Influence of Admixture and Paleolithic Range Contractions on Current European Diversity Gradients

Arenas, Miguel; François, Olivier; Currat, Mathias; Ray, Nicolas; Excoffier, Laurent Georges Louis

How to cite

ARENAS, Miguel et al. Influence of Admixture and Paleolithic Range Contractions on Current European Diversity Gradients. In: Molecular biology and evolution, 2013, vol. 30, n° 1, p. 57–61. doi: 10.1093/molbev/mss203

This publication URL: https://archive-ouverte.unige.ch/unige:24758
Publication DOI: 10.1093/molbev/mss203

Influence of Admixture and Paleolithic Range Contractions on Current European Diversity Gradients

Miguel Arenas,*[1,2] Olivier François,[3,4] Mathias Currat,[5] Nicolas Ray,[6,7] and Laurent Excoffier[1,2]

[1]Computational and Molecular Population Genetics Lab, Institute of Ecology and Evolution, University of Bern, Bern, Switzerland
[2]Swiss Institute of Bioinformatics, Lausanne, Switzerland
[3]Laboratoire Techniques de l'Ingénierie Médicale et de la Complexité, Faculty of Medicine, University Joseph Fourier, Grenoble, France
[4]Institute of Technology, Centre National de la Recherche Scientifique, La Tronche, France
[5]Anthropology, Genetics and Peopling History Lab, Anthropology Unit – Department of Genetics & Evolution, University of Geneva, Geneva, Switzerland
[6]EnviroSPACE Laboratory, Institute for Environmental Sciences, University of Geneva, Geneva, Switzerland
[7]Forel Institute, University of Geneva, Geneva, Switzerland

***Corresponding author**: E-mail: miguel.arenasbusto@iee.unibe.ch; miguelmmmab@gmail.com.

Associate editor: Ryan Hernandez

Letter

Abstract

Cavalli-Sforza and Edwards (Analysis of human evolution. 1963. In: Geerts SJ, editor. Genetics today: Proceedings of the 11th International Congress of Genetics, The Hague, The Netherlands. New York: Pergamon. p. 923–993.) initiated the representation of genetic relationships among human populations with principal component (PC) analysis (PCA). Their study revealed the presence of a southeast–northwest (SE–NW) gradient of genetic variation in current European populations, which was interpreted as the result of the demic diffusion of early neolithic farmers during their expansion from the near east. However, this interpretation has been questioned, as PCA gradients can occur even when there is no expansion and because the first PC axis is often orthogonal to the expansion axis. Here, we revisit PCA patterns obtained under realistic scenarios of the settlement of Europe, focusing on the effects of various levels of admixture between paleolithic and neolithic populations, and of range contractions during the last glacial maximum (LGM). Using extensive simulations, we find that the first PC (PC1) gradients are orthogonal to the expansion axis, but only when the expansion is recent (neolithic). More ancient (paleolithic) expansions alter the orientation of the PC1 gradient due to a spatial homogenization of genetic diversity over time, and to the exact location of LGM refugia from which re-expansions proceeded. Overall we find that PC1 gradients consistently follow an SE–NW orientation if there is a large paleolithic contribution to the current European gene pool, and if the main refuge area during the last ice age was in the Iberian Peninsula. Our study suggests that an SE–NW PC1 gradient is compatible with little genetic impact of neolithic populations on the current European gene pool, and that range contractions have affected observed genetic patterns.

Key words: range expansion, range contraction, admixture, principal component analysis, European settlement, last glacial maximum.

Introduction

Principal component (PC) analysis (PCA) was introduced by Cavalli-Sforza and Edwards (1963) to study the genetic diversity of European populations, and it remains a useful technique to analyze genetic diversity (e.g., Jakobsson et al. 2008; Novembre and Stephens 2008) as it nicely summarizes information embedded in large genetic data sets (Patterson et al. 2006; Novembre and Ramachandran 2011). Cavalli-Sforza et al. (1994) used PCA to summarize allele-frequency data from worldwide human populations and found gradients of genetic variation from Africa, which were interpreted as the results of past range expansions. In Europe, gradients were observed along a southeast (SE)–northwest (NW) axis (Menozzi et al. 1978; Piazza et al. 1995) as shown in figure 1C. They were interpreted as the consequence of a demic diffusion process of neolithic farmers from the near east, during which paleolithic hunter–gatherer populations were replaced with little or even any admixture (Ammerman and Cavalli-Sforza 1984; Sokal et al. 1991; Diamond and Bellwood 2003). However, the interpretation of PC gradients remains problematic because several factors influence their shape and their direction (Novembre and Stephens 2008; François et al. 2010; Novembre and Stephens 2010). In particular, Novembre and Stephens (2008) showed that PC gradients arise at equilibrium under isolation-by-distance models, thus without requiring any expansion. Additionally, François et al. (2010) showed that demographic and spatial population expansions do not necessarily lead to a PC gradient along the expansion axis, but more often along an axis orthogonal to the expansion direction. The authors explain this result as a consequence of allele surfing (e.g., Edmonds et al. 2004; Klopfstein et al. 2006; Excoffier and Ray 2008), which creates radiating geographic sectors of low diversity

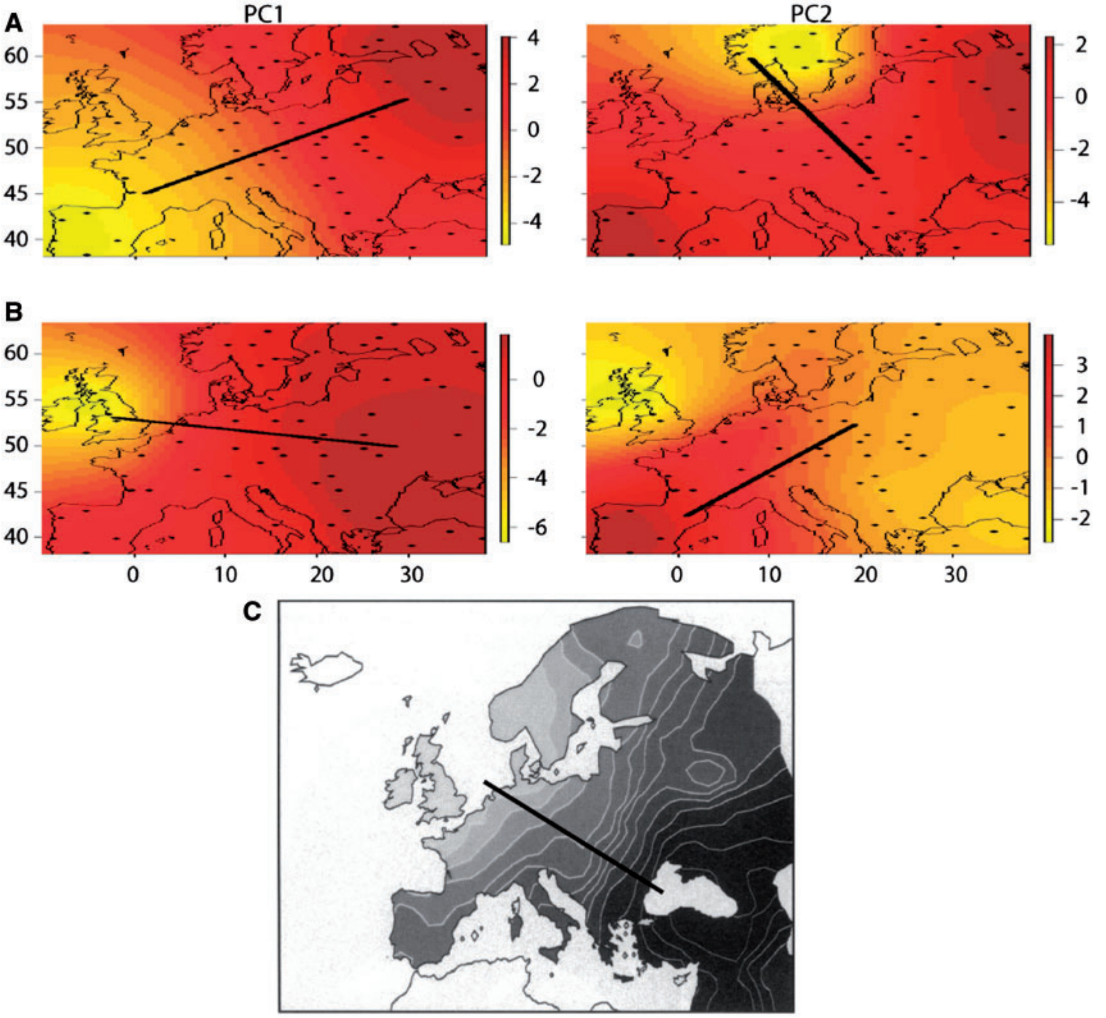

Fig. 1. SNP data PC maps for pure neolithic and pure paleolithic range expansions from Middle East. (*A*) Recent neolithic expansion from the Middle East. The PC1 gradient has a SW–NE orientation. (*B*) Old paleolithic expansion from the Middle East. The PC1 gradient has an E–W orientation with an isolation of the British Isles. Note that the PC scales are different in (*A*) and (*B*): neolithic expansions lead to more spatial genetic variation, probably due to the presence of sectors resulting from allele surfing. (*C*) Original PC1 map inferred from Piazza et al. (1995) (© 1995 National Academy of Sciences, USA) with a superimposed line connecting positive and negative PC1 centroids.

(Hallatschek et al. 2007), which are progressively eroded by short-range migration creating gradients of diversity (Excoffier and Ray 2008). They also showed that levels of admixture between neolithic and resident paleolithic populations could affect the orientation of PC gradients (François et al. 2010), but the relative contributions of paleolithic and neolithic populations to the current European gene pool is still highly debated (e.g., Chikhi et al. 1998; Richards et al. 2000; Barbujani and Chikhi 2006).

Another factor that has been ignored, but which might have deeply influenced European genetic patterns, is the last ice age that occurred 29–13 kya (Straus 1991). European hunter–gatherer groups thus probably went through a range contraction during the last ice age (see e.g., Straus 1991, and references therein) and re-expanded north to recolonize new suitable areas after the last glacial maximum (LGM) (Barbujani and Bertorelle 2001).

Here, we extend the spatially explicit simulations performed by François et al. (2010) and investigate a much larger array of possible scenarios of human settlement in Europe. We study pure paleolithic and neolithic expansions from the near east, as well as scenarios with varying levels of admixture among these populations. Importantly, we also explore the effect of the ice age by simulating range contractions toward southern Europe, followed by a refugial isolation period and a posterior re-expansion. The effects of active migrations toward the south during the range contraction (see Arenas et al. 2012) are also studied.

In keeping with previous results (François et al. 2010), the simulation of single nucleotide polymorphism (SNP) data with minimum allele frequency (MAF) >3% generates first PC (PC1) gradients that are often perpendicular to the expansion axis in the case of a recent (neolithic) expansion (figs. 1*A* and 2*A*). PC2 and PC3 maps usually highlight the British Isles and Scandinavia, due to their geographic isolation (fig. 1*A*). For pure paleolithic expansions, NW–SE and E–W gradients are usually observed in PC1 (figs. 1*B* and 2*A*), and PC2 and PC3 maps usually highlight the Iberian Peninsula (see an illustrative example in fig. 1*B*). The difference in PC1 gradient observed after paleolithic and neolithic

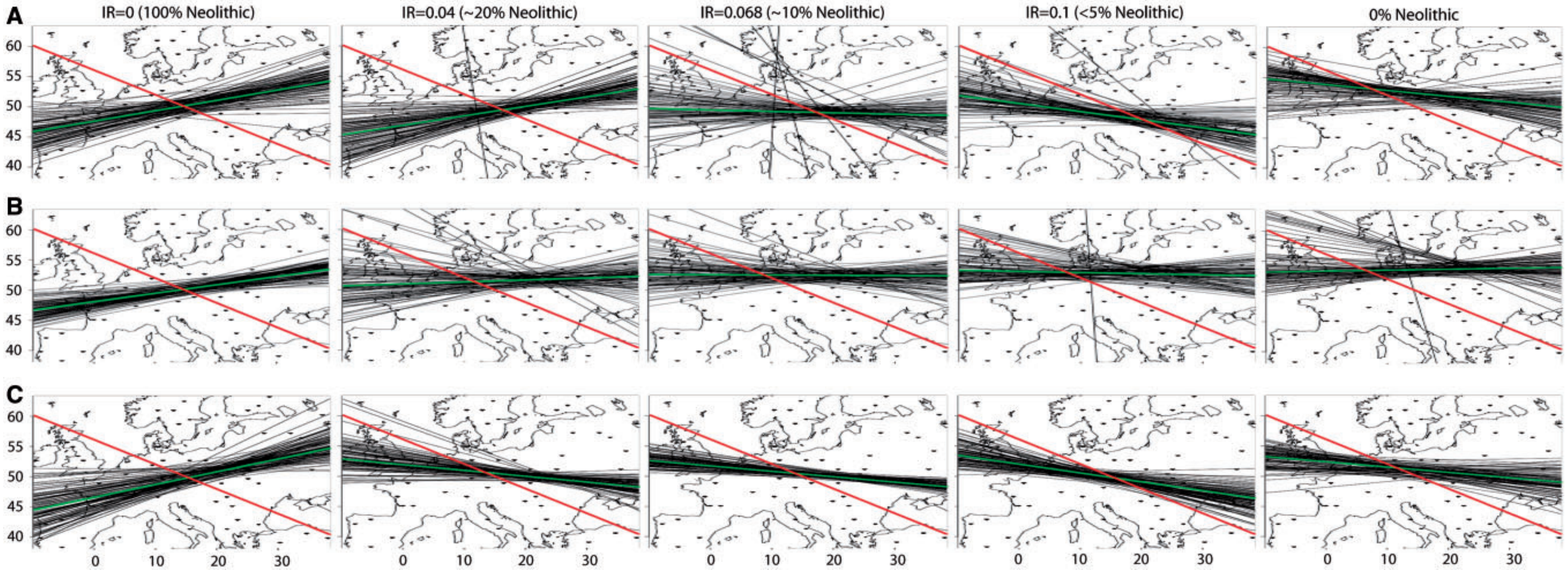

Fig. 2. Orientation of PC1 gradients in scenarios of double expansions from the Middle East with admixture between neolithic and paleolithic populations. The black lines represent PC gradient orientations (linking centroids of negative and positive PC1 coordinates, see Materials and Methods and fig. 1) for 100 replicates of the same scenario. The green line is the median of slopes and intercepts among replicates. The red line is the PC gradient orientation derived from the map of Piazza et al. (1995) shown in fig. 1C. (*A*) Pure range expansions. (*B*) Expansion-range contraction–re-expansion of the paleolithic populations, with a refuge area covering all southern Europe and active migrations to the south during the range expansion. (*C*) Same as (*B*), but with a refuge area restricted to the Iberian Peninsula. Scenarios with 0% neolithic correspond to a pure paleolithic expansion. IR, local interbreeding rate between paleolithic and neolithic populations. Gradient angles as a function of the proportions of total variance explained by PC1 are shown in supplementary fig. S6*A*, Supplementary Material online.

expansions seems due to the homogenization of genetic diversity after the paleolithic expansion, as a longer period of short-range migrations erases sectors formed during the expansion. This result is compatible with simulations showing that in absence of admixture, the age of the expansion is the main factor affecting the gradient axis (supplementary fig. S1, Supplementary Material online).

Consequently, levels of admixture between paleolithic and neolithic populations strongly influence the orientation of PC1 gradients (fig. 2*A*; supplementary fig. S3*A*, Supplementary Material online). With a neolithic contribution >20%, PC1 gradients remain SW–NE like for pure neolithic expansion, but as the paleolithic contribution increases, the direction of the PC1 gradient changes and becomes progressively NW–SE. Note that with limited (5%) neolithic admixture, the PC1 gradient has an even steeper NW–SE orientation than with a pure paleolithic expansion, which may be due to the existence of a gradient of admixture along the expansion axis (see Currat et al. 2008). Note also that a scenario of pure cultural diffusion (i.e., when the carrying capacity of paleolithic populations increases progressively from the Middle East, without involving any movement of neolithic populations) leads to PC gradients similar to those obtained under a pure paleolithic expansion (supplementary fig. S5, Supplementary Material online).

For unascertained SNP data (MAF = 0), PC1 gradients are randomly distributed (supplementary fig. S2, Supplementary Material online), probably because in this case most mutations have very low frequencies and are geographically restricted (François et al. 2010). Simulations performed with an intermediate MAF (1.5%), thus allowing for more recent mutations, also increases the variance of PC1 axis orientation (supplementary fig. S2, Supplementary Material online). Note that simulated microsatellite (STR) data show gradients similar to those obtained from SNP data with MAF >3% (supplementary fig. S2, Supplementary Material online).

Range contractions and posterior re-expansions (supplementary fig. S7, Supplementary Material online) have also a major impact on the PC1 gradient axes. When southern Europe is considered as a single large refugium, PC1 maps show E–W gradients (fig. 2*B*; supplementary fig. S3*B*, Supplementary Material online), but when the LGM refugium is restricted to the Iberian Peninsula, PC1 maps show steeper NW–SE gradients (fig. 2*C*; supplementary fig. S3*C*, Supplementary Material online). PC1 gradients are thus approximately perpendicular to the direction of the re-expansion axis, but this effect is only visible if the paleolithic contribution is large (>80%, fig. 2; supplementary fig. S3, Supplementary Material online) because the contraction and re-expansion episodes occur during the paleolithic (supplementary fig. S7, Supplementary Material online). Note that an absence of active migrations toward refuge areas during range contractions does not alter the mean PC gradients observed on figures 2*B* and *C* (compare with supplementary fig. S3*B* and *C*, Supplementary Material online), and that similar results are observed in scenarios with more recent postglacial re-expansions (14 kya, supplementary fig. S4, Supplementary Material online).

Our simulations thus show that range contractions during the LGM and admixture between neolithic and paleolithic humans have drastic effects on PC1 gradients, and suggest that very large levels of paleolithic ancestry are necessary to produce SE–NW PC1 gradients similar to those previously documented in Europe (e.g., Menozzi et al. 1978).

Note that we have not attempted here to quantitatively compare our results with the existing European large-scale SNP data, because the geographic orientation of the main axis of genetic differentiation is still controversial. Indeed, unlike the original analyses of Cavalli-Sforza, Lao et al. (2008) found an S–N axis of differentiation. The POPRES data set (Nelson et al. 2008) shows almost the same axis, but it is influenced by the presence of a few West Asian samples. Moreover, the main PCA axis changes drastically when Finns are included, showing that the main PCA axis might also be sensitive to the exact location and genetic differentiation of sampled populations.

In conclusion, even though it is difficult to infer history from PC components (Novembre and Stephens 2008), our simulation results show that a PC1 SE–NW cline is not compatible with a major contribution of neolithic populations into the gene pool of current Europeans, but with a major LGM refuge area for paleolithic populations in the Iberian peninsula, in line with previous mtDNA inferences (Pereira et al. 2005). They also stress the importance of the last ice age for the proper modeling of human evolution in Europe or in other continents. Our results have also implications for other species that underwent range contractions and posterior re-expansions, and show the necessity to properly locate these refugia to correctly interpret observed patterns of diversity.

Materials and Methods

Simulations of paleolithic and neolithic expansions were performed with the program SPLATCHE2 (Ray et al. 2010) using the same settings as François et al. (2010) (see supplementary material, Supplementary Material online). In addition, we have included the simulation of a range contraction toward southern Europe followed by a refugial isolation period and a re-expansion toward the North as outlined in supplementary figure S7, Supplementary Material online. The range contraction period started 25.5 kya and ended 21 kya, and consisted in a series of 26 progressive contraction events of ≈ 170 years each, during which a row of demes located in the most northern area became uninhabitable by setting its carrying capacity to zero (see Arenas et al. 2012). Additionally, the range contraction was simulated with either isotropic or anisotropic migration. For the latter, we imposed on the northern edge a strong migration toward the south ($m = 0.37$) and weak migration towards other directions ($m = 0.01$), such that hunter–gatherers had a larger probability to move toward refuge areas, as if they were sensitive to an environmental gradient. In all cases, migrations remained isotropic in the range core. After the contraction, populations remained in the refuge areas for 3,000 years, and could re-expand 18 kya or alternatively 14 kya (supplementary fig. S4, Supplementary Material online) from the refuge areas.

For all coalescent simulations, samples of 20 (haploid) individuals were collected from 60 locations as described in Francois et al. (2010). A data set of 100 SNP loci was simulated, either without ascertainment bias or by conditioning on a global MAF larger than 0.03 or 0.015; another dataset consisted in 100 STR loci simulated under a strict stepwise mutation model with mutation rate 5×10^{-4} per generation per locus. A total of 100 simulations were performed for each demographic scenario.

PCA was performed with the "prcomp" function of the R statistical package (further details are given in the supplementary material, Supplementary Material online). PC1 orientation was then assessed by connecting the geographical centroids of the positive and negative PC1 coordinates. This procedure gives a good indication of the gradients for PC1 (fig. 1), but is not reliable for other PCs where more complex PC surfaces might occur (Novembre and Stephens 2008).

Supplementary Material

Supplementary material and figures S1–S7 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank John Novembre and two anonymous reviewers for their helpful comments. This work was supported by grants EU/FP7 No. 226852 (SCALES) and Swiss NSF 3100-126074 to L.E. M.C. received financial support FROM the Swiss NSF (grants no. 31003A—127465) and the Marie Curie initial training network BEAN.

References

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

Arenas M, Ray N, Currat M, Excoffier L. 2012. Consequences of range contractions and range shifts on molecular diversity. *Mol Biol Evol.* 29:207–218.

Barbujani G, Bertorelle G. 2001. Genetics and the population history of Europe. *Proc Natl Acad Sci U S A.* 98:22–25.

Barbujani G, Chikhi L. 2006. Population genetics: DNAs from the European neolithic. *Heredity* 97:84–85.

Cavalli-Sforza LL, Edwards AWF. 1963. Analysis of human evolution. In: Geerts SJ, editor. Genetics today: proceedings of the 11th international congress of genetics, The Hague, The Netherlands. New York: Pergamon. p. 923–993.

Cavalli-Sforza LL, Menozzi P, Piazza A. 1994. The history and geography of human genes. Princeton (NJ): Princeton University Press.

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

Currat M, Ruedi M, Petit RJ, Excoffier L. 2008. The hidden side of invasions: massive introgression by local genes. *Evolution* 62:1908–1920.

Diamond J, Bellwood P. 2003. Farmers and their languages: the first expansions. *Science* 300:597–603.

Edmonds CA, Lillie AS, Cavalli-Sforza LL. 2004. Mutations arising in the wave front of an expanding population. *Proc Natl Acad Sci U S A.* 101:975–979.

Excoffier L, Ray N. 2008. Surfing during population expansions promotes genetic revolutions and structuration. *Trends Ecol Evol.* 23: 347–351.

François O, Currat M, Ray N, Han E, Excoffier L, Novembre J. 2010. Principal component analysis under population genetic models of range expansion and admixture. *Mol Biol Evol.* 27:1257–1268.

Hallatschek O, Hersen P, Ramanathan S, Nelson DR. 2007. Genetic drift at expanding frontiers promotes gene segregation. *Proc Natl Acad Sci U S A.* 104:19926–19930.

Jakobsson M, Scholz SW, Scheet P, et al. (20 co-authors). 2008. Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* 451:998–1003.

Klopfstein S, Currat M, Excoffier L. 2006. The fate of mutations surfing on the wave of a range expansion. *Mol Biol Evol.* 23:482–490.

Lao O, Lu TT, Nothnagel M, et al. (33 co-authors). 2008. Correlation between genetic and geographic structure in Europe. *Curr Biol.* 18: 1241–1248.

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

Nelson MR, Bryc K, King KS, et al. (23 co-authors). 2008. The Population Reference Sample, POPRES: a resource for population, disease, and pharmacological genetics research. *Am J Hum Genet.* 83:347–358.

Novembre J, Ramachandran S. 2011. Perspectives on human population structure at the cusp of the sequencing era. *Annu Rev Genomics Hum Genet.* 12:245–274.

Novembre J, Stephens M. 2008. Interpreting principal component analyses of spatial population genetic variation. *Nat Genet.* 40: 646–649.

Novembre J, Stephens M. 2010. Response to Cavalli-Sforza interview [Human Biology 82(3):245–266 (June 2010)]. *Hum Biol.* 82:469–470.

Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLoS Genet.* 2:e190.

Pereira L, Richards M, Goios A, et al. (13 co-authors). 2005. High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. *Genome Res.* 15:19–24.

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

Ray N, Currat M, Foll M, Excoffier L. 2010. SPLATCHE2: a spatially explicit simulation framework for complex demography, genetic admixture and recombination. *Bioinformatics* 26:2993–2994.

Richards M, Macaulay V, Hickey E, et al. (37 co-authors). 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet.* 67:1251–1276.

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

Straus LG. 1991. Southwestern Europe at the last glacial maximum. *Curr Anthropol.* 32:189–199.