Letter

Influence of admixture and Paleolithic range contractions on current European diversity gradients

Miguel Arenas*#, Olivier François[§], Mathias Currat°, Nicolas Ray[£] and Laurent Excoffier*#

* Computational and Molecular Population Genetics Lab (CMPG), Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, 3012 Berne, Switzerland.
 # Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland.
§ Laboratoire Techniques de l'Ingénierie Médicale et de la Complexité, Faculty of Medicine, University Joseph Fourier, Grenoble.
Institute of Technology, Centre National de la Recherche Scientifique UMR5525, La Tronche, France.
° Anthropology, Genetics and Peopling history Lab (AGP), Anthropology Unit - Department of Genetics & Evolution, University of Geneva, Rue Gustave-Revilliod 12, 1227 Geneva, Switzerland.
£ EnviroSPACE Laboratory, Climate Change and Climatic Impacts, Institute for Environmental Sciences, University of Geneva, 7 Route de Drize, Battelle D, 1227 Carouge, Switzerland.

Email addresses:

MA: miguel.arenasbusto@iee.unibe.ch
OF: olivier.francois@imag.fr
MC: mathias.currat@unige.ch
NR: nicolas.ray@unige.ch
LE: laurent.excoffier@iee.unibe.ch

Corresponding author:

Miguel Arenas
Computational and Molecular Population Genetics Lab (CMPG)
Institute of Ecology and Evolution
University of Bern
Baltzerstrasse 6
CH-3012 Bern, Switzerland
E-mail address: miguel.arenasbusto@iee.unibe.ch
Phone: +41 31 631 45 13
Fax: +41 31 631 48 88

1

Abstract

Cavalli-Sforza and colleagues (1963) initiated the representation of genetic relationships among human populations with principal component 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 a 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 a 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.

Principal component analysis (PCA) has been introduced by Cavalli-Sforza *et* al. (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, Price, and Reich 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, Piazza, and Cavalli-Sforza 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, Oden, and Wilson 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, Lillie, and Cavalli-Sforza 2004; Klopfstein, Currat, and Excoffier 2006; Excoffier and Ray 2008), which creates radiating geographic sectors of low diversity (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 (see 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 towards southern Europe, followed by a refugial isolation period and a posterior re-expansion. The effects of active migrations towards 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 SNP data with minimum allele frequency (MAF) > 3% generates PC1 gradients that are often perpendicular to the expansion axis in the case of a recent (Neolithic) expansion (Figure 1A and 2A). PC2 and PC3 maps usually highlight the British Isles and Scandinavia, due to their geographic isolation (see Figure 1A). For pure Paleolithic expansions, NW-SE and E-W gradients are usually observed in PC1 (Figure 1B and 2A), and PC2 and PC3 maps usually highlight the Iberian Peninsula (see an illustrative example in Figure 1B). The difference in PC1 gradient observed after Paleolithic and Neolithic 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 (Figure S1).

Consequently, levels of admixture between Paleolithic and Neolithic populations strongly influence the orientation of PC1 gradients (Figure 2A and S3A). 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 (see Figure S5).

For unascertained SNP data (MAF=0), PC1 gradients are randomly distributed (Figure S2), probably because in this case most mutations have very low frequencies and are geographically restricted (Francois 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 (see Figure S2). Note that simulated microsatellite (STR) data show gradients similar to those obtained from SNP data with MAF>3% (see Figure S2).

Range contractions and posterior re-expansions (Figure S7) 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 (Figures 2B and S3B), but when the LGM refugium is restricted to the Iberian Peninsula, PC1 maps show steeper NW-SE gradients (Figures 2C and S3C). 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%,

Figures 2 and S3) because the contraction and re-expansion episodes occur during the Paleolithic (see Figure S7). Note that an absence of active migrations towards refuge areas during range contractions does not alter the mean PC gradients observed on Figure 2B and 2C (compare with Figures S3B and S3C), and that similar results are observed in scenarios with more recent post-glacial re-expansions (14Kya, Figure S4).

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, Piazza, and Cavalli-Sforza 1978).

Note that we have not attempted here to quantitatively compare our results to existing European large-scale SNP data, since 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 a 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 past 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.

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 online supplementary material). In addition, we have included the simulation of a range contraction towards southern Europe followed by a refugial isolation period and a re-expansion towards the North as outlined in Figure S7. 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 towards the south ($m = 0.37$) and weak migrations towards other directions ($m = 0.01$), such that hunter-gatherers had a larger probability to move towards 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 (see Figure S4) 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 minor allele frequency (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). 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 (see Figure 1), but is not

reliable for other PCs where more complex PC surfaces might occur (Novembre and

Stephens 2008).

Supplementary Material

Supplementary figures S1-S7 are available at Molecular Biology and Evolution online

(http://www.mbe.oxfordjournals.org/).

Acknowledgments

Literature Cited

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

Arenas M, Ray N, Currat M, and Excoffier L. 2012. Consequences of Range Contractions and Range Shifts on Molecular Diversity. Mol Biol Evol. 29:207-218.

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

Barbujani G, and Chikhi L. 2006. Population genetics: DNAs from the European Neolithic. Heredity. 97:84-85.

Cavalli-Sforza LL, and Edwards AWF. 1963. Analysis of human evolution. Pp. 923-993 *in* Geerts SJ, ed. Genetics today: Proceedings of the 11th International Congress of Genetics. New York: Pergamon, The Hague, The Netherlands

Cavalli-Sforza LL, Menozzi P, and Piazza A. 1994. The history and geography of human genes. Princeton University Press, Princeton, New Jersey.

Chikhi L, Destro-Bisol G, Bertorelle G, Pascali V, and 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, and Excoffier L. 2008. The hidden side of invasions: massive introgression by local genes. Evolution. 62:1908-1920.

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

Edmonds CA, Lillie AS, and 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, and 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, and 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, and 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, Gibbs JR, VanLiere JM, Fung HC, Szpiech ZA, Degnan JH, Wang K, Guerreiro R, Bras JM, Schymick JC, Hernandez DG, Traynor BJ, Simon-Sanchez J, Matarin M, Britton A, van de Leemput J, Rafferty I, Bucan M, Cann HM, Hardy JA, Rosenberg NA, and Singleton AB. 2008. Genotype, haplotype and copy-number variation in worldwide human populations. Nature. 451:998-1003.

Klopfstein S, Currat M, and 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, Junge O, Freitag-Wolf S, Caliebe A, Balascakova M, Bertranpetit J, Bindoff LA, Comas D, Holmlund G, Kouvatsi A, Macek M, Mollet I, Parson W, Palo J, Ploski R, Sajantila A, Tagliabracci A, Gether U, Werge T, Rivadeneira F, Hofman A, Uitterlinden AG, Gieger C, Wichmann HE, Ruther A, Schreiber S, Becker C, Nurnberg P, Nelson MR, Krawczak M, and Kayser M. 2008. Correlation between genetic and geographic structure in Europe. Curr Biol. 18:1241-1248.

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

Nelson MR, Bryc K, King KS, Indap A, Boyko AR, Novembre J, Briley LP, Maruyama Y, Waterworth DM, Waeber G, Vollenweider P, Oksenberg JR, Hauser SL, Stirnadel HA, Kooner JS, Chambers JC, Jones B, Mooser V, Bustamante CD, Roses AD, Burns DK, Ehm MG, and Lai EH. 2008. The Population Reference Sample, POPRES: a resource for population, disease, and pharmacological genetics research. Am J Hum Genet. 83:347-358.

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

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

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

Pereira L, Richards M, Goios A, Alonso A, Albarran C, Garcia O, Behar DM, Golge M, Hatina J, Al-Gazali L, Bradley DG, Macaulay V, and Amorim A. 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, and 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, and 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, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, Villems R, Thomas M, Rychkov S, Rychkov O, Rychkov Y, Golge M, Dimitrov D, Hill E, Bradley D, Romano V, Cali F, Vona G, Demaine A, Papiha S, Triantaphyllidis C, Stefanescu G, Hatina J, Belledi M, Di Rienzo A, Novelletto A, Oppenheim A, Norby S, Al-Zaheri N, Santachiara-Benerecetti S, Scozari R, Torroni A, and Bandelt HJ. 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet. 67:1251-1276.

Sokal RR, Oden NL, and 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.

Figures

Figure 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.

Figure 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 Methods and Figure 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-reexpansion 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 corresponds 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 Figure S6A.
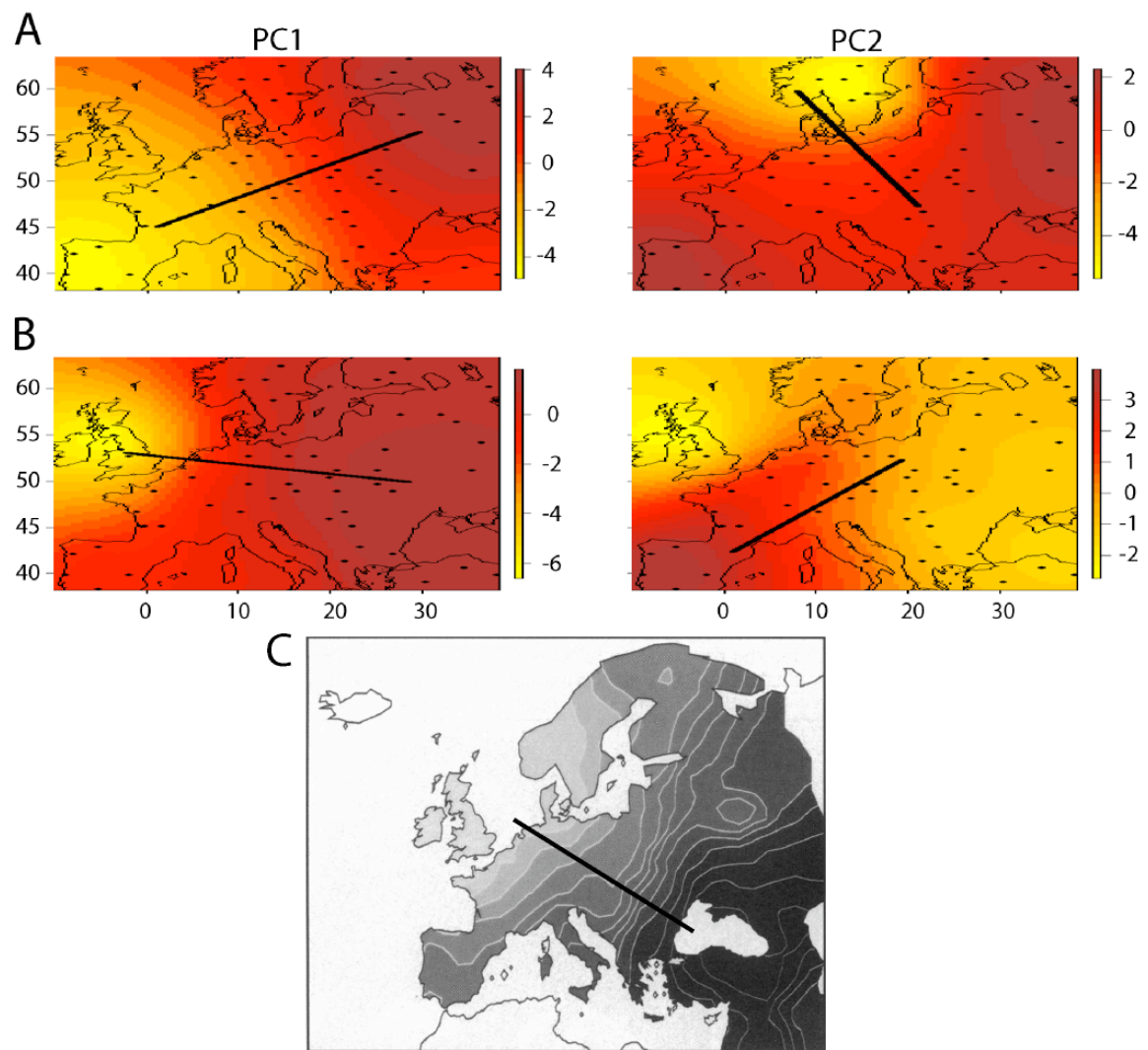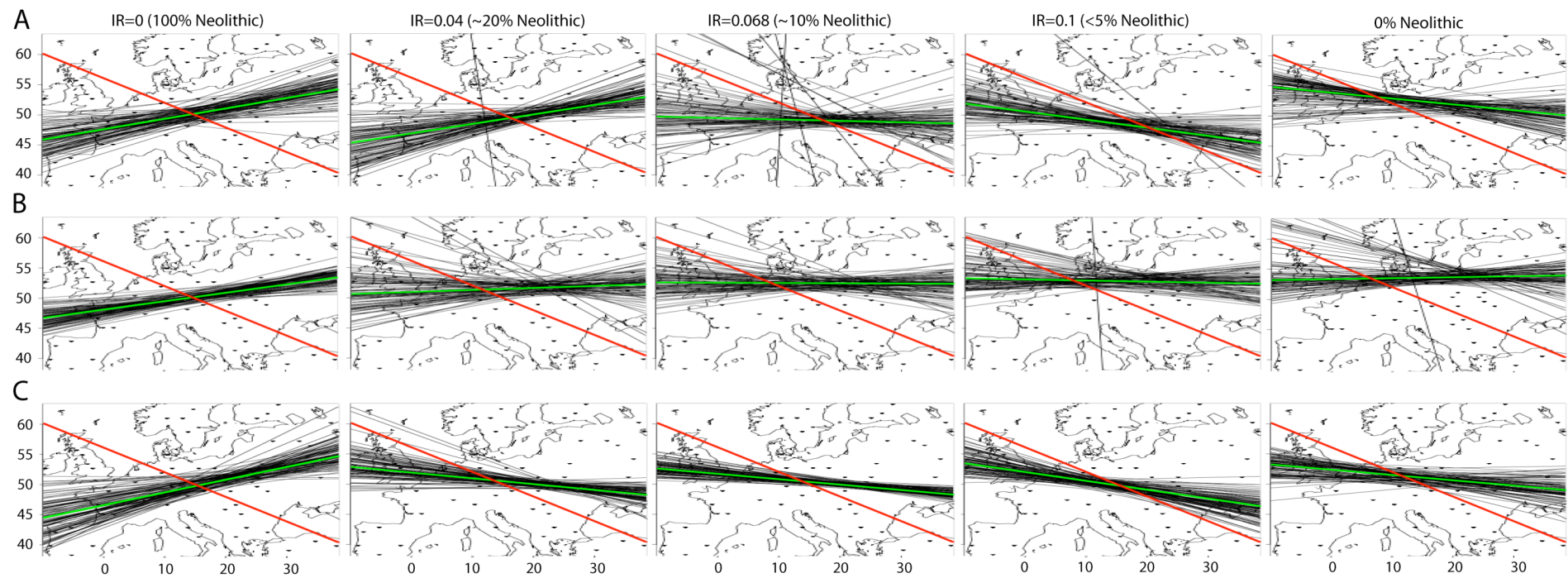
Figure 1.

Figure 2.