

SCIENTIFIC REPORTS



OPEN

Selection scan reveals three new loci related to high altitude adaptation in Native Andeans

Vanessa C. Jacovas¹, Cainã M. Couto-Silva², Kelly Nunes², Renan B. Lemes², Marcelo Z. de Oliveira³, Francisco M. Salzano¹, Maria Cátira Bortolini¹ & Tábita Hünemeier²

The Andean Altiplano has been occupied continuously since the late Pleistocene, ~12,000 years ago, which places the Andean natives as one of the most ancient populations living at high altitudes. In the present study, we analyzed genomic data from Native Americans living a long-time at Andean high altitude and at Amazonia and Mesoamerica lowland areas. We have identified three new candidate genes - *SP100*, *DUOX2* and *CLC* - with evidence of positive selection for altitude adaptation in Andeans. These genes are involved in the *TP53* pathway and are related to physiological routes important for high-altitude hypoxia response, such as those linked to increased angiogenesis, skeletal muscle adaptations, and immune functions at the fetus-maternal interface. Our results, combined with other studies, showed that Andeans have adapted to the Altiplano in different ways and using distinct molecular strategies as compared to those of other natives living at high altitudes.

Along their great expansion, humans have inhabited almost all environments in the five continents. Among several harsh environments that were occupied, the highlands are probably the ones that needed more adaptations for survival¹. At least in three geographically distinct locations have this evolutionary adaptation been studied: Andean Altiplano (South America), Himalaya (China/Tibet, Asia) and Semien Mountain (northern Ethiopia, Africa) Plateaus. Andes have been peopled continuously since the late Pleistocene, ~12,000 yBP² while the time of settlement and permanent occupation of both Tibet and Ethiopia remain a topic of debate, varying widely^{3,4}. Despite some uncertainties in the permanent occupation dating, it is certain that humans have inhabited these regions of hostile climates for thousands of years.

Several physiologic factors are associated with living at high altitude ($\geq 2,500$ meters where only 75% of the oxygen available at sea level occurs; http://www.altitude.org/air_pressure.php), including adaptations for high ultraviolet radiation index, thermal amplitude, and changes in the pulmonary capacity due to hypoxia^{5,6}. High altitude leads to a rapid physiologic/adaptive response in individuals from lowlands; however, prolonged exposure to environmental-related factors might have harmful outcomes. Remarkable features such as increased pulmonary function, hypoxia tolerance, and increased hemoglobin levels have been observed in Andean populations⁷. How such adaptations took place is still not clear, and just a few genes have been associated with the high altitude adaptation phenotype in human populations⁸⁻¹³.

Interestingly, the set of genes presenting signs of natural selection changes according to high altitude, indicating that under an analogous selective pressure, different genetic solutions have emerged. For instance, genomic scans for selection have revealed at least 40 candidate genes related to the Hypoxia Inducible Factor (HIF), such as *EPAS1* in populations from Tibet, *EGLN1* in Andeans and Tibetans and *THRB* and *ARNT2* in Ethiopians^{8,14-18}. The populations from the Andean plateau also presented signs of natural selection in other genes, such as *BRINP3*, *NOS2*, and *TBX5*, involved in the nitric oxide pathway (NOS) and related to cardiovascular health¹². In addition, Jacovas *et al.*¹⁹ using the candidate gene approach inferred that a combination of some derived and ancestral alleles of *USP7*, *LIF* and *MDM2* genes, all three in the *TP53* pathway, could have been essential for the successful establishment of Native American populations in the Andean highlands.

¹Genetics Departament, Biosciences Institute, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil.

²Department of Genetics and Evolutionary Biology, Biosciences Institute, University of São Paulo, São Paulo, SP, Brazil. ³Vale do Rio dos Sinos University, São Leopoldo, RS, Brazil. Vanessa C. Jacovas, Cainã M. Couto-Silva and Kelly Nunes contributed equally. Correspondence and requests for materials should be addressed to M.C.B. (email: maria.bortolini@ufrgs.br) or T.H. (email: hunemeier@usp.br)

SNP	Allele		Gene	Position	PBS	XPEHH			
	Ancestral	Derived				Andean vs. Mesoamerican	p-value	Andean vs. Amazonian	p-value
rs13411586	C*	T	<i>SP100</i>	230988046	0.5846	2.3789	0.0037	2.1703	0.0065
rs9678342	C*	T	<i>SP100</i>	230991955	0.5547	2.3193	0.0044	2.1356	0.0071
rs7582700	T*	C	<i>SP100</i>	231024349	0.4644	2.2704	0.0050	2.1074	0.0076
rs7039618	A	G*	<i>TMEM38B</i>	107497627	0.3618	0.0842	0.3312	0.5672	0.1458
rs3817141	T*	C	<i>TMEM38B</i>	107507950	0.3906	0.0255	0.3099	0.6205	0.1351
rs10978213	G*	A	<i>TMEM38B</i>	107511706	0.3618	0.0235	0.3092	0.6171	0.1358
rs10816302	A*	G	<i>TMEM38B</i>	107526354	0.3835	0.0937	0.2697	0.6664	0.1264
rs10978240	A	G*	<i>TMEM38B</i>	107575093	0.3923	0.0764	0.2753	0.6307	0.1331
rs1046778	T	C*	<i>AS3MT</i>	104651474	0.3124	0.5023	0.5118	0.4008	0.4631
rs269866	G*	A	<i>DUOX2</i>	43181698	0.6185	2.0599	0.0086	2.5865	0.0021
rs440191	A	G*	<i>CLC</i>	44913483	0.3039	1.6207	0.0234	0.3166	0.2046

Table 1. Population Branch Statistic (PBS) individual values and Cross-Population Extended Haplotype Homozygosity (XP-EHH) for all SNPs found under selection in Native Andean populations. Ancestral and derived alleles according to the 1000 Genomes data. *Putative selected alleles.

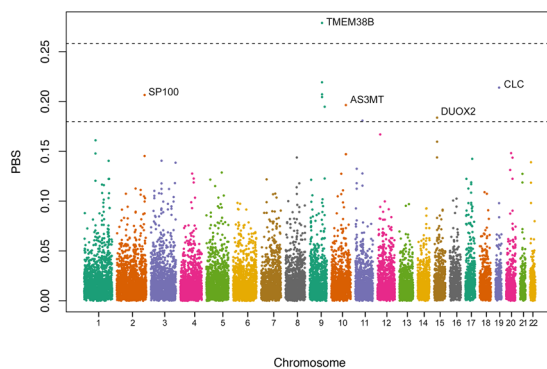


Figure 1. Average PBS values in windows of 20 SNPs, using a step size of 5 SNPs. The 99.5th and 99.9th percentiles of the empirical distribution are shown as black dashed horizontal lines. Names of genes associated with the highest peaks are shown.

Since different investigations pointed to distinct sets of genes involved in high altitude adaptation, more studies are necessary to fully understand the different genetic landscapes present in highland populations around the world. In the present study, we compared genomic data from Native American populations living for a long-time at high altitude (Andean Altiplano) with those living at lowlands (Amazon and Mesoamerica), with the purpose of expanding our knowledge about the genetic repertoire responsible for the successful human colonization of the Andes.

Results

Natural selection analysis. Population Branch Statistic (PBS) values were estimated for each individual SNP. To avoid spurious results due to single SNPs, windows of 20 SNPs were used to estimate the mean PBS values for a given region. Then, we checked the outliers' peaks, above the 99.5th and 99.9th percentiles, to identify in each outlier window the SNPs with the highest PBS value and assigned the gene to which it belonged (or the nearest gene). Based on this approach, five candidate genes were identified: *SP100* (SP100 Nuclear Antigen), *TMEM38B* (Transmembrane Protein 38B), *AS3MT* (Arsenite 3 Methyltransferase), *DUOX2* (Dual Oxidase 2) and *CLC* (Charcot-Leyden Crystal Galectin, also known as Galectin-10) (Table 1 and Fig. 1). Among these candidate genes, *AS3MT* and *TMEM38B* have been identified in previous scans for natural selection in Andeans^{13,20}.

Neutral coalescent simulations indicated that these deviations were statistically significant (p-values ranging between 0.03 and 0.0001; Fig. 2, Table S1), consistent with the action of positive selection as opposed to genetic drift in increasing the frequency of the putative selected alleles at all five tested loci. In addition, we applied the Cross-Population Extended Haplotype Homozygosity (XP-EHH) test to the same regions. The XP-EHH results also show significant differences between the Andean and Mesoamerican groups in three SNPs (rs13411586, rs9678342, rs7582700) of *SP100* and one SNP (rs269866) of *DUOX2* (Table 1). These SNPs, which are under putative selection in the PBS analysis with the most extreme values (0.46 to 0.62), also present significant XP-EHH values ≥ 2 in both Andean vs Mesoamerican and Andean vs Amazonian groups.

The observed allele density provided by the iHS test showed a notable Gaussian distribution pattern for all three groups (Fig. S1), with homozygosity decaying according to the distance from the focal markers.

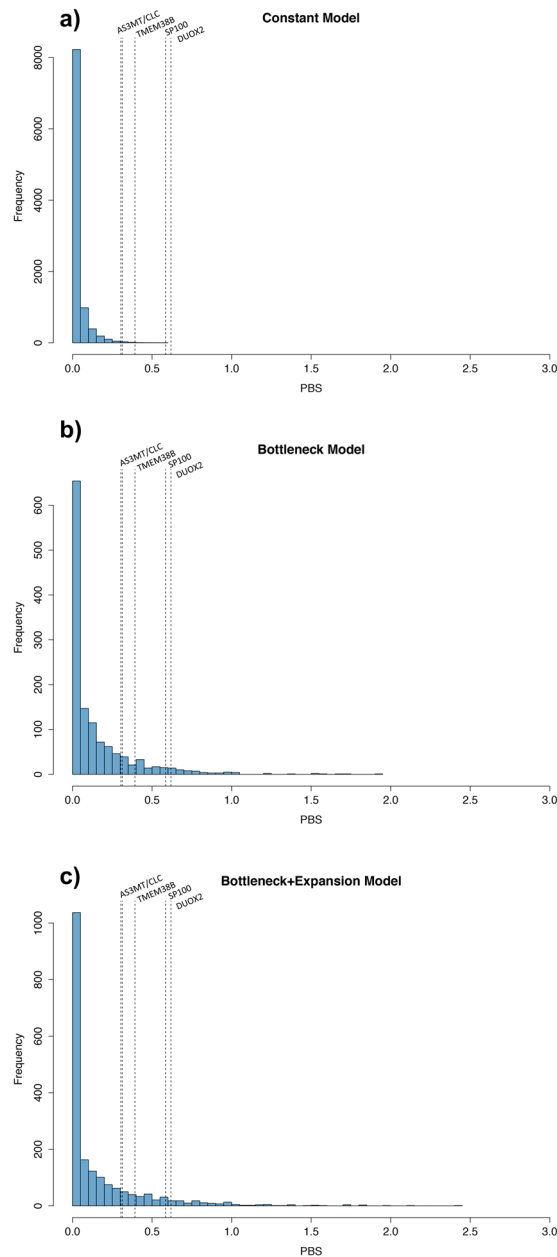


Figure 2. Distribution of 10,000 simulated PBS values under three neutral coalescent models. **(a)** Constant population model. **(b)** Population bottleneck model; and **(c)** Population bottleneck followed by expansion model. The dashed line represents the top observed PBS SNP values in the empirical datasets.

It should be noted that the distribution of alleles C (rs13411586, *SP100*), G (rs269866, *DUOX2*) and G (rs440191, *CLC*), which presented the highest PBS values (Table 1), showed their highest values in areas of very high Andean altitudes (Table 2 and Fig. 3).

Bootstrap simulations indicated that in all instances the 95% confidence interval of allele frequencies in lowlanders does not include the average values observed for populations living in high altitudes (>4000 m above sea level) (Fig. S2), suggesting that the differences found in allele frequencies between population groups might be caused by a non-random evolutionary process.

Effects of putatively selected alleles on gene expression. Homozygotes for the *DUOX2* putatively selected allele (rs269866 G) presented a slight increase in the expression of the *DUOX2* protein (Fig. S3). Multiple testing across tissues showed significant expression of this protein in thyroid (m-value = 1.0), lungs (m-value = 0.996) and aorta artery (0.996) (Fig. S4). Homozygotes for the rs13411586 (*SP100*) putatively selected allele (C) presented an increase in the expression of the *SP100* protein in skeletal muscles (Fig. S5). Multiple testing across tissues showed significant expression of this protein in skeletal muscle (m-value = 1.0) and testis (m-value = 0.971) (Fig. S6). There is no information available about the *CLC* gene expression profile.

Population (n)	<i>DUOX2</i> G allele (rs269866)	<i>SP100</i> C allele (rs13411586)	<i>CLC</i> G allele (rs440191)
Mesoamerican Lowland (<2,500 m.)			
Total (153)	0.068*	0.045*	0.128*
South American (Andean) Highland (≥4,000 m.)			
Total (63)	0.420*	0.397*	0.452*
South American (Amazonian) Lowland (<2,500 m.)			
Total (106)	0.048*	0.053*	0.142*

Table 2. Frequencies of the putatively selected alleles in the populational groups. *Weighted average.

Discussion

We identified five loci under positive selection in Andean Native populations. Two of them were previously described: *AS3MT* was found to be under positive selection in Colla Andeans systematically exposed to arsenic water²⁰ while *TMEM38B* reduced the negative effects of polycythemia (elevated hematocrit or decreased plasma volume) at high altitudes¹³. Three other genes, *SP100*, *DUOX2*, and *CLC* were identified for the first time in a high-altitude context in the present study. These genes are part of the *TP53* pathway, already indicated as a potential candidate to be under natural selection in high altitude populations^{19,21}.

SP100 is a single-copy gene in the human genome that produces several alternatively spliced Sp100 protein isoforms known as modulators of the p53 activity²². We found three SNPs in the *SP100* gene with high and significant PBS values, as well as significant XP-EHH values when Andeans were compared to others. One of these SNPs, rs13411586, is differentially expressed in humans; our prediction showed that individuals homozygous for the putatively selected allele (C) have increased Sp100 production.

Interestingly, we also identified that the *SP100* gene is differentially expressed in skeletal muscles (Fig. S3). Studies have revealed that a member of the HIF pathway, HIF-1, plays an important role in the regulation of oxygen homeostasis, which includes the physiological skeletal and heart muscle adaptations in situations of oxygen reduction due to muscular effort^{23–25} and ischemic cardiomyopathy, respectively²⁶. Exposure to high altitude leads to reduced muscle mass and performance (e.g. lower work capacity and standing fatigue), except when one is evolutionarily adapted to it^{27–29}.

HIF-1 protects cell-survival during low oxygen supply, while p53 promotes genome cell-death under hypoxia. The reason for these apparently antagonistic roles can be in the difference of the oxygen quantity available; in a normal condition, both p53 and HIF-1 levels are low, but in mild hypoxia, the p53 level remains low, whereas the HIF-1 level increases, protecting cells still relatively healthy from destruction. In severe hypoxia, p53 accumulation promotes the repression or degradation of anti-apoptotic proteins like HIF-1, inducing apoptosis of the damaged cells^{30–32}. Sp100 is known as a modulator of the p53 activity²² and under tissue hypoxia due to ischemia, it is downregulated, leading to genomic instability²⁶. The Andean population presents high allele C (rs13411586) frequency (Table 2), which in homozygosis increase Sp100 production according to our prediction test. Our result suggests an evolutionary solution to keep Sp100 at an adequate level in an environment with a constant low oxygen level. Furthermore, it is possible to speculate that there is an intricate balance in the level of expression of the *SP100*, *TP53* and *HIF-1* genes under hypoxia, considering both short (reversible physiological and metabolic adaptations) and long-term evolutionary adaptation scenarios.

DUOX2, expressed in epithelial cells of various tissues including nasal and lung, participates in the hydrogen peroxide (H_2O_2) pathway, which is required in the final steps of thyroid hormones production. It is also involved in Reactive Oxygen Species (ROS), a byproduct of the normal oxygen metabolism even under normal physiologic conditions³³. However, different stressor conditions can increase the ROS production, i.e. high-altitude exposure (hypoxia and UV exposure), and pathological conditions such as cancer³⁴. Salmeen *et al.*³⁵ provided evidence that *DUOX2* plays a role in a p53-dependent checkpoint mechanism for cell cycle entry.

In vitro and *in vivo* experiments showed that oxidative stress and generation of ROS caused by *DUOX2* over-expression, in both hypoxia and hyperoxia, contribute to inflammation, carcinogenesis and cell death^{36–42}. For instance, a functional study³⁶ showed that under hyperoxia conditions, mutant mice for *DUOX2* had significant lower acute lung injuries induced by hyperoxia. This finding pointed to the importance of these proteins in the response to changes of oxygen concentration in the environment. Another study³⁷ found that chickens submitted to hypoxia (>3,000 m) had increased activity of *DUOX/NOX* proteins, indicating the physiological role of these enzymes in the process of adaptation to oxidative stress.

Our results on the expression of the *DUOX2* putatively selected allele G (high PBS values and significant XP-EHH value > 2; Table 1) also pointed to higher levels of protein expression in humans, mainly in the lungs and arteries. It is noteworthy that ROS contributes to inflammation in the vessel walls. Kim & Byzowa⁴³ demonstrated that ROS has an important role in angiogenesis, a process of new blood vessel growth. Angiogenesis is a key event in the physiological response to hypoxia and therefore might have a role in the adaptation to high altitude in long-term residents, especially in individuals with excessive erythropoiesis (like those found in the Chronic Mountain Sickness [CMS] phenotype), to compensate a plausible change in microcirculation^{44,45}.

SNP rs440191 is located at the 3'UTR region of *CLC*, and the putatively selected allele G is in complete linkage disequilibrium with the *CLC* rs395892 G allele in the Mexican population⁴⁶. The latter is associated with eosinophil and basophil counts⁴⁷, while rs440191 has so far been investigated just in approaches assessing allergic susceptibilities⁴⁸. Gene expression queries did not show any significant eQTL related to this polymorphism, preventing any prediction of tissue-specific expression.

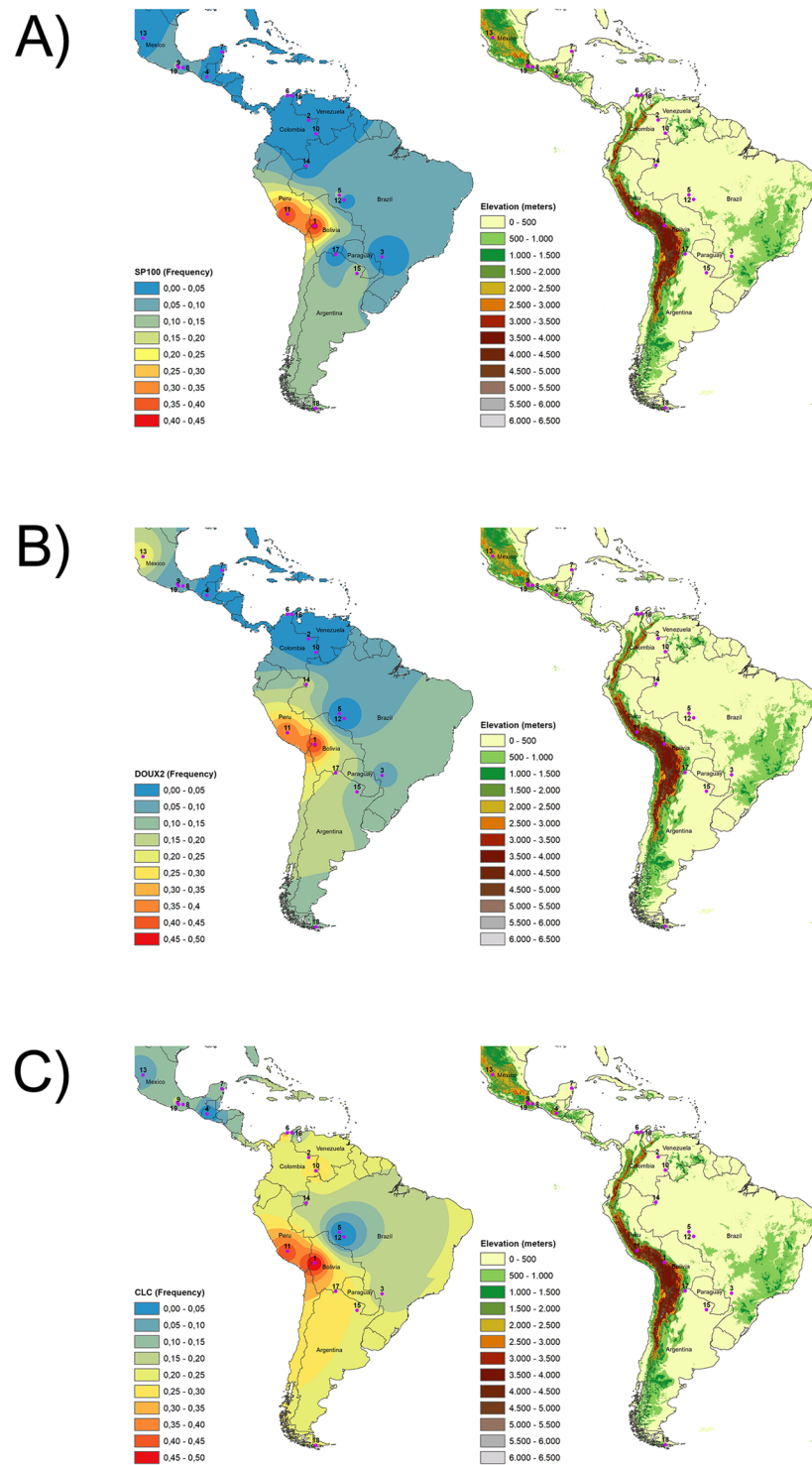


Figure 3. (a) rs13411586_C (*SP100*). (b) rs269866_G (*DUOX2*) and (c) rs440191_A (*CLC*) allele frequency distributions according to altitude. Populations ($n \geq 3$): 1. Aymara, 2. Guahibo, 3. Guarani, 4. Kaqchikel, 5. Karitiana, 6. Kogi, 7. Maya, 8. Mixe, 9. Mixtec, 10. Piapoco, 11. Quechua, 12. Surui, 13. Tepehuano, 14. Ticuna, 15. Toba, 16. Wayuu, 17. Wichi, 18. Yaghan and 19. Zapotec.

CLC (galectin-10) is still a poorly studied gene when compared to other members of the functionally polyvalent galectin family. It is recognized as a lysophospholipase expressed in eosinophils and basophils, although some authors identified it just as an enzyme that interacts with lysophospholipases⁴⁹. The only functional study regarding this protein showed that hypoxia increases eosinophil accumulation and *CLC* production in humans, concomitant with a delay in constitutive apoptosis, antagonizing the normal pro-apoptotic effect of agents that normally induce eosinophil apoptosis⁵⁰.

Regulation by the p53 transcription factor seems to be important in the galectin family genes' expression. For instance, the galectin-3 gene has a binding site for p53, and p53 increases the transcription of paralogous galectin-7^{51–53}. Altered expression of galectin genes, including *CLC*, was implicated in cancer emergence and progression, highlighting the role of the galectins in cell proliferation via cell death programs⁵⁴.

Investigations with galectin paralogs have shown that galectin-1 in the first term ovine gestation placenta prevented inflammatory processes that harm the fetus⁵⁵, while galectin-13, which has the highest homology to *CLC*, is a member of the group of the so-called “pregnancy-related proteins”, due to its special immune functions at the feto-maternal interface^{56,57}. These fundamental cell functions, already described for humans and other placental mammals, may indicate the path that connects our *CLC* findings and the selection pressure in the Andean hostile climate.

In conclusion, our results pointed to a complex adaptation that occurred in Andean natives, which involved the *CLC*, *SP100* and *DUOX2* genes, not previously correlated in contexts of long-time adaptation to high altitudes. We also reinforced the role of the *TP53* pathway at least for the adaptation to the Andean environmental stresses. Combined with other studies, and incorporating the present one, it is clear that Andeans have adapted to the Altiplano in different ways and using distinct molecular strategies than those of other natives living at high altitude.

Methods

Populations. We analyzed 213,987 SNPs determined with Illumina 610quad from 63 Native Americans living at extreme high altitude ($\geq 4,000$ m; 63% of the oxygen available at sea level; http://www.altitude.org/air_pressure.php) and 259 living at lowland areas ($< 2,500$ m), data previously published by Reich *et al.*⁵⁸. Highlanders included Aymara and Quechua Andeans, while lowlanders were represented by 25 populations from the Mesoamerican and South American lowlands. Details about these populations, sample sizes and allelic frequencies are given in Table S2. Additional information, including ethical authorizations for evolutionary and anthropological studies, can be found in the primary publication⁵⁸.

Population Branch Statistic (PBS) analysis. PBS determinations were performed between pair of populations, using Andean and Amazonian populations as sister groups and Mesoamericans as an outgroup. The analysis was carried out as described by Yi *et al.*⁵⁹, with only the polymorphic SNPs in at least two of the populations being considered. From the genetic distances (F_{ST}) between the three population groups examined, PBS measures if there are alleles with extreme frequencies in the Andean group as compared to the other two. Under a scenario of genetic drift only, we expect that Andeans and Amazonians will be more similar genetically than both compared to Mesoamericans. If, however, there has been local adaptation, we should detect genes that have been targeted by selection in Andeans. PBS values were estimated for both individual SNPs and windows of 20 SNPs overlapped in five SNPs. The empiric distribution of PBS values, with a 99.5th threshold, was used to determine signals of positive selection (more details in Amorim *et al.*⁶⁰).

Demographic simulations. To verify the significance of the observed positive selection signals we simulated different demographic models, according to reported historical population data and inferred effective population sizes. We adapted the models described by Valverde *et al.*¹¹, to account for the divergence between Mesoamericans, Andeans and Amazonians. Assuming that the American continent was peopled beginning at 15,000 yBP, the Andes colonized by 12,000 yBP and the Amazon by 10,000 yBP, and based on N_e estimated by Valverde *et al.*¹¹, we simulated the three demographic models proposed by them: (a) Constant Model: N_e of 7,000 individuals with constant size in all populations throughout history; (b) Bottleneck Model: N_e 8,000 in Mesoamerica, 4,000 in Andes and 2,000 in Amazon; and (c) Bottleneck + Expansion Model: model b with bottlenecks reducing the effective size of all populations by 50% in the last 10,000 years followed by a sharp expansion in the last 8,000 years. Simulations were performed in the MS program⁶¹ with 10,000 replicates for each demographic scenario.

Linkage disequilibrium analysis. We also used three linkage disequilibrium-based methods: extended haplotype homozygosity (EHH)⁶², integrated haplotype score (iHS)⁶³, and cross-population extended haplotype homozygosity (XP-EHH)⁶⁴. These approaches adopt the same core principle, that an advantageous allele under a hard sweep rise in frequency — carrying its neighbor alleles and therefore promoting homozygosity extension — quickly enough that recombination is not able to break down the haplotype. EHH statistics calculate the homozygosity rate from a core region (putative allele under selection) to the neutral scenery, *i.e.* the probability that any two randomly chosen chromosomes will be identical by descent, from the core region to a distance x . iHS evaluates the EHH considering both ancestral and derived alleles, and XP-EHH is used to calculate EHH/iHS between populations, therefore controlling for local variation. These tests are complementary; while iHS is better for detecting incomplete sweeps, XP-EHH has more power to detect sweeps near fixation⁶⁵. Both measurements and significance were calculated through the ‘rehh’ R package⁶⁶.

Geographical analysis. To evaluate the variants spatial distribution, weighted inverse distance interpolation (IDW) was used to determine cell values using a weighted linear combination of a set of sample points. Weight is a function of the inverse distance⁶⁷. The maps were made with the ArcGis 10.5 software and the cartographic base was georeferenced to the World Geodetic System (WGS84).

Bootstrap Simulations. To verify whether the allele frequencies of the candidate variants under selection are significantly different among extreme high ($> 4,000$ m) and lowland ($< 4,000$ m) populations, we obtained the 95% confidence intervals of the average allele frequency of the lowland populations by means of 10,000 computer-assisted bootstrap simulations with replacement, considering a sample as having the same size and

genotypic proportions observed in the real one. The average allele frequencies from high and lowland populations were obtained by weighing the observed frequencies according to their sample sizes.

Analysis of gene expression. We used the Genotype-Tissue Expression Portal (GTEx; <https://www.gtexportal.org/home/>) to evaluate possible associations between each of the candidate alleles with highest differentiation and gene expression across human tissues looking for evidence of quantitative trait loci (eQTLs). The m-value is the posterior probability that an eQTL effect exists in each tissue tested in the cross-tissue meta-analysis. The m-value ranges between 0 and 1 (m-values > 0.9 mean that the tissue is predicted to have an eQTL effect).

References

- Espinoza-Navarro, O., Díaz, J., Rodríguez, H. & Moreno, A. Effects of altitude on anthropometric and physiological patterns in Aymara and non-Aymara population between 18 and 65 years in the province of Parinacota Chile (3.700 masl). *Int. J. Morphol* **29**, 34–40 (2011).
- Rademaker, K. *et al.* Paleoindian settlement of the high-altitude Peruvian Andes. *Science* **346**, 466–469 (2014).
- Aldenderfer, M. Peopling the Tibetan plateau: Insights from archaeology. *High Alt. Med. Biol.* **12**, 141–147 (2011).
- Lu, D. *et al.* Ancestral origins and genetic history of Tibetan highlanders. *Am. J. Hum. Genet.* **99**, 580–594 (2016).
- Moore, L. G. Human genetic adaptation to high altitude. *High Alt. Med. Biol.* **2**, 257–279 (2001).
- Beall, C. M. Two routes to functional adaptation: Tibetan and Andean high-altitude natives. *Proc. Natl. Acad. Sci.* **104**, 8655–8660 (2007).
- Bigham, A. W. & Lee, F. S. Human high-altitude adaptation: Forward genetics meets the HIF pathway. *Genes Dev.* **28**, 2189–2204 (2014).
- Scheinfeldt, L. B. *et al.* Genetic adaptation to high altitude in the Ethiopian highlands. *Genome Biol.* **13**, R1, <https://doi.org/10.1186/gb-2012-13-1-r1> (2012).
- Huerta-Sánchez, E. *et al.* Genetic signatures reveal high-altitude adaptation in a set of ethiopian populations. *Mol. Biol. Evol.* **30**, 1877–1888 (2013).
- Simonson, T. S., Huff, C. D., Witherspoon, D. J., Prchal, J. T. & Jorde, L. B. Adaptive genetic changes related to haemoglobin concentration in native high-altitude Tibetans. *Exp. Physiol.* **100**, 1263–1268 (2015).
- Valverde, G. *et al.* A novel candidate region for genetic adaptation to high altitude in Andean populations. *PLoS One* **10**, e0125444 (2015).
- Fehren-Schmitz, L. & Georges, L. Ancient DNA reveals selection acting on genes associated with hypoxia response in pre-Columbian Peruvian highlanders in the last 8500 years. *Sci. Rep.* **6**, 23485 (2016).
- Crawford, J. E. *et al.* Natural selection on genes related to cardiovascular health in high-altitude adapted Andeans. *Am. J. Hum. Genet.* **101**, 752–767 (2017).
- Beall, C. M. *et al.* Natural selection on EPAS1 (HIF2) associated with low hemoglobin concentration in Tibetan highlanders. *Proc. Natl. Acad. Sci.* **107**, 11459–11464 (2010).
- Bigham, A. *et al.* Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. *PLoS Genet.* **6** (2010).
- Simonson, T. S. *et al.* Genetic evidence for high-altitude adaptation in Tibet. *Science* **329**, 72–75 (2010).
- Peng, Y. *et al.* Genetic variations in tibetan populations and high-altitude adaptation at the Himalayas. *Mol. Biol. Evol.* **28**, 1075–1081 (2011).
- Xu, S. *et al.* A genome-wide search for signals of high-altitude adaptation in tibetans. *Mol. Biol. Evol.* **28**, 1003–1011 (2011).
- Jacovas, V. C. *et al.* Genetic variations in the TP53 pathway in native americans strongly suggest adaptation to the high altitudes of the Andes. *PLoS One* **10**, e0137823 (2015).
- Eichstaedt, C. A. *et al.* Positive selection of AS3MT to arsenic water in Andean populations. *Mutat. Res. Mol. Mech. Mutagen.* **780**, 97–102 (2015).
- Eichstaedt, C. A. *et al.* The Andean adaptive toolkit to counteract high altitude maladaptation: genome-wide and phenotypic analysis of the Collas. *PLoS One* **9**, e93314 (2014).
- Berscheminski, J. *et al.* Sp100A is a tumor suppressor that activates p53-dependent transcription and counteracts E1A/E1B-55K-mediated transformation. *Oncogene* **35**, 3178–3189 (2016).
- Semenza, G. L. Regulation of mammalian O₂ homeostasis by hypoxia-inducible Factor 1. *Annu. Rev. Cell Dev. Biol.* **15**, 551–578 (1999).
- Vogt, M., Billeter, R. & Hoppeler, H. Effect of hypoxia on muscular performance capacity: Living low-training high. *Ther. Umsch.* **60**, 419–424 (2003).
- Lindholm, M. E. & Rundqvist, H. Skeletal muscle hypoxia-inducible factor-1 and exercise. *Exp. Physiol.* **101**, 28–32 (2016).
- Herrer, I. *et al.* Gene expression network analysis reveals new transcriptional regulators as novel factors in human ischemic cardiomyopathy. *BMC Med. Genomics* **8**, 14 (2015).
- Marconi, C., Marzorati, M. & Cerretelli, P. Work capacity of permanent residents of high altitude. *High Alt. Med. Biol.* **7**, 105–115 (2006).
- Brutsaert, T. D. Do high-altitude natives have enhanced exercise performance at altitude? *Appl. Physiol. Nutr. Metab.* **33**, 582–592 (2008).
- Counter, S. A., Buchanan, L. H., Ortega, F., Jacobs, A. B. & Laurell, G. Assessment of the brainstem-mediated stapedius muscle reflex in Andean children living at high altitudes. *High Alt. Med. Biol.* **18**, 37–45 (2017).
- Schmid, T., Zhou, J. & Brüne, B. HIF-1 and p53: communication of transcription factors under hypoxia. *J. Cell. Mol. Med.* **8**, 423–431 (2004).
- Obacz, J., Pastorekova, S., Vojtesek, B. & Hrstka, R. Cross-talk between HIF and p53 as mediators of molecular responses to physiological and genotoxic stresses. *Mol. Cancer* **12**, 93 (2013).
- Zhou, C.-H., Zhang, X.-P., Liu, F. & Wang, W. Modeling the interplay between the HIF-1 and p53 pathways in hypoxia. *Sci. Rep.* **5**, 13834 (2015).
- Devasagayam, T. P. A. *et al.* Free radicals and antioxidants in human health: current status and future prospects. *J. Assoc. Physicians India* **52**, 794–804 (2004).
- Gupta, S. C. *et al.* Upsides and downsides of reactive oxygen species for cancer: the roles of reactive oxygen species in tumorigenesis, prevention, and therapy. *Antioxid. Redox Signal.* **16**, 1295–1322 (2012).
- Salmeen, A., Park, B. O. & Meyer, T. The NADPH oxidases NOX4 and DUOX2 regulate cell cycle entry via a p53-dependent pathway. *Oncogene* **29**, 4473–4484 (2010).
- Kim, M.-J. *et al.* Dual oxidase 2 in lung epithelia is essential for hyperoxia-induced acute lung injury in mice. *Antioxid. Redox Signal.* **21**, 1803–18 (2014).
- Bautista-Ortega, J., Cortes-Cuevas, A., Ellis, E. A. & Ruiz-Feria, C. A. Supplemental L-arginine and vitamins e and C preserve xanthine oxidase activity in the lung of broiler chickens grown under hypobaric hypoxia. *Poult. Sci.* **93**, 979–988 (2014).

38. Fletcher, N. M. *et al.* Nicotinamide adenine dinucleotide phosphate oxidase is differentially regulated in normal myometrium versus leiomyoma. *Reprod. Sci.* **21**, 1145–1152 (2014).
39. Dias-Freitas, F., Metelo-Coimbra, C. & Roncon-Albuquerque, R. Molecular mechanisms underlying hyperoxia acute lung injury. *Respir. Med.* **119**, 23–28 (2016).
40. Min, H. J. *et al.* ROS-dependent HMGB1 secretion upregulates IL-8 in upper airway epithelial cells under hypoxic condition. *Mucosal Immunol.* **10**, 685–694 (2017).
41. Lin, S.-C. *et al.* High immunoreactivity of DUOX2 is associated with poor response to preoperative chemoradiation therapy and worse prognosis in rectal cancers. *J. Cancer* **8**, 2756–2764 (2017).
42. MacFie, T. S. *et al.* DUOX2 and DUOXA2 form the predominant enzyme system capable of producing the reactive oxygen species H₂O₂ in active ulcerative colitis and are modulated by 5-aminosalicylic acid. *Inflamm. Bowel Dis.* **20**, 514–524 (2014).
43. Kim, Y. & Byzova, T. V. Oxidative stress in angiogenesis and vascular disease. *Blood* **123**, 625–632 (2015).
44. Ge, R.-L. *et al.* B-type natriuretic peptide, vascular endothelial growth factor, endothelin-1, and nitric oxide synthase in chronic mountain sickness. *Am. J. Physiol. Heart Circ. Physiol.* **300**, H1427–33 (2011).
45. Buroker, N. E. *et al.* AKT3, ANGPTL4, eNOS3, and VEGFA associations with high altitude sickness in Han and Tibetan Chinese at the Qinghai-Tibetan Plateau. *Int. J. Hematol.* **96**, 200–213 (2012).
46. The 1000 Genomes Project. <http://www.nature.com/articles/nature09534>.
47. Astle, W. J. *et al.* The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* **167**, 1415–1429 (2016).
48. Bryborn, M., Halldén, C., Säll, T. & Cardell, L. O. CLC - a novel susceptibility gene for allergic rhinitis? *Allergy* **65**, 220–228 (2010).
49. Ackerman, S. J. *et al.* Charcot-leyden crystal protein (galectin-10) is not a dual function galectin with lysophospholipase activity but binds a lysophospholipase inhibitor in a novel structural fashion. *J. Biol. Chem.* **277**, 14859–14868 (2002).
50. Porter, L. M. *et al.* Hypoxia causes IL-8 secretion, Charcot Leyden crystal formation, and suppression of corticosteroid-induced apoptosis in human eosinophils. *Clin. Exp. Allergy* **47**, 770–784 (2017).
51. Popyak, K., Xia, Y., Zweier, J. L., Kinzler, K. W. & Vogelstein, B. A model for p53-induced apoptosis. *Nature* **389**, 300–305 (1997).
52. Raimond, J., Rouleux, F., Monsigny, M. & Legrand, A. The second intron of the human galectin-3 gene has a strong promoter activity down-regulated by p53. *FEBS Lett.* **363**, 165–169 (1995).
53. Cooper, D. N. W. Galectinomics: finding themes in complexity. *Biochim. Biophys. Acta - Gen. Subj.* **1572**, 209–231 (2002).
54. Gopalan, V. *et al.* The expression profiles of the galectin gene family in colorectal adenocarcinomas. *Hum. Pathol.* **53**, 105–113 (2016).
55. Iglesias, M. M., Rabinovich, G. A., Ivanovic, V., Sotomayor, C. & Wolfenstein-Todel, C. Galectin-1 from ovine placenta. *Amino-acid sequence, physicochemical properties and implications in T-cell death.* *Eur. J. Biochem.* **252**, 400–407 (1998).
56. Than, N. G. *et al.* Functional analyses of placental protein 13/galectin-13. *Eur. J. Biochem.* **271**, 1065–1078 (2004).
57. Su, J. *et al.* Galectin-13, a different prototype galectin, does not bind β -galactosides and forms dimers via intermolecular disulfide bridges between Cys-136 and Cys-138. *Sci. Rep.* **8**, 980 (2018).
58. Reich, D. *et al.* Reconstructing native American population history. *Nature* **488**, 370–374 (2012).
59. Yi, X. *et al.* Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* **329**, 75–78 (2010).
60. Amorim, C. E. *et al.* Genetic signature of natural selection in first Americans. *Proc. Natl. Acad. Sci.* **114**, 2195–2199 (2017).
61. Hudson, R. R. Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* **18**, 337–338 (2002).
62. Sabeti, P. C. *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832–837 (2002).
63. Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS Biol.* **4**, e72 (2006).
64. Sabeti, P. C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).
65. Vitti, J. J., Grossman, S. R. & Sabeti, P. C. Detecting natural selection in genomic data. *Annu. Rev. Genet.* **47**, 97–120 (2013).
66. Gautier, M. & Vitalis, R. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* **28**, 1176–1177 (2012).
67. Watson, D. F. & Philip, G. M. A refinement of inverse distance weighted interpolation. *Geoprocessing* **2**, 315–327 (1985).

Acknowledgements

We thank Eduardo Gameiro for the technical support. This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP (Grant 15/26875-9; TH), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES (Grant 1645581; KN) and Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq.

Author Contributions

M.C.B. and T.H. conceived the study. T.H. designed the analyses. C.M.C.-S. and K.N. performed the demographic and selection analyses. V.C.J. compiled the environmental variables and populational data. R.B.L. performed the bootstrap simulations. M.Z.O. carried out the geographical analysis. F.M.S., M.C.B. and T.H. wrote the manuscript with inputs from the other authors.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-31100-6>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018