The genetic prehistory of the Andean highlands 7,000 Years BP though European contact

Authors: John Lindo[1,2*], Randall Haas[3], Courtney Hofman[4], Mario Apata[5], Mauricio Moraga[5], Ricardo Verdugo[5], James T. Watson[6], Carlos Viviano Llave[7], David Witonsky[2], Enrique Vargas Pacheco[9], Mercedes Villena[9], Rudy Soria[9], Cynthia Beall[10], Christina Warinner[4,8], John Novembre[2], Mark Aldenderfer[11*], Anna Di Rienzo[2*]

Affiliations:

[1]Department of Anthropology, Emory University, Atlanta, GA 30322 USA

[2]Department of Human Genetics, University of Chicago, Chicago, IL 60637 USA

[3]Department of Anthropology, University of California, Davis, CA 95616 USA

[4]Department of Anthropology, University of Oklahoma, Norman, OK 73019 USA

[5]Programa de Genética Humana, Universidad de Chile, Santiago, Chile

[6]Arizona State Museum and School of Anthropology, University of Arizona, Tucson, AZ 85721 USA

[7]Peruvian Register of Professional Archaeologists, Peru

[8]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany 07745

[9]Respiratory Department, Instituto Boliviano de Biología de Altura, La Paz, Bolivia

[10]Department of Anthropology, Case Western University, Cleveland, Ohio 44106 USA

[11]School of Social Sciences, Humanities, and Arts, University of California Merced, Merced, California, USA 95343

*Correspondence to: jlindo@emory.edu, maldenderfer@ucmerced.edu, dirienzo@bsd.uchicago.edu

Abstract: The peopling of the Andean highlands above 2500m in elevation was a complex process that included cultural, biological and genetic adaptations. Here we present a time series of ancient whole genomes from the Andes of Peru, dating back to 7,000 calendar years before present (BP), and compare them to 64 new genome-wide genetic variation datasets from both high and lowland populations. We infer three significant features: a split between low and high elevation populations that occurred between 9200-8200 BP; a population collapse after European contact that is significantly more severe in South American lowlanders than in highland populations; and evidence for positive selection at genetic loci related to starch digestion and plausibly pathogen resistance after European contact. Importantly, we do not find selective sweep signals related to known components of the human hypoxia response, which may suggest more complex modes of genetic adaptation to high altitude.

One Sentence Summary: Ancient DNA from the Andes reveals a complex picture of human adaptation from early settlement to the colonial period.

Main Text: The Andean highlands of South America have long been considered a natural laboratory for the study of genetic adaptation of humans (*1*), yet the genetics of Andean highland populations remain poorly understood. People likely entered the highlands shortly after their arrival on the continent (*2, 3*), and while some have argued that humans lived permanently in the Central Andean highlands by 12,000 BP (*4*), other research indicates that permanent occupation began between 9500-9000 BP (*5-7*). Regardless of when the peopling of high elevation environments began, selective pressure on the human genome was likely strong due to challenging environmental factors but also social processes such as the intensification of subsistence resources and residential sedentism (*8*), which in turn promoted the development of agricultural economies, social inequality, and relatively high population densities across much of the highlands. European contact initiated an array of economic, social, and pathogenic changes (*9, 10*). Although it is known that the peoples of the Andean highlands experienced population contraction after contact (*11*), its extent is debated from both archaeological and ethnohistorical perspectives, especially concerning the size of the indigenous populations at initial contact (*12, 13*).

To address hypotheses regarding the population history of Andean highlanders and their genetic adaptations, we collected a time series of ancient whole genomes from individuals in the Lake Titicaca region of Peru (Fig. S1, Table 1). The series represents three different cultural periods, which include individuals from (A) Rio Uncallane, a series of cave crevice tombs dating to ~1,800 BP and used by fully sedentary agriculturists, (B) Kaillachuro, a ~3,800 year old site marked by the transition from mobile foraging to agropastoralism and residential sedentism, and (C) Soro Mik'aya Patjxa (SMP), an 8000-6500 year old site inhabited by residentially mobile hunter-gatherers (*14*). We then compare the genomes of these ancient individuals to 25 new genomes and 39 new genome-wide SNP datasets generated from two modern indigenous populations: the Aymara of highland Bolivia and the Huilliche-Pehuenche of coastal lowland Chile. The Aymara are an agropastoral people who have occupied the Titicaca basin for at least 2,000 years (*15*). The Huilliche-Pehuenche are traditionally hunter-gatherers from the southern coastal forest of Chile (*16*).

To explore the population history of the Andean highlands, we first assess the genetic affinities of the prehistoric individuals and compare them to modern South Americans as well as other ancient Native Americans. Second, we construct a demographic model that estimates the timing of the lowland-highland population split as well as the population collapse following European contact, and finally, we explore evidence for genetic changes associated with selective pressures associated with the permanent occupation of the highlands, the intensification of tuber usage, and the impact of European-borne diseases.

Results

Samples and sequencing. We performed high-throughput shotgun sequencing of extracted DNA samples from seven individuals from three archaeological periods with proportions of endogenous DNA ranging from 2.60% to 44.8%. The samples span in age from 6,800 to 1,400 cal BP (Table S1). The samples from the Rio Uncallane site showed the highest endogenous content and five were chosen for frequency-based analyses and sequenced to an average read depth of 4.74x (Table 1). The oldest sample is from the SMP site, SMP5. The sample was directly dated to ~6,800 years BP and exhibited the highest endogenous DNA proportion among burials from the same site (3.15%) and was sequenced to an average read depth of 3.92x. The

sample from the Kaillachuro site, at about 3,800 years BP (culturally dated), was sequenced to an average read depth of 1.44x. All samples exhibited low contamination rates (1-4%) and DNA damage patterns consistent with degraded DNA (Table S2, Fig. S2).

We also generated new data from two contemporary populations: 24 Aymara individuals from Ventilla, Bolivia, which is near Lake Titicaca (Fig. S1), were sequenced to an average read depth of 4.36x (Table S3), and an additional individual was sequenced to a read depth of 31.20x (*see SI Appendix*). The modern Huilliche-Pehuenche (n=39) from the south of Chile (Fig. S1) were genotyped on an Axiom LAT1 Array and imputed utilizing the 1,000 genome phase 3 (*17*) and the Aymara whole genome sequence data as a reference panel. Analysis using the program *ADMIXTURE* (*18*) revealed that all the modern samples had trivial amounts of non-indigenous ancestry (less than 5%).

Genetic relationship between ancient and modern individuals. We employed outgroup f_3 statistics to assess the shared genetic ancestry among the ancient individuals and 156 worldwide populations (*19*). C/T and G/A polymorphic sites were removed from the data set to guard against the most common forms of post-mortem DNA damage (*20*). Outgroup f_3 statistics of a worldwide dataset demonstrate that all seven ancient individuals from all three time periods display greater affinity with Native American groups than with other worldwide populations (Fig. 1). Ranked outgroup f_3 statistics suggest that the ancient individuals from all three time periods tend to share greatest affinity with Andean living groups (Fig. 1A).

Principal components analysis reveals a tight clustering of the Rio Uncallane and Kaillachuro (K1) individuals, which overlays with modern Andean populations (Fig. 2). The oldest individual, SMP5, places closer to the intersect between the modern Andean groups of the Quechua and the Aymara. To further elucidate the relationship amongst the ancient Andean individuals and modern populations from South America, we examined maximum likelihood trees inferred with *TreeMix* (*21*). We observe that individuals from all three time periods form a sister group to the modern Andean high-altitude population, the Aymara (Bolivia) (Fig. 3A-C). The connection between the modern and ancient Andeans continues with the ADMIXTURE-based cluster analysis (Fig. 3D). Both the Rio Uncallane and K1 trace their genetic ancestry to a single component (shown in red), which is also shared by the modern Andean populations of the Quechua and the Aymara (Fig. 3D). SMP5 traces most of his ancestry to the same component, but also exhibits a component (shown in brown) found in Siberian populations, specifically the Yakut. USR1, a 12,000 year old individual from Alaska hypothesized to be part of the ancestral population related to all South American populations (*22*), also shares this component, as do previously published ancient individuals from North America (i.e., Anzick-1, Saqqaq and Kennewick).

Demographic model. Given that the above analyses provide an inference of genetic affinity in the Andes extending over roughly 4,000 years BP, and perhaps up to 7,000 years, we explored models relevant to the timing of the first permanent settlements in the Andes, as well as the extent of the population collapse occurring after European contact in the 1500s (Fig. 4). We utilized a composite likelihood method (Fastsimcoal2 (*23*)), which was informed by the site frequency spectra (SFS) of multiple populations, including the Rio Uncallane (n=5) (high altitude, pre-European contact), the Aymara (n=25) (high altitude, post-European contact), and the Huilliche-Pehuenche (n=39) (lowland population, post-European contact). We infer a split between low and high-altitude populations in the Andes occurring at 8,750 years (95% CI[8200, 9250]). We also infer the magnitude of the population collapse in the Andes after European contact to be a 27% reduction in effective population size and occurring 425 years ago (95% CI[400, 450]). The model also

inferred the collapse experienced by the Mixe (included in the model for the split between Northern and Southern Hemisphere populations) and the Huilliche-Pehuenches after European contact, which were much more severe. We infer a 94% (95% CI[0.94-0.96]) collapse for the Mixe and a comparable 96% (95% CI[0.95-0.96]) collapse for the Huilliche-Pehuenche. The collapse difference between the high and low altitude populations from the Andes region is significant (Huilliche-Pehuenche vs Aymara: $P < 0.0001$ (95% CI[0.66-0.71]), chi-square test).

However, we were concerned with the heterogenous nature of the data used in the model, since the Huilliche-Pehuenche data was imputed from a SNP array. Therefore, we calculated heterozygosity on each population using an overlapping set of ascertained SNPs from the San. We found a similar pattern of effective population size changes, using heterozygosity as a proxy, as we did with the demographic model (Supplementary text, Fig. S3).

Adaptive allele frequency divergence. Next, we turn to detecting the impact of natural selection. Permanent settlement of high altitude locales in the Andes likely required adaptations to the stress of hypobaric hypoxia, such as those detected in Tibetan, Nepali, and Ethiopian highland populations (*24-27*). Several genetic studies have also been performed on modern populations of the Andes (*28-31*). The availability of genetic variation data from an ancient Andean population, i.e., the Rio Uncallane samples, offers new opportunities for investigating adaptations in this region of the Americas. More specifically, the ancient data allows us to test for adaptations in a population that was not exposed to the confounding effects associated with the European contact, i.e., the demographic bottleneck and the exposure to additional strong selective pressures due to the introduction of new pathogens (*32, 33*). Therefore, we test for high allele frequency divergence resulting from adaptations to the Andean high altitude environment, utilizing a whole genome approach with the five Rio Uncallane individuals (~1800 BP), and we contrast them with a lowland population, the Huilliche-Pehuenche from Chile, who likely inhabited the region for several thousand years before the arrival of the Spanish (*34*). We utilized the Population Branch Statistic (PBS) (*27*) to scan for alleles that reached high frequency in the Rio Uncallane relative to the Huilliche-Pehuenche. The Han Chinese from the 1,000 Genomes Project (*17*) were utilized as the third comparative population. The single nucleotide polymorphisms (SNPs) showing the most extreme PBS score and, therefore, the highest differentiation in the ancient Andeans, corresponds to a region near *MGAM* (Fig. 5A, Table S4). *MGAM* is an intestinal enzyme associated with starch digestion (*35*) and could possibly be linked to the intensification of tuber use and ultimately the transition to agriculture (*8, 36*). The second strongest signal came from the gene *DST,* which encodes a cytoskeleton linker protein active in neural and muscle cells (*37*). *DST* has been associated with differential expression under hypoxic conditions (*38, 39*) and cardiovascular health (*40, 41*).

The second test for selection targeted hypotheses concerning the adaptation to European-introduced pathogens (Fig. 5B). The PBS was conducted with the Aymara (post-European contact), the Rio Uncallane (pre-European contact), and the Han Chinese (*17*). The strongest selection signal comes from SNPs in the vicinity of *CD83*, which encodes an immunoglobulin receptor and is involved in a variety of immune pathways, including T-cell receptor signaling (*42-44*). The gene has also been shown to be upregulated in response to the vaccinia virus infection (*45*), the virus used for smallpox vaccination. In addition, the differentiated SNPs associated with the *CD83* signal show a variety of chromatin alterations in immune cells, including monocytes, T-cells, and natural killer (NK) cells (*46*), suggesting that they may be functional. Lastly, a gene in

the top 0.01% of the scan, *IL-36R,* codes for an interleukin receptor that has been associated with the skin inflammatory pathway and vaccinia infection (*47*).

Discussion

South America is thought to have been populated relatively soon after the first human entry into the Americas, some 15,000 years ago (*3*). However, steep environmental gradients in western South America would have posed substantial challenges to population expansion. Among the harshest of these environments is the Andean highlands, which boasts frigid temperatures, low partial pressure of oxygen, and intense UV radiation. Despite this, however, humans eventually spread throughout the Andes and occupied them permanently. Archaeological evidence suggests that hunter-gathers entered the highlands as early as 12,000 years BP (*4*), with permanent occupation beginning around 9,000 years BP (*5-7*). The evidence presented here indicates genetic affinity between populations from different time periods in the high elevation Lake Titicaca region from at least 3,800 years BP, and possibly 7,000 years BP. This affinity extends to the present high-altitude Andean communities of the Aymara and Quechua.

Although our samples do not extend beyond 7,000 years BP, we were able to model the initial entry into the Andes after the split between North and South American groups. Our model, utilizing a 1.25×10^{-8} mutation rate, shows a correlation with archaeological evidence regarding the split between North and South groups occurring nearly 14,750 years ago (95% CI[14,225, 15,775]), which agrees with the oldest known site in South America of Monte Verde in southern Chile (~14,000 years BP) (*3*). The date for the split between low and high-altitude populations was inferred to 8,750 years (95% CI[8,200, 9,250]), which is younger than previously reported by a study utilizing modern genomes alone (*48*). This date provides a terminus ante quem time frame for the origins of adaptations known in modern highland populations.

We also present evidence for genes that may have been under selective pressure caused by environmental stressors in the Andes. Interestingly, none of our most extreme signals for positive selection were related to the hypoxia pathway. Instead, we find differentiated SNPs in the *DST* gene*,* which has been linked to the proper formation of cardiac muscle in mice (*40, 41*). Furthermore, the *DST* intronic SNP that was most differentiated (rs149112613) shows histone modifications associated with blood and the right ventricle of the heart (*46*). This correlates to Andean highlanders tending to have enlarged right ventricles associated with moderate pulmonary hypertension (*1*). This finding also parallels hypotheses proposed by Crawford et al. (*49*) that Andeans may have adapted to high altitude hypoxia via cardiovascular modifications.

The most extreme signal may represent adaptations to an agricultural subsistence and diet. The top ranked gene, *MGAM*, is associated with starch digestion (*50*). The associated high frequency SNPs in the ancient Andean population (Table S4), exhibit chromatin marks in cells from the gastro-intestinal tract (Fig 5A). The variant may be highly differentiated between the ancient Andeans and the lowlanders (the Huilliche-Pehuenche) due to differences in subsistence strategies. The Huilliche-Pehuenche are traditionally hunter-gatherers, with archaeological evidence suggesting that their ancestors have been practicing this mode of subsistence for thousands of years in the region before European contact in the 1500s (*51*). In contrast, the Andes is one of the oldest New World centers for agriculture, which included starch-rich plants such as maize (~4000 years BP) (*52*) and the potato (~3400 Years BP) (*8*). Selection acting on the *MGAM* gene in the ancient Andeans may represent an adaptive response to greater reliance upon starchy domesticates. Recent archaeological findings based on dental wear patterns and microbotanical remains similarly

suggest that intensive tuber processing and thus selective pressures for enhanced starch digestion began at least 7000 years ago (*8, 36*). Furthermore, we see a similar signal (top 0.01%) when we contrast the hunter-gatherers from Brazil (Karitiana/Surui, sequence data (*53*)) with the ancient Andeans, as well as with the Aymara vs. the Huilliche-Pehuenche and the Karitiana/Surui. One further note, we did not detect *amylase* high copy number in the ancient Andes population before European contact, suggesting a different evolutionary path for starch digestion in the Andes when contrasted with Europeans (*54*).

Selection with respect to the environment in the Andes is not limited to the ancient past. In 1532, the environment radically changed with the arrival of the Spanish (*55*). Not only were long-standing states and social organizations disrupted, but the environment itself was altered with the arrival of European-introduced pathogens, which may have preceded the arrival of the Spanish via trade routes (*55*). Some of the most devastating epidemics were related to smallpox, occurring in the 1500s and 1600s (*56*). These combined factors are thought to have decimated the local populations (*56*). We infer the population decline in the Andes, utilizing the ancient and modern Andeans, and found the decline in effective population size to be 27% (95% CI[0.23, 0.34]). This is a modest decline compared to archaeological and historical estimates, which reached upwards of 90% of the total population (*13*). In contrast, the model infers a much more severe collapse for lowland Andean populations of the Huilliche-Pehuenche, exceeding 90%. We also explored alternative scenarios to make sure the model was not biasing the inferred collapse and found that the estimated population size reduction remained significantly less severe in the highland compared to the lowland populations (see Supplement). Although we did not have pre-Contact ancient samples for these populations in Chile to inform the model, the large difference suggests that high-altitude populations may have suffered a less intense decline compared to the more easily accessible populations near sea level in the Andes region. This is also supported by long-lasting warfare in the Chilean lowland region with the Spanish that lasted well into the 19th Century (*55*).

Our data also show that the populations in the modern Andes have high genetic affinity with the ancient populations preceding European contact. Although a strict continuity test (*57*) that does not allow for recent gene flow was not significant (see Supplement), modern Andeans are likely the descendants of the people that suffered the epidemics described in historical texts. In the Andes, missionary reports suggest that disease may have arrived before formal Spanish contact in 1532 (*58*) and that the first epidemics were likely caused by smallpox (*55*). We infer that selection acted within the past 500 years on the immune response making it likely that modern Andeans descend from the survivors of these epidemics. The selection scan along the branch of the modern Andeans, contrasted with the ancient group, revealed the strongest signal to be associated with an immune gene connected to smallpox, *CD83* (*45*). The second most highly differentiated SNPs were in the vicinity of *RPS29*, which codes for a ribosomal protein, and is involved in viral mRNA translation and metabolism, including that of influenza (*59*). Another top gene, *IL-36R* (rank #36), is thought to have evolved alternative cytokine signaling to compensate for viruses, such as pox viruses, that can evade the immune system (*60*). Furthermore, the top SNP associated with *IL-36R* (rs1117797) exhibits a QTL signal associated with *IL18RAP*, a gene involved in mediating the immune response to the vaccinia virus (*61*). The relative strength of these signals and the role played by the associated genes, may indicate that selection favored alleles that directly affected the pathogenicity of the diseases encountered by the ancestors of the epidemic survivors.

In conclusion, human adaptation to the Andean highlands involved a variety of factors and was complicated by the arrival of Europeans and the drastic changes that followed. Despite harsh

environmental factors, the Andes were populated relatively early after entry into the continent. The adaptive traits necessary for permanent occupation may have been selected for in a relatively short amount of time, on the order of a few thousand years. Given the multifaceted nature of the adaptation, we are not surprised to find genetic affinity in the populations of the Andes dating to at least 4,000 years BP, and possibly extending to 7,000 years BP.

References and Notes:

1. C. M. Beall, Adaptation to High Altitude: Phenotypes and Genotypes. *Annual Review Anthropology*. **43**, 251–272 (2014).

2. M. Aldenderfer, Modelling plateau peoples: the early human use of the world's high plateaux. *World Archaeology*. **38**, 357–370 (2007).

3. T. D. Dillehay *et al.*, Monte Verde: seaweed, food, medicine, and the peopling of South America. *Science*. **320**, 784–786 (2008).

4. K. Rademaker *et al.*, Paleoindian settlement of the high-altitude Peruvian Andes. *Science*. **346**, 466–469 (2014).

5. R. Haas *et al.*, Humans permanently occupied the Andean highlands by at least 7 ka. *R Soc Open Sci*. **4**, 170331 (2017).

6. D. Chala-Aldana *et al.*, Investigating mobility and highland occupation strategies during the Early Holocene at the Cuncaicha rock shelter through strontium and oxygen isotopes. *Journal of Archaeological Science: Reports* (2017), doi:10.1016/j.jasrep.2017.10.023.

7. M. S. Aldenderfer, *Montane foragers: Asana and the south-central Andean archaic* (University of Iowa Press, Iowa City, 1998).

8. C. U. Rumold, M. S. Aldenderfer, Late Archaic-Early Formative period microbotanical evidence for potato at Jiskairumoko in the Titicaca Basin of southern Peru. *Proc. Natl. Acad. Sci. U.S.A*. **113**, 13672–13677 (2016).

9. A. Ramenofsky, Native American disease history: past, present and future directions. *World Archaeology*. **35**, 241–257 (2010).

10. E. Zubrow, The depopulation of native America. *Antiquity*. **64**, 754–765 (1990).

11. S. A. Wernke, T. M. Whitmore, Agriculture and Inequality in the Colonial Andes: A Simulation of Production and Consumption Using Administrative Documents. *Hum Ecol*. **37**, 421–440 (2009).

12. N. D. Cook, *Demographic Collapse* (Cambridge University Press, 2004).

13. H. Dobyns, An Appraisal of Techniques with a New Hemispheric Estimate. *Current Anthropology*. **7**, 395–416 (1966).

14. W. R. Haas, C. V. Llave, Hunter-gatherers on the eve of agriculture: investigations at Soro Mik'aya Patjxa, Lake Titicaca Basin, Peru, 8000–6700 BP. *Antiquity*. **89**, 1297–1312 (2015).

15. D. L. Browman, Titicaca Basin archaeolinguistics: Uru, Pukina and Aymara AD 750–1450. *World Archaeology*. **26**, 235–251 (2010).

16. N. Lira, The Maritime Cultural Landscape of Northern Patagonia. *J Mari Arch*. **12**, 199–221 (2017).

17. 1000 Genomes Project Consortium *et al.*, A global reference for human genetic variation. *Nature*. **526**, 68–74 (2015).

18. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. **19**, 1655–1664 (2009).

19. M. Raghavan *et al.*, Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science*. **349**, aab3884 (2015).

20. A. W. Briggs *et al.*, Patterns of damage in genomic DNA sequences from a Neandertal. *PNAS*. **104**, 14616–14621 (2007).

21. J. K. Pickrell, J. K. Pritchard, Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genet*. **8**, e1002967 (2012).

22. J. V. Moreno-Mayar *et al.*, Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature*. **488**, 370–9 (2018).

23. L. Excoffier, I. Dupanloup, E. Huerta-Sánchez, V. C. Sousa, M. Foll, Robust Demographic Inference from Genomic and SNP Data. *PLoS Genet*. **9**, e1003905–17 (2013).

24. C. M. Beall *et al.*, Natural selection on EPAS1 (HIF2alpha) associated with low hemoglobin concentration in Tibetan highlanders. *Proc. Natl. Acad. Sci. U.S.A*. **107**, 11459–11464 (2010).

25. E. Huerta-Sánchez *et al.*, Genetic signatures reveal high-altitude adaptation in a set of ethiopian populations. *Mol Biol Evol*. **30**, 1877–1888 (2013).

26. C. Jeong *et al.*, Admixture facilitates genetic adaptations to high altitude in Tibet. *Nat Commun*. **5**, 3281 (2014).

27. X. Yi *et al.*, Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*. **329**, 75–78 (2010).

28. A. Bigham *et al.*, Identifying Signatures of Natural Selection in Tibetan and Andean Populations Using Dense Genome Scan Data. *PLoS Genet*. **6**, e1001116 (2010).

29. L. Fehren-Schmitz, L. Georges, Ancient DNA reveals selection acting on genes associated with hypoxia response in pre-Columbian Peruvian Highlanders in the last 8500 years. *Scientific Reports*. **6**, srep23485 (2016).

30. C. A. Eichstaedt *et al.*, Evidence of Early-Stage Selection on EPAS1 and GPR126 Genes in Andean High Altitude Populations. *Scientific Reports*. **7**, 13042 (2017).

31. G. Valverde *et al.*, A Novel Candidate Region for Genetic Adaptation to High Altitude in Andean Populations. *PLoS ONE*. **10**, e0125444 (2015).

32. J. Lindo *et al.*, A time transect of exomes from a Native American population before and after European contact. *Nat Commun*. **7**, 13175 (2016).

33. J. Lindo *et al.*, Patterns of Genetic Coding Variation in a Native American Population before and after European Contact. *Am. J. Hum. Genet.* **102**, 806–815 (2018).

34. G. A. Brito, Antiguas Culturas Del Norte Chico. *Museo Chileno de Arte Precolombino*, 15–29 (2001).

35. C. Pontremoli *et al.*, Natural Selection at the Brush-Border: Adaptations to Carbohydrate Diets in Humans and Other Mammals. *Genome Biol Evol*. **7**, 2569–2584 (2015).

36. J. T. Watson, R. Haas, Dental evidence for wild tuber processing among Titicaca Basin foragers 7000 ybp. *American Journal of Physical Anthropology*. **164**, 117–130 (2017).

37. M. Safran *et al.*, GeneCards Version 3: the human gene integrator. *Database (Oxford)*. **2010** (2010), doi:10.1093/database/baq020.

38. G. P. Elvidge *et al.*, Concordant Regulation of Gene Expression by Hypoxia and 2-Oxoglutarate-dependent Dioxygenase Inhibition. *J. Biol. Chem.* **281**, 15215–15226 (2006).

39. M. T. Ghorbel *et al.*, Transcriptomic analysis of patients with tetralogy of Fallot reveals the effect of chronic hypoxia on myocardial gene expression. *The Journal of Thoracic and Cardiovascular Surgery*. **140**, 337–345.e26 (2010).

40. G. Dalpé *et al.*, Dystonin-Deficient Mice Exhibit an Intrinsic Muscle Weakness and an Instability of Skeletal Muscle Cytoarchitecture. *Developmental Biology*. **210**, 367–380 (1999).

41. J. G. Boyer, K. Bhanot, R. Kothary, C. Boudreau-Larivière, Hearts of Dystonia musculorum Mice Display Normal Morphological and Histological Features but Show Signs of Cardiac Stress. *PLoS ONE*. **5**, e9465 (2010).

42. Y. Fujimoto *et al.*, CD83 expression influences CD4+ T cell development in the thymus. *Cell*. **108**, 755–767 (2002).

43. M. Lechmann, S. Berchtold, J. Hauber, A. Steinkasserer, CD83 on dendritic cells: more than just a marker for maturation. *Trends Immunol.* **23**, 273–275 (2002).

44. N. Scholler *et al.*, Cutting edge: CD83 regulates the development of cellular immunity. *J. Immunol.* **168**, 2599–2602 (2002).

45. S. Agrawal, S. Gupta, A. Agrawal, Vaccinia virus proteins activate human dendritic cells to induce T cell responses in vitro. *Vaccine*. **27**, 88–92 (2009).

46. L. D. Ward, M. Kellis, HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–4 (2012).

47. A. M. Foster *et al.*, IL-36 Promotes Myeloid Cell Infiltration, Activation, and Inflammatory Activity in Skin. *J. Immunol.* **192**, 6053–6061 (2014).

48. D. N. Harris *et al.*, Evolutionary genomic dynamics of Peruvians before, during, and after the Inca Empire. *PNAS*. **115**, E6526–E6535 (2018).

49. J. E. Crawford *et al.*, Natural Selection on Genes Related to Cardiovascular Health in High-Altitude Adapted Andeans. *Am. J. Hum. Genet.* **101**, 752–767 (2017).

50. B. L. Nichols *et al.*, The maltase-glucoamylase gene: common ancestry to sucrase-isomaltase with complementary starch digestion activities. *PNAS*. **100**, 1432–1437 (2003).

51. G. Guarda, *Nueva historia de Valdivia* (Ediciones Universidad Católica de Chile, Valdivia, 2001).

52. L. Perry *et al.*, Early maize agriculture and interzonal interaction in southern Peru. *Nature*. **440**, 76–79 (2006).

53. S. Mallick *et al.*, The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature*. **538**, 201–206 (2016).

54. G. H. Perry *et al.*, Diet and the evolution of human amylase gene copy number variation. *Nat. Genet.* **39**, 1256–1260 (2007).

55. N. D. Cook, W. G. Lovell, *Secret judgments of God: Old world disease in colonial Spanish America.* (University of Oklahoma Press, Norman, 2001).

56. N. D. Cook, *Born to die: disease and New World conquest, 1492-1650.* (Cambridge, 1998).

57. J. G. Schraiber, Assessing the Relationship of Ancient and Modern Populations. *Genetics*. **208**, 383–398 (2018).

58. L. Roberts, Disease and death in the New World. *Science*. **246**, 1245–1247 (1989).

59.   F. Belinky *et al.*, PathCards: multi-source consolidation of human biological pathways. *Database (Oxford)*. **2015**, bav006–bav006 (2015).

60.   L. E. Jensen, Interleukin-36 cytokines may overcome microbial immune evasion strategies that inhibit interleukin-1 family signaling. *Sci. Signal.* **10**, eaan3589 (2017).

61.   I. H. Haralambieva *et al.*, Common SNPs/Haplotypes in IL18R1 and IL18 Genes Are Associated With Variations in Humoral Immunity to Smallpox Vaccination in Caucasians and African Americans. *J Infect Dis*. **204**, 433–441 (2011).

62.   M. Rasmussen *et al.*, The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature*. **506**, 225–229 (2014).

63.   M. Rasmussen *et al.*, Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature*. **463**, 757–762 (2010).

64.   M. Rasmussen *et al.*, The ancestry and affiliations of Kennewick Man. *Nature*. **57**, 787–10 (2015).

65.   J. H. Marcus, J. Novembre, Visualizing the geography of genetic variants. *Bioinformatics*. **33**, 594–595 (2017).

66.   A. G. Hogg *et al.*, SHCal13 Southern Hemisphere Calibration, 0–50,000 Years cal BP. *Radiocarbon*. **55**, 1889–1903 (2016).

67.   A. Parnell, Bchron: Radiocarbon Dating, Age-Depth Modelling, Relative Sea Level Rate Estimation, and Non-Parametric Phase Modelling. (2016), (available at https://CRAN.R-project.org/package=Bchron).

68.   C. O. Lovejoy, Dental wear in the Libben population: Its functional pattern and role in the determination of adult skeletal age at death. *American Journal of Physical Anthropology*. **68**, 47–56 (1985).

69.   M. Stuiver, H. Polach, Reporting of 14C data. *Radiocarbon*. **19**, 355–363 (1977).

70.   J. Dabney *et al.*, Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15758–15763 (2013).

71.   M. Meyer, M. Kircher, Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harb Protoc*. **2010**, pdb.prot5448–pdb.prot5448 (2010).

72.   J. Lindo *et al.*, Ancient individuals from the North American Northwest Coast reveal 10,000 years of regional genetic continuity. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4093–4098 (2017).

73. M. Schubert, S. Lindgreen, L. Orlando, AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes*. **9**, 395 (2016).

74. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. **25**, 1754–1760 (2009).

75. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **25**, 2078–2079 (2009).

76. H. Jónsson, A. Ginolhac, M. Schubert, P. L. F. Johnson, L. Orlando, mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*. **29**, 1682–1684 (2013).

77. Q. Fu *et al.*, Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature*. **514**, 445–449 (2014).

78. K. Prüfer *et al.*, The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*. **505**, 43–49 (2014).

79. A. McKenna *et al.*, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

80. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics 2014 15:1*. **15**, 356 (2014).

81. P. Danecek *et al.*, The variant call format and VCFtools. *Bioinformatics*. **27**, 2156–2158 (2011).

82. B. N. Howie, P. Donnelly, J. Marchini, A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLoS Genet*. **5**, e1000529 (2009).

83. M. Rueda, A. Torkamani, SG-ADVISER mtDNA: a web server for mitochondrial DNA annotation with data from 200 samples of a healthy aging cohort. *BMC Bioinformatics 2014 15:1*. **18**, 16080 (2017).

84. E. Tamm *et al.*, Beringian Standstill and Spread of Native American Founders. *PLoS ONE*. **2**, e829 (2007).

85. G. Renaud, V. Slon, A. T. Duggan, J. Kelso, Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 47 (2015).

86. A. L. Price *et al.*, Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).

87.   M. Sikora *et al.*, Population Genomic Analysis of Ancient and Modern Genomes Yields New Insights into the Genetic Ancestry of the Tyrolean Iceman and the Genetic Structure of Europe. *PLoS Genet*. **10**, e1004353 (2014).

88.   A. A. Behr, K. Z. Liu, G. Liu-Fang, P. Nakka, S. Ramachandran, Pong: fast analysis and visualization of latent clusters in population genetic data. *bioRxiv*, 1–15 (2015).

89.   B. S. Weir, C. C. Cockerham, Estimating f-statistics for the analysis of population structure. *Evolution*. **38**, 1358–1370 (1984).

90.   L. Arbiza, E. Zhong, A. Keinan, NRE: a tool for exploring neutral loci in the human genome. *BMC Bioinformatics 2014 15:1*. **13**, 301 (2012).

91.   S. Gravel *et al.*, Reconstructing Native American Migrations from Whole-Genome and Whole-Exome Data. *PLoS Genet*. **9**, e1004023 (2013).

92.   X. Wang *et al.*, CNVcaller: highly efficient and widely applicable software for detecting copy number variations in large populations. *Gigascience*. **6**, 1–12 (2017).

93.   D. Carpenter, L. M. Mitchell, J. A. L. Armour, Copy number variation of human AMY1 is a minor contributor to variation in salivary amylase expression and activity. *Hum. Genomics*. **11**, 2 (2017).

94.   M. Livi Bacci, The Depopulation of Hispanic America after the Conquest. *Population and Development Review*. **32**, 199–232 (2006).

Data and materials availability: Ancient and modern DNA sequences are available from NCBI Sequence Read Archive, accession no. PRJNA470966. The Huilliche-Pehuenche SNP data will

be available via a data access agreement with Ricardo Verdugo at the Universidad de Chile. **Competing interests:** The authors declare no competing financial interests.

Permits: Archaeological data recovery at SMP and international export of artifacts were carried out under Peruvian Ministry of Culture permit no. 064-2013-DGPA-VMPCIC/MC and 138-2015-VMPCIC/MC. Multiple permits were issued by the Peruvian Instituto Nacional de Cultura for research at Kaillachuro and the Rio Uncallane sites from 1994-2002.
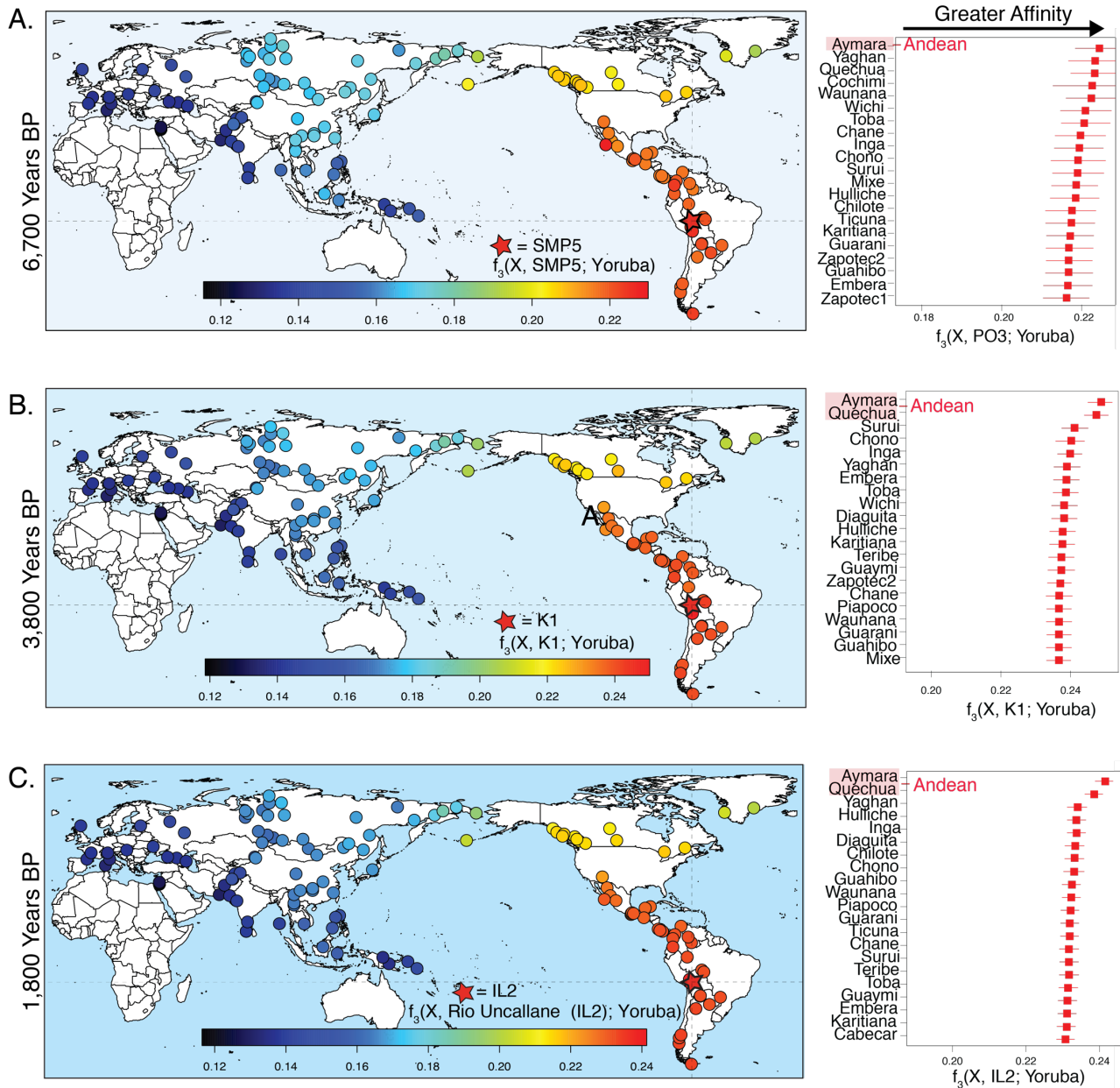
Supplementary Materials:

Materials and Methods

Figures S1-S5

Tables S1-S6

References (*59-93*)

Figure 1. f_3 statistics. (Left) Heat map represents the outgroup f_3 statistics estimating the amount of shared genetic drift between the ancient Andean individuals and each of 156 contemporary populations since their divergence with the African Yoruban population. (Right) Ranked f_3 statistics showing the greatest affinity of the ancient Andeans with respect to 45 indigenous populations of the Americas.
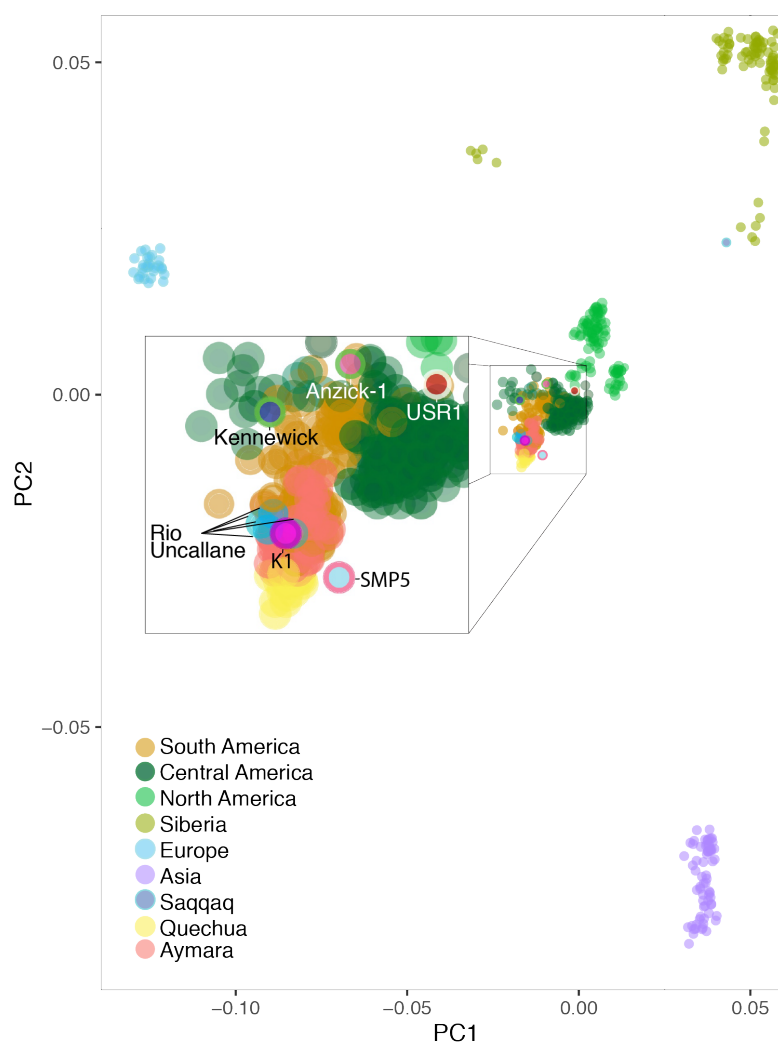
Figure 2. Principal components analysis. Principal components analysis projecting the ancient Andeans (IL2, IL3, IL4, IL5, IL7, K1, SMP5), USR1 (*22*) (Alaska), Anzick-1 (*62*) (Montana), Saqqaq (*63*) (Greenland), and Kennewick Man (*64*) (Washington) onto a set of non-African populations from Raghavan et al. (*19*), with Native American populations masked for nonnative ancestry.
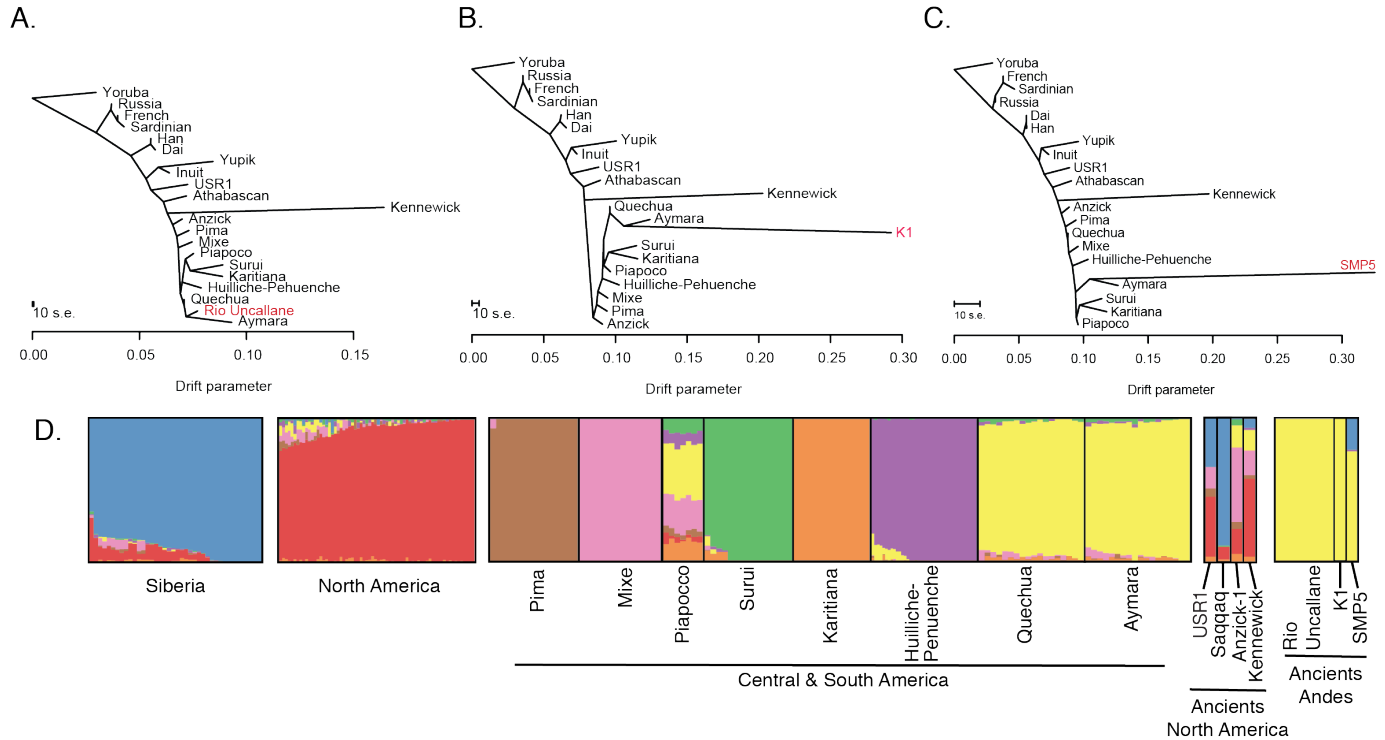
Figure 3. Genetic affinity of ancient Andeans to global and regional indigenous populations. (A-C) Maximum likelihood trees generated by *TreeMix* (*21*) using whole-genome sequencing data from the Simons Genome Project (*53*). (D) Cluster analysis generated by ADMIXTURE (*18*) for a set of indigenous populations from Siberia, the Americas, Anzick-1, Kennewick, Saqqaq, USR1, and the ancient Andeans. The number of displayed clusters is *K* = 8, which was found to have the best predictive accuracy given the lowest cross-validation index value. Native American populations masked for nonnative ancestry.
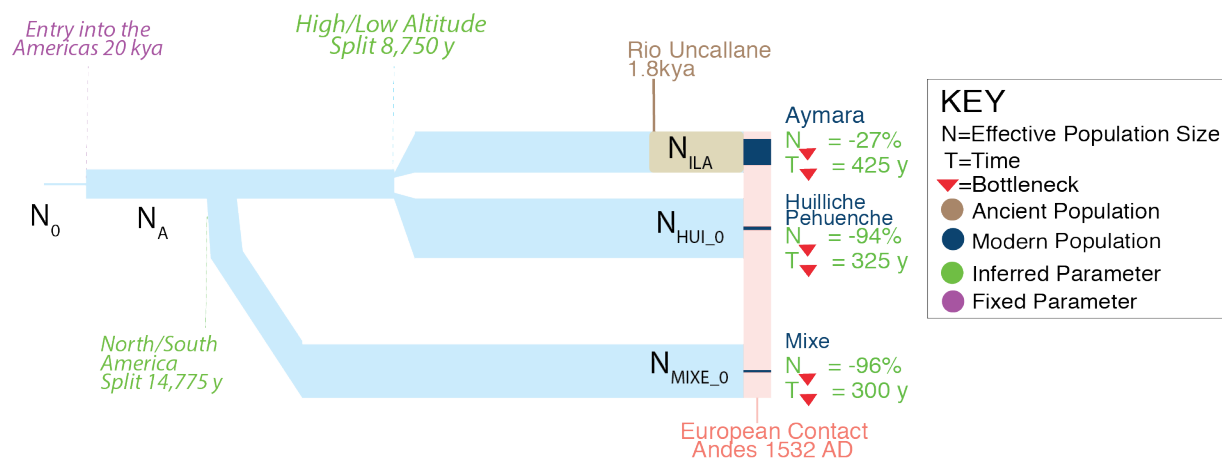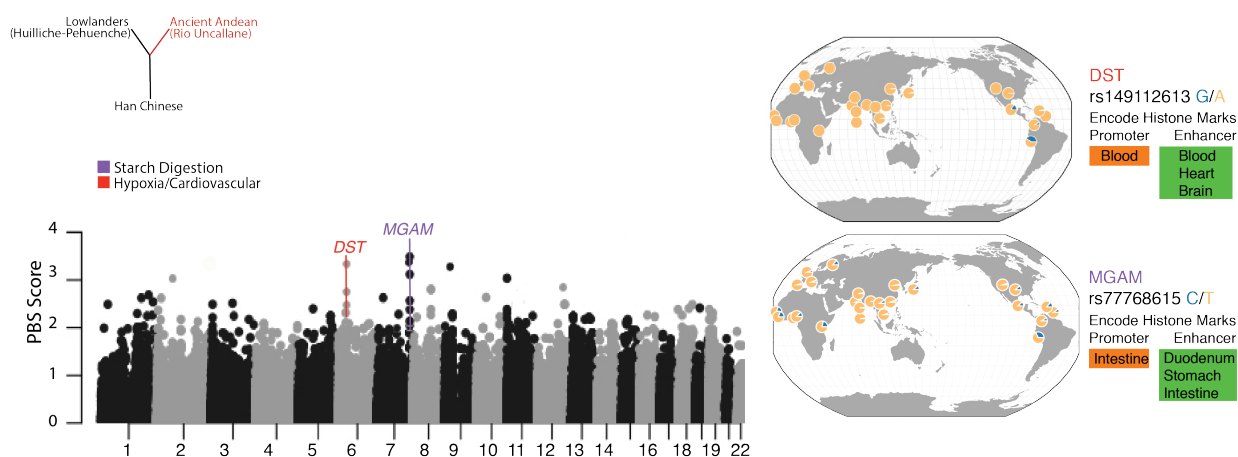
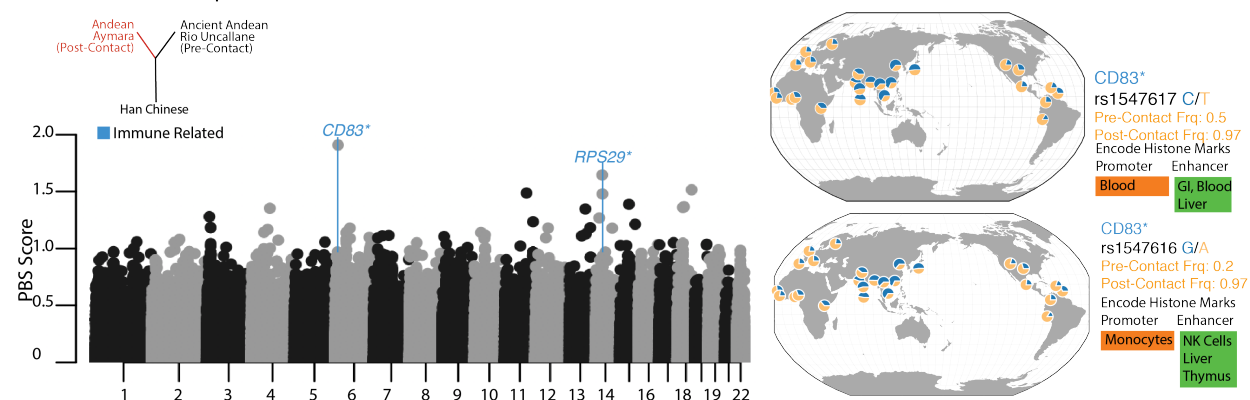**Figure 4. Demographic model of the Andes.**

5

Figure 5. Selection scans. (A) PBS scans testing for differentiation along the Ancient Andean branch, with respect to the lowland population of the Huilliche-Pehuenche. (B) PBS scan testing for differentiation along the modern Aymara branch (post-European contact), with respect to the pre-contact ancient population of the Rio Uncallane. Selected global SNP frequencies (*65*) and histone modifications (*46*) are shown to the right, which are relevant to the strongest selection signals. *Indicates closest gene.

Table 1. Ancient Sample Sequencing results

ID	[14]C BP	mode cal. BP[2]	95% cal. BP[4]	mtDNA Haplo-group	Sex[1]	Site	Endogenous DNA	Mapped Reads	Average Read Depth	Whole Genome Coverage
IL2	Not dated	NA	NA	D1	XX	Rio Uncallane	38.6%	193316050	6.51	0.99
IL3	1830±20	1711	1783-1616	B2	XX	Rio Uncallane	29.9%	175348009	5.89	0.92
IL4	1630±20	1514	1529-1426	C1b	XY	Rio Uncallane	17.1%	67544346	2.31	0.75
IL5	1805±20	1703	1722-1611	C1c	XX	Rio Uncallane	25.2%	88991036	3.34	0.94
IL7	1885±20	1806	1825-1730	C1b	XY	Rio Uncallane	44.9%	184150720	5.62	0.98
K1	Insufficient collagen	NA	NA	C1b	XX	Kaillachuro	2.60%	12815292	1.44	0.26
SMP5	6030±20	6822	6884-6757	C1c	XY	Soro Mik'aya Patjxa	3.15%	3655632	3.92	0.08

[1]Sex was determined from sequence reads, using the method described in Skoglund et al. (30).
[2]Calibrated 95% range, SHcal13 (*66, 67*).

5

10