



Consequences of diverse evolutionary processes on american genetic gradients of modern humans

Catarina Branco^{1,2,3,4} · Miguel Velasco⁵ · Macarena Benguigui⁵ · Mathias Currat^{6,7} · Nicolas Ray^{8,9} · Miguel Arenas^{1,2,3,5}

Received: 26 April 2018 / Revised: 2 July 2018 / Accepted: 3 July 2018 / Published online: 18 July 2018
© The Genetics Society 2018

Abstract

European genetic gradients of modern humans were initially interpreted as a consequence of the demic diffusion of expanding Neolithic farmers. However, recent studies showed that these gradients may also be influenced by other evolutionary processes such as population admixture or range contractions. Genetic gradients were observed in the Americas, although their specific evolutionary causes were not investigated. Here we extended the approach used to study genetic gradients in Europe to analyze the influence of diverse evolutionary scenarios on American genetic gradients. Using extensive computer simulations, we evaluated the impact of (i) admixture between expansion waves of modern humans, (ii) the presence of ice-sheets during the last glacial maximum (LGM) and (iii) long-distance dispersal (LDD) events, on the genetic gradients (detected by principal component analysis) of the entire continent, North America and South America. The specific simulation of North and South America showed that genetic gradients are usually orthogonal to the direction of range expansions—either expansions from Bering or posterior re-expansions to recolonize northern regions after ice sheets melting—and we suggest that they result from *allele surfing* processes. Conversely, our results on the entire continent show a northwest-southeast gradient obtained with any scenario, which we interpreted as a consequence of isolation by distance along the long length of the continent. These findings suggest that distinct genetic gradients can be detected at different regions of the Americas and that subcontinent regions present gradients more sensible to evolutionary and environmental factors (such as LDD and the LGM) than the whole continent.

Introduction

Genetic diversity among populations across a landscape can sometimes be distributed along gradients. Genetic diversity

gradients can be observed using principal component analysis (PCA), as this approach can summarize information from large genetic datasets (Novembre and Stephens 2008). For instance, Cavalli-Sforza and Edwards (1963) applied PCA on gene frequencies to study the European genetic diversity of modern humans. They found a PC1 gradient (genetic gradient observed with the first principal component) along a southeast-northwest (SE-NW) axis of Europe,

Electronic supplementary material The online version of this article (<https://doi.org/10.1038/s41437-018-0122-x>) contains supplementary material, which is available to authorized users.

✉ Miguel Arenas
marenas@uvigo.es

¹ Department of Biochemistry, Genetics and Immunology, University of Vigo, Vigo, Spain

² Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP), Porto, Portugal

³ Instituto de Investigação e Inovação em Saúde (i3S), Universidade do Porto, Porto, Portugal

⁴ Department of Genetics and Biotechnology, University of Trás-os-Montes and Alto Douro (UTAD), Vila Real, Portugal

⁵ Centre for Molecular Biology “Severo Ochoa”, Consejo Superior de Investigaciones Científicas (CSIC), Madrid, Spain

⁶ Anthropology, Genetics and Peopling History Lab, Department of Genetics & Evolution -Anthropology Unit, University of Geneva, Geneva, Switzerland

⁷ Institute of Genetics and Genomics in Geneva (IGE3), University of Geneva, Geneva, Switzerland

⁸ EnviroSPACE Lab, Institute for Environmental Sciences, University of Geneva, Geneva, Switzerland

⁹ Institute of Global Health, University of Geneva, Geneva, Switzerland

a pattern which was later confirmed by posterior studies (Menozzi et al. 1978; Piazza et al. 1995; Sokal et al. 1989). This finding was explained as an effect of demic diffusion of initial Neolithic farmers during their range expansion from the Near East (Chikhi et al. 2002; Sokal and Menozzi 1982), together with the replacement of Paleolithic hunter-gatherer populations with little or without admixture (Currat and Excoffier 2005; Sokal et al. 1991). However, posterior studies asserted that the orientation of PC gradients could be influenced by additional processes. Novembre and Stephens (2008) showed that PC gradients can even be obtained without any range expansion. Next, François et al. (2010) and Arenas et al. (2013) demonstrated that PC gradients can be orthogonal to the direction of the expansion as a consequence of *allele surfing* (Edmonds et al. 2004; Excoffier and Ray 2008), where surfing mutations occurring on the wave of advance of the range expansion generate highly differentiated genetic sectors (this process is more visible with small population sizes and low migration rates (Klopfstein et al. 2006), especially after recent expansion where sectors were not yet removed through genetic homogenization (Excoffier and Ray 2008)). Moreover, the level of admixture between Neolithic and Paleolithic populations can also alter the orientation of PC gradients (Arenas et al. 2013; François et al. 2010), although, in practice, there is not yet a consensus concerning the real level of this admixture in Europe (Barbujani and Chikhi 2006; Chikhi et al. 1998; Currat and Excoffier 2005; Skoglund et al. 2014). In addition, Paleolithic range contractions resulting from the last glacial maximum (LGM) period (Straus 1991) and posterior re-expansions (Barbujani and Bertorelle 2001) could also influence these genetic gradients (Arenas et al. 2013; Arenas et al. 2014).

All above cited studies focused on European populations and little is known about genetic gradients in other continents. Regarding the Americas, genetic gradients or clines derived from real data commonly present a clear northwest-southeast (NW–SE) orientation (Cavalli-Sforza et al. 1993; Cavalli-Sforza et al. 1994). Concerning North America, most of studies also showed a NW–SE orientation based on extensive genetic data (Cavalli-Sforza et al. 1994; Salas et al. 2009) and based on the geographic distribution of linguistic families and subfamilies (Cavalli-Sforza et al. 1994). In South America, first genetic variation analysis by O'Rourke and Suarez (1986) showed a lack of geographic structure, although this analysis was based on only two summary statistics computed from a very limited number of populations. Cavalli-Sforza et al. (1994), based on the presence of linguistic families and subfamilies and on genetic data, showed a northeast-southwest (NE–SW) genetic gradient. A similar gradient based on mitochondrial DNA haplogroup patterns was identified by Salas et al. (2009).

Here we studied the influence of diverse evolutionary processes and environmental factors on PC gradients observed in the Americas, which to our knowledge was not studied so far. In the Americas, we identified some processes or factors that could affect PC gradients (Supplementary Table S1; Supplementary Material): (1) the presence of ice sheets during the LGM, (2) population admixture between different expanding population waves, and (3) migration with long-distance dispersal (LDD) events.

(1) As a consequence of the LGM, North America presented two large ice sheets (Laurentide and Cordilleran) that could affect the entry of modern humans in North America (Marshall et al. 2002; Ray and Adams 2001). Hence, two entry routes to the Americas have been proposed and highly discussed, a coastal route following North Pacific coastlines and an inland route (ice-free corridor) at the eastern side of the Rocky Mountains (Bodner et al. 2012; Fagundes et al. 2008; Pedersen et al. 2016). In addition, ice sheets could have induced temporal ice-free refugia and posterior expansions to colonize northern regions after melting (Rogers et al. 1991).

(2) There is clear evidence of more than one founding populations in the Americas (Reich et al. 2012; Skoglund et al. 2015) with complex admixture among them during the Lithic stage (e.g., Reich et al. 2012; Skoglund et al. 2015).

(3) A few studies suggested that the expansion of modern humans through the world could involve long-distance dispersal (LDD) events (Ray and Excoffier 2010), for example, traveling by boats (Balme 2013). In this regard, a recent study on the settlement of Eurasia by modern humans found that evolutionary scenarios accounting for LDD better fitted real data than evolutionary scenarios without LDD (Alves et al. 2016).

Following previous works (Arenas et al. 2013; François et al. 2010), we performed spatially explicit computer simulations to explore the influence of the cited factors (i.e., ice sheets during LGM, population admixture and LDD) on the observed American genetic gradients. We separately analyzed the entire continent, North America and South America to examine the influence of these processes at different scales.

Materials and methods

We performed spatially-explicit computer simulations with the program *SPLATCHE2* (Ray et al. 2010) to mimic the settlement of the Americas under diverse evolutionary and environmental scenarios. Usually spatially-explicit models provide more realistic simulations than models ignoring the geography of the landscape (Benguigui and Arenas 2014; Ray and Excoffier 2009). Basically, *SPLATCHE2* simulates

genetic data following three main steps: (i) A forward in time simulation of the evolutionary history of the entire population accounting for spatial and demographic information, (ii) a coalescent reconstruction of the evolutionary history of a sample, which is embedded in the previously obtained history of the entire population and (iii) a simulation of molecular evolution over the coalescent genealogy of the sample to generate the genetic data of the samples (Arenas 2012; Yang 2006).

Using geographic information systems (GIS) we generated a two-dimensional lattice of 94×144 demes covering the Americas (3,815 are inland), where each deme corresponds to $100 \times 100 \text{ km}^2$ (Supplementary Figure S13; Supplementary Material). An equal-area projection was used for the Americas to minimize deme area distortion. We considered a deme located at the Bering Strait to start the colonization. Next, demes exchange migrants with neighboring demes under a two-dimensional stepping-stone migration model (Kimura and Weiss 1964). Some parameters related to the evolution of modern humans (i.e., generation time and growth rate) followed previous works based on Eurasian populations (Alves et al. 2016; Arenas et al. 2013; François et al. 2010; Pimenta et al. 2017), although we increased the migration rate to simulate a colonization of the Americas rapid enough to fit with the current knowledge (details for all the applied parameters are shown in Supplementary Material). Indeed, the population size at the onset of the colonization of the Americas followed previous works based on American populations (Fagundes et al. 2018; Gravel et al. 2013; Hey 2005; Kitchen et al. 2008; Mulligan et al. 2004) (details in Supplementary Material).

We simulated several scenarios by combining diverse processes (Supplementary Table S1, Supplementary Material) that we hypothesized to be influential on current American genetic gradients:

(1) Colonization of the landscape by one or by two populations. The exact number and timing of expansions on the Americas remains uncertain and under debate. Following some studies, here we simulated an initial colonization wave (hereafter called "first" expansion) from western of current Alaska at 18 kya (Dillehay 2009). Under scenarios with admixture, an additional Amerindian wave (hereafter called "second" expansion) was simulated from the same origin at 11 kya (beginning the Holocene) (Forster et al. 1996) (Supplementary Figures S14–S25, Supplementary Material). These two waves represent possible expansion waves of Amerindians (Dillehay 2009; Forster et al. 1996). Under two populations we explored several levels of admixture (contribution of the second population to the final genetic pool: 100%, $\approx 12\%$, $\approx 6\%$, $< 5\%$; further details in Supplementary Material). Under two populations with admixture, the second population replaces the first

population due to a competitive advantage (higher carrying capacity) but genetic lineages from the first population could still be present at different amount in the final genetic pool according to the level of admixture (regulated by the interbreeding rate IR). If there is not admixture ($IR = 0$), then the second population still replaces the first population due to competition, but the first population does not contribute genetically because there is no interbreeding between them. Under two populations, genetic samples are always taken in the second population.

(2) Colonization of the landscape considering and ignoring ice sheets derived from the LGM. Under this scenario we simulated two ice sheets in North America by setting the carrying capacity of the demes covered by ice to zero (Arenas et al. 2012) from 18 kya to 10 kya, according to estimates of the duration of ice sheets, frozen grounds and subsequent inundations (Bodner et al. 2012) (Supplementary Figures S16–S19 and S22–S25). Under ice-sheet scenarios, we allowed north to south coastal and/or inland corridors of 1–2 demes (100–200 km) width. We assumed an absence of ice sheets in South America at the arrival of modern humans to this region. The Andean Mountains presented an ice sheet during the LGM (Ray and Adams 2001) but it is assumed that first modern humans arrived there after the LGM.

(3) Colonization of the landscape considering and ignoring LDD. LDD events were simulated with the model developed by Ray and Excoffier (2010) (further details in Supplementary Material) and according to a LDD distribution estimated from human data (Novembre et al. 2005). Next, we applied a LDD proportion of 0.05 (Alves et al. 2016; Mona et al. 2014) and we considered 10 demes (1,000 km) as a maximum distance of dispersal per generation (Alves et al. 2016) (Figures S15, S19, S21 and S25).

For each scenario (Supplementary Table S1), we performed a total of 100 simulations. Each simulation generated genetic samples of 20 (haploid) individuals in 91 sampling locations spatially distributed along the entire continent (Supplementary Figure S13, Supplementary Material). Following Arenas et al. (2013), each individual included 100 independent SNP loci by conditioning on a global minor allele frequency (MAF) larger than 0.03 to avoid getting too many rare regionally private alleles (Arenas et al. 2013; François et al. 2010). Indeed, we simulated another dataset with 100 STR loci evolved under a strict stepwise mutation model with mutation rate 5×10^{-4} per generation per locus (Arenas et al. 2013).

PCA was computed with the "prcomp" function of the R software environment. A PC gradient was included on the PC1 map of each analyzed dataset by connecting the geographical centroids of the positive and negative coordinates and next, we computed the median over the 100 PC1 gradients generated under each evolutionary scenario. The

goals of applying PCA rather than traditional summary statistics are that PCA considers global patterns of genetic variation caused by multiple factors (ancestral expansions and admixture, ancestral range contractions, irregular migration, irregular sampling, among others) (François et al. 2010; Novembre and Stephens 2008) and allows extrapolation for inferences in non-sampled regions (Arenas et al. 2013; François et al. 2010).

Results

Genetic gradients over the entire American continent are not influenced by the studied population admixture, LGM or LDD

Genetic data simulated along the entire continent generated a NW–SE PC1 gradient that was persistent to the studied presence or absence of ice sheets, LDD events, level of population admixture and genetic marker (Fig. 1 and Supplementary Figures S1–S4, Supplementary Material). This gradient follows the axis of population expansions and corresponds to the gradient commonly observed in real data (Fig. 1). The second principal component (PC2), and subsequent PCs, describes regions with some isolation (Arenas et al. 2013; François et al. 2010). Our inferred PC2 maps for the entire continent showed several regions under this situation (Supplementary Figure S5, Supplementary Material): Alaska, the Labrador Peninsula, Central America and Patagonia. Additionally, in absence of admixture, the second expansion showed some isolation of regions located at the northeast of Brazil.

Genetic gradients over North America are influenced by the studied population admixture, LGM and LDD

In contrast to the changeless pattern generated in the entire continent, PC1 gradients obtained in North America varied under different combinations of considering or ignoring ice sheets, population admixture and LDD.

With the first expansion only, the absence of ice sheets resulted in NE–SW gradients (Fig. 2a, right column), while the presence of ice sheets (under either the coastal, the inland or both corridors) leads to different gradients with NW–SE orientation (Fig. 2b–d, right column).

Similarly, when two populations are considered we found that PC1 gradients depend on the amount of admixture between them and the presence or absence of ice sheets (Fig. 2). If ice sheets are not considered, PC1 gradients for lack or any level of admixture present a NE–SW orientation (Fig. 2a). By contrast, if ice sheets are considered (under either the coastal, the inland or both corridors), the second

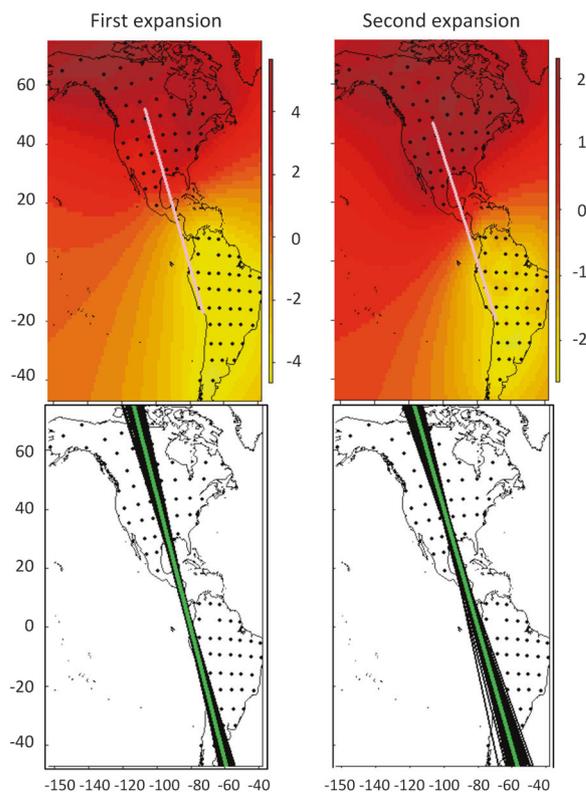


Fig. 1 Genetic gradients derived from pure first and second Paleolithic range expansions from western Alaska along the American continent. (Left) Illustrative example of PC1 map obtained from a simulated old single first expansion. (Right) Illustrative example of PC1 map obtained from a simulated recent pure second expansion. (Below) PC1 maps for single first and second expansions derived from 100 computer simulations. Each black line represents the PC1 gradient (connecting positive and negative centroids) from an independent computer simulation. The green line represents the median of slopes and intercepts among the 100 simulations. These gradients can be compared with gradients derived from real data (Cavalli-Sforza et al. 1993; Cavalli-Sforza et al. 1994)

expansion alone (without admixture) also generates a PC1 gradient with orientation NE–SW (Fig. 2b–d, left column), while under any level of contribution of the first population to the final genetic pool we found PC1 gradients with orientation NW–SE (Fig. 2b–d, right columns). Similar results are obtained under different levels of carrying capacity (Supplementary Figure S6; Supplementary Material).

The consideration of LDD generated PC1 gradients that differ between considering a single second expansion or any level of admixture between the two expanding populations with or without ice sheets (Supplementary Figure S8, Supplementary Material). Interestingly, the scenario with admixture presented an opposite orientation with (NW–SE) or without LDD (NE–SW) when ice sheets are not considered (Fig. 2a). Additionally, we noted that LDD increased the variance of PC1 gradients among simulations

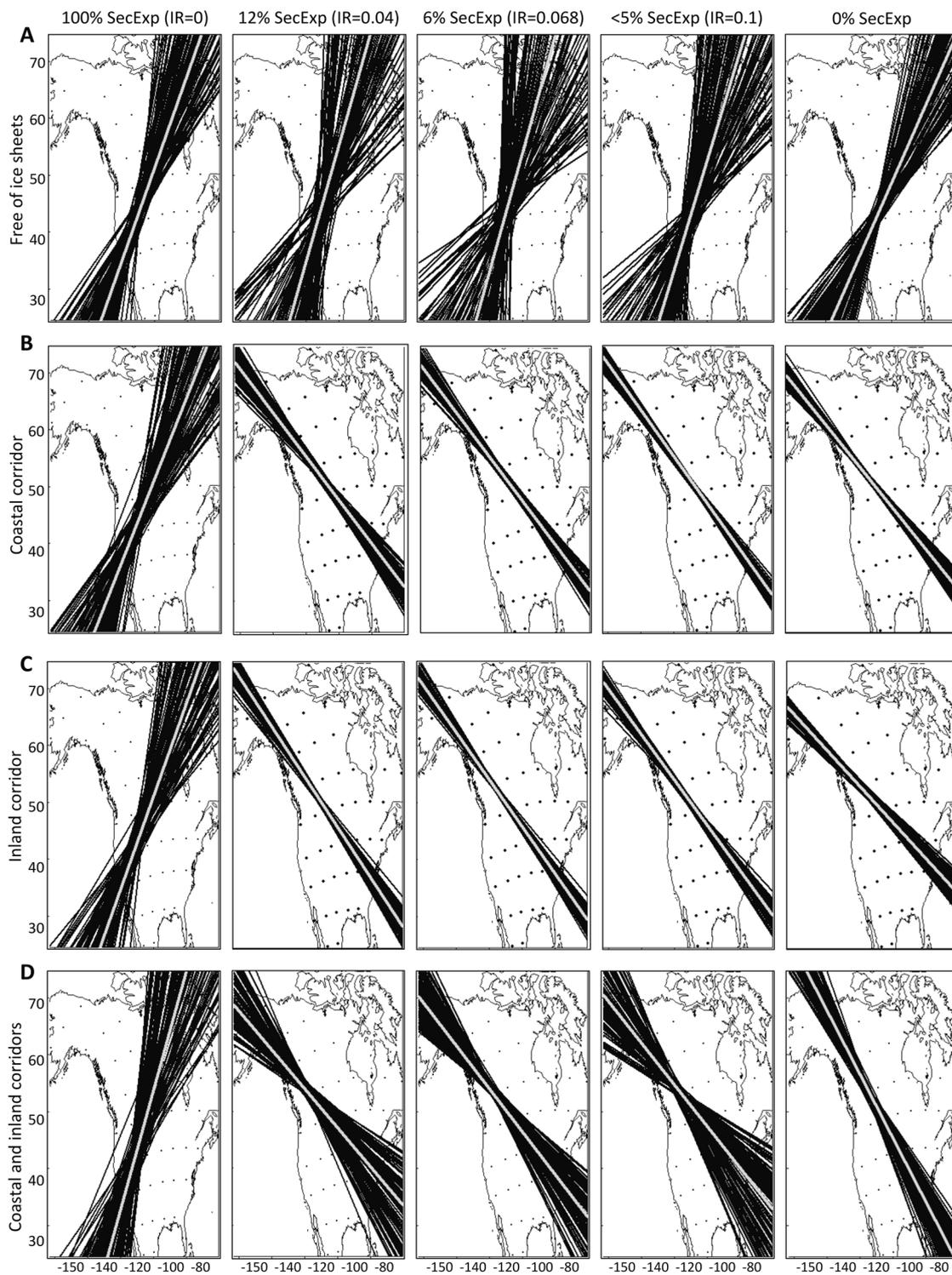


Fig. 2 PC1 gradients of North America considering different levels of admixture and with or without ice sheets. Each black line represents the PC1 gradient orientation (connecting positive and negative centroids) from an independent computer simulation. The green line represents the median of slopes and intercepts among the 100 simulations. Scenario of 0% of second expansion (SecExp) indicates a single first expansion. IR: local Interbreeding Rate between both populations.

a Range expansions without ice sheets. **b** Range expansions with ice sheets considering the coastal corridor. **c** Range expansions with ice sheets considering the inland corridor. **d** Range expansions with ice sheets considering both coastal and inland corridors. Diverse illustrative examples of these PC1 gradients are shown in the Supplementary Figure S7

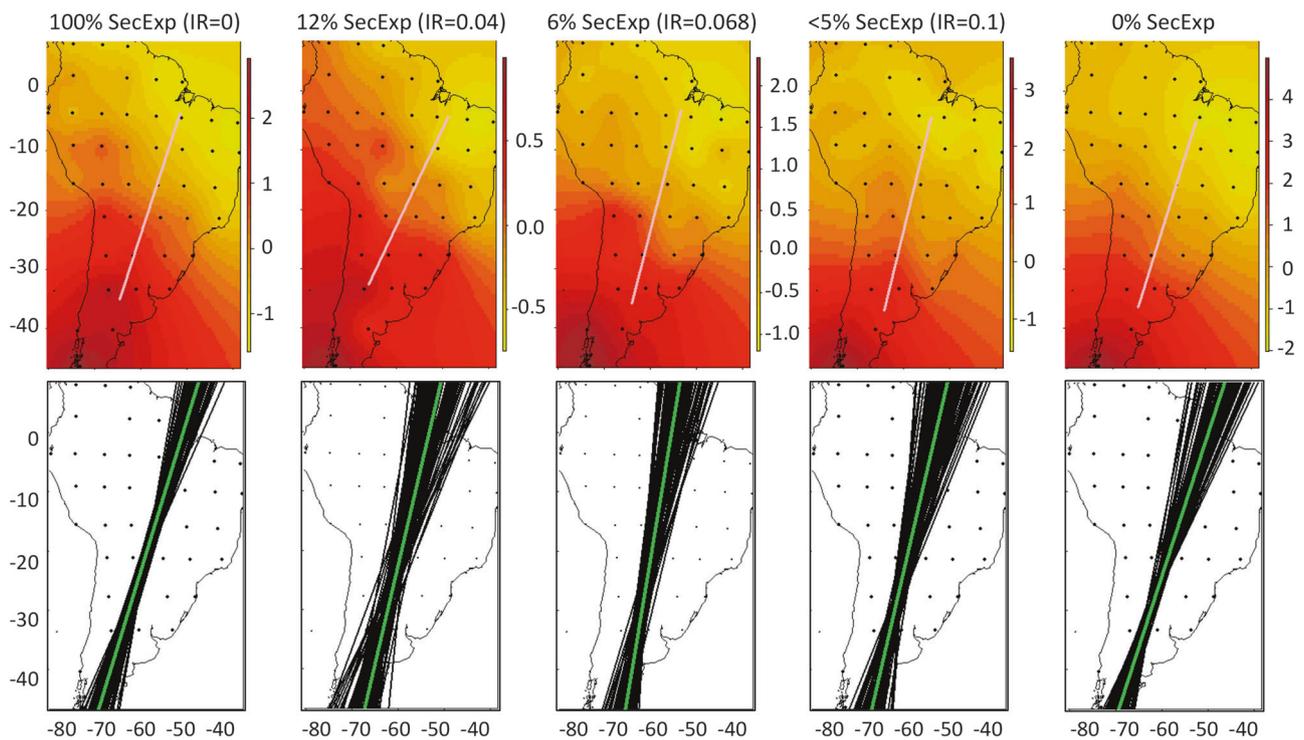


Fig. 3 PC1 gradients of South America considering different levels of population admixture. (Above) Illustrative example of PC1 maps obtained from a single simulation. (Below) PC1 maps derived from 100 computer simulations. Each black line represents the PC1 gradient orientation (connecting positive and negative centroids) from an

independent computer simulation. The green line represents the median of slopes and intercepts among the 100 simulations. Scenario of 0% of second expansion (SecExp) indicates a single first expansion. IR: local Interbreeding Rate between both populations

(Supplementary Figure S8). The gradient commonly observed in real data (NW–SE) (Cavalli-Sforza et al. 1994) may thus only be reproduced by scenarios including the presence of ice sheets with any contribution from the first expansion or alternatively, scenarios with LDD and any contribution from the first expansion, the latter with more variance.

Concerning PC2 maps, we found that three regions appear isolated: Alaska, the Labrador Peninsula and Central America (Figure S9, Supplementary Material).

Genetic gradients over South America are not influenced by the studied population admixture or LDD

All studied evolutionary scenarios, including different levels of population admixture and presence or absence of LDD, generated NE–SW PC1 gradients along South America (Fig. 3 and Supplementary Figures S10–S11, Supplementary Material). This gradient is almost orthogonal to the axis of expansion and corresponds to the one commonly observed with real data. Estimated PC2 maps highlighted north (Central America), south (Patagonia) and east (northeast of Brazil) regions of South America (Supplementary Figure S12, Supplementary Material).

Discussion

Influence of evolutionary and environmental factors on genetic gradients

Our results showed contrasting PC1 gradients depending on the studied geographic scale. We suggest that this finding can be derived from the conjunction of different evolutionary factors at different spatial scales.

Genetic gradients over the entire continent followed a NW–SE orientation (along the general orientation of the population expansions) and were not influenced by any studied evolutionary process or environmental factor (presence or absence of ice sheets, population admixture, presence or absence of LDD). We suggest that this persistent gradient is mainly due to an isolation by distance (IBD) process that could be induced by the long NW–SE distance of the continent, which is in agreement with Kanitz et al. (2018). Although serial founder effects (SFE) can also generate a genetic diversity gradient following the orientation of the range expansion (Barbujani et al. 1995), our inferred gradient was invariable to LDD events (which can lead to range expansions with different orientations) and to expansion origins in other locations of the landscape (i.e., east of South America, results not shown). For these two

reasons, we favor IBD over SFE due to population expansion.

By contrast, PC1 gradients in North America only were much more sensible to the underlined evolutionary processes and environmental factors. Ignoring ice sheets leads to PC1 gradients with NE–SW orientation, almost orthogonal to the direction of the range expansion, which is invariable with the level of population admixture. We explain these gradients by an “allele surfing” process considering that these expansions are very recent (18–11 kya), with dates relatively close to the onset of the Neolithic range expansion in Europe (10 kya), where *allele surfing* would have prevailed in absence of admixture with previous populations (Arenas et al. 2013; François et al. 2010), and therefore without enough time to homogenize genetic sectors. Next, scenarios allowing a contribution from the first population combined to the presence of ice sheets, present PC1 gradients with NW–SE orientation. One explanation is that this orientation is due to an *allele surfing* process during the first expansion, which has a large influence. Because the first expansion is restricted to a coastal corridor due to the presence of ice sheets, it thus does not result in the creation of sectors and to orthogonal differentiation. An alternative explanation is that the NW–SE gradient was generated during a posterior range expansion to colonize northern regions that become habitable after melting of the ice sheets (i.e., Supplementary Figures S16–S19 and S22–S25). Scenarios with LDD presented a larger variance of PC1 gradients among simulations probably because the direction of LDD events is chosen randomly (Alves et al. 2016; Ray and Excoffier 2010). Interestingly, while single second expansions with LDD showed a PC1 gradient with NE–SW orientation, scenarios with population admixture presented PC1 gradients with NW–SE orientation. Our explanation is that LDD tends to homogenize genetic diversity over the space and thus tend to blur the gradients (Mona et al. 2014). Thus, while the single second expansion under LDD could still keep some genetic signatures of *allele surfing*, admixed populations (either with or without ice sheets), which have more ancestral genetic information derived from the first expansion (and therefore LDD operated during a longer time period), may have had time to remove genetic sectors through LDD leading to gradients that reflect IBD.

PC1 gradients in South America presented a NE–SW orientation that was invariable with the admixture and the presence or absence of LDD and similar to the gradient found with real data. We suggest that these gradients are derived from an *allele surfing* process because this range expansion is more recent and thus sectors derived from *allele surfing* can still be present in the current population.

Our estimated PC2 gradients proposed several regions with some genetic isolation that could be simply explained

with the geography (i.e., geographic isolation) of the landscape.

Fitting simulated genetic gradients with genetic gradients derived from real data

Here we did not attempt to quantitatively contrast our inferences with the American genetic gradients derived from real data. Actually, the orientation of the main gradient in the Americas is similar among studies. However, qualitative comparisons with genetic gradients obtained from real data show that our PC1 gradient along the entire continent, independently of the simulated evolutionary scenario, is a good approximation of the gradient commonly observed with real data (Fig. 1) (Cavalli-Sforza et al. 1993; Cavalli-Sforza et al. 1994; Salas et al. 2009). Consequently, this result indicates that all studied evolutionary scenarios similarly fit the real data and that PC1 gradients do not bring information to distinguish between the studied processes. Next, our estimated PC2 maps presented isolated regions (i.e., the Labrador Peninsula and Alaska, which could be caused by geographic isolation) that also were described in PC2 maps derived from real data (Cavalli-Sforza et al. 1993; Cavalli-Sforza et al. 1994).

Concerning North America, NW–SE (Cavalli-Sforza et al. 1993; Cavalli-Sforza et al. 1994; Salas et al. 2009) gradients were found for real data. Our results indicate that such a gradient can be obtained from scenarios that consider ice sheets and/or LDD with any contribution from the first expansion (Fig. 2, Supplementary Figures S6 and S8).

In South America, genetic gradients derived from real data disagree. First studies showed a lack of geographic structure (O’Rourke and Suarez 1986). Later, Cavalli-Sforza et al. (1994) found a NE–SW genetic gradient based on genetic data [see also (Salas et al. 2009)] and on the geographical location of a variety of linguistic families and subfamilies, which could be favored by the Andes and that is similar to our inferred PC1 gradient (Fig. 3). Here we found that the orientation of the PC1 gradient is invariable with the level of population admixture or the presence/absence of LDD. All studied scenarios therefore similarly fit with real genetic gradients and no particular scenario can be preferred. It means the PC1 gradient is not informative to distinguish between the studied evolutionary scenarios. Indeed, Cavalli-Sforza et al. (1993; 1994) and Salas et al. (2009) showed that the northeast region of this continent might present some differentiation, which was also identified with our PC2 map (Supplementary Figure S12). Indeed, Cavalli-Sforza et al. (1994) also identified Patagonia as an isolated region in their PC3–PC5 maps estimated from real data, which was also found in our PC2 maps derived from simulated data.

In agreement with previous studies (Benguigui and Arenas 2014; Novembre and Stephens 2008), we conclude that spatially-explicit computer simulations followed by PCA can be useful to study the influence of evolutionary and environmental factors on genetic gradients by exploring the range of scenarios leading to the observed patterns. We conclude that the genetic gradient of modern humans over the entire American continent—following the direction of the range expansion—was not influenced by any studied factor (presence of ice sheets generated during the LGM, population admixture and migration including LDD events) under the considered assumptions (parameters and models) and we suggest that this gradient was rather caused by IBD than by just successive founder events. Indeed, we suggest that the long length of this continent played an important role in the genetic gradient displayed by the PC1, favoring the genetic contribution of IBD. Of course, the expansions over the whole American continent were likely to generate *allele surfing* and because they are recent, the signal could still be there at the subcontinent scale, as found in European expansions (Arenas et al. 2013; François et al. 2010). However, in addition to that, the long N–S distance of this continent can impose an IBD process increasing genetic differentiation between northern and southern populations, which thus blur the effect of any evolutionary or environmental factor investigated in this study.

Nevertheless, when exploring only North America, our genetic gradients were usually orthogonal to the direction of the most recent range expansion (either initial expansion from Bering or expansions to colonize regions previously covered by ice-sheets with contribution of the first population to the final genetic pool), and consequently we suggest that they could be caused mostly by *allele surfing*. In South America, the orientation of the simulated genetic gradient—orthogonal to the direction of the range expansion—was also not influenced by the level of population admixture or LDD, suggesting the effect of *allele surfing*.

In all, we suggest that PC1 gradients differ among different landscapes (north, south or entire continent) because the underlined evolutionary forces are constrained by features of the landscape. Future studies could consider more complex scenarios to refine our results, namely by counting for present and ancestral vegetation maps and local geographical barriers such as relevant rivers and mountain ranges. Concerning the modeling of human evolution, multiple expanding waves could be envisaged to reflect the influence of European populations regarding admixture (Hunley and Healy 2011) and demographics (Lindo et al. 2016; O’Fallon and Fehren-Schmitz 2011) (here we focus on indigenous Americans), more realistic demographic parameters (i.e., migration rate and carrying capacity varying with time) and selection will also be

useful. Hopefully these scenarios will be incorporated into future spatially-explicit computer simulators.

Acknowledgements MA was initially supported by the fellowship “Juan de la Cierva” JCI-2011-10452 from the Spanish Government and by the FCT Starting Grant IF/00955/2014 from the Portuguese Government, and finally supported by the fellowship “Ramón y Cajal” RYC-2015-18241 from the Spanish Government.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Alves I, Arenas M, Currat M, Sramkova Hanulova A, Sousa VC, Ray N et al. (2016) Long-distance dispersal shaped patterns of human genetic diversity in Eurasia. *Mol Biol Evol* 33:946–958
- Arenas M (2012) Simulation of molecular data under diverse evolutionary scenarios. *PLoS Comput Biol* 8:e1002495
- Arenas M, Francois O, Currat M, Ray N, Excoffier L (2013) Influence of admixture and paleolithic range contractions on current European diversity gradients. *Mol Biol Evol* 30:57–61
- Arenas M, Mona S, Trochet A, Sramkova Hanulova A, Currat M, Ray N et al (2014) The scaling of genetic diversity in a changing and fragmented world. In: Henle K, Potts SG, Kunin WE, Matsinos YG, Similä J, Pantis JD et al (eds) *Scaling in Ecology and Biodiversity Conservation*. Pensoft Publishers, Sofia, p 55–60.
- 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
- Balme J (2013) Of boats and string: the maritime colonisation of Australia. *Quat Int* 285:68–75
- Barbujani G, Bertorelle G (2001) Genetics and the population history of Europe. *Proc Natl Acad Sci USA* 98:22–25
- Barbujani G, Chikhi L (2006) Population genetics: DNAs from the European neolithic. *Heredity* 97:84–85
- Barbujani G, Sokal RR, Oden NL (1995) Indo-European origins: a computer-simulation test of five hypotheses. *Am J Phys Anthropol* 96:109–132
- Benguigui M, Arenas M (2014) Spatial and temporal simulation of human evolution. *Methods, frameworks and applications. Curr Genom* 15:245–255
- Bodner M, Perego UA, Huber G, Fendt L, Rock AW, Zimmermann B et al. (2012) Rapid coastal spread of First Americans: novel insights from South America’s Southern Cone mitochondrial genomes. *Genome Res* 22:811–820
- Cavalli-Sforza LL, Edwards AWF (1963) Analysis of human evolution. In: Geerts SJ (ed). *Genetics today: proceedings of the 11th international congress of genetics, The Hague, The Netherlands*. New York: Pergamon. Vol. 3, p. 923–933.
- Cavalli-Sforza LL, Menozzi P, Piazza A (1993) Demic expansions and human evolution. *Science* 259:639–646
- Cavalli-Sforza LL, Menozzi P, 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, Barbujani G (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, Nichols RA, Barbujani G, Beaumont MA (2002) Y genetic data support the Neolithic demic diffusion model. *Proc Natl Acad Sci USA* 99:11008–11013

- Currat M, Excoffier L (2005) The effect of the Neolithic expansion on European molecular diversity. *Proc Biol Sci* 272:679–688
- Dillehay TD (2009) Probing deeper into first American studies. *Proc Natl Acad Sci USA* 106:971–978
- Edmonds CA, Lillie AS, Cavalli-Sforza LL (2004) Mutations arising in the wave front of an expanding population. *Proc Natl Acad Sci USA* 101:975–979
- Excoffier L, Ray N (2008) Surfing during population expansions promotes genetic revolutions and structuration. *Trends Ecol Evol* 23:347–351
- Fagundes NJ, Kanitz R, Eckert R, Valls AC, Bogo MR, Salzano FM et al. (2008) Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am J Hum Genet* 82:583–592
- Fagundes NJR, Tagliani-Ribeiro A, Rubicz R, Tarskaia L, Crawford MH, Salzano FM et al. (2018) How strong was the bottleneck associated to the peopling of the Americas? New insights from multilocus sequence data. *Genet Mol Biol* 41:206–214
- Forster P, Harding R, Torroni A, Bandelt HJ (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* 59:935–945
- 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
- Gravel S, Zakharia F, Moreno-Estrada A, Byrnes JK, Muzzio M, Rodriguez-Flores JL et al. (2013) Reconstructing Native American migrations from whole-genome and whole-exome data. *PLoS Genet* 9:e1004023
- Hey J (2005) On the number of New World founders: a population genetic portrait of the peopling of the Americas. *PLoS Biol* 3:e193
- Hunley K, Healy M (2011) The impact of founder effects, gene flow, and European admixture on native American genetic diversity. *Am J Phys Anthropol* 146:530–538
- Kanitz R, Guillot EG, Antoniazza S, Neuenschwander S, Goudet J (2018) Complex genetic patterns in human arise from a simple range-expansion model over continental landmasses. *PLoS ONE* 13:e0192460
- Kimura M, Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* 49:561–576
- Kitchen A, Miyamoto MM, Mulligan CJ (2008) A three-stage colonization model for the peopling of the Americas. *PLoS One* 3:e1596
- 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
- Lindo J, Huerta-Sanchez E, Nakagome S, Rasmussen M, Petzelt B, Mitchell J et al. (2016) A time transect of exomes from a Native American population before and after European contact. *Nat Commun* 7:13175
- Marshall SJ, James TS, Clarke GKC (2002) North American ice sheet reconstructions at the last glacial maximum. *Quat Sci Rev* 21:175–192
- Menozi P, Piazza A, Cavalli-Sforza L (1978) Synthetic maps of human gene frequencies in Europeans. *Science* 201:786–792
- Mona S, Ray N, Arenas M, Excoffier L (2014) Genetic consequences of habitat fragmentation during a range expansion. *Heredity* 112:291–299
- Mulligan CJ, Hunley K, Cole S, Long JC (2004) Population genetics, history, and health patterns in native americans. *Annu Rev Genom Hum Genet* 5:295–315
- Novembre J, Galvani AP, Slatkin M (2005) The geographic spread of the CCR5 Delta32 HIV-resistance allele. *PLoS Biol* 3:e339
- Novembre J, Stephens M (2008) Interpreting principal component analyses of spatial population genetic variation. *Nat Genet* 40:646–649
- O’Fallon BD, Fehren-Schmitz L (2011) Native Americans experienced a strong population bottleneck coincident with European contact. *Proc Natl Acad Sci USA* 108:20444–20448
- O’Rourke DH, Suarez BK (1986) Patterns and correlates of genetic variation in South Amerindians. *Ann Hum Biol* 13:13–31
- Pedersen MW, Ruter A, Schweger C, Friebe H, Staff RA, Kjeldsen KK et al. (2016) Postglacial viability and colonization in North America’s ice-free corridor. *Nature* 537:45–49
- 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 USA* 92:5836–5840
- Pimenta J, Lopes AM, Comas D, Amorim A, Arenas M (2017) Evaluating the Neolithic expansion at both shores of the Mediterranean sea. *Mol Biol Evol* 34:3232–3242
- Ray N, Adams JM (2001) A GIS-based vegetation map of the world at the last glacial maximum (25,000–15,000 BP). *Internet Archaeol* 11.<https://doi.org/10.11141/ia.11.2>
- 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
- Ray N, Excoffier L (2009) Inferring past demography using spatially explicit population genetic models. *Hum Biol* 81:141–157
- Ray N, Excoffier L (2010) A first step towards inferring levels of long-distance dispersal during past expansions. *Mol Ecol Resour* 10:902–914
- Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N et al. (2012) Reconstructing native American population history. *Nature* 488:370–374
- Rogers RA, Rogers LA, Hoffmann RS, Martin LD (1991) Native American biological diversity and the biogeographic influence of ice age refugia. *J Biogeogr* 18:623–630
- Salas A, Lovo-Gomez J, Alvarez-Iglesias V, Cerezo M, Lareu MV, Macaulay V et al. (2009) Mitochondrial echoes of first settlement and genetic continuity in El Salvador. *PLoS One* 4:e6882
- Skoglund P, Mallick S, Bortolini MC, Chennagiri N, Hunemeier T, Petzl-Erler ML et al. (2015) Genetic evidence for two founding populations of the Americas. *Nature* 525:104–108
- Skoglund P, Malmstrom H, Omrak A, Raghavan M, Valdiosera C, Gunther T et al. (2014) Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. *Science* 344:747–750
- Sokal RR, Harding RM, Oden NL (1989) Spatial patterns of human gene frequencies in Europe. *Am J Phys Anthropol* 80:267–294
- Sokal RR, Menozzi P (1982) Spatial autocorrelations of HLA frequencies in Europe support demic diffusion of early farmers. *Am Nat* 119:1–17
- 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
- Yang Z (2006) *Computational molecular evolution*. Oxford University Press, Oxford, England