Inferring Continental Ancestry of Argentineans from Autosomal, Y-Chromosomal and Mitochondrial DNA

Daniel Corach[1*], Oscar Lao[2], Cecilia Bobillo[1], Kristiaan van Der Gaag[3], Sofia Zuniga[3], Mark Vermeulen[2], Kate van Duijn[2], Miriam Goedbloed[2], Peter M. Vallone[4], Walther Parson[5], Peter de Knijff[3] and Manfred Kayser[2]

[1]*Servicio de Huellas Digitales Genéticas and Cátedra de Genética y Biología Molecular, Faculty of Pharmacy and Biochemistry, University of Buenos Aires, Argentina*
[2]*Department of Forensic Molecular Biology, Erasmus University Medical Center Rotterdam, Rotterdam, The Netherlands*
[3]*Department of Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands*
[4]*Biochemical Science Division, National Institute of Standards and Technology, Gaithersburg, Maryland, USA*
[5]*Institute of Legal Medicine, Innsbruck Medical University, Austria*

Summary

We investigated the bio-geographic ancestry of Argentineans, and quantified their genetic admixture, analyzing 246 unrelated male individuals from eight provinces of three Argentinean regions using ancestry-sensitive DNA markers (ASDM) from autosomal, Y and mitochondrial chromosomes. Our results demonstrate that European, Native American and African ancestry components were detectable in the contemporary Argentineans, the amounts depending on the genetic system applied, exhibiting large inter-individual heterogeneity. Argentineans carried a large fraction of European genetic heritage in their Y-chromosomal (94.1%) and autosomal (78.5%) DNA, but their mitochondrial gene pool is mostly of Native American ancestry (53.7%); instead, African heritage was small in all three genetic systems (<4%). Population substructure in Argentina considering the eight sampled provinces was very small based on autosomal (0.92% of total variation was between provincial groups, p = 0.005) and mtDNA (1.77%, p = 0.005) data (none with NRY data), and all three genetic systems revealed no substructure when clustering the provinces into the three geographic regions to which they belong. The complex genetic ancestry picture detected in Argentineans underscores the need to apply ASDM from all three genetic systems to infer geographic origins and genetic admixture. This applies to all worldwide areas where people with different continental ancestry live geographically close together.

Keywords: Argentina, genetic ancestry, autosomal SNPs, Y-SNPs, mtDNA

Introduction

Establishing reliable genetic knowledge about bio-geographic ancestry, the degree of admixture and the extent of population substructure is of relevance mostly in the field of epidemiological studies, but can also be useful in the forensic context and is additionally interesting from a historical point of view.

Argentineans are usually considered as a population of strict European ancestry; however, historical evidence suggests that this may not be the case. For instance, a census in 1869 showed that the population of Argentina involving 1,756,000 people was composed of "criollos": individuals of European descent born outside the original European countries, "mestizos": individuals of mixed Native American and European descent, "zambos": individuals of mixed Native American and African descent, and "mulatos": individuals of mixed European and African descent, in addition to people of assumed unmixed African and Native American ancestry. It can be assumed that three major admixture episodes happened during Argentinean population history. The first involved Native Americans and Western Europeans (mostly Spaniards) and started

*Corresponding author: Prof. Dr. Daniel Corach, Servicio de Huellas Digitales Genéticas and Cátedra de Genética y Biología Molecular, Faculty of Pharmacy and Biochemistry, University of Buenos Aires, Argentina. Junin 956, 1113-Buenos Aires ARGENTINA. Tel: 00541149648281; Fax: 00541149648282; E-mail: dcorach@ffyb.uba.ar

soon after the first arrival of the Spanish conquistadores in the 16th century. The second admixture episode additionally included West-Africans, and began after African slaves were first introduced to the territory in the late 16th, century with constant influx until 1810. Finally, a third major admixture period involved two sides: the already admixed Argentinean population consisting of non-mixed or already mixed Native American, West-European (mostly Spanish) and West-African individuals on one hand, and on the other a large number of Europeans who entered Argentina between 1856 and 1930. These late European immigrants mostly came from Italy and Spain (over 76%), followed by French, Polish, Russians, and Germans (approx. 13%), representing a major migration wave of over 5,700,000 immigrants entering the country, of which 3,000,000 settled down in Argentina. Notably, in 1914 more than one third of Argentineans were born outside the country (Martìnez Sarasola, 2005). Much more recently, people from various countries in Africa, Asia, and Europe as well as people of various ancestries from neighbouring South American countries entered Argentina, adding to the human diversity of the country.

Although some expectations about the continental background of Argentineans can be formulated based on census information, such information does not provide quantitative estimates on the degree of admixture in the contemporary Argentinean population. Molecular genetics offers suitable tools to investigate bio-geographic ancestry in detail including the detection and quantification of admixture proportions. Some information has been published about the genetic make-up of Argentineans either employing markers from autosomal DNA (Sala et al., 1998, 1999; Marino et al., 2006a,b,c; Seldin et al., 2007), from the non-recombining part of the Y-chromosome (NRY) (Kayser et al., 1997; Corach et al., 2001; Kayser et al., 2001; Marino et al., 2007), or, to a lesser degree, also from mitochondrial DNA (mtDNA) (Ginther et al., 1993; Corach et al., 1997; Bobillo et al., 2009), but reliable inferences on bio-geographic ancestry are limited. Also, the combined analysis of uni-parentally and bi-parentally inherited markers in the same individuals has rarely been done (Martínez Marignac et al., 2004; Corach et al., 2006; Salas et al., 2008). Therefore we analyzed 249 unrelated males from eight provinces of three Argentinean regions by means of DNA markers from the autosomal, Y-chromosomal and mitochondrial parts of the human genome suitable for detecting continental origins. In addition to genetic markers, we also used paternal surnames as culturally-transmitted markers, to further extend our bio-geographic ancestry analyses in the Argentinean population, similar to one previous study on Columbians (Bedoya et al., 2006). It should be noted that in Argentina, in contrast to most South American countries, only paternal, and not maternal, family names are used.

Materials and Methods

Samples

Samples were collected at the Servicio de Huellas Digitales Genéticas (DNA Fingerprinting Service), School of Pharmacy and Biochemistry, University of Buenos Aires, Argentina. Sampling took place during the period 2005–2007 and included unrelated male volunteer donors, who participated in paternity testing and signed written consent statement forms, approved by the local Ethical Committee. Personal information was treated anonymously. Blood samples were obtained by finger puncture and spotted onto FTA paper. DNA extraction was performed following the manufacturer's protocol (FTA, Whatman, www.whatman.com/DNACollection.aspx). Initially, 249 samples were ascertained, however one individual with a Japanese surname and two others with Middle Eastern surnames were excluded, leaving 246 samples from individuals with European and Native American surnames in the study. Sample came from eight provinces from three geographical regions of the country (Fig. 1): Formosa (AFO, N = 11), Chaco (ACA, N = 1), Misiones (AMI, N = 28) and Corrientes (ACO, N = 21) from the north-eastern Argentinean region (N = 61); Santa Fe (ASF, N = 3) and Buenos Aires (ABS, N = 150) from the central Argentinean region (N = 153), as well as Río Negro (ARN, N = 31) and Chubut (ACH, N = 1) from the southern Argentinean region (N = 32). Sample size was chosen as an approximation to match the relative contribution of each region to the entire Argentinean population (National Institute of Statistics and Censuses, INDEC 2001 www.indec.mecon.ar) to achieve an approximate representation of the Argentinean population. Although this resulted in some of the provinces having too small sample sizes, all samples were used, especially in the framework of the regional approach and when considering Argentina as a whole population. At the time of sampling the most likely geographic origin of the surname of all the participants was recorded. However, for the purpose of this study no record of the surname itself but only its geographic origin was used to assure anonymous treatment. In particular, individual surnames were inspected for likely European (comprising Belgium, Croatia, France, Germany, Greece, Italy, Netherlands, Poland, Portugal, Spain, Russia, United Kingdom and Ukraine), African, Asian, Middle East and Native American origin. Geographical origin of the non-Native American surnames was assessed according to Hanks & Hodges (1989). Surnames of Native American origin were identified using linguistic knowledge, mainly by detecting linguistic elements within the surnames of known Amerindian origins such as Mupudungun, Diaguita, etc. In addition, we ascertained from the Human Genome Diversity Project –Centre d'Etudes d'Polumorphismes Humains (HGDP-CEPH: http://www.cephb.fr/en/hgdp/diversity.php/) samples of those individuals that represent the closest geographic relatives of the most likely true parental populations for Argentineans (Native South Americans, diverse Europeans, and West Sub-Saharan Africans) given existing knowledge of the Argentinean history (Rock, 1987). These were in particular 29 French, 24 Basques, 47 Italians comprising 11 from Bergamo, 28 from Sardinia, and 8
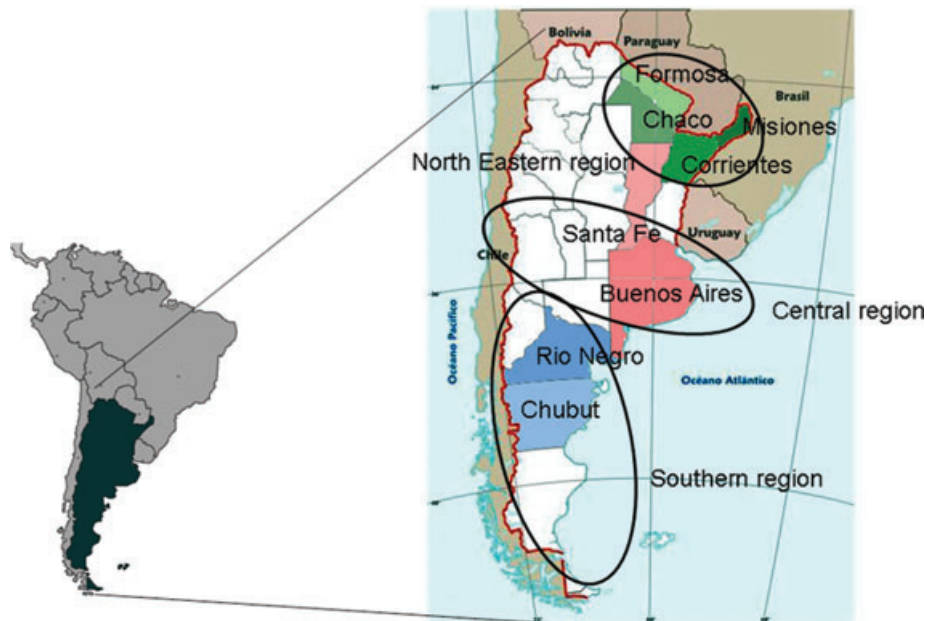
Figure 1 Map of Argentina indicating the provinces and regions sampled.

from Tuscany, 16 Orcadian Islanders from Great Britain, 25 Russians and 17 from Adygei in the Caucasus region, (i.e., all 158 Europeans included in HGDP). Also sampled were 23 Karitiana and 21 Surui from Brazil, (i.e., all 44 Native South-Americans included in HGDP) as well as 24 Mandenka from Senegal and 25 Yoruba from Nigeria (i.e., all 49 West Sub-Saharan Africans included in HGDP).

Genotyping

Autosomal DNA

Twenty-four autosomal SNPs were ascertained from a pool of 62 pre-selected markers by applying a genetic algorithm for maximising the amount of non-redundant continental ancestry information per single marker as described in Lao et al. (2006). Fifty six of these markers had been previously ascertained from a dataset of >10,000 SNPs generated using the Affymetrix GeneChip® Human Mapping 10K Array Xba131 (Mapping 10K array) in the Y Chromosome Consortium (YCC) cell line panel. They comprised the most promising continental ancestry markers ascertained from the YCC dataset when applied to the HGDP-CEPH samples as described elsewhere (Lao et al., 2006; Kersbergen et al., 2009). The remaining six markers were ascertained from pigmentation candidate genes and had shown a strong continental population differentiation in the HGDP-CEPH samples as described in Lao et al. (2007). SNP ascertainment was focused to enrich for four continental ancestry components: Sub-Saharan Africa, East-Asia, Eurasia, and Native America. The following SNPs were used here: rs1876482, rs2179967, rs1048610, rs1371048, rs1478785, rs1369290, rs952718, rs1405467, rs1344870, rs1391681, rs1461227, rs1907702,

rs2052760, rs714857, rs721352, rs722869, rs926774, rs1448484, rs1667751, rs1858465, rs1465648, rs16891982, rs1808089, rs3843776. Genotyping was performed in two multiplex SNaP-shot reactions based on the principle of primer extension as will be described in detail elsewhere (Vallone et al. in preparation). SNP genotypes were scored by two independent analysts and finally reviewed by a third one.

Mitochondrial DNA

The entire control region (CR) of human mtDNA from nucleotide positions 16024 to 576 was sequenced following EMPOP recommendations as described in detail elsewhere (Brandstätter et al., 2007; updated in Parson & Bandelt, 2007). The sequences were aligned to the revised Cambridge Reference Sequence (rCRS; Andrews et al., 1999) using Sequencher vs. 4.8 (GeneCodes, Ann Arbor, MI, USA), following updated nomenclature guidelines (Bandelt & Parson, 2008). All samples were evaluated twice by two independent analysts and results were compared using in-house software and finally reviewed by a third analyst. Furthermore, particular coding-region SNPs were analysed using a modified version (Amry et al. in preparation) of a previously published assay (Alvarez-Iglesias et al., 2007) and by direct sequence analysis to detail the haplogroup affiliations in cases where the CR sequences did not provide sufficient information for reliable haplogroup designation. In particular, samples belonging to Asian and Native American lineages were SNP analysed at positions 12468 (hg A2c); 6755 (hg B2b); 1888 (hg C1c); 7697 (hg C1d); 8383, 8419, 9431 (hg D4c2), and 5319 (hg D4c2a), whereas samples belonging to West Eurasian lineages were analysed at positions 7028, 2706 (hg H); 9716 (hg K2); 14798 (hg J1c); 4580 (hg V); and 12308, 12372 (hg U).

Y-chromosome DNA

Variation of the non-recombining part of the human Y-chromosome (NRY) was identified by means of 44 NRY-SNPs in total. Twenty four NRY-SNPs were genotyped in all samples (including: SRY 1532, M91, M168, M145, M174, 12f2, M96, M213, M201, M69, M52, M170, M172, M9, M20, M106, M214, Tat, M175, M45, MEH2, M207, M269, and M124). Aiming to maximise continental differentiation of haplogroup origins we additionally genotyped 20 NRY-SNPs on subsets of samples based on the results from the 24 SNP analyses. M3 was genotyped on samples with the derived allele of MEH2, M242 among samples of haplogroup P(xQ1a, R), and eighteen additional SNPs among samples identified as belonging to haplogroup E (M33, P2, M2, M154, M191, M215, M35, M78, V12, M224, V32, V13, V22, M81, M123, M281, V6, and M75). A single multiplex SNaPshot assay using the principle of primer extension was designed for a core set of 24 NRY-SNPs. Primer sequences for M45, M52, M170, M172, M173, M175, M213, 12f2 and SRY1532 were taken from the literature (Sanchez et al., 2003). For the remaining NRY markers, reference sequences for each locus were obtained from the BLAST human genome database (http://www.ncbi.nlm.nih.gov/blast/) and PCR-primers were designed for fragments ranging from 70 to 225 bp in length using Primer 3 v.0.2 (http://frodo.wi.mit.edu) with default settings. Lengths of designed primers ranged from 19 to 27 nucleotides, primers with five or more bases at the 3′ end complementary to part of another primer in the multiplex were discarded or redesigned to avoid primer-dimer formation. Amplicon sequences were checked with BLAST for sequence homology in the human genome (all primer information can be found in Supporting Table S3). Extension primers were designed using Assay Design Software Version 1.0.6 (Biotage, Uppsala, Sweden). Primers with four or more bases at the 3′ end complementary to part of another primer in the multiplex were discarded or redesigned to avoid non-specific primer-extension. To achieve different fragment length differences the multiplex primer lengths were altered by adding a piece of a "neutral" sequence or a poly-C tail as described by Sanchez et al. (2003). Each primer pair was first validated in a singleplex PCR containing 0.5 ng template DNA from a selection of samples (including a female control), 1 × PCR buffer containing 1.5 mM MgCl^2 (Applied Biosystems, Foster City, CA, USA), 100 μM of each dNTP (GE, the Netherlands)) 0.4 μM of each desalted primer (Biolegio Nijmegen, the Netherlands) and 0.6 units of AmpliTaq Gold® DNA polymerase (Applied Biosystems). In the final multiplex PCR, 0.5 ng template DNA was amplified in a 12.5 μl reaction volume containing 1 × PCR buffer, 6.5 mM total MgCl$_2$, 200 μM of each dNTP and 2.5 units of AmpliTaq Gold® DNA polymerase. During multiplex validation primer concentrations were adjusted (0.1–0.4 μM) to achieve optimally balanced signal intensity for all markers. All initial PCRs were performed in a GeneAmp 9700 thermal cycler (Applied Biosystems) with an initial denaturation at 94°C for 10 min followed by 35 cycles of 30 s at 94°C, 30 s at 60°C, 30 s at 72°C and a final extension for 5 min at 72°C. To eliminate excess primers and dNTPs, 2 μl ExoSAP-IT® (USB, Affymetrix, Cleveland,

USA) was added and incubated at 37°C for 30 min, followed by a final enzyme inactivation at 80°C for 15 min. Extension reactions were performed in a 5 μl reaction volume using 1 μl purified PCR product, 2.5 μl of SNaPshot multiplex Ready Reaction Mix (Applied Biosystems) and 0.4 μM primer (HPLC or PAAGE purified). During multiplex validation primer concentrations were optimized (0.06–0.5 μM) for balanced signal intensity. All reactions were performed using a GeneAmp 9700 thermal cycler with a initial denaturation at 96°C for 2 min, followed by 25 cycles of 10 s at 96°C, 5 s at 50°C and 30 s at 60°C. To eliminate unincorporated ddNTPs 1.25 μl SAP® -reagent (USB) was added incubated at 37°C for 1 hour. SAP was inactivated by incubation at 75°C for 15 min. 2 μl of the SAP-treated extension product was analysed with an ABI3100 Genetic Analyzer using a 36 cm capillary array, polymer POP4 and Genescan 120 LIZ as internal size standard. Data were analyzed using GeneMapper ID v3.2.1 software (Applied Biosystems). After background subtraction and colour separation, peaks were sorted into bins according to sizes by comparison to the internal size standard. An Excel-sheet was used to transfer exported allele tables and for automatic haplogroup assignments.

Statistical analyses

NRY haplogroups were derived from genotyping of Y-SNPs using the marker phylogeny as described elsewhere (Karafet et al., 2008). Mitochondrial DNA haplogroups were inferred from sequence data of the complete CR with the additional information of coding SNPs if needed (see above). The geographic origin of the haplogroups was assumed from published NRY (Bortolini et al., 2003; Jobling & Tyler-Smith, 2003; Luis et al., 2004; Semino et al., 2004; Cruciani et al., 2007) and mtDNA data (Richards et al., 1998; Macaulay et al., 1999; Finnilä et al., 2001; Kivisild et al., 2006; Kong et al., 2006; Achilli et al., 2008; Behar et al., 2008) and the Argentinean samples were grouped accordingly. STRUCTURE (Pritchard et al., 2000) was performed by doing 50,000 burnings and retaining the next 50,000 Monte Carlo-Markov Chain runs for final analyses. Three parental populations were assumed and frequencies were updated according to the observed frequencies in these three populations. Ten different replicates were performed and convergence, mixing and reproducibility of the different runs were checked. Multi dimensional scaling (MDS) plots were obtained using SPSS 15.0 (SPSS for Windows, Rel. 15.0.1. 2006. SPSS Inc., Chicago, Illinois, USA). Distruct 1.1 (Rosenberg, 2004) was used to tune the output from STRUCTURE and Grapher 7 (http://www.goldensoftware.com) was used to perform a ternary plot of the most likely amount of ancestry of each parental population per individual. Additional analyses were carried out with the most likely proportion of Native American ancestry estimated by STRUCTURE. Similarities in the amount of Native American ancestry between clusters of individuals (regional assignment, geographic origin of the surname of each individual) were tested by means of Kruskal-Wallis and Mann-Whitney tests computed with SPSS 15.0. Similarities in the proportion of geographic ancestry for mtDNA and NRY were tested by means of Fisher exact test as implemented in SPSS 15.0. In order to

quantify how much of the genetic variation of the autosomal markers was explained under particular individual assignments, an AMOVA analysis (Excoffier et al., 1992) was computed. Two different individual clustering scenarios were considered: i) regional assignment and ii) surname assignment.

Results

Genetic Ancestry and Admixture of Argentineans

Autosomal DNA

Individual clustering analyses were performed using genetic data from 24 autosomal ancestry-sensitive SNPs in 246 Argentineans from eight provinces and three regions of the country (Fig. 1). We additionally typed these SNPs in 158 Europeans (all European HGDP-CEPH samples), 44 Native South Americans (Brazilian Karitiana and Surui HGDP-CEPH samples) and 49 West Sub-Saharan Africans (Mandenka and Yoruba HGDP-CEPH samples). The latter three groups were included as parental populations in the STRUCTURE analysis together with the Argentinean data to assess the degree of continental geographic ancestry in the Argentinean samples. These parental reference samples have been specifically ascertained because the respective groups are geographically most closely related to the expected true parental populations for Argentineans (Rock, 1987) from the global reference data of the HGDP-CEPH samples available to us.

A MDS analysis performed with a matrix of identity-by-state distances between pairs of individuals and two dimensions revealed that the three parental populations showed similar genetic distances between them, indicating similar power to detect genetic ancestry of all three parental groups in the Argentineans (Fig. 2). As evident from the plot, most of the Argentinean samples clustered with or closest to Europeans, some appeared between Europeans and Native Americans indicating some degree of genetic admixture between these two groups, three samples clustered close to Native Americans, and no Argentinean sampled appeared close to Africans (Fig. 2). In a STRUCTURE analysis using the same three parental populations we observed a similar pattern, European ancestry for most Argentinean samples, but also a considerable fraction of Native American ancestry in a number of Argentinean samples (Fig. 3). Overall across Argentinean samples, the mean ancestry components as revealed from the STRUCTURE analysis were 78.6% (95% confidence interval (CI) ranging from 31.5% to 96.6%) for European, 17.3% (95% CI from 1.5% to 63.8%) for Native American, and 4.2% (95% CI ranging from 1.1% to 19.0%) for West–African ancestry (Table 1).
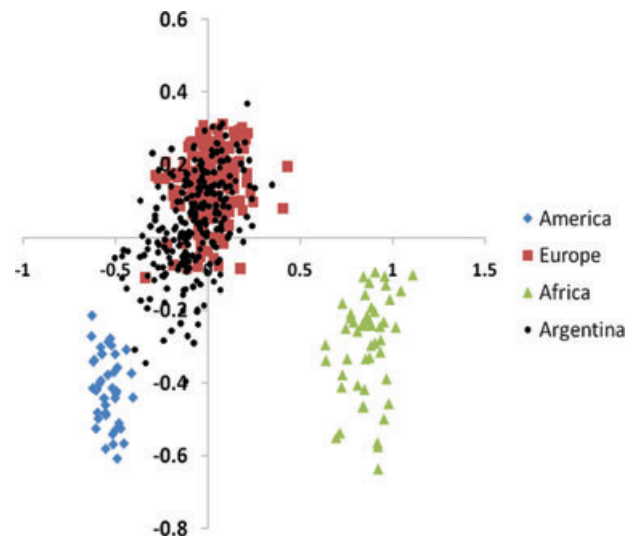

Figure 2 MDS analysis plot of the Argentinean samples together with three assumed parental populations: Europeans, Native South Americans and West Sub-Saharan Africans.

Furthermore, a large admixture heterogeneity between individuals was observed especially involving Native American and European components, ranging from ∼0% of Native American component and ∼90% of European ancestry to ∼80% of Native American ancestry and ∼5% of European component (see Supporting Fig. S1). Both ancestry components are strongly negatively correlated in the Argentinean samples (slope = −1.154, p value slope = < 2e–16, Pearson correlation r-squared = 0.887, p value < 2.2e–16 of a linear regression with the logit values of Native American and European ancestry). In contrast, African ancestry is poorly correlated either with the amount of European ancestry (Pearson r-squared: 0.043, p value = 0.00061) or Native American ancestry component (Pearson correlation r-squared: 0.0035, p value = 0.354) in the Argentinean samples. AMOVA revealed that grouping of Argentineans according to the eight provinces they were sampled from explained a very small proportion of 0.92% (p = 0.00489) of the total autosomal genetic variation. No population substructure was detected when clustering individuals according to the three geographic regions to which these provinces belong (0.5% of variation between groups, p value = 0.264).

Y-chromosomal and mitochondrial DNA

Investigation of NRY-SNPs ascertained to maximise the detection of continental geographic ancestry revealed 19 NRY haplogroups in the Argentinean samples (Supporting Table S1 with their most likely continent of origin indicated). Across Argentinean samples the overall ancestry components as revealed from NRY data were 94.1% for European,
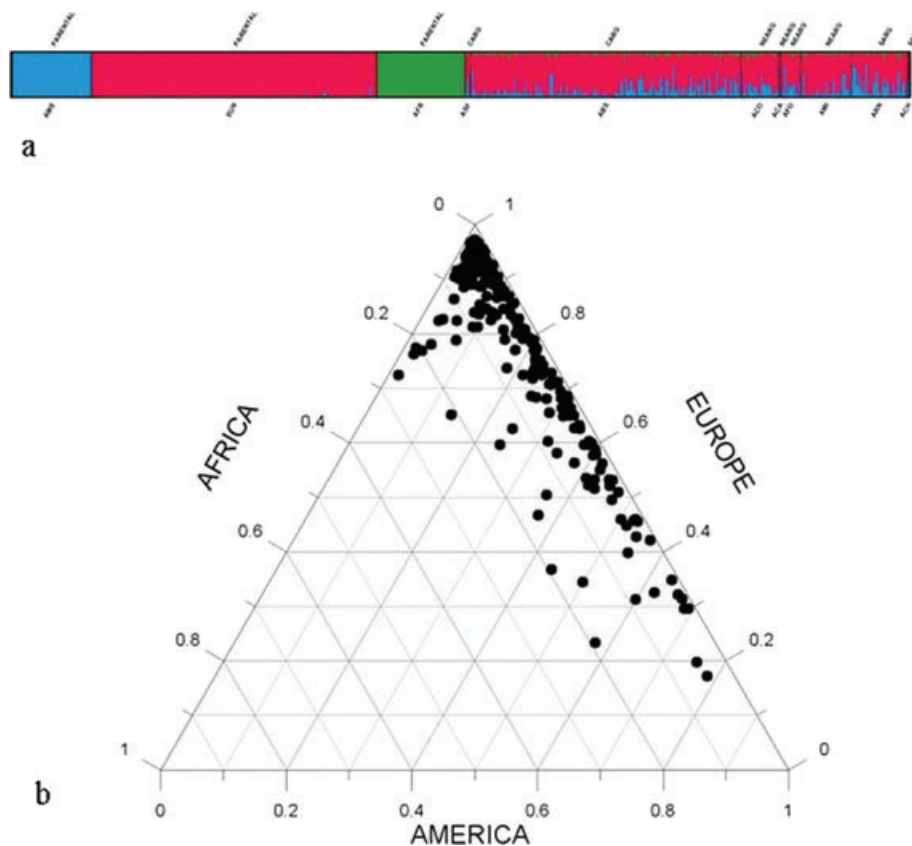
Figure 3 Ancestry estimation, based on STRUCTURE 2.0, of the Argentinean samples assuming three parental populations: Native South Americans (AME), Europeans (EUR) and West Sub-Saharan Africans (AFR) compared with Argentineans from eight provinces (see Material and Methods for acronym references). a) District-based bar plot where each bar represents an individual and the three first blocks the putative parental populations, b) ternary plot in which each dot represents an Argentinean individual and in each vertex a putative parental population.

Table 1 Mean amount of continental autosomal ancestry (standard deviations) of Argentineans from each of the provinces and geographic regions based on STRUCTURE analysis

Region	Province*/region	N	Native S-American	European	W Ss-African
North-Eastern	ACA	1	67.3	31.4	1.3
	ACO	21	17.61 (12.25)	77.45 (11.94)	4.96 (5.14)
	AFO	11	21.93 (19.72)	74.74 (19.56)	3.34 (3.29)
	AMI	28	13.35 (17.02)	82.40 (18.02)	4.25 (4.39)
	SubTotal_Northern	61	17.25 (17.27)	78.48 (17.45)	4.28 (4.44)
Central	ABS	150	15.03 (16.18)	80.80 (16.39)	4.18 (4.06)
	ASF	3	18.87 (21.47)	77.47 (22.70)	3.63 (1.27)
	SubTotal_Central	153	15.10 (16.21)	80.73 (16.45)	4.17 (4.03)
Southern	ARN	31	27.64 (23.04)	68.51 (24.59)	3.84 (4.47)
	ACH	1	30.7	66.4	2.9
	SubTotal_Southern	32	27.74 (22.67)	68.45 (24.19)	3.82 (4.4)
Total		246	17.28 (17.84)	78.57 (18.24)	4.15 (4.17)

*ABS = Buenos Aires, ACA = Chaco, ACO = Corrientes, AFO = Formosa, AMI = Misiones, ARN = Rio Negro, ASF = Santa Fe

Table 2 Percentage of Y-chromosomal and mtDNA continental ancestry of Argentineans from three geographic regions

	Native American	European	African
*mtDNA**			
North-Eastern (61)	67.2	31.1	1.6
Central (153)	45.7	52.9	1.3
Southern (32)	65.6	28.1	3.1
Total (246)	53.7	44.3	2.0
*NRY-DNA**			
North-Eastern (61)	2.5	95.8	1.7
Central (153)	4.8	94.7	0.5
Southern (32)	10.9	87.5	1.6
Total (246)	4.9	94.1	0.9

*for NRY data the expected ancestry proportions given the frequency of each haplogroup in the three continents is shown since for some NRY haplogroups no single continent of origin could be assigned (see Supporting Table S1 for details), whereas for mtDNA haplogroups single continents of ancestry could be unequivocally assigned and were used here.
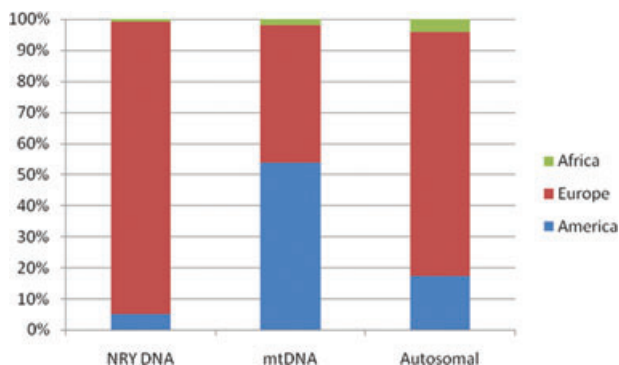

Figure 4 Proportions of Native American, European and West Sub-Saharan African ancestry of the Argentinean samples as inferred from NRY and mtDNA haplogroups and autosomal ASDM typing results analyzed by STRUCTURE.

4.9 % for Native American, and 0.9% for African ancestry (Table 2, Fig. 4). An AMOVA using NRY-SNP haplogroup data revealed no evidence of population substructure when considering the eight provinces from which the Argentineans were sampled (-1.57% of variation between groups, p value = 0.92), and also not when clustering the provinces according to the three geographic regions to which they belong (2.2% of variation between groups, p value = 0.33).

From the mtDNA data we identified 59 different mtDNA haplogroups among the 246 Argentineans (Supporting Table S2 with their most likely continent of origin indicated). Resulting overall mtDNA-based continental ancestry components were estimated at 44.3% European, 53.7% Native

S–American, and 2.0% African (Table 2, Fig. 4). An AMOVA using mtDNA haplogroups showed that grouping of Argentineans according to their eight sampling provinces explained a small amount, 1.77% (p = 0.00489), of the total mtDNA variation, but no substructure was observed when clustering the provinces according to the three regions to which they belong (-0.17% of variation between groups, p value = 0.37). A Kruskal–Wallis test performed with the autosomal amount of Amerindian ancestry and clustering the individuals according to the geographic origin of their mtDNA was strongly statistically significant ($\chi^2 = 71.64$, p value 2.77e–016). A similar result was observed when the geographic origin of the NRY and mtDNA were taken into account together ($\chi^2 = 82.26$, p value 2.8e–016).

The proportions of continental ancestry estimated from NRY and mtDNA data in the Argentinean samples were statistically significantly different from each other (Fisher exact test = 158.78, two tail p-value = 4.89e–036) even when considering the ancestry of both loci at the individual level (Fisher exact test = 22.07; two tail p value = 0.016). In order to test whether the observed proportion of ancestry in the autosomal markers could produce the observed ancestry estimations in mtDNA and NRY, we assumed that the observed ancestry proportions of mtDNA and NRY were the outcome of a multinomial distribution with success probability of each ancestry class given by that estimated in the autosomal markers. A statistically significant p value was observed when comparing the proportions of autosomal ancestry with the proportions observed in mtDNA (p = 4.65e–39), as well as when comparing with NRY (p = 7.96e–11). This suggests that it is quite unlikely that the same amounts of admixture observed in the autosomal markers could by sampling chance produce the observed amounts of admixture that has been detected in the two sex-linked loci.

Distribution of Native American Genetic Ancestry Among Argentinean Paternal Surnames

An AMOVA using the autosomal genetic data revealed that a small amount, 1.27% (p < 0.0005), of the total autosomal genetic variation is explainable by grouping the Argentinean samples according to the inferred geographic origin of the sample donor's surnames. We further tested whether the density distribution of the proportion of the Native American autosomal ancestry was different depending on the geographic origin of the surname of each individual (see Supporting Fig. S2). Notably, there were individuals with 90% European genetic ancestry carrying surnames of Native American origin as well as other individuals with a Spanish surname but 80% of Native American genetic ancestry. A

Kruskal-Wallis test performed with the amount of Native American genetic ancestry and the geographic origin of the European surnames was statistically significant (Kruskal-Wallis test p = 5.1e–005). Statistically significant differences (Fisher exact test p value = 0.002) were also observed in the continental NRY ancestry depending on the continental ancestry of the surname. Whereas 96% of the individuals with European surnames carried European Y-chromosomes, 50% of the samples from individuals with Amerindian surnames had European Y chromosomes.

Discussion

Although often considered the European people of South-America, historical records demonstrate that contemporary Argentineans are the result of genetic admixture processes involving three continental contributors: Native Americans, Western Africans and Europeans. However, robust genetic admixture estimates using markers from uni- and bi-parentally inherited parts of the genome, as well as transmitted cultural markers such as the surname, in the same set of samples had not yet been conducted. Hence, we employed a wide range of DNA markers from the autosomes, the Y-chromosome and from mtDNA that are ancestry-sensitive for the three continental groups putatively involved in the Argentinean population history in those individuals collected throughout the country. Examples from other parts of the world, such as the Pacific (Kayser et al., 2006; Kayser et al., 2008), have shown that analysing both uni-parental as well as the bi-parental parts of the genome is essential as the different parts of the genome can reveal different geographic ancestry components, providing important insights into different aspects of human population history. Evolutionary studies, association mapping, disease-risk prediction and forensic analysis are some of the research fields in which genetic admixture estimates are relevant (Sans, 2002; Liu et al., 2005; Yang et al., 2005; Wang et al., 2008).

Our investigation focusing on the Argentinean population with a tri-parental model revealed a structure with varying genetic proportions depending on the genetic system analysed. Clustering analyses performed on data from autosomal ancestry-sensitive SNPs – providing asomewhat approximate representation of the bi-parentally inherited part of the genome – demonstrated a major European component (overall 78.5%) in the pooled Argentinean sample, whereas the Native American component (overall 17.3%) was lower but considerable, and the African component was very small (overall 4.1%). Very similar values were obtained previously by Seldin et al. (2007) in a smaller sample of smaller geographic distribution than was analysed by us. Notably, their results were achieved with 54 more autosomal ancestry-sensitive SNPs

than were applied in the present study, illustrating that our 24 ASM-SNPs may contain more continental ancestry information. Slightly lower proportions of European ancestry, together with slightly higher proportions of Native American and African ancestry were observed for individuals with European surnames from La Plata City, but only five ancestry-sensitive markers were used (Martínez-Marignac et al., 2004). However, despite this general trend in the genetic variation, it should be noted that our analyses revealed a considerable genetic heterogeneity at the individual level, also noticed before in a smaller number of Argentinean samples (Seldin et al., 2007). About 40% of the sampled individuals attained more than 90% of European ancestry, but the Native American proportion was as high as 80% albeit this was observed only in a single individual. Such findings reflect the dynamics of the recent demographic history of the extant Argentinean population, indicating that individuals retaining a higher proportion of European ancestry could be descendants from the recent newcomers to the Argentinean population, whereas those with a higher amount of Native American admixture could be descendants from the first contact between the European and Amerindian populations, starting approximately 500 years ago. Much lower European, together with much higher Native American, ancestry proportions than were observed on average here were detected in three Argentinean groups by Wang et al. (2008) using 751 autosomal short-tandem repeat (STR) polymorphisms, and the discrepancies may in part be related to differences between the samples used.

When considering lineage-specific genetic markers, a different picture was observed, depending on whether these were maternally or paternally inherited. Based on NRY data the overall European component was very high (overall 94.1%); in particular 1.2 times higher than has been established from autosomal markers, whereas the Native American component (overall 4.9%) was very low, in particular 3.5 times lower than from autosomal DNA. The African proportion (overall 0.9%) was about 4.7 times lower than was detected with autosomal DNA. So far, studies focused on the amount of male lineage ancestry in the Argentinean population have only analysed the presence of Native American motifs, characterised by the mutation M3, depicting a C to T transition at locus DYS199 (Underhill et al., 1996). Our results corroborate the limited proportion of Native American ancestry observed by Salas et al. (2008) using NRY-STRs in samples from Cordoba (Fisher exact test p value = 0.178). However, our results showed statistically significant differences (Fisher exact test p value 1.819e–040) with the Native American proportions observed in individuals with European surnames from the city of La Plata (Martínez-Marignac et al., 2004) and with those observed by Corach et al. (2006) using samples from three different geographic regions (Fisher exact test p value 1.302e–006). These discrepancies might be

explainable by the different sampling ascertainment schema of the samples used. In particular, samples used in the present study came from private paternity testing afforded by the participants themselves, whereas those from the Corach et al. (2006) study came from forensic case work. A socio-economic bias associated with bio-geographic ancestry could explain the differences of the ancestry estimations in the two studies. The previous and current results together suggest that individuals with a Native American Y chromosome may be more associated to forensic case work, which could imply that the combination of socio-economic factors and bio-geographical ancestry may be acting as a confounding factor.

In contrast to autosomal and NRY data, the analysis of mtDNA haplogroups identified Native American ancestry as a major component (overall 53.7%), which was 3.1 times higher than detected with autosomal markers and 11 times higher than with NRY markers. Consequently, the European contribution detected with mtDNA (overall 44.3%) was 1.8 times lower than with autosomal markers and 2.1 times lower than with NRY markers. The African mtDNA proportion (overall 2.0%) was about half of what we detected based on autosomal data but about twice that estimated for NRY DNA. Previous mtDNA-based estimates of Native American ancestry in Argentineans are somewhat similar (Martínez Marignac et al., 2004; Fisher exact test p value when compared with the current study = 0.389; Corach et al., 2006; Fisher exact test p value 0.355). The latter would imply that the geographic ancestry of mtDNA is independent from the assumed socio-economic differences in forensic versus paternity testing sampling in contrast to the NRY findings.

The presence of large differences in the continental ancestry proportions detected with uni-parentally inherited markers, as we also found here in Argentineans, seems to be a common observation within the Latin American countries (Batista dos Santos et al., 1999; Martínez Marignac et al., 2004; Bertoni et al., 2005, Campos-Sánchez et al., 2006). Typically, the proportion of Native American ancestry is low for male lineages but high for female ones, whereas European ancestry is high for male lineages but low for female ones, as a consequence of sex-mediated ancestry differences in the admixture history of the population. In our case, soon after the first European contact with the current territory of Argentina, the original population was dramatically affected by a severe reduction of the Native American male population, as a result of the conquest. Consequently, an increased proportion of offspring from European males and Native American women were born, due in part to low European female population size and the reproductive preponderance of the European invaders. This situation might reflect the political decisions by the Spanish crown for implementing a strategy for population growth and colonial occupation of the invaded territories. Additional social factors, also in subsequent periods, limited

Native Amerindian male genetic flow into the admixed population.

It is interesting to note that although the number of West-Africans who were introduced to the territory of contemporary Argentina by European slave traders between 1580 and 1813 was large (i.e. 100,000 for the La Plata and Boliva region Rawley & Behrendt, 2005), and slavery was abolished in 1853, the African component we detected in our Argentinean samples was very low with any of the three genetic systems applied. This result indicates a low degree of African admixture in the general Argentinean population, which is different to North American countries such as the United States where African admixture components in European Americans are up to about one third (Kittles et al., 2003). Hence, a stronger social barrier may have existed in Argentina, resulting in a lower number of offspring from parents of West-African and European (or mixed European-Native American) ancestry than that which occurred in North-America. Moreover, these results could also suggest the presence of demographic pressures against individuals carrying a large West African ancestry in Argentina. In particular, the "freedom of wombs" law (children of slaves were born free) established by the Constituent Assembly in 1813, might have stimulated slave masters to sell their female slaves outside the country, reducing the matrilineages of African ancestry. It should be noted that slave trading was prohibited since 1813 but slavery was abolished in 1853. In addition, Argentineans of African descendent were recruited as soldiers during the independence war between 1810 and 1818, and in the war of the Triple Alliance (Argentina, Uruguay and Brazil against Paraguay) between 1864 and 1870, with high mortality rates. These wars reduced the number of African males, hence reducing African patrilineages. Finally, the impact of cholera (1861 and 1864) and yellow fever (1871) epidemics especially affected the poorest parts of Argentinean society, which included most of the people of African descent.

In addition to using genetic diversity for quantifying biogeographic ancestry and admixture, we have also analysed the surnames of the sample donors as a paternally inherited social marker of ancestry. Our results demonstrate a trend between the geographic origin of the surnames and the amount of autosomal Native Amerindian ancestry, particularly with Spanish surnames, which may be explainable in the context of European arrival times. Male Spanish conquers were the first to meet with the aboriginal Amerindian population and therefore had most time to admix with them, whereas additional non-Spanish European sub-populations arrived much later in Argentina, with less time for establishing admixture. However, we also detected some outliers with higher proportions of European ancestry and Amerindian surname and vice versa, which may be explained by random genetic drift or illegitimate paternities. Kidnapping of people of

European descent by Native Americans, as happened especially during the first half of 19th century, may be another explanation for this phenomenon. These captives were forced to live in Native American communities and not only adopted their lifestyle but also received Amerindian names; however, the frequency of such events was too low (Martinez Sarasola, 2005) to explain our observation. On the contrary, European surnames associated with large Amerindian genetic components may potentially reflect a different scenario: most of the Native Americans lacked a composite name and may have been given Spanish surnames from e.g. encomenderos, administrative officers and clergymen. In addition some people of Native American descent might have decided to change their surnames into European ones in order to avoid social discrimination. An alternative explanation would be adoptions of Native Americans by people of European ancestry.

In conclusion, we found that the contemporary Argentinean population carries a major European ancestry component in their autosomal DNA, and even more so in their Y-chromosomal gene pool, in agreement with prior expectancies. In contrast, most of the Argentinean mitochondrial gene pool was of Native American ancestry. The African genetic ancestry of Argentineans was very low based on all three genetic systems, which is remarkable given the large number of West African slaves brought to the region. Differences between the amounts of European and Native American ancestry components detected using paternal NRY and maternal mtDNA markers are in line with the sex-biased admixture history involving predominantly European men and Native American women at least in the early periods of European contact. In the Argentinean population only very small amounts of population substructure in respect of the eight sampled provinces were observed based on autosomal and mtDNA data (none based on NRY data), and no substructure was detected with all three genetic systems when clustering the provinces into the three geographic regions to which they belong. The complex genetic ancestry picture revealed in our study underscores the need for the combined use of ancestry-sensitive markers of both uni-parental as well as of bi-parental transmission in order to obtain more accurate inferences of bio-geographic ancestry which is relevant in epidemiological, historical and forensic studies. This is also important in other South-American countries that underwent similar events of sex-biased admixture; but, moreover, is also relevant in all other countries where people of different continental origin live in close geographic proximity (such as the United States of America) allowing continental admixture to have occurred. Our data also show that using surnames as a proxy for ancestry inferences (or "ethnic affiliation") might be misleading; however their use may supplement genetic information arising from ancestry sensitive markers and may provide interesting insights into the social behaviour of a population.

References

Achilli, A., Perego, U. A., Bravi, C. M., Coble, M. D., Kong, Q.-P., Woodward, S. R., Salas, A., Torroni, A. & Bandelt, H. J. The phylogeny of the four pan-American mtDNA haplogroups: implications for evolutionary and disease studies. *PLoS One* **3**, e1764.

Alvarez-Iglesias, V., Jaime, J. C., Carracedo, A. & Salas, A. (2007) Coding region mitochondrial DNA SNPs: targeting East Asian and Native American haplogroups. *Forensic Sci Int Genet* **1**, 44–55.

Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M. & Howell, N. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* **23**, 147.

Bandelt, H. J. & Parson, W. (2008) Consistent treatment of length variants in the human mtDNA control region: a reappraisal. *Int J Legal Med* **122**, 11–21.

Batista dos Santos, S. E., Rodrigues, J. D., Ribeiro-dos-Santos, A. K. & Zago, M. A. (1999) Differential contribution of indigenous men and women to the formation of an urban population in the Amazon region as revealed by mtDNA and Y-DNA. *Am J Phys Anthropol* **109**, 175–180.

Bedoya, G., Montoya, P., García, J., Soto, I., Bourgeois S., Carvajal, L., Labuda, D., Alvarez, V., Ospina, J., Hedrick, P. W. & Ruiz-Linares, A. (2006) Admixture dynamics in Hispanics: a shift in the nuclear genetic ancestry of a South American population isolate. *Proc Natl Acad Sci USA* **103**, 7234–7239.

Behar, D. M., Villems, R., Soodyall, H., Blue-Smith, J., Pereira, L., Metspalu, E., Scozzari, R., Makkan, H., Tzur, S., Comas, D., Bertranpetit, J., Quintana-Murci, L., Tyler-Smith, C., Wells, R. S., Rosset, S. & Genographic Consortium. (2008) The dawn of human matrilineal diversity. *Am J Hum Genet* **82**, 1130–1140.

Bertoni, B., Jin, L., Chakraborty, R. & Sans, M. (2005) Directional mating and a rapid male population expansion in a hybrid Uruguayan population. *Am J Hum Biol* **17**, 801–808.

Bobillo, M. C., Zimmermann, B., Sala, A., Huber, G., Röck, A., Bandelt, H. J., Corach, D. & Parson, W. (2009) Amerindian mitochondrial DNA haplogroups predominate in the population of Argentina: towards a first nationwide forensic mitochondrial DNA sequence database. *Int J Legal Med*, in press.

Bortolini, M. C., Salzano, F. M., Thomas, M. G., Stuart, S., Nasanen, S. P., Bau, C. H., Hutz, M. H., Layrisse, Z., Petzl-Erler, M. L., Tsuneto, L. T., Hill, K., Hurtado, A. M., Castro-de-Guerra, D., Torres, M. M., Groot, H., Michalski, R., Nymadawa, P., Bedoya, G., Bradman, N., Labuda, D. & Ruiz-Linares, A. (2003)

Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am J Hum Genet* **73**, 524–539.

Brandstätter, A., Niederstätter, H., Pavlic, M., Grubwieser, P. & Parson, W. (2007) Generating population data for the EMPOP database – an overview of the mtDNA sequencing and data evaluation processes considering 273 Austrian control region sequences as example. *Forensic Sci Int* **166**, 164–175.

Campos-Sánchez, R., Barrantes, R., Silva, S., Escamilla, M., Ontiveros, A., Nicolini, H., Mendoza, R., Munoz, R. & Raventos, H. (2006) Genetic structure analysis of three Hispanic populations from Costa Rica, Mexico, and the southwestern United States using Y-chromosome STR markers and mtDNA sequences. *Hum Biol* **78**, 551–563.

Corach, D., Penacino, G., Sala, A., Iannucci, N., Bernardi, P., Doretti, M., Fondebrider, L., Ginarte, A., Inchaurregui, A., Somigliana, C., Turner, S. & Hagelberg, E. (1997) Additional approaches to DNA typing of skeletal remains: the search for missing persons killed during the last dictatorship in Argentina. *Electrophoresis* **18**, 1608–1612.

Corach, D., Filgueira Risso, L., Marino, M., Penacino, G. & Sala, A. (2001) Routine Y-STR typing in forensic casework. *Forensic Sci Int* **118**, 131–135.

Corach, D., Marino, M. & Sala, A. (2006) Relevant genetic contribution of Amerindian to the extant population of Argentina. In: *Progress in Forensic Genetics 11, International Congress Series*, Vol. 1288 (eds. A. Amorim, F. Corte-Real & N. Morling), pp. 130–132. The Netherlands: Elsevier.

Cruciani, F., La Fratta, R., Trombetta, B., Santolamazza, P., Sellitto, D., Colomb, E. B., Dugoujon, J. M., Crivellaro, F., Benincasa, T., Pascone, R., Moral, P., Watson, E., Melegh, B., Barbujani, G., Fuselli, S., Vona, G., Zagradisnik, B., Assum, G., Brdicka, R., Kozlov, A., Efremov, G. D., Coppa, A., Novelletto, A. & Scozzari, R. (2007) Tracing past human male movements in northern/eastern Africa and western Eurasia: new clues from Y-chromosomal haplogroups E-M78 and J-M12. *Mol Biol Evol* **24**, 1300–1311.

Excoffier, L., Smouse, P. E. & Quattro, J. M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491.

Finnilä, S., Lehtonen, M. S. & Majamaa, K. (2001) Phylogenetic network for European mtDNA. *Am J Hum Genet* **680**, 1475–1484.

Ginther, C. H., Corach, D., Penacino, G., Rey, J. A., Carnese, F. R., Hutz, M. H., Anderson, A., Just, J., Salzano, F. M. & King, M-C. (1993) Genetic variation among the Mapuche Indians from the Patagonian region of Argentina: mitochondrial sequence variation and allele frequencies of several nuclear gene. In: *DNA Fingerprinting: state of the Science* (eds. S. D. J. Pena, R. Chakraborty, J. T. Epplen & A. J. Jeffreys), pp. 211–219, Berlin: Brinkäuser.

Hanks, P. & Hodges, F. (1989) *A dictionary of surnames.* Oxford: Oxford University Press.

Jobling, M. A. & Tyler-Smith, C. (2003) The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* **4**, 598–612.

Karafet, T. M., Mendez, F. L., Meilerman, M. B., Underhill, P. A., Zegura, S. L. & Hammer, M. F. (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* **18**, 830–838.

Kayser, M., Caglia, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., Honda, K., Jobling, M., Krawczak, M., Leim, K., Meuser, S., Meyer, E., Oesterreich, W., Pandya, A., Parson, W., Penacino, G.,

Perez-Lezaun, A., Piccinini, A., Prinz, M., Schmitt, C., Schneider, P., Szibor, R., Teifel-Greding, J., Weichhold, G., de Knijff, P. & Roewer, L. (1997) Evaluation of Y-chromosomal STRs: a multicenter study. *Int J Legal Med* **110**, 125–133.

Kayser, M., Krawczak, M., Excoffier, L., Dieltjes, P., Corach, D., Pascali, V., Gehrig, C., Bernini, L. F., Jespersen, J., Bakker, E., Roewer, L. & de Knijff, P. (2001) An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. *Am J Hum Genet* **68**, 990–1018.

Kayser, M., Brauer, S., Cordaux, R., Casto, A., Lao, O., Zhivotovsky, L. A., Moyse-Faurie, C., Rutledge, R. B., Schiefenhoevel, W., Gil, D., Lin, A. A., Underhill, P. A., Oefner, P. J., Trent, R. J. & Stoneking, M. (2006) Melanesian and Asian origins of Polynesians: mtDNA and Y-chromosome gradients across the Pacific. *Mol Biol Evol* **23**, 2234–2244.

Kayser, M., Lao, L., Saar, K., Brauer, S., Wang, X., Nürnberg, P., Trent, R. J. & Stoneking, M. (2008) Genome-wide analysis indicates more Asian than Melanesian ancestry of Polynesians. *Am J Hum Genet* **82**, 194–198.

Kersbergen, P., van Duijn, K., Kloosterman, A. D., den Dunnen, J. T., Kayser, M. & de Knijff, P. (2009) Developing a set of ancestry-sensitive DNA markers reflecting continental origins of humans. *BMC Genetics*, in press.

Kittles, R. A. & Weiss, K. M. (2003) Race, ancestry, and genes: implications for defining disease risk. *Annu Rev Genomics Hum Genet* **4**, 33–67.

Kivisild, T., Shen, P., Wall, D. P., Do, B., Sung, R., Davis, K., Passarino, G., Underhill, P. A., Scharfe, C., Torroni, A., Scozzari, R., Modiano, D., Coppa, A., de Knijff, P., Feldman, M., Cavalli-Sforza, L. L. & Oefner, P. J. (2006) The role of selection in the evolution of human mitochondrial genomes. *Genetics* **172**, 373–387.

Kong, Q. P., Bandelt, H. J., Sun, C., Yao, Y. G., Salas, A., Achilli, A., Wang, C. Y., Zhong, L., Zhu, C. L., Wu, S. F., Torroni, A. & Zhang, Y. P. (2006) Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations. *Hum Mol Genet* **15**, 2076–2086.

Lao, O., van Duijn, K., Kersbergen, P., de Knijff, P. & Kayser, M. (2006) Proportioning whole-genome single-nucleotide-polymorphism diversity for the identification of geographic population structure and genetic ancestry. *Am J Hum Genet* **78**, 680–690.

Lao, O., de Gruijter, J. M., van Duijn, K., Navarro, A. & Kayser, M. (2007) Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms. *Ann Hum Genet* **71**, 354–369.

Liu, N., Chen, L. & Wang, S. (2005) Comparison of single-nucleotide polymorphism and microsatellites in inference of population structure. *BMC Genet* **6**, S26–S30.

Luis, J. R., Rowold, D. J., Regueiro, M., Caeiro, B., Cinnioğlu, C., Roseman, C., Underhill, P. A., Cavalli-Sforza, L. L. & Herrera, R. J. (2004) The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. *Am J Hum Genet* **74**, 532–544.

Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonné-Tamir, B., Sykes, B. & Torroni, A. (1999) The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* **64**, 232–249.

Marino, M., Sala, A. & Corach, D. (2006a) Population genetic analysis of 15 autosomal STRs loci in the central region of Argentina. *Forensic Sci Int* **161**, 72–77.

Marino, M., Sala, A. & Corach, D. (2006b) Genetic analysis of the populations from Northern and Mesopotamian provinces of Argentina by means of 15 autosomal STRs. *Forensic Sci Int* **160**, 224–230.

Marino, M., Sala, A. & Corach, D. (2006c) Genetic attributes of 15 autosomal STRs in the population of two patagonian provinces of Argentina. *Forensic Sci Int* **160**, 84–88.

Marino, M., Sala, A. & Corach, D. (2007) Genetic attributes of the YHRD minimal haplotype in 10 provinces of Argentina. *Forensic Sci Int Genet* **1**, 129–133.

Martínez Marignac, V. L., Bertoni, B., Parra, E. J. & Bianchi, N. O. (2004) Characterization of admixture in an urban sample from Buenos Aires, Argentina, using uniparentally and biparentally inherited genetic markers. *Hum Biol* **76**, 543–557.

Martinez Sarasola, C. (2005) *Nuestros paisanos los indios*. Buenos Aires: Emece Ed.

Parson, W. & Bandelt, H. J. (2007) Extended guidelines for mtDNA typing of population data in forensic science. *Forensic Sci Int Genet* **1**, 13–19.

Pritchard, J. K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.

Richards, M. B., Macaulay, V. A., Bandelt, H. J. & Sykes, B. C. (1998) Phylogeography of mitochondrial DNA in western Europe. *Ann Hum Genet* **62**, 241–260.

Rawley, J. A. & Behrendt S. D. (2005) *The transatlantic slave trade: a history, revised*. Lincoln and London: University of Nebraska Press.

Rock, D. (1987) *Argentina, 1516-1987: from Spanish colonization to Alfonsín*. Berkeley and Los Angeles: University of California Press.

Rosenberg, N. A. (2004) Distruct: a program for the graphical display of population structure. *Mol Ecol Notes* **4**, 137–138.

Sala, A., Penacino, G. & Corach, D. (1998) Comparison of allele frequencies of eight STR loci from Argentinean Amerindian and European populations. *Hum Biol* **70**, 937–947.

Sala, A., Penacino, G., Carnese, R. & Corach, D. (1999) Reference database of hypervariable genetic markers of Argentina: applications for molecular anthropology and forensic casework. *Electrophoresis* **208**, 1733–1739.

Salas, A., Jaime, J. C., Alvarez-Iglesias, V. & Carracedo, A. (2008) Gender bias in the multiethnic genetic composition of central Argentina. *J Hum Genet* **53**, 662–674.

Sanchez, J. J., Børsting, C., Hallenberg, C., Buchard, A., Hernandez, A. & Morling, N. (2003) Multiplex PCR and minisequencing of SNPs–a model with 35 Y chromosome SNPs. *Forensic Sci Int* **137**, 74–84.

Sans, M., Weimer, T. A., Franco, M. H., Salzano, F. M., Bentancor, N., Alvarez, I., Bianchi, N. O. & Chakraborty, R. (2002) Unequal contributions of male and female gene pools from parental populations in the African descendants of the city of Melo, Uruguay. *Am J Phys Anthropol* **118**, 33–44.

Seldin, M. F., Tian, C., Shigeta, R., Scherbarth, H. R., Silva, G., Belmont, J. W., Kittles, R., Gamron, S., Allevi, A., Palatnik, S. A., Alvarellos, A., Paira, S., Caprarulo, C., Guillerón, C., Catoggio, J., Prigione, C., Berbotto, G. A., García, M. A., Perandones, C. E., Pons-Estel, B. A. & Alarcon-Riquelme, M. E. (2007) Argentine population genetic structure: large variance in Amerindian contribution. *Am J Phys Anthropol* **132**, 455–462.

Semino, O., Magri, C., Benuzzi, G., Lin, A. A., Al-Zahery, N., Battaglia, V., Maccioni, L., Triantaphyllidis, C., Shen, P., Oefner, P. J., Zhivotovsky, L. A., King, R., Torroni, A., Cavalli-Sforza, L. L., Underhill, P. A. & Santachiara-Benerecetti, A. S. (2004) Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am J Hum Genet* **74**, 1023–1034.

Underhill, P. A., Jin, L., Zemans, R., Oefner, P. J. & Cavalli-Sforza, L. L. (1996) A pre-Colombian Y chromosome-specific transition and its implications for human evolutionary history. *Proc Natl Acad Sci USA* **93**, 196–200.

Wang, S., Ray, N., Rojas, W., Parra, M. V., Bedoya, G., Gallo, C., Poletti, G., Mazzotti, G., Hill, K., Hurtado, A. M., Camrena, B., Nicolini, H., Klitz, W., Barrantes, R., Molina, J. A., Freimer, N. B., Bortolini, M. C., Salzano, F. M., Petzl-Erler, M. L., Tsuneto, L. T., Dipierri, J. E., Alfaro, E. L., Bailliet, G., Bianchi, N. O., Llop, E., Rothhammer, F., Excoffier, L. & Ruiz-Linares, A. (2008) Geographic patterns of genome admixture in Latin American mestizos. *PLoS Genet* **4**, e1000037.

Yang, N., Li, H., Criswell, L. A., Gregersen, P. K., Alarcon-Riquelme, M. E., Kittles, R., Shigeta, R., Silva, G., Patel, P. I., Belmont, J. W. & Seldin, M. F. (2005) Examination of ancestry and ethnic affiliation using highly informative diallelic DNA markers: application to diverse and admixed populations and implications for clinical epidemiology and forensic medicine. *Hum Genet* **118**, 382–392.

Supporting Information

Additional Supporting Information may be found in the online version of the article:

Table S1 Frequencies of NRY haplogroups in Argentineans from each of the provinces and geographic regions and their assumed continental region of origins.

Table S2 Frequencies of mtDNA haplogroups in Argentineans from each of the provinces and geographic regions and their assumed continental region of origins.

Table S3 Sequence information and reaction details of minisequencing assays.

Figure S1 Histogram of the proportion of Native American autosomal ancestry estimated by means of STRUCTURE in the Argentinean population.

Figure S2 Boxplot of the Native American autosomal genetic ancestry according to geographic origin of surnames

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organised for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.