Nature Genet.* **26**:144–146.
- Luikart, G., P. R. England, D. Tallmon, S. Jordan, and P. Taberlet. 2003. The power and promise of population genomics: From genotyping to genome typing. *Nature Rev. Genet.* **4**:981–994.
- Macaulay, V., M. Richards, E. Hickey, E. Vega, F. Cruciani, V. Guida, R. Scozzari, B. Bonne-Tamir, B. Sykes, and A. Torroni. 1999. The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am. J. Hum. Genet.* **64**:232–249.
- Menozi, P., A. Piazza, and L. L. Cavalli-Sforza. 1978. Synthetic maps of human gene frequencies in Europeans. *Science* **201**:786–792.
- Mishmar, D., E. Ruiz-Pesini, P. Golik et al. (13 co-authors). 2003. Natural selection shaped regional mtDNA variation in humans. *Proc. Natl. Acad. Sci. USA* **100**:171–176.
- Nichols, R. 2001. Gene trees and species trees are not the same. *Trends Ecol. Evol.* **16**:358–364.
- Otte, M. 2000. The history of European populations as seen by archaeology. Pp 41–44 in C. Renfrew and H. Boyle, eds. *Archaeogenetics: DNA and the population prehistory of Europe*. McDonald Institute for Archaeological Research, Cambridge.
- Pääbo, S. 1999. Human evolution. *Trends. Cell. Biol.* **9**:13–16.
- Pinhasi, R., R. A. Foley, and M. Mirazòn Lahr. 2000. Spatial and temporal patterns in the Mesolithic-Neolithic archaeological record of Europe. Pp 45–56 in C. Renfrew and H. Boyle, eds. *Archaeogenetics: DNA and the population prehistory of Europe*. McDonald Institute for Archaeological Research, Cambridge.
- Rando, J. C., F. Pinto, A. M. Gonzalez, M. Hernandez, J. M. Larruga, V. M. Cabrera, and H. J. Bandelt. 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.
- Relethford, J. A. 2001. Absence of regional affinities of Neandertal DNA with living humans does not reject multiregional evolution. *Am. J. Phys. Anthropol.* **115**:95–98.
- Renfrew, C. 1987. *Archaeology and language: the puzzle of Indo-European origins*. Jonathan Cape, London.
- Richards, M., H. Corte-Real, P. Forster, V. Macaulay, H. Wilkinson-Herbots, A. Demaine, S. Papiha, R. Hedges, H. J. Bandelt, and B. Sykes. 1996. Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am. J. Hum. Genet.* **59**:185–203.
- Richards, M., V. Macaulay, E. Hickey et al. (34 co-authors). 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet.* **67**:1251–1276.
- Richards, M., V. Macaulay, A. Torroni, and H. J. Bandelt. 2002. In search of geographical patterns in European mitochondrial DNA. *Am. J. Hum. Genet.* **71**:1168–1174.
- Romualdi, C., D. Balding, I. S. Nasidze, G. Risch, M. Robichaux, S. T. Sherry, M. Stoneking, M. A. Batzer, and G. Barbujani. 2002. Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. *Genome Res.* **12**:602–612.
- Rosenberg, N. A., J. K. Pritchard, J. L. Weber, H. M. Cann, K. K. Kidd, L. A. Zhivotovsky, and M. W. Feldman. 2002. Genetic structure of human populations. *Science* **298**:2381–2385.
- Rosser, Z. H., T. Zerjal, M. E. Hurles et al. (63 co-authors). 2000. Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am. J. Hum. Genet.* **67**:1526–1543.
- Sajantila, A., P. Lahermo, T. Anttinen et al. (13 co-authors). 1995. Genes and languages in Europe: an analysis of mitochondrial lineages. *Genome Res.* **5**:42–52.
- Seielstad, M. T., E. Minch, and L. L. Cavalli-Sforza. 1998. Genetic evidence for a higher female migration rate in humans. *Nat. Genet.* **20**:278–280.
- Semino, O., G. Passarino, A. Brega, M. Fellous, and S. Santachiara-Benerecetti. 1996. A view of the Neolithic demic diffusion in Europe through two Y chromosome-specific markers. *Am. J. Hum. Genet.* **59**:964–968.
- Semino, O., G. Passarino, P. J. Oefner et al. (17 co-authors). 2000. The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science* **290**:1155–1159.
- Simoni, L., F. Calafell, D. Pettener, J. Bertranpetit, and G. Barbujani. 2000. Geographic patterns of mtDNA diversity in Europe. *Am. J. Hum. Genet.* **66**:262–278.
- Simoni, L., P. Gueresi, D. Pettener, and G. Barbujani. 1999. Patterns of gene flow inferred from genetic distances in the Mediterranean region. *Hum. Biol.* **71**:399–415.

- Sokal, R. R., N. L. Oden, M. S. Rosenberg, and D. DiGiovanni. 1997. Ethnohistory, genetics, and cancer mortality in Europeans. *Proc. Natl. Acad. Sci. USA* **94**:12728–12731.
- Sokal, R. R., N. L. Oden, and C. Wilson. 1991. Genetic evidence for the spread of agriculture in Europe by demic diffusion. *Nature* **351**:143–145.
- Sokal, R. R., and F. J. 1995. *Biometry*. W. H. Freeman and Company, New York.
- Stringer, C. 2003. Out of Ethiopia. *Nature* **423**:692–694.
- Stumpf, M. P., and D. B. Goldstein. 2001. Genealogical and evolutionary inference with the human Y chromosome. *Science* **291**:1738–1742.
- Terwilliger, J. D., S. Zollner, M. Laan, and S. Pääbo. 1998. Mapping genes through the use of linkage disequilibrium generated by genetic drift: 'Drift mapping' in small populations with no demographic expansion. *Hum. Hered.* **48**: 138–154.
- Torrioni, A., H. J. Bandelt, V. Macaulay et al. (33 co-authors). 2001. A signal, from human mtDNA, of postglacial recolonization in Europe. *Am. J. Hum. Genet.* **69**:844–852.
- Underhill, P. A., P. Shen, A. A. Lin et al. (24 co-authors). 2000. Y chromosome sequence variation and the history of human populations. *Nat. Genet.* **26**:358–361.
- Vernesi, C., S. Fuselli, L. Castri, G. Bertorelle, and G. Barbujani. 2002. Mitochondrial diversity in linguistic isolates of the Alps: a reappraisal. *Hum. Biol.* **74**:725–730.
- Wilson, J. F., D. A. Weiss, M. Richards, M. G. Thomas, N. Bradman, and D. B. Goldstein. 2001. Genetic evidence for different male and female roles during cultural transitions in the British Isles. *Proc. Natl. Acad. Sci. USA* **98**:5078–5083.
- Wiuf, C. 2001. Do  $\Delta F508$  heterozygotes have a selective advantage? *Genet. Res.* **78**:41–47.
- Zvelebil, M. 1986. Mesolithic prelude and Neolithic revolution. Pp 5–16 in M. Zvelebil, ed. *Hunters in transition: Mesolithic societies of temperate Eurasia and their transition to farming*. Cambridge University Press, Cambridge.

David Goldstein, Associate Editor

Accepted March 5, 2004