Annals of human genetics

doi: 10.1111/j.1469-1809.2012.00712.x

Apolipoprotein E/C1/C4/C2 Gene Cluster Diversity in Two Native Andean Populations: Aymaras and Quechuas

Magdalena Gayà-Vidal[1,2], Georgios Athanasiadis[1], Robert Carreras-Torres[1], Marc Via[1], Esther Esteban[1], Mercedes Villena[3], René Vasquez[4], Jean-Michel Dugoujon[2] and Pedro Moral[1]*

[1]*Biologia Animal, Facultat de Biologia, Universitat de Barcelona, Barcelona 08028, Spain*
[2]*Laboratoire AMIS, Université de Toulouse III-CNRS UMR 5288, Toulouse 31073, France*
[3]*Instituto Boliviano de Biología de Altura, La Paz, 591, Bolivia*
[4]*Instituto Boliviano de Biología de Altura, Potosí, 591, Bolivia*

Summary

The *APOE/C1/C4/C2* gene cluster presents high relevance in lipid metabolism and, therefore, has important epidemiological implications. Here, we study for the first time the variation patterns of 25 polymorphisms (10 short tandem repeats, STRs, and 15 single nucleotide polymorphismas, SNPs) in two native Andean samples from Bolivia (45 Aymaras and 45 Quechuas) as well as one European sample (n = 41) as external reference. We estimated diversity parameters, linkage disequilibrium patterns, population structure, and possible selective effects. In general, diversity was low and could be partly attributed to selection (probably due to its physiological importance), since the *APOE/C1/C4/C2* region was highly conserved compared to the flanking genes in both Bolivians and Europeans. Moreover, the lower gene diversity in Bolivians compared to Europeans for some markers might indicate different demographic histories. Regarding the *APOE* isoforms, in addition to $\varepsilon 3$ (94%) and $\varepsilon 4$ (5%), isoform $\varepsilon 2$ (1%) was also detected in Bolivians. In relation to previous hypotheses, our results support that genetic drift or founder effects rather than selection for increased cholesterol absorption are the main factors that have shaped the distribution of *APOE* isoforms observed in South America.

Keywords: *APOE/C1/C4/C2* gene cluster, *APOE*, Native Americans, Andean region, South America, high-altitude populations

Introduction

Apolipoproteins play an important role in lipid metabolism. One of the most studied ones is apolipoprotein E (ApoE), involved in lipoprotein metabolism and lipid transport. The *APOE* gene is located in chromosome 19q13.2, closely linked to the *APOC1*, *APOC4*, and *APOC2* genes and forming the *APOE/C1/C4/C2* gene cluster, which spans about 48 kb. ApoE, C1, and C2 proteins are constituents of chylomicrons, very low-density lipoproteins and high-density lipoproteins (HDL). ApoC4 plays an important role in the metabolism of circulating lipids as an activator of lipoprotein lipase (Ken-Dror et al., 2010).

Three common isoforms with different physiological properties have been described for the ApoE: $\varepsilon 2$, $\varepsilon 3$, and $\varepsilon 4$. These isoforms are determined by two SNPs (rs429358 and rs7412 in the coding residues 112 and 158; Hanis et al., 1991). In comparison with the most frequent variant (*APOE* $\varepsilon 3$), *APOE* $\varepsilon 4$ has been associated with higher total and LDL cholesterol levels and increased risk of cardiovascular and Alzheimer's diseases (Song et al., 2004; Lucatelli et al., 2011). Other polymorphisms, such as –219G→T, located in the *APOE* promoter, have also been associated with cardiovascular, Alzheimer's, and Parkinson's diseases (Tycko et al., 2004; Xin et al., 2010).

To a lesser extent, variation in the other components of the *APOE/C1/C4/C2* gene cluster has also been described, but its role in the risk of complex diseases is not clear (Kamboh et al., 2000). For instance, the most studied polymorphism in the *APOC1* gene is the insertion of the CGTT tetranucleotide at position –317 of the promoter that has a negative effect on *APOC1* transcription levels (Xu et al., 1999). In recent years, detailed studies of the variation in the whole cluster have been performed, mainly in populations of

*Corresponding author: Pedro Moral, Dpt. Biologia Animal-Antropologia, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal, 645, Barcelona 08028, Spain. Tel: 34 934021461; Fax: 34 934035740; E-mail: pmoral@ub.edu

European origin, although most of them have been carried out in epidemiological surveys (Klos et al., 2008; Ken-Dror et al., 2010).

To date, studies in Native Americans have mainly reported *APOE* isoform frequencies (Crews et al., 1993; Marin et al., 1997; Scacchi et al., 1997; de Andrade et al., 2000; Jaramillo-Correa et al., 2001; Demarchi et al., 2005). An additional study provided data for two polymorphisms in the *APOC1* and *APOC2* genes in five Brazilian samples (de Andrade et al., 2002). Moreover, there are genome-wide SNP data available for several Native American groups from the Human Genome Diversity Project (HGDP)-Centre d'Etude du Polymorphisme Humain (CEPH) project, although sample sizes are very small ($n < 15$; Jakobsson et al., 2008; Li et al., 2008).

In light of the above, this paper explores for the first time the variation of 25 polymorphisms in two Native Andean populations from Bolivia. Furthermore, in order to provide an external group for comparison, a European sample from Spain has also been included. Twenty-four of these markers lie within a 108 kb-long region including the *APOE/C1/C4/C2* gene cluster as well as three flanking genes (*PVRL2*, *TOMM40, CLPTM1*).

The high-altitude habitat of the two Bolivian populations adds a special anthropological and genetic interest to this study. It is well known that high-altitude populations are adapted to hypoxia with morphological and physiological particularities (Baker, 1969; Beall, 2007a, b). Some of these adaptations are metabolic and hormonal changes such as lower blood glucose, better and faster glucose uptake in peripheral tissues, higher insulin sensitivity, greater hyperglycaemic hormone secretion and caloric expense, and a lesser use of triglycerides in energy production due to lower oxygen concentration. High-altitude Andeans also present lower prevalence of high blood pressure and diabetes mellitus (Seclén et al., 1999; Mohanna et al., 2006). Several studies have reported a particular lipid profile in Andean groups compared to sea-level inhabitants. This profile includes high levels of plasma triglycerides, low HDL, and hypercholesterolemia, also observed in Tibetans (Mohanna et al., 2006; Sherpa et al., 2011). However, there is no agreement on the lipid profile of altitude Andeans with previous studies that reported lower total and LDL cholesterol levels, higher HDL and lower triglycerides levels (Piedra et al., 1981; Bellido et al., 1993). Variation on these phenotypes may be influenced by different factors such as genetics, chronic hypoxia, lifestyle, and diet.

The main goal of this work is to provide the first genetic data on Native Andean populations for the *APOE/C1/C4/C2* gene cluster. Due to the functional nature of some of the studied polymorphisms, a good knowledge of their variation patterns in different populations is of great value for both population and epidemiological purposes. As an additional step, we used the data to explore aspects of both population structure in South America (contrasting with previously published data) and the evolutionary history of the *APOE/C1/C4/C2* genomic region.

Materials and Methods

Samples

Two Native American samples from the main native linguistic groups in Bolivia (Aymaras and Quechuas) were collected by the *Instituto Boliviano de Biología de Altura* (IBBA) in Bolivia. Aymara samples came from the area near Lake Titicaca (Lat: 16°34′S, Lon: 68°23′W, Elevation: 4000 m), whereas Quechua samples came from the Potosí department (Lat: 19°13′S, Lon: 65°49′W, Elevation: 3244 m), both regions being located in the Andean Altiplano. The study included a total of 90 unrelated individuals of both sexes (45 for each population sample). A more detailed description of these samples has been presented elsewhere (Gayà-Vidal et al., 2010; 2011). For a broader context, a European sample from Northeast Spain (Girona, Catalonia) was also included in the analysis. The European sample consisted of 41 unrelated individuals of both sexes that had their four grandparents born in the region. All subjects gave their informed consent and the study was approved by the ethical committees from the involved institutions (IBBA and University of Barcelona).

Polymorphisms

The 25 studied markers included 15 SNPs and nine STRs located in the genomic region between positions 50063546 and 50172127 of chromosome 19 (NCBI36/hg18), containing the *APOE/C1/C4/C2* gene cluster, and one STR located at 134.2 kb downstream from this region. Marker ID reference and their relative location on the genomic region are shown in Figure 1.

SNPs were selected according to different criteria: functional importance, coverage of the gene cluster flanking regions, and formation of compound SNPSTR markers (Mountain et al., 2002). *APOE* isoforms were coded by the rs429358 and rs7412 genotypes. Ten STRs were selected after scanning the *APOE/C1/C4/C2* region with the UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway) and the SNPSTR database (Mountain et al., 2002). Details on the analysed SNPs and STRs are described in Tables 1 and 2.

Genotyping

With the exception of rs7412, all SNPs were typed with the iPLEX Gold assay on the Sequenom MassARRAY Platform

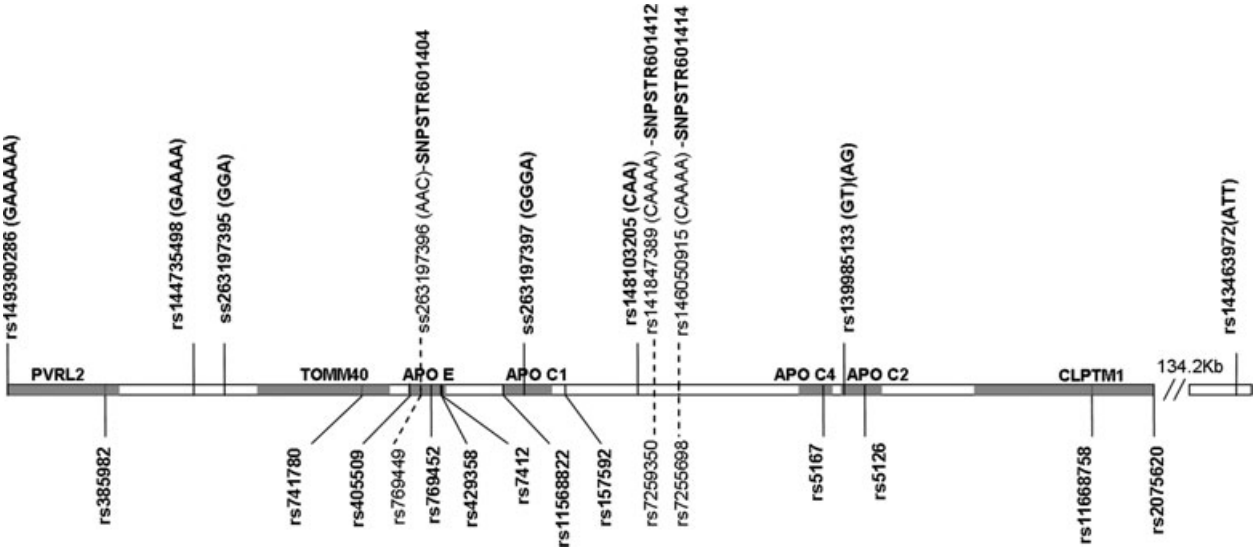

Figure 1 Markers examined and their relative location. Grey boxes represent the genes. STRs are in the upper part of the scheme and SNPs in the bottom part. Dashed lines represent SNPSTR markers.

(Sequenom, San Diego, CA, USA). SNP rs7412 was genotyped by Real-Time PCR (RT-PCR), using a TaqMan SNP genotyping assay protocol (Applied Biosystems, Foster City, CA, USA) and a total volume of 5 μl per well. RT-PCR fluorescence measurements and data collection were carried out on an ABI PRISM 7900 HT Sequence Detection System (Applied Biosystems).

STRs were genotyped by PCR amplification using standard conditions in a total volume of 10 μl (protocol available upon request). One of the oligonucleotides from each pair of primers was modified at the 5′ end with 6-FAM or HEX fluorescent dye. The diluted PCR products were analysed in an Applied Biosystems 3130 Genetic Analyser using the GeneScan 500 ROX size standard and genotypes were determined using the ABI Prism GeneMapper v3.0 software (Applied Biosystems). Positive and negative controls were included in all experiments, and the exact number of repeats was verified by direct sequencing.

Statistical Analysis

Allele frequencies and heterozygosities were calculated with Genetix v4.05.2 (Belkhir et al., 1998) and Hardy-Weinberg equilibrium was assessed with Arlequin v3.1 (Excoffier et al., 2005). Pairwise linkage disequilibrium (LD) for biallelic markers was quantified by the r^2 statistic using Haploview v.4.1 (Barrett et al., 2005). Also, the Black and Krafsur test (Black & Krafsur, 1985) and the LD test (10,000 permutations) were performed with Genetix v4.05.2. The statistical significance of the nonrandom distribution of each pair of loci was evaluated by Fisher's exact test with Genepop v4.0 (Raymond & Rousset, 1995; Rousset, 2008). Haplotype phase was inferred with PHASE v2.1 (Stephens et al., 2001; Stephens & Donnelly, 2003).

To check for possible selective pressure acting on the *APOE/C1/C4/C2* gene cluster, we followed two independent strategies: (i) we performed the long-range haplotype (LRH) test (Sabeti et al., 2002) and (ii) we compared Fst values between neutral and potentially functional loci.

The LRH test detects recent positive selection by investigating the decay of LD between a predefined core haplotype that includes a locus of interest and a number of genotyped markers at variable distances. To do so, two different statistics were calculated using the Sweep software program (P.C. Sabeti; unpublished data): the extended haplotype heterozygosity (EHH) and the relative extended haplotype heterozygosity (REHH). EHH represents the probability that two randomly chosen chromosomes carrying the core haplotype of interest are homozygous at all markers for the entire interval from the core region to a certain distance. Core haplotypes that have unusually high EHH (long-range LD) and a high population frequency indicate the presence of a mutation that increased its frequency in the population faster than expected under neutral evolution. Moreover, the REHH test controls for recombination rate variation across the region (Sabeti et al., 2002). In our study, only SNPs were considered and the core haplotype consisted of the two SNPs determining the *APOE* isoforms (rs429358 and rs7412).

In addition, Fst values were calculated for all loci by a locus-by-locus analysis of molecular variance (AMOVA) with Arlequin v3.1. The Fst values of the potentially functional loci [i.e. the two SNPs determining the *APOE* isoforms (rs429358

Table 1 Allele frequencies and heterozygosities for the SNPs.

Gene location	SNPSTR	SNPs	Aymara	H	Quechua	H	Europeans	H
PVRL2 intron		**rs385982**	(45)	0.533	(44)	0.364	(41)	0.342
		G	0.38		0.34		0.22	
		T	0.62		0.66		0.78	
TOMM40 intron		**rs741780**	(45)	0.511	(44)	0.455	(41)	0.585
		A	0.32		0.32		0.51	
		G	0.68		0.68		0.49	
APOE promoter		**rs405509**	(39)	0.436	(39)	0.461	(41)	0.537
		A	0.68		0.56		0.41	
		C	0.32		0.44		0.59	
APOE intron 2	601404	**rs769449**	(43)	0.070	(42)	0.119	(41)	0.098
		A	0.03		0.06		0.05	
		G	0.97		0.94		0.95	
APOE exon 3		**rs769452** (T/C)	(38)	-	(38)	-	41	-
		T	1		1		1	
APOE exon 4, determining the APOE isoforms		**rs429358**	(39)	0.103	(38)	0.105	(39)	0.103
		C	0.05		0.04		0.08	
		T	0.95		0.95		0.92	
		rs7412	(45)	0.022	(44)	-	(41)	0.195
		C	0.99		1		0.88	
		T	0.01		0		0.12	
APOC1 promoter		**rs11568822**	(45)	0.067	(43)	0.023	(41)	0.415
		-	0.94		0.99		0.77	
		CGTT	0.06		0.01		0.23	
Between *APOC1* and *APOC4*		**rs157592** (G/T)	(41)	-	(41)	-	(41)	-
		G	1		1		1	
	601412	**rs7259350** (C/T)	(45)	-	(44)	-	(41)	-
		C	1		1		1	
	601414	**rs7255698** (C/G)	(45)	-	(44)	-	(41)	-
		C	1		1		1	
APOC4 exon 3		**rs5167**	(45)	0.511	(44)	0.568	(41)	0.537
		G	0.32		0.35		0.37	
		T	0.68		0.65		0.63	
APOC2 exon 4		**rs5126** (A/C)	(39)	-	(40)	-	(41)	-
		A	1		1		1	
CLPTM1 introns		**rs11668758**	(45)	0.467	(44)	0.546	(41)	0.439
		C	0.68		0.66		0.68	
		T	0.32		0.34		0.32	
		rs2075620	(39)	0.410	(39)	0.513	(41)	0.439
		A	0.69		0.67		0.68	
		G	0.31		0.33		0.32	
		Average H*		0.313		0.315		0.369

Numbers inside brackets represent the number of individuals genotyped. SNP IDs are in bold.
H, observed heterozygosity.
*Considering the 10 polymorphic loci.

and rs7412), the SNPs located in the *APOE* and *APOC1* promoters (rs405509 and rs11568822), and the missense SNP of *APOC4* (rs5167)] were compared to those of the neutral ones with the nonparametric Mann-Whitney test using R (http://www.r-project.org). In the absence of selection, no statistically significant differences are expected between the Fst distributions of neutral versus potentially functional loci.

Population differentiation between samples was evaluated using an exact G test (Genepop v4; Raymond & Rousset, 1995) for the allele frequencies of the 25 markers.

Table 2 Allele frequencies and heterozygosities for the STRs.

Gene location	SNPSTR	STRs	Aymara	H	Quechua	H	Europeans	H
PVRL2 intron	601395[a]	**rs149390286**	(42)	0.429	(38)	0.395	(39)	0.308
		$(TTTTTC)_6$	0.69		0.72		0.85	
		$(TTTTTC)_7$	0.31		0.28		0.15	
Between *PVRL2* and *TOMM40*	601400[a]	**rs144735498**	(43)	-	(31)	-	(36)	0.083
		(TTTTC)7	0		0		0.04	
		$(TTTTC)_8$	1		1		0.96	
		ss263197395	(28)	-	(24)	-	(39)	-
		$(GGA)_{10}$	1		1		1	
APOE intron2	601404	**ss263197396**	(24)	-	(29)	-	(38)	-
		$(TTG)_{13}$	1		1		1	
APOC1 intron 3	601408[a]	**ss263197397**	(19)	-	(20)	-	(10)	-
		$(GGGA)_9$	1		1		1	
Between *APOC1* and *APOC4*		**rs148103205**	(42)	0.024	(44)	-	(41)	0.024
		$(AAC)_7$	0		0		0.01	
		$(AAC)_9$	0.99		1		0.99	
		$(AAC)_{10}$	0.01		0		0	
	601412	**rs141847389**	(45)	0.267	(44)	0.205	(41)	0.268
		$(CAAAA)_3$	0		0		0.04	
		$(CAAAA)_4$	0.18		0.15		0.12	
		$(CAAAA)_5$	0.82		0.85		0.84	
	601414	**rs146050915**	(35)	-	(36)	-	(40)	-
		$(TTTTG)_5$	1		1		1	
APOC2 intron 1	601416[a]	**rs139985133**	(44)	0.750	(38)	0.763	(41)	0.829
		$(TG)(AG)_{17}$	0.44		0.40		0.16	
		$(TG)(AG)_{19}$	0.01		0		0	
		$(TG)(AG)_{21}$	0.01		0		0.09	
		$(TG)(AG)_{24}$	0.01		0		0	
		$(TG)(AG)_{25}$	0.18		0.18		0.04	
		$(TG)(AG)_{26}$	0.02		0		0	
		$(TG)(AG)_{27}$	0		0.01		0.05	
		$(TG)(AG)_{28}$	0.02		0.14		0.32	
		$(TG)(AG)_{29}$	0.27		0.22		0.13	
		$(TG)(AG)_{30}$	0.01		0.04		0.16	
		$(TG)(AG)_{31}$	0		0		0.05	
		$(TG)(AG)_{32}$	0.01		0		0	
		$(TG)(AG)_{33}$	0		0		0.01	
134.2 Kb 3′	601432[a]	**rs143463972**	(43)	0.093	(38)	0.053	(41)	0.439
		$(ATT)_8$	0.05		0.03		0.34	
		$(ATT)_9$	0.95		0.97		0.66	
		Average H*		0.313		0.283		0.325

Numbers inside brackets represent the number of individuals genotyped. STR IDs are in bold.
H, observed heterozygosity.
*Considering the polymorphic loci.
[a]For these SNPSTR markers, only the STR was genotyped, not the SNP.

Comparisons were performed at different levels depending on the available data from other Native American groups: (i) three loci (rs405509, rs5167, rs11668758) in five populations from the HGDP-CEPH Project: Karitiana (14 individuals) and Surui (8) from Brazil, Piapoco and Curripaco from Colombia (7), and Maya (21) and Pima (14) from Mexico; (ii) one locus (rs11568822) in five Brazilian samples (de Andrade et al., 2002): WaiWai (29), Xavante (31), Gaviao (29), Zoró (30), and Surui (24); and (iii) *APOE* isoforms in 37 Native South American populations. In addition, population structure was explored considering the 39 populations for the *APOE* isoforms through a hierarchical AMOVA with Arlequin v3.1, using geographical criteria (tropical forest, southern corn– according to Demarchi et al. (2005)– and Andes).

Table 3 APOE isoform genotypes in the two Bolivian and the European samples.

	Genotypes						Alleles		
Population	$\varepsilon2/\varepsilon2$	$\varepsilon2/\varepsilon3$	$\varepsilon2/\varepsilon4$	$\varepsilon3/\varepsilon3$	$\varepsilon3/\varepsilon4$	$\varepsilon4/\varepsilon4$	$\varepsilon2$	$\varepsilon3$	$\varepsilon4$
Aymaras (n = 39)	0	0	1	35	3	0	0.013	0.936	0.051
Quechuas (n = 38)	0	0	0	34	4	0	0	0.947	0.053
Bolivians (n = 77)	0	0	1	69	7	0	0.007	0.942	0.052
Europeans (n = 39)	1	8	0	25	4	1	0.128	0.795	0.077

Results

Allele Frequencies and Heterozygosity

Five out of 15 SNPs were monomorphic, three of which formed SNPSTR markers (Table 1). All polymorphic SNPs were in Hardy-Weinberg equilibrium. In the two Bolivian samples, six polymorphic loci presented diversity values from 0.36 to 0.57 and four (*APOE* and *APOC1* SNPs) presented heterozygosity <0.12 (minor allele frequencies ≤6%). Average heterozygosity across the 10 polymorphic SNPs was slightly higher in the European sample than in Bolivians (Table 1), although differences were not statistically significant.

Variation in the STRs was also remarkably low: five markers were monomorphic, four were biallelic (one of them was triallelic in Europeans), and only one was multiallelic with ten (Aymaras), six (Quechuas), and nine (Europeans) different alleles (Table 2). Heterozygosity values ranged from 0.024 to 0.750 and average values were slightly lower in Bolivians than in Europeans, but without significant differences. The most notable differences were found for STR rs143463972, with Europeans presenting five and eight times higher heterozygosity than Quechuas and Aymaras, respectively.

It is important to mention that the $(TG)_n(AG)_m$ STR was initially thought to be a $(TG)_n$ microsatellite according to different databases. However, after sequencing five individuals, variation in the number of repeats in the contiguous $(AG)_m$ was observed. The alleles of 17, 19, 24, 26, and 29 repeats presented different dinucleotide combinations according to the $(TG)_{9/12/17/18/21}(AG)_{7/8}$ pattern, indicating that it is a compound STR.

No significant differences in allele frequencies of any marker were found between the two Bolivian populations after applying the Bonferroni correction. Differences between each Bolivian sample and the European population were significant for two STRs (rs139985133, $P < 0.0001$; rs143463972, $P < 0.0001$). Significant differences were also observed between Quechuas and Europeans for two SNPs (rs7412, $P < 0.001$; rs11568822, $P < 0.0001$), and between Aymaras and Europeans for one locus (rs405509, $P < 0.001$). These SNPs also showed differences between Bolivians pooled together and Europeans.

As for the *APOE* isoforms (Table 3), the most common genotype in Bolivians was the homozygote for *APOE* $\varepsilon3$ (90%), followed by $\varepsilon3/\varepsilon4$ (9%), and $\varepsilon2/\varepsilon4$ (only 1%). Thus, the $\varepsilon3$ allele presented a very high frequency (∼94%), followed by $\varepsilon4$ (∼5%) and $\varepsilon2$ (0.7%; only one allele was detected in the Aymara sample). In Europeans, genotype frequencies appeared slightly different ($\varepsilon3/\varepsilon3 = 64\%$; $\varepsilon2/\varepsilon3 = 20.5\%$, $\varepsilon3/\varepsilon4 = 10\%$, $\varepsilon4/\varepsilon4 = 2.6\%$, and $\varepsilon2/\varepsilon2 = 2.6\%$), but differences were not significant.

Linkage Disequilibrium and Haplotypes

LD patterns between all pairs of biallelic markers (SNPs and STRs) for the pooled Bolivian and the European samples are shown in Figure 2. For Bolivians, the analysis revealed moderate to high LD ($r^2 \geq 0.35$, Fisher's exact test, $P < 0.001$) among markers in the 5′ flanking region (*PVRL2* and *TOMM40* genes). Similarly, complete linkage was observed in the 3' flanking region between rs11668758 and rs2075620 (*CLPTM1*), and very high LD between these *CLPTM1* markers and rs5167 (*APOC4*; $P < 0.001$), and rs139985133 STR (*APOC2*; $P < 0.0001$). In contrast, a clear break was observed in the LD between *APOC1* and *APOC4*. LD between *APOE* and *APOC1* SNPs was heterogeneous, probably due to the low variability of some markers. In fact, rs11568822 (*APOC1*) presented moderate LD with rs429358 ($P < 0.0001$) and no LD with rs7412 ($P > 0.05$). Within the *APOE* region, it is interesting to note the lack of LD between rs405509 in the promoter and the other *APOE* SNPs ($P > 0.05$) as well as between rs429358 and rs7412, determining the *APOE* isoforms, only 138 bp apart.

The LD analysis was also performed considering the *APOE* isoforms instead of the two determinant SNPs, since this is the method used in previous studies. Strong LD between the *APOE* isoforms and *APOE* rs769449 ($r^2 = 0.71$, $P < 0.0001$) was found. Moderate LD was observed in the pooled Bolivians between the *APOE* isoforms and *APOC1* rs11568822 ($r^2 = 0.61$; $P < 0.0001$), although it was almost complete in Aymaras ($r^2 = 0.99$, $P < 0.0001$), but low in Quechuas ($r^2 = 0.24$, $P > 0.05$). High LD was also observed between the *APOE* isoforms and the STR rs148103205 in Aymaras ($r^2 = 0.97$, $P = 0.026$).

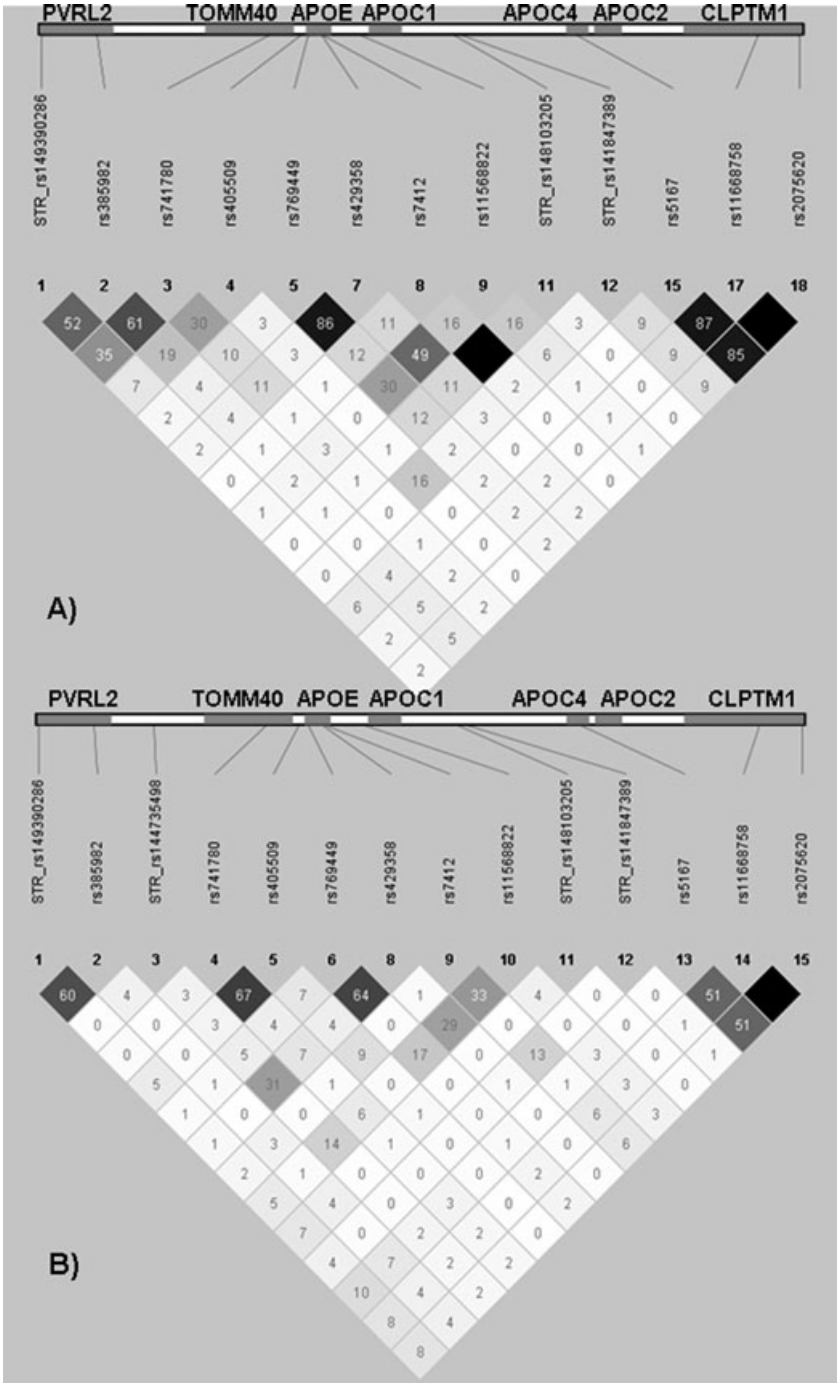

Figure 2 LD patterns from Haploview in Bolivians (A) and Europeans (B) for biallelic markers. The colour scheme represents r^2 values (white: $r^2 = 0$, black: $r^2 = 1$, shades of grey: $0 < r2 < 1$). Numbers refer to r^2 values (%), an empty cell being 100% ($r^2 = 1$). To include the STR rs141847389 in Europeans, the allele $(CAAAA)_3$ was excluded for this analysis.

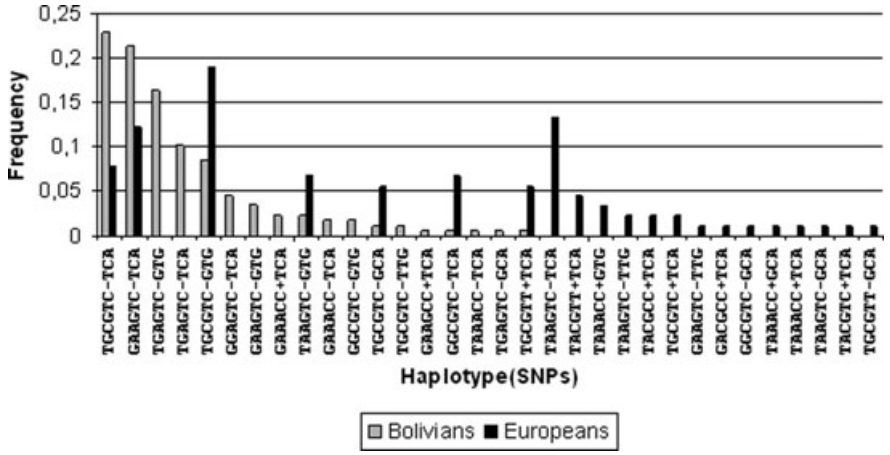

Figure 3 Frequency of the estimated SNP-haplotypes in the pooled Bolivian samples and in the European sample.

In comparison, LD patterns in Europeans were, in general, similar to those of Bolivians (Fig. 2B). However, weaker LD was observed in the 5′ end of the studied region, with high LD only within *PVRL2* (STR rs149390286 and rs385982; $r^2 = 0.61$, $P < 0.0001$), as well as in the 3' end between *CLPTM1* and *APOC4* ($r^2 = 0.51$, $P < 0.0001$), and between *CLPTM1* and *APOC2* ($P = 0.04$). More homogenous LD values were observed between the *APOC1* and *APOE* loci ($r^2 = 0.33$, $P < 0.001$).

Figure 3 shows the estimated SNP haplotypes in the pooled Bolivian sample and in Europeans. Eighteen and 21 different haplotypes were found in Bolivians and Europeans, respectively, but the two samples only shared seven of these haplotypes. In Bolivians, the four most common haplotypes accounted for 71% of the total frequency whereas in Europeans only for 52%. There was no coincidence in the ranking of the most frequent haplotypes between Bolivians and Europeans. Detailed frequencies are presented in Table S1.

Potential Role of Natural Selection: Long-Range Haplotype Tests and Fst Comparisons

The LRH test (Fig. S1) showed that the three core haplotypes presented a similar LD decay pattern in Europeans, which did not indicate the presence of selection. In Bolivians, the EHH and REHH patterns were not clear. The most frequent haplotype (94%, TC; corresponding to isoform $\varepsilon 3$) presented a higher EHH than the minor core haplotype CC (corresponding to the ancestral $\varepsilon 4$) on the right part of the graph, but not on the left part. High EHH (EHH = 1) was observed near the core for the most frequent haplotype. In both Europeans and Bolivians, the drastic LD decay occurring 3′ when the distance increases, probably reflects the LD break between *APOC1* and *APOC4*.

Fst values revealed a higher degree of between-population differentiation for the functional SNPs (0.036, $P < 0.0001$) than for the neutral ones (0.014, $P < 0.01$; Table 4). However,

Table 4 Fst values for the neutral and functional variants considering: (i) all samples (Bolivians and Europeans) and (ii) the Bolivian samples alone.

	Fst	
Polymorphism	Bolivians & Europeans	Bolivians
STR rs149390286	0.023 NS	0
rs385982	0.019 NS	0
STR rs144735498	0.029 NS	-
rs741780	0.039*	0
rs405509	0.058**	0.015 NS
rs769449	0	0
rs429358	0	0
rs7412	0.085****	0
rs11568822	0.124****	0.018 NS
STR rs148103205	0	0.001 NS
STR rs141847389	0	0
rs5167	0	0
STR rs139985133	0.072****	0.002 NS
rs11668758	0	0
rs2075620	0	0
STR rs143463972	0.230****	0
Global Fst_neutral_SNPs	0.014*	0
Global Fst_functional_SNPs	0.036****	0.001 NS
Global Fst_SNPs	0.021**	0
Global Fst_markers (SNPs & STRs)	0.034****	0

****$P < 0.0001$, ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$.
NS, nonsignificant.
Functional markers are in bold.

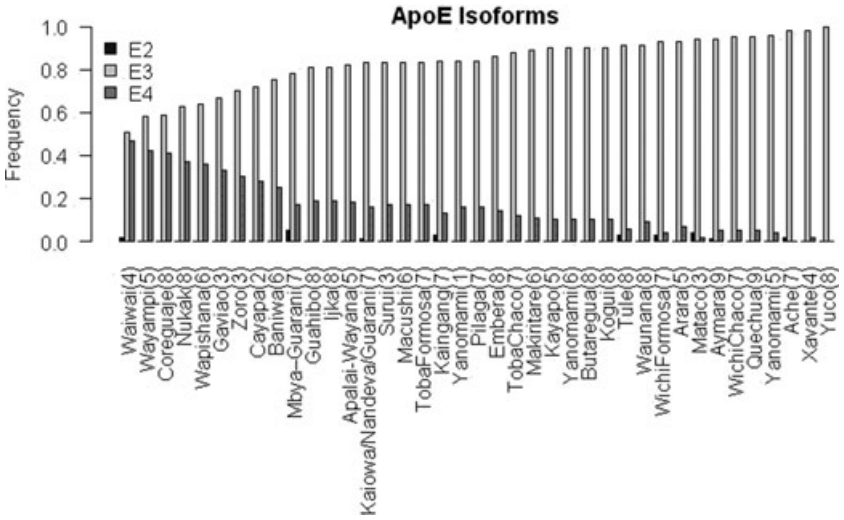

Figure 4 APOE isoform frequencies in Native South Americans. Numbers refer to references: 1, Crews et al., 1993; 2, Scacchi et al., 1997; 3, de Andrade et al., 2000; 4, de Andrade et al., 2002; 5, Marin et al., 1997; 6, JV Nel (de Andrade et al., 2000); 7, Demarchi et al., 2005; 8, Jaramillo-Correa et al., 2001; 9, present work.

no significant differences were present when comparing Fst values of neutral with functional markers according to the Mann-Whitney test ($P > 0.05$). Among the functional loci, the highest Fst values were observed for markers on the *APOE* and *APOC1* promoters and rs7412, ranging from 0.06 to 0.12. The global Fst value was higher when considering both SNP and STR markers than when considering only SNPs, probably due to the contribution of the STR rs143463972 at 134.2 kb 3′ of the region, which presented the highest Fst value of all markers.

Variation in Native American Samples

The exact test of population differentiation based on the available data from three loci in American samples from the HGDP revealed that Bolivian samples presented significant differences from Piapoco and Curripaco (Colombia), Kariatiana (Brazil), and Pima (Mexico), mainly due to differences in rs11668758 and rs5167. Differences were not significant between Bolivians and Surui (Brazil) or Mayas (Mexico). The fact that some of these populations presented very low sample sizes, especially Piapoco-Curripaco and Surui, should be taken into account. The comparison based on rs11568822 (*APOC1*) in seven South American samples, showed a high among-population differentiation (Fst = 0.45, $P < 0.0001$). The frequency of the deletion allele ranged from 0.77 (Gaviao) to 0.99 (Quechuas). Quechuas did not present significant differences from Xavante (also with a high frequency of the deletion– 98%), whereas Aymaras did not present any significant differences from Xavante, Zoró, or Gaviao.

APOE $\varepsilon 3$ is the most frequent *APOE* isoform in all 39 South American samples, with frequencies ranging from 0.51 to 1 (Fig. 4). On the contrary, the absence of *APOE* $\varepsilon 2$ is a common feature among South American populations, present at very low frequencies (from 0.01 to 0.05) only in eight out of the 39 samples. On the other hand, there is high variation in the frequency of *APOE* $\varepsilon 4$ (from 0 to 0.47). Considering the 39 South American samples, *APOE* isoform variation between populations was significant (Fst = 0.07, $P < 0.001$) and the exact test of population differentiation revealed significant differences between the two Bolivian samples and nine other samples (Cayapa, Gaviao, Zoro, WaiWai, Wayampi, Wapishana, Baniwa, Coreguaje, and Nukak). No structure was revealed by the hierarchical AMOVA in South America (Fst = 0.083, Fsc = 0.059, Fct = 0.025; $P < 0.05$).

Table 5 shows the *APOE* isoforms/*APOC1* haplotypes observed in Native South American samples. The most common haplotype was $\varepsilon 3$/deletion (–) followed by $\varepsilon 4$/+, but at different frequencies. The other three haplotypes were present only in some Native American samples. The two Bolivian samples presented one of the highest frequencies of the $\varepsilon 3$/– haplotype (~95%), right after Xavante. It is important to note that the second most frequent haplotype in the Aymaras was $\varepsilon 4$/+, whereas in Quechuas it was $\varepsilon 4$/–.

Table 5 APOE isoforms/APOC1 promoter haplotype frequencies and standard deviations in Native Americans and Europeans.

Haplotype	Wai-Wai*	Xavante*	Gaviao*	Zoró*	Surui*	Aymara	Quechua	Bolivians	All Native Am.	Europeans
ε2/–	0	0	0	0	0	0	0	0	0	0.011 ± 0.004
ε2/+	0	0	0	0	0	0.011 ± 0.005	0	0.006 ± 0.003	0.001	0.100 ± 0.004
ε3/–	0.55 ± 0.07	0.98 ± 0.02	0.67 ± 0.06	0.62 ± 0.07	0.83 ± 0.05	0.944 ± 0.002	0.953 ± 0.009	0.949 ± 0.004	0.790	0.766 ± 0.005
ε3/+	0.02 ± 0.03	0	0.04 ± 0.03	0.16 ± 0.05	0	0	0	0	0.031	0.022 ± 0.009
ε4/–	0.27 ± 0.06	0	0.09 ± 0.04	0.09 ± 0.02	0	0	0.035 ± 0.009	0.017 ± 0.004	0.071	0
ε4/+	0.16 ± 0.05	0.02 ± 0.02	0.20 ± 0.05	0.13 ± 0.05	0.17 ± 0.05	0.044 ± 0.005	0.011 ± 0.006	0.028 ± 0.003	0.104	0.100 ± 0.009
Total N	29	31	29	30	24	45	44	89	232	45

For the *APOC1* promoter (rs11568822), "–" corresponds to the deletion and "+" to the CGTT insertion.
*Data from Andrade et al. (2002).

Discussion

This work studies the variability of the genomic region encompassing the *APOE/C1/C4/C2* gene cluster in two Andean populations from Bolivia. As such, it provides novel data on the genetic characterisation of Native American populations. The 25 analysed markers constitute the first data not only on Andean samples, but also, for most markers, on Native Americans in general, let alone the first data for 9 out of the 10 STRs examined in this genomic region.

In general, our results revealed low genetic variation in the Andean samples for both sets of markers as well as for the *APOE* isoforms. However, the differences in marker diversity according to their location, as well as the allele frequency comparisons between Bolivian and European samples, suggests that some of this low diversity in Bolivia could be explained as a result of selection, while another part would reflect the demographic history of Native American populations.

Our analysis suggests that selective factors may have kept variability low in the middle of the studied genomic region (from *TOMM40* to *APOC2*). In this middle region, 8 out of 12 SNPs present very low gene diversity both in Native Americans and Europeans (Table 1). A similar observation stands true for the STRs. Seven out of 10 markers with a more central position were monomorphic or showed very low diversity both in Bolivians and Europeans (Table 2).

On the contrary, the SNPs with the highest diversities were those located at the edges of the studied region, both in Bolivians and Europeans, as well as for most of the HapMap populations (data not shown). This is reflected in the fact that the most common haplotypes differ mainly at the edges (Fig. 3, Table S1), while the core formed by the *APOE* and *APOC1* polymorphisms remains practically invariable. As a result, the *APOE* ε3 isoform and the deletion in the *APOC1* promoter are the most common alleles worldwide. A similar situation is observed for the STRs, with rs149390286 at the 5′ end being one of the most diverse STR in Bolivians and Europeans, and rs143463972 at the 3′ end showing high diversity, although only in Europeans. This high conservation of the *APOE/C1/C4/C2* region is probably related to its biological importance.

At this point, we should mention that the high diversity of the compound STR $(TG)_n(AG)_m$ rs139985133 is remarkable. Although it presented significant differences in allele frequencies, the range of repeats detected in Bolivians was quite similar to that found in our European population and in a French sample (Zouali et al., 1999). The sequencing of five individuals in this study confirmed a previous report that diversity at this locus is attributable to variation at both $(TG)_n$ and $(AG)_m$ motifs (Fornage et al., 1992).

The fact that the EHH test did not show signs of selection in Europeans and the results were not clear in the Bolivian populations (Fig. S1) could reflect that this method may not be sensitive enough to detect selection in cases where selection is not recent enough. Long haplotypes as a signature of positive selection persist for relatively short periods (<30,000 years; Sabeti et al., 2006). The fact that the most frequent isoform worldwide (ε3) is not the ancestral one (ε4) is consistent with an increase of the ε3 allele frequency prior to the major population expansions that accompanied the spread of anatomically modern humans <100,000 years ago (Fullerton et al., 2000). These authors found that the APOE locus was one of the less variable genes examined so far, and proposed a reduction in variation associated with the rise in frequency of an advantageous mutation. Reduction in genetic diversity is a signature that may be used to detect positive selection up to <250,000 years (Sabeti et al., 2006).

As for the Fst comparisons, although no significant differences were detected between the functional and neutral Fst distributions according to the Mann-Whitney test, which is admittedly a rather conservative test, it is worth noting that the variation at the functional loci was more than double that of the neutral ones. In addition, the highest Fst values within the gene cluster region were observed in the APOE and APOC1 promoters and rs7412, which have a clear physiological role.

In any case, it is important to note that selection is likely not detectable in many instances, such as if the selection advantage is too small or selection acts on an allele that is already at an intermediate frequency in the population (Sabeti et al., 2006).

On the other hand, the lower gene diversity observed in the Andean populations compared to Europeans for several markers could be related to the particular demographic history of Native Americans, where drift and founder effects might have had an important impact. In particular, the STR rs143463972 located 134.2 kb downstream of the *APOE/C1/C4/C2* genomic region presented higher heterozygosity in Europeans than in Bolivians. Also, Bolivians presented a lower number of SNP-based haplotypes, some of them with extreme frequencies compared to Europeans. Even though the LD pattern in Bolivians presented several similarities to those from HapMap and other studies (Klos et al., 2008; Ken-Dror et al., 2010), such as the presence of an LD block at the 3′ of the region and a break in LD between *APOC1* and *APOC4*, the higher degree of LD values in Bolivians compared to Europeans is consistent with a more recent origin and/or a bottleneck during the first settlement of the Americas and particularly of South America.

It is worth mentioning that the absence of LD between rs405509 (in the *APOE* promoter) and the *APOE* isoform alleles is in agreement with a recent study (Ken-Dror et al., 2010), but in contrast to other reports that found partial LD (Fullerton et al., 2000; Heijmans et al., 2002). It is likely that these controversial results are associated with the use of different LD measures (D′ or r^2) to some extent.

The difference in the LD pattern between the *APOE* isoforms and *APOC1* observed in Aymaras (almost complete) and Quechuas (low) confirms the existence of variation in the distribution of these haplotypes previously observed in Native Americans (de Andrade et al., 2002). The almost complete LD in Aymaras, with the $\varepsilon 3/-$, $\varepsilon 4/+$, and $\varepsilon 2/+$ arrangements, and the presence of $\varepsilon 4/-$ in Quechuas (Table 5) might be due to the fact that the Aymara population has been more isolated than the Quechuas (Gayà-Vidal et al., 2011).

As for the variation among Native Americans considering three loci, it is important to note that no differences were observed between Bolivians and Mayas. This similarity between Andean and Mesoamerican populations has also been observed for other systems such as autosomal microsatellites (Wang et al., 2008). According to these authors, this similarity was compatible with an early Pacific coastal colonisation, although it could also be interpreted as recent gene flow along the coast. Comparisons based on rs11568822 (*APOC1*) as well as on the *APOE* isoforms-*APOC1* haplotypes in seven South American samples revealed a relatively high variation within South America, probably due to differences in the population history of some of these populations (de Andrade et al., 2002). Finally, comparisons of *APOE* isoforms across South America did not show a clear pattern of variation. The heterogeneity observed was mainly due to differences on the *APOE* $\varepsilon 4$ frequencies. The two Bolivian samples of this work present a high prevalence of *APOE* $\varepsilon 3$ and, consequently, a very low frequency of *APOE* $\varepsilon 4$, within the variation range observed in South Americans. The fact that one individual from the Aymara sample was a carrier of an *APOE* $\varepsilon 2$ allele is important in the frame of the controversial question about the presence or absence of *APOE* $\varepsilon 2$ in Native South Americans. Most Native South American populations studied so far (30 out of 39) do not have this allele and some have it at very low frequencies. *APOE* $\varepsilon 2$ is present in Native Americans either because it was present in the founding populations at a very low frequency (and lost in some groups) or due to admixture with non-Native Americans (de Andrade et al., 2000). The fact that the Aymara sample presents a very low degree of admixture according to different markers (Gayà-Vidal et al., 2011) supports its presence in the founding population, although more data are necessary for such an affirmation.

Concerning the *APOE* $\varepsilon 4$ distribution in Native Americans, Corbo and Scacchi (1999) suggested that this allele could have been favoured by increasing cholesterol absorption in populations with low-cholesterol diets. Conversely, de Andrade et al. (2000) suggested that founder effects, isolation, and genetic drift, rather than natural selection, could have played a major role in determining *APOE* $\varepsilon 4$ frequencies in Native Americans. The *APOE* $\varepsilon 4$ frequency of the two Bolivian samples is among the lowest (0.05) in South America and their diet, typical of high-altitude populations, is mainly based on carbohydrates and low intake of animal fats and proteins (Caen et al., 1974). Moreover, their meat consumption comes mainly from camelids, which are characterised by low cholesterol content (Saadoun & Cabrera, 2008). Thus, our data seem to support a genetic drift explanation rather than a selection effect. Nevertheless, in the case of our Bolivian samples, the role of the *APOE* has to be understood in the context of the adaptations to a high-altitude environment.

Conclusion

This paper illustrates that the study of the variation in the *APOE/C1/C4/C2* region in the general population is useful for both population and epidemiological purposes. Our results showed a high degree of conservation of the *APOE/C1/C4/C2* gene cluster, which may be due to its physiological importance. Nevertheless, certain differences detected between the Bolivian samples and Europeans would reflect their different demographic histories. The *APOE* isoform frequencies in Bolivians were within the range observed for South Americans, providing new insights into some controversial issues. The low frequency of *APOE* $\varepsilon 4$ in Andeans suggests that its frequency in South Americans would be mainly due to genetic drift and founder effects. More samples should be studied to reach more robust conclusions.

Acknowledgements

We are grateful to all the donors for providing DNA samples and to the people who contributed to the collection of samples. This research has been supported by the Ministerio de Ciencia y Tecnología (CGL2005–3391, CGL2008–03955) and the Generalitat de Catalunya (2005-SGR00252). The work of MGV has been financed by an FI grant from the "Departament d'Educació i Universitats de la Generalitat de Catalunya i el Fons Social Europeu." The authors declare that they have no conflict of interest.

References

Baker, P. T. (1969) Human adaptation to high altitude. *Science* **163**, 1149–1156.

Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005) Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265.

Beall, C. M. (2007a) Two routes to functional adaptation: Tibetan and Andean high-altitude natives. *Proc Natl Acad Sci USA* **104** (Suppl 1), 8655–8660.

Beall, C. M. (2007b) Detecting natural selection in high-altitude human populations. *Respir Physiol Neurobiol* **158**, 161–171.

Belkhir, K., Borsa, P., Goudet, J., Chikhi, L., & Bonhomme, F. (1998) *GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations*. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000. Montpellier, France: Université de Montpellier II.

Bellido, D., Barragán, M., Aguilar, M., & Colque, N. (1993) Perfil lipídico de la población de Viacha. *Acta Andina* **2**, 28–29.

Black, W. C. & Krafsur, E. S. (1985) A FORTRAN program for the calculation and analysis of two-locus linkage disequilibrium coefficients. *Theor Appl Genet* **70**, 491–496.

Caen, J. P., Drouet, L., Bellanger, R., Michel, H., & Henon, P. (1974) Thrombosis, platelet behaviour, fibrinolytic activity and diet on the Andes Plateau. Dietary fats and thrombosis. *Haemostasis* **2**, 13–20.

Corbo, R. M. & Scacchi, R. (1999) Apolipoprotein E (APOE) allele distribution in the world. Is APOE*4 a 'thrifty' allele? *Ann Hum Genet* **63**, 301–310.

Crews, D. E., Kamboh, M. I., Mancilha-Carvalho, J. J., & Kottke, B. (1993) Population genetics of apolipoprotein A-4, E, and H polymorphisms in Yanomami Indians of northwestern Brazil: Associations with lipids, lipoproteins, and carbohydrate metabolism. *Hum Biol* **65**, 211–224.

de Andrade, F. M., Coimbra, C. E. Jr., Santos, R. V., Goicoechea, A., Carnese, F. R., Salzano, F. M., & Hutz, M. H. (2000) High heterogeneity of apolipoprotein E gene frequencies in South American Indians. *Ann Hum Biol* **27**, 29–34.

de Andrade, F. M., Ewald, G. M., Salzano, F. M., & Hutz, M. H. (2002) Lipoprotein lipase and APOE/APOC-I/APOC-II gene cluster diversity in native Brazilian populations. *Am J Hum Biol* **14**, 511–518.

Demarchi, D. A., Salzano, F. M., Altuna, M. E., Fiegenbaum, M., Hill, K., Hurtado, A. M., Tsunetto, L. T., Petzl-Erler, M. L., & Hutz, M. H. (2005) APOE polymorphism distribution among Native Americans and related populations. *Ann Hum Biol* **32**, 351–365.

Excoffier, L., Laval, G., & Schneider, S. (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinform Online* **1**, 47–50.

Fornage, M., Chan, L., Siest, G., & Boerwinkle, E. (1992) Allele frequency distribution of the (TG)n(AG)m microsatellite in the apolipoprotein C-II gene. *Genomics* **12**, 63–68.

Fullerton, S. M., Clark, A. G., Weiss, K. M., Nickerson, D. A., Taylor, S. L., Stengård, J. H., Salomaa, V., Vartiainen, E., Perola, M., Boerwinkle, E., & Sing, C. F. (2000) Apolipoprotein E variation at the sequence haplotype level: Implications for the origin and maintenance of a major human polymorphism. *Am J Hum Genet* **67**, 881–900.

Gayà-Vidal, M., Dugoujon, J. M., Esteban, E., Athanasiadis, G., Rodríguez, A., Villena, M., Vasquez, R., & Moral, P. (2010) Autosomal and X chromosome Alu insertions in Bolivian Aymaras and Quechuas: Two languages and one genetic pool. *Am J Hum Biol* **22**, 154–162.

Gayà-Vidal, M., Moral, P., Saenz-Ruales, N., Gerbault, P., Tonasso, L., Villena, M., Vasquez, R., Bravi, C. M., & Dugoujon, J. M. (2011) mtDNA and Y-chromosome diversity in Aymaras and Quechuas from Bolivia: Different stories and special genetic traits of the Andean Altiplano populations. *Am J Phys Anthropol* **145**, 215–230.

Hanis, C. L., Hewett-Emmett, D., Douglas, T. C., Bertin, T. K., & Schull, W. J. (1991) Effects of the Apolipoprotein E polymorphism on levels of lipids, lipoproteins, and apolipoproteins among Mexican-Americans in Starr County, Texas. *Arterioscler Thromb* **1**, 362–370.

Heijmans, B. T., Slagboom, P. E., Gussekloo, J., Droog, S., Lagaay, A. M., Kluft, C., Knook, D. L., & Westendorp, R. G. (2002) Association of APOE epsilon2/epsilon3/epsilon4 and promoter gene variants with dementia but not cardiovascular mortality in old age. *Am J Med Genet* **107**, 201–208.

Jakobsson, M., Scholz, S. W., Scheet, P., Gibbs, J. R., VanLiere, J. M., Fung, H. C., Szpiech, Z. A., Degnan, J. H., Wang, K., Guerreiro, R., Bras, J. M., Schymick, J. C., Hernandez, D. G., Traynor, B. J., Simon-Sanchez, J., Matarin, M., Britton, A., van de Leemput, J., Rafferty, I., Bucan, M., Cann, H. M., Hardy, J. A., Rosenberg, N. A., & Singleton, A. B. (2008) Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* **451**, 998–1003.

Jaramillo-Correa, J. P., Keyeux, G., Ruiz-Garcia, M., Rodas, C., & Bernal, J. (2001) Population genetic analysis of the genes APOE, APOB (3′VNTR) and ACE in some black and Amerindian communities from Colombia. *Hum Hered* **52**, 14–33.

Kamboh, M. I., Aston, C. E., & Hamman, R. F. (2000) DNA sequence variation in human apolipoprotein C4 gene and its effect on plasma lipid profile. *Atherosclerosis* **152**, 193–201.

Ken-Dror, G., Talmud, P. J., Humphries, S. E., & Drenos, F. (2010) *APOE/C1/C4/C2* gene cluster genotypes, haplotypes and lipid levels in prospective coronary heart disease risk among UK healthy men. *Mol Med* **16**, 389–399.

Klos, K., Shimmin, L., Ballantyne, C., Boerwinkle, E., Clark, A., Coresh, J., Hanis, C., Liu, K., Sayre, S., & Hixson, J. (2008) *APOE/C1/C4/C2* hepatic control region polymorphism influences plasma APOE and LDL cholesterol levels. *Hum Mol Genet* **17**, 2039–2046.

Li, J. Z., Absher, D. M., Tang, H., Southwick, A. M., Casto, A. M., Ramachandran, S., Cann, H. M., Barsh, G. S., Feldman, M., Cavalli-Sforza, L. L., & Myers, R. M. (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–1104.

Lucatelli, J. F., Barros, A. C., Silva, V. K., Machado, Fda. S., Constantin, P. C., Dias, A. A., Hutz, M. H., & de Andrade, F. M. (2011) Genetic influences on Alzheimer's disease: Evidence of interactions between the genes APOE, APOC1 and ACE in a sample population from the South of Brazil. *Neurochem Res* **36**, 1533–1539.

Marin, G. B., Tavella, M. H., Guerreiro, J. F., Santos, S. E. B., & Zago, M. A. (1997) Absence of the E2 allele of apolipoprotein in Amerindians. *Braz J Genet* **20**, 741–743.

Mohanna, S., Rossana, B., & Seclén, S. (2006) Lipid profile, waist circumference, and body mass index in a high altitude population. *High Alt Med Biol* **7**, 245–255.

Mountain, J. L., Knight, A., Jobin, M., Gignoux, C., Miller, A., Lin, A. A., & Underhill, P. A. (2002) SNPSTRs: Empirically derived, rapidly typed, autosomal haplotypes for inference of population history and mutational processes. *Genome Res* **12**, 1766–1772.

Piedra, A., Marticorena, E., & Muro, M. (1981) Lípidos en individuos normales de altura y del mar a propósito del proceso de envejecimiento en la altura. *Arch Biol Andina* **11**, 220–225.

Raymond, M. & Rousset, F. (1995) GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *J Heredity* **86**, 248–249.

Rousset, F. (2008) Genepop′007: A complete reimplementation of the Genepop software for Windows and Linux. *Mol Ecol Resources* **8**, 103–106.

Saadoun, A. & Cabrera, M. C. (2008) A review of the nutritional content and technological parameters of indigenous sources of meat in South America. *Meat Sci* **80**, 570–581.

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., Richter, D. J., Schaffner, S. F., Gabriel, S. B., Platko, J. V., Patterson, N. J., McDonald, G. J., Ackerman, H. C., Campbell, S. J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., & Lander, E. S. (2002) Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832–837.

Sabeti, P. C., Schaffner, S. F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., Palma, A., Mikkelsen, T. S., Altrhuler, D., & Lander, E. S. (2006) Positive selection in the human lineage. *Science* **312**, 1614–1620.

Scacchi, R., Corbo, R. M., Rickards, O., Mantuano, E., Guevara, A., & De Stefano, G. F. (1997) Apolipoprotein B and E genetic polymorphisms in the Cayapa Indians of Ecuador. *Hum Biol* **69**, 375–382.

Seclén, S., Leey, J., Villena, A., Herrera, B., Penacho, C., Carrasco, A., & Vargas, R. (1999) Prevalencia de diabetes mellitus, hipertensión arterial, hipercolesterolemia y obesidad como factores de riesgo coronario y cerebrovascular en población adulta de la costa, sierra y selva del Perú. *Acta Med Peruana* **17**, 8–12.

Sherpa, L. Y., Deji, S. H., Stigum, H., Chongsuvivatwong, V., Luobu, O., Thelle, D. S., Nafstad, P., & Bjertness, E. (2011) Lipid profile and its association with risk factors for coronary heart disease in the highlanders of Lhasa, Tibet. *High Alt Med Biol* **12**, 57–63.

Song, Y., Stampfer, M. J., & Liu, S. (2004) Meta-analysis: Apolipoprotein E genotypes and risk for coronary heart disease. *Ann Intern Med* **141**, 137–147.

Stephens, M. & Donnelly, P. (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* **73**, 1162–1169.

Stephens, M., Smith, N. J., & Donnelly, P. (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* **68**, 978–989.

Tycko, B., Lee, J. H., Ciappa, A., Saxena, A., Li, C. M., Feng, L., Arriaga, A., Stern, Y., Lantigua, R., Shachter, N., & Mayeux, R. (2004) *APOE* and *APOC1* promoter polymorphisms and the risk of Alzheimer disease in African American and Caribbean Hispanic individuals. *Arch Neurol* **61**, 1434–1439. Erratum in: *Arch Neurol* 2005 62, 332.

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.

Xin, X. Y., Ding, J. Q., & Chen, S. D. J. (2010) Apolipoprotein E promoter polymorphisms and risk of Alzheimer's disease: Evidence from meta-analysis. *Alzheimers Dis* **19**, 1283–1294.

Xu, Y., Berglund, L., Ramakrishnan, R., Mayeux, R., Ngai, C., Holleran, S., Tycko, B., Leff, T., & Shachter, N. S. (1999) A common Hpa I RFLP of apolipoprotein C-I increases gene transcription and exhibits an ethnically distinct pattern of linkage disequilibrium with the alleles of apolipoprotein E. *J Lipid Res* **40**, 50–58.

Zouali, H., Faure-Delanef, L., & Lucotte, G. (1999) Chromosome 19 locus apolipoprotein C-II association with multiple sclerosis. *Mult Scler* **5**, 134–136.

Supporting Information

Additional supporting information may be found in the online version of this article:

Figure S1 Long-range haplotype analysis considering the SNPs.

Table S1 Frequency (±SE) of the estimated haplotypes in the two different world regions.

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 reorganised 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.

Received: 23 January 2012
Accepted: 10 April 2012