

## ANTHROPOLOGY

# The genetic prehistory of the Andean highlands 7000 years BP though European contact

John Lindo<sup>1,2\*</sup>, Randall Haas<sup>3</sup>, Courtney Hofman<sup>4,5</sup>, Mario Apata<sup>6</sup>, Mauricio Moraga<sup>6,7</sup>, Ricardo A. Verdugo<sup>6,8\*</sup>, James T. Watson<sup>9</sup>, Carlos Viviano Llave<sup>10</sup>, David Witonsky<sup>2</sup>, Cynthia Beall<sup>11</sup>, Christina Warinner<sup>4,5,12</sup>, John Novembre<sup>2</sup>, Mark Aldenderfer<sup>13\*</sup>, Anna Di Rienzo<sup>2\*</sup>

The peopling of the Andean highlands above 2500 m 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 7000 calendar years before the present (BP), and compare them to 42 new genome-wide genetic variation datasets from both highland and lowland populations. We infer three significant features: a split between low- and high-elevation populations that occurred between 9200 and 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. 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.

## INTRODUCTION

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 before the present (BP) (4), other research indicates that permanent occupation began between 9500 and 9000 BP (5, 6). Regardless of when the peopling of high-elevation environments began, selective pressure on the human genome was likely strong—not only because of challenging environmental factors, but also because social processes such as intensification of subsistence resources and residential sedentism (7) 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 (8). Although it is known that the peoples of the Andean highlands experienced population contraction after contact (8), its extent is debated from both archeological and ethnohistorical perspectives, especially concerning the size of the indigenous populations at initial contact (9).

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 and Table 1). The series represents three different cultural periods, which include individuals from (i) Rio Uncallane, a series of cave crevice tombs dating to ~1800 BP and used by fully sedentary agriculturists, (ii) Kaillachuro, a ~3800-year-old site marked by the transition from mobile foraging to agropastoralism and residential sedentism, and (iii) Soro Mik'aya Patjxa (SMP), an 8000- to 6500-year-old site inhabited by residentially mobile hunter-gatherers (10). We then compare the genomes of these ancient individuals to 25 new genomes and 39 new genome-wide single-nucleotide polymorphism (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 2000 years (11). The Huilliche-Pehuenche are traditionally hunter-gatherers from the southern coastal forest of Chile (12).

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 and 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 last, 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 archeological periods with proportions of endogenous DNA ranging from 2.60 to 44.8%. The samples span in age from 6800 to 1400 calendar years BP (table S1). The samples from the Rio Uncallane site showed the highest endogenous content; five were chosen for frequency-based analyses and sequenced to an average read depth of 4.74×

<sup>1</sup>Department of Anthropology, Emory University, Atlanta, GA 30322, USA. <sup>2</sup>Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA. <sup>3</sup>Department of Anthropology, University of California, Davis, CA 95616, USA. <sup>4</sup>Department of Anthropology, University of Oklahoma, Norman, OK 73019, USA. <sup>5</sup>Laboratories of Molecular Anthropology and Microbiome Research, University of Oklahoma, Norman, OK 73019, USA. <sup>6</sup>Programa de Genética Humana, ICBM, Universidad de Chile, Santiago, Chile. <sup>7</sup>Departamento de Antropología, Universidad de Chile, Santiago, Chile. <sup>8</sup>Departamento de Oncología Básico-Clinica, Universidad de Chile, Santiago, Chile. <sup>9</sup>Arizona State Museum and School of Anthropology, University of Arizona, Tucson, AZ 85721, USA. <sup>10</sup>Peruvian Register of Professional Archaeologists, Peru. <sup>11</sup>Department of Anthropology, Case Western University, Cleveland, OH 44106, USA. <sup>12</sup>Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena 07745, Germany. <sup>13</sup>School of Social Sciences, Humanities, and Arts, University of California, Merced, Merced, CA 95343, USA.

\*Corresponding author. Email: jlindo@emory.edu (J.L.); raverdugo@u.uchile.cl (R.A.V.); maldenderfer@ucmerced.edu (M.A.); dirienzo@bsd.uchicago.edu (A.D.R.)

**Table 1. Ancient sample sequencing results.** mtDNA, mitochondrial DNA; NA, not applicable.

ID	<sup>14</sup> C BP	Mode calibrated BP*	95% calibrated BP*	mtDNA Haplogroup	Sex <sup>†</sup>	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	SMP	3.15	3655632	3.92	0.08

\*Calibrated 95% range, SHcal13 (65, 66).

†Sex was determined from sequence reads.

(Table 1). The oldest sample is from the SMP site SMP5. The sample was directly dated to ~6800 years BP, exhibited the highest endogenous DNA proportion among burials from the same site (3.15%), and sequenced to an average read depth of 3.92×. The sample from the Kaillachuro site, at about 3800 years BP (culturally dated), was sequenced to an average read depth of 1.44×. All samples exhibited low contamination rates (1 to 4%) and DNA damage patterns consistent with degraded DNA (table S2 and 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 4.36× (table S3), and an additional individual was sequenced to a read depth of 31.20× (see the Supplementary Materials). The modern Huilliche-Pehuenche (13) ( $n = 39$ ) from the south of Chile (fig. S1) were genotyped on an Axiom LAT1 Array and imputed using the 1000 Genomes Project phase 3 (14) and the Aymara whole-genome sequence data as a reference panel. Analysis using the program ADMIXTURE (15) revealed that all the modern samples had trivial amounts of nonindigenous ancestry (less than 5%).

### Genetic relationship between ancient and modern individuals

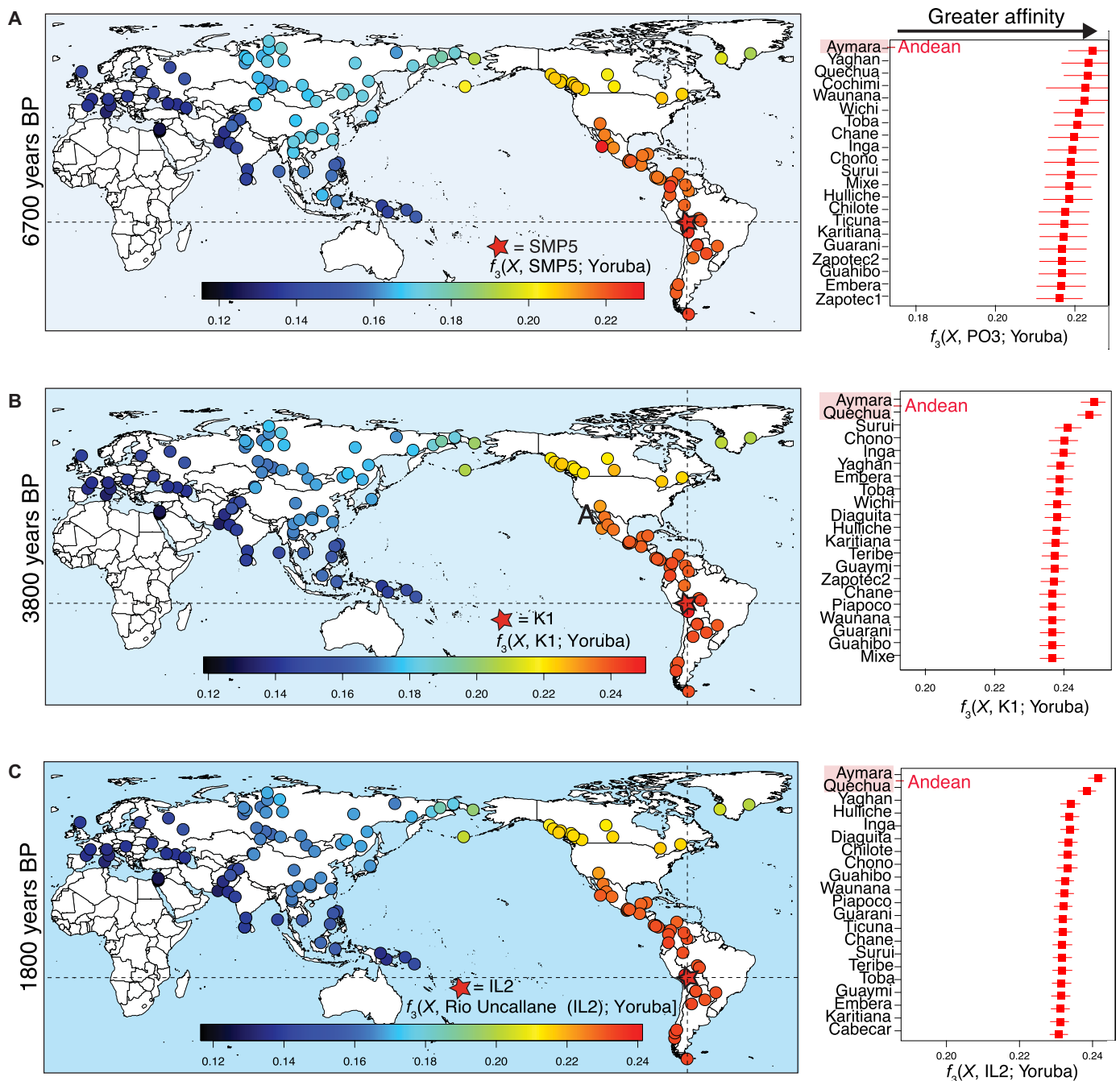
We used outgroup  $f_3$  statistics to assess the shared genetic ancestry among the ancient individuals and 156 worldwide populations (16). C/T and G/A polymorphic sites were removed from the dataset to guard against the most common forms of postmortem DNA damage (17). Outgroup  $f_3$  statistics of 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 the 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 among the ancient Andean individuals and modern populations from South America, we examined maximum likelihood trees inferred with TreeMix (18). We observe that individuals from all three time periods form a sister group to the modern Andean high-altitude population, the Aymara (Bolivia; Fig. 3, A to 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 not only traces most of this 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 (19), 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 4000 years BP, and perhaps up to 7000 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 used a composite likelihood method [Fastsimcoal2 (20)], 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 8750 years [95% confidence interval (CI), 8200 to 9250]. We also infer the magnitude of the population collapse in the Andes after European contact to be a

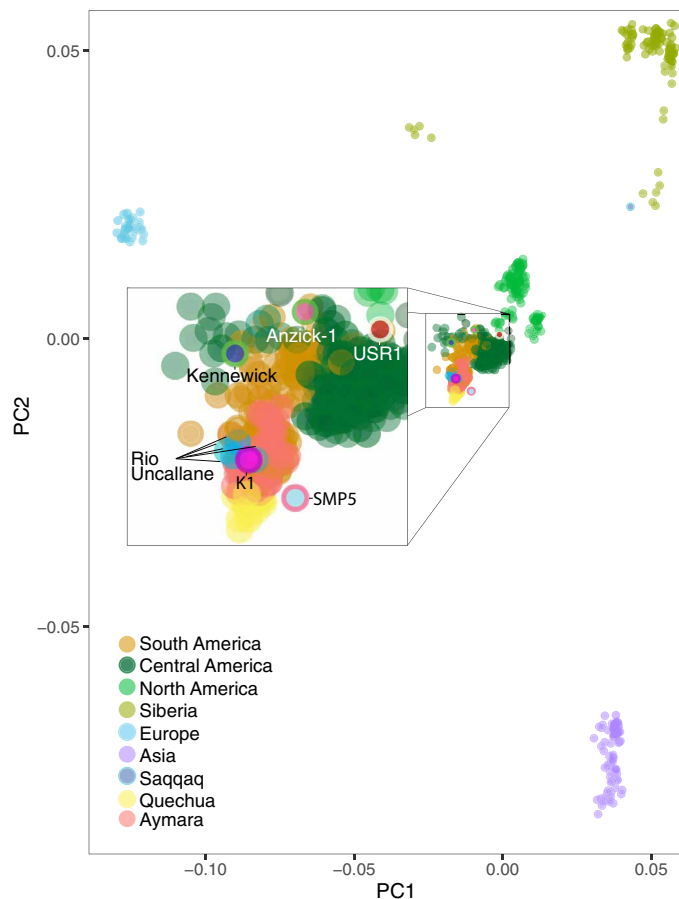


**Fig. 1.  $f_3$  statistics.** (A to C) Left: Heat map represents the outgroup  $f_3$  statistics estimating the amount of shared genetic drift between the ancient Andean individuals and each of 156 contemporary populations since their divergence with the African Yoruban population. Right: Ranked  $f_3$  statistics showing the greatest affinity of the ancient Andeans with respect to 45 indigenous populations of the Americas.

27% reduction in effective population size and occurring 425 years ago (95% CI, 400 to 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-Pehuenche after European contact, which were much more severe. We infer a 94% (95% CI, 0.94 to 0.96) collapse for the Mixe and a comparable 96% (95% CI, 0.95 to 0.96) collapse for the Huilliche-

Pehuenche. The collapse difference between high- and low-altitude populations from the Andes region is significant [Huilliche-Pehuenche versus Aymara,  $P < 0.0001$  (95% CI, 0.66 to 0.71),  $\chi^2$  test].

However, we were concerned with the heterogeneous nature of the data used in the model, since the Huilliche-Pehuenche data were imputed from an SNP array. Therefore, we calculated heterozygosity on each population using an overlapping set of ascertained SNPs



**Fig. 2. Principal components analysis.** Principal components analysis projecting the ancient Andeans (IL2, IL3, IL4, IL5, IL7, K1, and SMP5), USR1 (Alaska) (19), Anzick-1 (Montana) (61), Saqqaq (Greenland) (62), and Kennewick Man (Washington) (63) onto a set of non-African populations from Raghavan *et al.* (16), with Native American populations masked for non-native ancestry. PC, principal component.

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 Materials and Methods and 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 (21–24). Several genetic studies have also been performed on modern populations of the Andes (25–28). 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 allow 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 (29). Therefore, we test for high allele frequency divergence resulting from adaptations to the Andean high-altitude environment, using a whole-genome approach with the five Rio Uncallane individuals

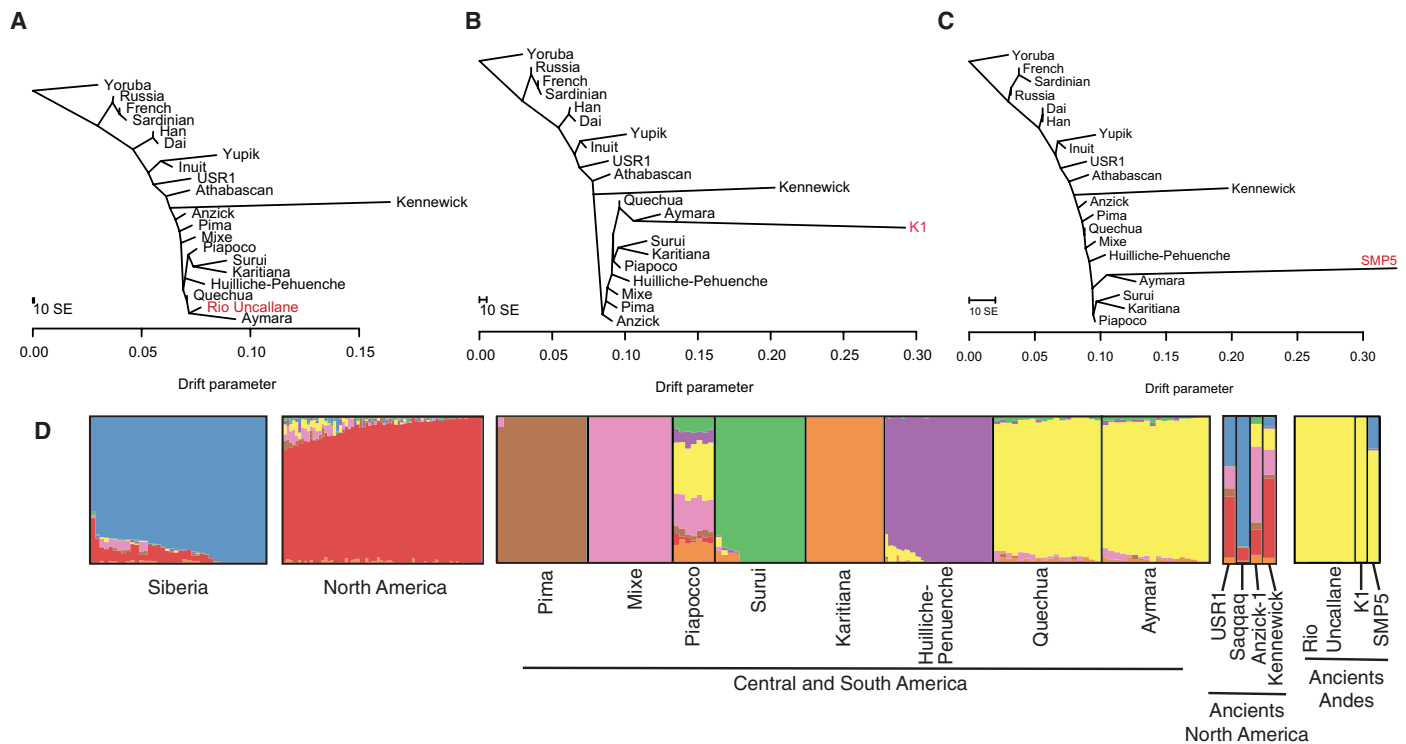
(~1800 BP), and we contrast them with a lowland population, the Huilliche-Pehuenche from Chile, who likely inhabited the region for several thousand years before the arrival of the Spanish (30). We used the Population Branch Statistic (PBS) (24) to scan for alleles that reached high frequency in the Rio Uncallane relative to the Huilliche-Pehuenche. The Han Chinese from the 1000 Genomes Project (14) were used as the third comparative population. The SNPs showing the most extreme PBS score and, therefore, the highest differentiation in the ancient Andeans correspond to a region near *MGAM* (Fig. 5A and table S4). *MGAM* is an intestinal enzyme associated with starch digestion (31) and could possibly be linked to the intensification of tuber use and ultimately the transition to agriculture (7, 32). The second strongest signal came from the gene *DST*, which encodes a cytoskeleton linker protein active in neural and muscle cells (33). *DST* has been associated with differential expression under hypoxic conditions (34) and cardiovascular health (35).

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 (14). 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 (36). The gene has also been shown to be up-regulated in response to the vaccinia virus infection (37), 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 cells (38), suggesting that they may be functional. Last, 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 (39).

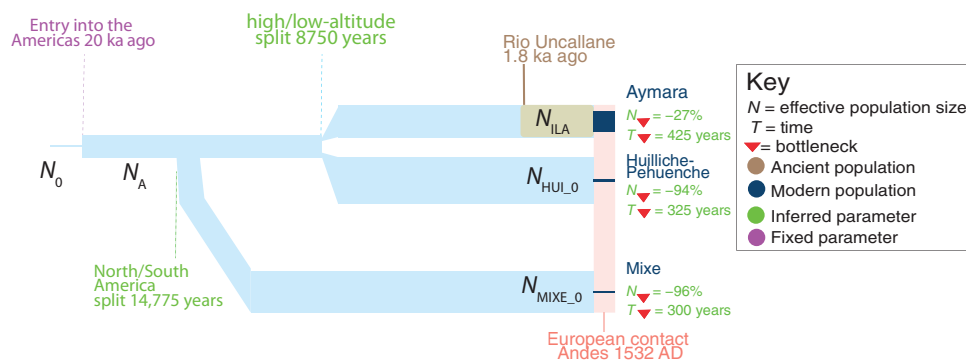
### DISCUSSION

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 are the Andean highlands, which boast frigid temperatures, low partial pressure of oxygen, and intense ultraviolet radiation. Despite this, however, humans eventually spread throughout the Andes and occupied them permanently. Archeological evidence suggests that hunter-gatherers entered the highlands as early as 12,000 years BP (4), with permanent occupation beginning around 9000 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 3800 years BP, and possibly 7000 years BP. This affinity extends to the present high-altitude Andean communities of the Aymara and Quechua.

Although our samples do not extend beyond 7000 years BP, we were able to model the initial entry into the Andes after the split between North and South American groups. Our model, using a mutation rate of  $1.25 \times 10^{-8}$ , shows a correlation with archeological evidence regarding the split between North and South groups occurring nearly 14,750 years ago (95% CI, 14,225 to 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 8750 years



**Fig. 3. Genetic affinity of ancient Andeans to global and regional indigenous populations.** (A to C) Maximum likelihood trees generated by TreeMix (18) using whole-genome sequencing data from the Simons Genome Diversity Project (46). (D) Cluster analysis generated by ADMIXTURE (15) 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 non-native ancestry. SE, standard error.



**Fig. 4. Demographic model of the Andes.** ILA, Rio Uncallane.

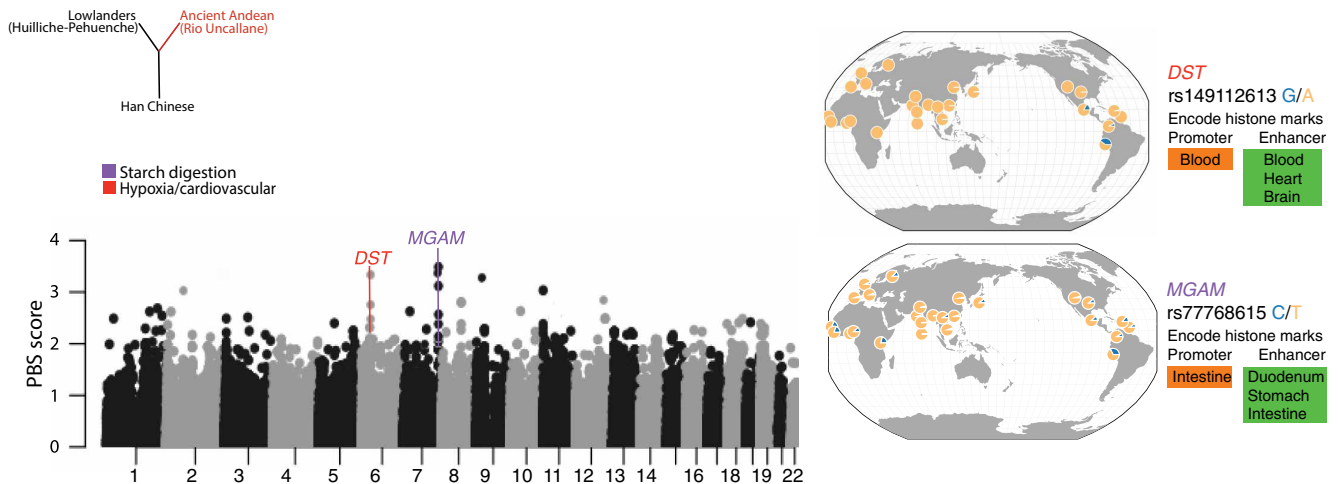
(95% CI, 8200 to 9250), which is younger than previously reported by a study using modern genomes alone (40). This date provides a terminus ante quem time frame for the origins of adaptations known in modern highland populations.

We also present evidence for genes that may have been under selective pressure caused by environmental stressors in the Andes. 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 (35, 41). Furthermore, the *DST* intronic SNP that was most differentiated (rs149112613) shows histone modifications associated with blood and the right ventricle of the heart (38). This correlates to Andean highlanders tending to have enlarged right

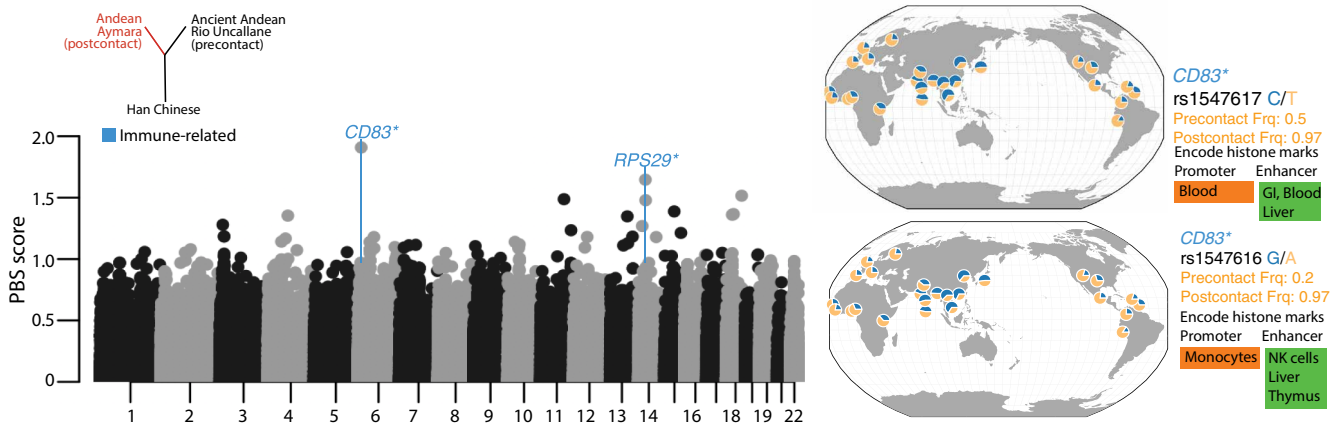
ventricles associated with moderate pulmonary hypertension (1). This finding also parallels hypotheses proposed by Crawford *et al.* (42) that Andeans may have adapted to high-altitude hypoxia via cardiovascular modifications.

The most extreme signal may represent adaptations to an agricultural subsistence and diet. The top-ranked gene, *MGAM*, is associated with starch digestion (43). The associated high-frequency SNPs in the ancient Andean population (table S4) exhibit chromatin marks in cells from the gastrointestinal tract (Fig. 5A). The variant may be highly differentiated between the ancient Andeans and the lowlanders (the Huilliche-Pehuenche) because of differences in subsistence strategies. The Huilliche-Pehuenche individuals are traditionally hunter-gatherers, with archeological evidence suggesting that their

**A Selection: High altitude**



**B Selection: European-borne disease**



**Fig. 5. Selection scans.** (A) PBS scans testing for differentiation along the Ancient Andean branch, with respect to the lowland population of the Huillliche-Pehuenche. (B) PBS scan testing for differentiation along the modern Aymara branch (post-European contact), with respect to the precontact ancient population of the Rio Uncallane. Selected global SNP frequencies (Frq) (64) and histone modifications (38) are shown on the right, which are relevant to the strongest selection signals. “\*” indicates closest gene. GI, gastrointestinal tract; NK, natural killer.

ancestors have been practicing this mode of subsistence for thousands of years in the region before European contact in the 1500s (44). 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) (45) and the potato (~3400 years BP) (7). Selection acting on the *MGAM* gene in the ancient Andeans may represent an adaptive response to greater reliance upon starchy domesticates. Recent archeological 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 (7, 32). Furthermore, we see a similar signal (top 0.01%) when we contrast the hunter-gatherers from Brazil [Karitiana/Surui, sequence data (46)] with the ancient Andeans, as well as with the Aymara versus the Huillliche-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 (47).

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 (29). Not only were long-standing states and social organizations disrupted, but also the environment itself was altered with the arrival of European-introduced pathogens, which may have preceded the arrival of the Spanish via trade routes (29). Some of the most devastating epidemics were related to smallpox, occurring in the 1500s and 1600s (8). These combined factors are thought to have decimated the local populations (8). We inferred the population decline in the Andes, using the ancient and modern Andeans, and found the decline in effective population size to be 27% (95% CI, 0.23 to 0.34). We also simulated DNA sequence data immediately before the collapse between the Rio Uncallane and the Aymara using a truncated model and found a reduction in average heterozygosity of 23% (see the Supplementary Materials). This is a modest decline compared to archeological and historical estimates, which reached upward of 90% of the total population (48). In contrast, the model infers a much more severe collapse for lowland Andean populations of the Huillliche-Pehuenche, exceeding 90%. We also explored alternative scenarios to make sure that 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 the Supplementary Materials). Although we did not have precontact 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 (29).

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 (49) that does not allow for recent gene flow was not significant (see the Supplementary Materials), 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 (50) and that the first epidemics were likely caused by smallpox (29). 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* (37). 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 those of influenza (51). 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 (52). Furthermore, the top SNP associated with *IL-36R* (rs1117797) exhibits a QTL (quantitative trait locus) signal associated with *IL18RAP*, a gene involved in mediating the immune response to the vaccinia virus (53). 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 marked 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 4000 years BP, and possibly extending to 7000 years BP.

## MATERIALS AND METHODS

### Community engagement

Although the communities in this study are not structured around tribal leadership, as in North America, all samples were collected by indigenous representatives and/or local researchers. For the Aymara population, the samples were collected by E. Vargas Pacheco and M. Villena, two representatives of the Aymara community. The Huilliche-Pehuenche samples correspond to a historical collection of modern DNA samples at the Program of Human Genetics of the University of Chile. All of the participants were healthy individuals, over 18 years of age, and gave their informed consent. The study was approved by the Ethics Committee of the Faculty of Medicine,

University of Chile. M.Ap. helped analyze the data while visiting at the University of Chicago.

### DNA extraction

We used standard ancient DNA extraction methods following stringent guidelines to work with ancient human remains and conducted these in a dedicated ancient DNA laboratory at the Laboratories of Molecular Anthropology and Microbiome Research at the University of Oklahoma, following Dabney *et al.* (54) with the following modifications. After decontaminating each tooth with bleach and removing the outer layer of cementum, the whole tooth was crushed. The resulting bone powder was predigested in 1 ml of 0.5 M EDTA for 15 min at room temperature (RT) to remove loose contaminants. The supernatant was then removed, and the sample was digested in additional 1 ml of 0.5 M EDTA at RT overnight. One hundred microliters of QIAGEN proteinase K was added to the samples the following day and incubated for 8 hours at 37°C. The supernatant was then removed, and an additional 0.5 ml of EDTA was added along with 50  $\mu$ l of proteinase K for 3 to 5 days, until the powder was no longer mineralized. Silica column-based extraction protocol was then followed as described in (54). DNA was extracted from molars belonging to individuals IL2, IL3, IL4, IL5, IL7, K1, and SMP5.

### Library construction

Libraries were constructed from 30  $\mu$ l of DNA extract as detailed by (55), with the following modifications. Blunt-end repair was performed with the NEBNext Quick DNA Library Prep Master Mix Set for 454 following the manufacturer's protocol and deactivating the *Bst* enzyme at 80°C for 20 min. The libraries were amplified with KAPA HiFi U+ kit, using 10  $\mu$ l of DNA, 25  $\mu$ l of KAPA ReadMix (2 $\times$ ), 11  $\mu$ l of H<sub>2</sub>O, 1  $\mu$ l of bovine serum albumin (2.5 mg/ml), and 1.5  $\mu$ l of the IS5 and IS7 primers (each). The thermocycler program was as follows: initial denature at 95°C for 5 min, (98°C for 20 s, 60°C for 15 s, and 72°C for 30 s) for *X* cycles, final extension at 72°C for 5 min, and hold at 4°C. Cycles were determined via quantitative polymerase chain reaction using 1  $\mu$ l of the preamplified library to determine the amplification curve.

### Whole-genome capture

Given the low endogenous DNA of K1 and SMP5, we performed a whole-genome capture using the myBaits whole-genome enrichment kit enhanced with protocol modifications recommended for the study of ancient DNA (56). The two enriched libraries were then pooled and sequenced (PE125) one lane of an Illumina HiSeq 4000, located at the University of Chicago Genomics Facility.

### Sample sequencing and SNP array

Twenty-four Aymara individuals underwent shotgun sequencing, multiplexing six individuals on eight lanes, on the Illumina HiSeq 4000, PE125, located at the University of Chicago Genomics Facility. The average coverage for the 24 individuals, after bioinformatics processing, was 4.36 (table S3). One additional Aymara was sequenced to high coverage, 30 $\times$ , using the HiSeq X (PE150) at Macrogen Labs (<http://macrogenlab.com>). Thirty-nine Huilliche-Pehuenche individuals were genotyped via the Axiom LAT1 Chip (Affymetrix, Santa Clara, CA), and 22 of these were previously published (13). The ancient individuals were sequenced with the HiSeq X (PE150), one lane each, at Macrogen Labs. For this round of sequencing, SMP5 and K1 were not enriched.

## Test for strict genetic continuity

Given that the above analyses provide an inference of temporal genetic affinity in the Andes from roughly 4000 years BP, and perhaps extending to 7000 years, we tested for strict continuity. Here, continuity is defined as a genetic affinity between two temporal populations without external gene flow. We performed the test for continuity using the method described in Schraiber (49). We find that when the Rio Uncallane individuals are tested against the 25 modern Aymara, the test fails to show strict continuity (230,649 SNPs:  $t_1 = 0.00443$ ,  $t_2 = 0.04189$ ;  $\log \text{LRT } P = -\infty$ ). However, this test might be inappropriate for the region, since it assumes no outside gene flow between populations. Several civilizations rose and fell in the region between the two time periods represented by the ancient and modern populations. These civilizations, such as the Wari and Incas, built elaborate trade networks throughout the Andes, and it is likely that the ancestral populations leading to the Aymara experienced regional admixture before European contact.

## Demographic history model

Parameters for the demographic model (Fig. 4) were inferred with FastSimCoal2 v. 2.6.0.2 (20). For the inferred values, 100 optimizations were run, taking the best-likelihood parameters from each. The inferences used a per-base per-generation mutation rate of  $1.25 \times 10^{-8}$  and joint derived SFS for the Mixe (46), Illave, Aymara, and Huilliche-Pehuenche, which were estimated using ANGSD (analysis of next-generation sequencing data) (57). This SFS contained monomorphic and polymorphic sites based on hg19 that were limited to regions thought to be neutrally evolving, which were identified using the NRE (neutral region explorer) tool (58). A parametric bootstrapping approach was used to construct the 95% CIs. Specific parameters are listed in the Supplementary Materials.

## Copy number variation analysis

We performed a copy number variation analysis using CNVcaller (59) on the Rio Uncallane ancient population (an agriculture-based culture, pre-European contact), using USR1 as an outgroup (an 11,500-year-old ancient individual from Alaska found to be ancestral to Native American populations). We found no evidence of *AMY1* high copy number in the Rio Uncallane, which may suggest a different evolutionary path for starch digestion in the Andes when compared to Europeans. Furthermore, we only found *AMY1C* to have two copies, while the other *AMY1* genes showed zero copies. Previous studies, focusing on European populations, found a range of 2 to 18 copies, with an average of six copies per person [see (60) and references therein].

## SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/4/11/eaau4921/DC1>

Supplementary Materials and Methods

Fig. S1. Location of ancient and modern samples.

Fig. S2. DNA damage patterns for the ancient individuals.

Fig. S3. Heterozygosity rate calculated with overlapping ascertained SNPs from the SAN panel of the Affymetrix Human Origin Array.

Fig. S4. Maximum likelihood trees with  $m = 1$  migration events.

Fig. S5. Cluster analysis generated by ADMIXTURE.

Table S1. Radiocarbon determinations on five individuals from the Illave region.

Table S2. Contamination estimates.

Table S3. Modern Aymara sequencing results.

Table S4. Positive selection candidates associated with the Andean highlands and pathogen response.

Table S5. Selection scan top signals.

Table S6. Demographical model parameters.

References (69–88)

## REFERENCES AND NOTES

1. C. M. Beall, Adaptation to high altitude: Phenotypes and genotypes. *Annu. Rev. Anthropol.* **43**, 251–272 (2014).
2. M. Aldenderfer, Modelling plateau peoples: The early human use of the world's high plateaux. *World Archaeol.* **38**, 357–370 (2007).
3. T. D. Dillehay, C. Ramirez, M. Pino, M. B. Collins, J. Rossen, J. D. Pino-Navarro, Monte Verde: Seaweed, food, medicine, and the peopling of South America. *Science* **320**, 784–786 (2008).
4. K. Rademaker, G. Hodgins, K. Moore, S. Zarrillo, C. Miller, G. R. Bromley, P. Leach, D. A. Reid, W. Y. Alvarez, D. H. Sandweiss, Paleoindian settlement of the high-altitude Peruvian Andes. *Science* **346**, 466–469 (2014).
5. D. Chala-Aldana, H. Bocherens, C. Miller, K. Moore, G. Hodgins, K. Rademaker, Investigating mobility and highland occupation strategies during the Early Holocene at the Cuncacha rock shelter through strontium and oxygen isotopes. *J. Archaeol. Sci. Rep.* (2017).
6. M. S. Aldenderfer, *Montane Foragers: Asana and the South-Central Andean Archaic* (University of Iowa Press, 1998).
7. 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).
8. N. D. Cook, *Born to Die: Disease and New World Conquest, 1492-1650* (Cambridge Univ. Press, 1998).
9. N. D. Cook, *Demographic Collapse* (Cambridge Univ. Press, 2004).
10. 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).
11. D. L. Browman, Titicaca Basin archaeolinguistics: Uru, Pukina and Aymara AD 750–1450. *World Archaeol.* **26**, 235–251 (2010).
12. N. Lira, The Maritime cultural landscape of northern Patagonia. *J. Mari. Arch.* **12**, 199–221 (2017).
13. C. de la Fuente, M. C. Ávila-Arcos, J. Galimany, M. L. Carpenter, J. R. Homburger, A. Blanco, P. Contreras, D. Cruz Dávalos, O. Reyes, M. San Roman, A. Moreno-Estrada, P. F. Campos, C. Eng, S. Huntsman, E. G. Burchard, A. S. Malaspinas, C. D. Bustamante, E. Willerslev, E. Lloper, R. A. Verdugo, M. Moraga, Genomic insights into the origin and diversification of late maritime hunter-gatherers from the Chilean Patagonia. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E4006–E4012 (2018).
14. 1000 Genomes Project Consortium, A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean, G. R. Abecasis, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
15. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
16. M. Raghavan, M. Steinrücken, K. Harris, S. Schiffels, S. Rasmussen, M. DeGiorgio, A. Albrechtsen, C. Valdiosera, M. C. Ávila-Arcos, A.-S. Malaspinas, A. Eriksson, I. Moltke, M. Metspalu, J. R. Homburger, J. Wall, O. E. Cornejo, J. V. Moreno-Mayar, T. S. Korneliusson, T. Pierre, M. Rasmussen, P. F. Campos, P. de Barros Damgaard, M. E. Allentoft, J. Lindo, E. Metspalu, R. Rodríguez-Varela, J. Mansilla, C. Henrickson, A. Seguin-Orlando, H. Malmström, T. Stafford Jr., S. S. Shringarpure, A. Moreno-Estrada, M. Karmin, K. Tambets, A. Bergström, Y. Xue, V. Warmuth, A. D. Friend, J. Singarayer, P. Valdes, F. Ballou, I. Lebonero, J. L. Vera, H. Rangel-Villalobos, D. Pettener, M. J. Mosher, D. Archer, J. Cybulski, C. P. E. Zollikofer, M. S. Ponce de León, C. I. Smith, V. Grimes, K.-A. Pike, M. Deal, B. T. Fuller, B. Arriaza, V. Standen, M. F. Luz, F. Ricaut, N. Guidon, L. Osipova, M. I. Voevodova, O. L. Posukh, O. Balonovsky, M. Lavryashina, Y. Bogunov, E. Khusnutdinova, M. Gubina, E. Balanovska, S. Fedorova, S. Litvinov, B. Malyarchuk, M. Derenko, M. J. Mosher, D. Archer, J. Cybulski, B. Petzelt, J. Mitchell, R. Worl, P. J. Norman, P. Parham, B. M. Kemp, T. Kivisild, C. Tyler-Smith, M. S. Sandhu, M. Crawford, R. Villems, D. G. Smith, M. R. Waters, T. Goebel, J. R. Johnson, R. S. Malhi, M. Jakobsson, D. J. Meltzer, A. Manica, R. Durbin, C. D. Bustamante, Y. S. Song, R. Nielsen, E. Willerslev, Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* **349**, aab3884 (2015).
17. A. W. Briggs, U. Stenzel, P. L. Johnson, R. E. Green, J. Kelso, K. Prüfer, M. Meyer, J. Krause, M. T. Ronan, M. Lachmann, S. Pääbo, Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 14616–14621 (2007).
18. J. K. Pickrell, J. K. Pritchard, Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
19. J. V. Moreno-Mayar, B. A. Potter, L. Vinner, M. Steinrücken, S. Rasmussen, J. Terhorst, J. A. Kamm, A. Albrechtsen, A.-S. Malaspinas, M. Sikora, J. D. Reuther, J. D. Irish, R. S. Malhi, L. Orlando, Y. S. Song, R. Nielsen, D. J. Meltzer, E. Willerslev, Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature* **553**, 203–207 (2018).
20. L. Excoffier, I. Dupanloup, E. Huerta-Sánchez, V. C. Sousa, M. Foll, Robust demographic inference from genomic and SNP data. *PLoS Genet.* **9**, e1003905 (2013).
21. C. M. Beall, G. L. Cavalleri, L. Deng, R. C. Elston, Y. Gao, J. Knight, C. Li, J. C. Li, Y. Liang, M. McCormack, H. E. Montgomery, H. Pan, P. A. Robbins, K. V. Shianna, S. C. Tam, N. Tsering, K. R. Veeramah, W. Wang, P. Wangdui, M. E. Weale, Y. Xu, Z. Xu, L. Yang, M. J. Zaman, C. Zeng, L. Zhang, X. Zhang, P. Zhaxi, Y. T. Zheng, Natural selection on EPAS1 (HIF2 $\alpha$ ) associated with low hemoglobin concentration in Tibetan highlanders. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 11459–11464 (2010).



22. E. Huerta-Sánchez, M. Degiorgio, L. Pagani, A. Tarekegn, R. Ekong, T. Antao, A. Cardona, H. E. Montgomery, G. L. Cavalleri, P. A. Robbins, M. E. Weale, N. Bradman, E. Bekele, T. Kivisild, C. Tyler-Smith, R. Nielsen, Genetic signatures reveal high-altitude adaptation in a set of Ethiopian populations. *Mol. Biol. Evol.* **30**, 1877–1888 (2013).
23. C. Jeong, G. Alkorta-Aranburu, B. Basnyat, M. Neupane, D. B. Witonsky, J. K. Pritchard, C. M. Beall, A. Di Rienzo, Admixture facilitates genetic adaptations to high altitude in Tibet. *Nat. Commun.* **5**, 3281 (2014).
24. X. Yi, E. Huerta-Sanchez, X. Jin, Z. X. P. Cuo, J. E. Pool, X. Xu, H. Jiang, N. Vinckenbosch, T. S. Korneliussen, H. Zheng, T. Liu, W. He, K. Li, R. Luo, X. Nie, H. Wu, M. Zhao, H. Cao, J. Zou, Y. Shan, S. Li, Q. Yang, Asan, P. Ni, G. Tian, J. Xu, X. Liu, T. Jiang, R. Wu, G. Zhou, M. Tang, J. Qin, T. Wang, S. Feng, G. Li, J. L. Huasang, W. Wang, F. Chen, Y. Wang, X. Zheng, Z. Li, Z. Bianba, G. Yang, X. Wang, S. Tang, G. Gao, Y. Chen, Z. Luo, L. Gusang, Z. Cao, Q. Zhang, W. Ouyang, X. Ren, H. Liang, H. Zheng, Y. Huang, J. Li, L. Bolund, K. Kristiansen, Y. Li, Y. Zhang, X. Zhang, R. Li, S. Li, H. Yang, R. Nielsen, J. Wang, J. Wang, Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* **329**, 75–78 (2010).
25. A. Bigham, M. Bauchet, D. Pinto, X. Mao, J. M. Akey, R. Mei, S. W. Scherer, C. G. Julian, M. J. Wilson, D. López Herráez, T. Brutsaert, E. J. Parra, L. G. Moore, M. D. Shriver, Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. *PLoS Genet.* **6**, e1001116 (2010).
26. 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. *Sci. Rep.* **6**, 23485 (2016).
27. C. A. Eichstaedt, L. Pagani, T. Antao, C. E. Inchley, A. Cardona, A. Mörseburg, F. J. Clemente, T. J. Sluckin, E. Metspalu, M. Mitt, R. Mägi, G. Hudjashov, M. Metspalu, M. Mormina, G. S. Jacobs, T. Kivisild, Evidence of early-stage selection on EPAS1 and GPR126 genes in Andean high altitude populations. *Sci. Rep.* **7**, 13042 (2017).
28. G. Valverde, H. Zhou, S. Lippold, C. de Filippo, K. Tang, D. López Herráez, J. Li, M. Stoneking, A novel candidate region for genetic adaptation to high altitude in Andean populations. *PLoS ONE* **10**, e0125444 (2015).
29. N. D. Cook, W. G. Lovell, *Secret Judgments of God: Old World Disease in Colonial Spanish America* (University of Oklahoma Press, 2001).
30. G. A. Brito, in *Antiguas Culturas Del Norte Chico* (Museo Chileno de Arte Precolombino, 2001), pp. 15–29.
31. C. Pontremoli, A. Mozzi, D. Forni, R. Cagliani, U. Pozzoli, G. Menozzi, J. Vertemara, N. Bresolin, M. Clerici, M. Sironi, Natural selection at the brush-border: Adaptations to carbohydrate diets in humans and other mammals. *Genome Biol. Evol.* **7**, 2569–2584 (2015).
32. J. T. Watson, R. Haas, Dental evidence for wild tuber processing among Titicaca Basin foragers 7000 ybp. *Am. J. Phys. Anthropol.* **164**, 117–130 (2017).
33. M. Safran, I. Dalah, J. Alexander, N. Rosen, T. I. Stein, M. Shmoish, N. Nativ, I. Bahir, T. Doniger, H. Krug, A. Sirota-Madi, T. Olender, Y. Golan, G. Stelzer, A. Harel, D. Lancet, GeneCards version 3: The human gene integrator. *Database* **2010**, baq020 (2010).
34. M. T. Ghorbel, M. Cherif, E. Jenkins, A. Mokhtari, D. Kenny, G. D. Angelini, M. Caputo, Transcriptomic analysis of patients with tetralogy of Fallot reveals the effect of chronic hypoxia on myocardial gene expression. *J. Thorac. Cardiovasc. Surg.* **140**, 337–345.e26 (2010).
35. 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).
36. 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).
37. 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).
38. 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–D934 (2012).
39. A. M. Foster, J. Baliwag, C. S. Chen, A. M. Guzman, S. W. Stoll, J. E. Gudjonsson, N. L. Ward, A. Johnston, IL-36 promotes myeloid cell infiltration, activation, and inflammatory activity in skin. *J. Immunol.* **192**, 6053–6061 (2014).
40. D. N. Harris, W. Song, A. C. Shetty, K. S. Levano, O. Cáceres, C. Padilla, V. Borda, D. Tarazona, O. Trujillo, C. Sanchez, M. D. Kessler, M. Galarza, S. Cristarano, H. Montejo, P. O. Flores-Villanueva, E. Tarazona-Santos, T. D. O'Connor, H. Guio, Evolutionary genomic dynamics of Peruvians before, during, and after the Inca Empire. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E6526–E6535 (2018).
41. G. Dalpé, M. Mathieu, A. Comtois, E. Zhu, S. Wasiak, Y. De Repentigny, N. Leclerc, R. Kothary, Dystonin-deficient mice exhibit an intrinsic muscle weakness and an instability of skeletal muscle cytoarchitecture. *Dev. Biol.* **210**, 367–380 (1999).
42. J. E. Crawford, R. Amaru, J. Song, C. G. Julian, F. Racimo, J. Y. Cheng, X. Guo, J. Yao, B. Ambale-Venkatesh, J. A. Lima, J. I. Rotter, J. Stehlik, L. G. Moore, J. T. Prchal, R. Nielsen, Natural selection on genes related to cardiovascular health in high-altitude adapted Andeans. *Am. J. Hum. Genet.* **101**, 752–767 (2017).
43. B. L. Nichols, S. Avery, P. Sen, D. M. Swallow, D. Hahn, E. Sterchi, The maltase-glucoamylase gene: Common ancestry to sucrase-isomaltase with complementary starch digestion activities. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1432–1437 (2003).
44. G. Guarda, *Nueva Historia de Valdivia* (Ediciones Universidad Católica de Chile, 2001).
45. L. Perry, D. H. Sandweiss, D. R. Piperno, K. Rademaker, M. A. Malpass, A. Umire, P. de la Vera, Early maize agriculture and interzonal interaction in southern Peru. *Nature* **440**, 76–79 (2006).
46. S. Mallick, H. Li, M. Lipson, I. Mathieson, M. Gymrek, F. Racimo, M. Zhao, N. Chennagiri, S. Nordenfelt, A. Tandon, P. Skoglund, I. Lazaridis, S. Sankararaman, Q. Fu, N. Rohland, G. Renaud, Y. Erlich, T. Willems, C. Gallo, J. P. Spence, Y. S. Song, G. Poletti, F. Balloux, G. van Driem, P. de Knijff, I. G. Romero, A. R. Jha, D. M. Behar, C. M. Bravi, C. Capelli, T. Hergiv, A. Moreno-Estrada, O. L. Posukh, E. Balanovska, O. Balanovsky, S. Karachanak-Yankova, H. Sahakyan, D. Toncheva, L. Yepiskoposyan, C. Tyler-Smith, Y. Xue, M. S. Abdullah, A. Ruiz-Linares, C. M. Beall, A. Di Rienzo, C. Jeong, E. B. Starikovskaya, E. Metspalu, J. Parik, R. Vilems, B. M. Henn, U. Hodoglugil, R. Mahley, A. Sajantila, G. Stamatoyannopoulos, J. T. S. Wee, R. Khusainova, E. Khusnutdinova, S. Litvinov, G. Ayodo, D. Comas, M. F. Hammer, T. Kivisild, W. Klitz, C. A. Winkler, D. Labuda, M. Bamshad, L. B. Jorde, S. A. Tishkoff, W. S. Watkins, M. Metspalu, S. Dryomov, R. Sukernik, L. Singh, K. Thangaraj, S. Pääbo, J. Kelso, N. Patterson, D. Reich, The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016).
47. G. H. Perry, N. J. Dominy, K. G. Claw, A. S. Lee, H. Fiegler, R. Redon, J. Werner, F. A. Villanea, J. L. Mountain, R. Misra, N. P. Carter, C. Lee, A. C. Stone, Diet and the evolution of human amylase gene copy number variation. *Nat. Genet.* **39**, 1256–1260 (2007).
48. H. F. Dobyns, An appraisal of techniques with a new hemispheric estimate. *Curr. Anthropol.* **7**, 395–416 (1966).
49. J. G. Schraiber, Assessing the relationship of ancient and modern populations. *Genetics* **208**, 383–398 (2018).
50. L. Roberts, Disease and death in the New World. *Science* **246**, 1245–1247 (1989).
51. F. Belinky, N. Nativ, G. Stelzer, S. Zimmerman, T. I. Stein, M. Safran, D. Lancet, PathCards: Multi-source consolidation of human biological pathways. *Database* **2015**, bav006 (2015).
52. L. E. Jensen, Interleukin-36 cytokines may overcome microbial immune evasion strategies that inhibit interleukin-1 family signaling. *Sci. Signal.* **10**, eaa3589 (2017).
53. I. H. Haralambieva, I. G. Ovsyannikova, N. Dhiman, R. B. Kennedy, M. O'Byrne, V. S. Pankratz, R. M. Jacobson, G. A. Poland, 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).
54. J. Dabney, M. Knapp, I. Glocke, M.-T. Gansauge, A. Weihmann, B. Nickel, C. Valdiosera, N. García, S. Pääbo, J.-L. Arsuaga, M. Meyer, 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).
55. M. Meyer, M. Kircher, Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, pdb.prot5448 (2010).
56. J. Lindo, A. Achilli, U. A. Perego, D. Archer, C. Valdiosera, B. Petzelt, J. Mitchell, R. Worl, E. J. Dixon, T. E. Fifield, M. Rasmussen, E. Willerslev, J. S. Cybulski, B. M. Kemp, M. DeGiorgio, R. S. Malhi, 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).
57. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).
58. L. Arbiza, E. Zhong, A. Keinan, NRE: A tool for exploring neutral loci in the human genome. *BMC Bioinformatics* **13**, 301 (2012).
59. X. Wang, Z. Zheng, Y. Cai, T. Chen, C. Li, W. Fu, Y. Jiang, CNVcaller: Highly efficient and widely applicable software for detecting copy number variations in large populations. *GigaScience* **6**, 1–12 (2017).
60. 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).
61. M. Rasmussen, S. L. Anzick, M. R. Waters, P. Skoglund, M. DeGiorgio, T. W. Stafford Jr., S. Rasmussen, I. Moltke, A. Albrechtsen, S. M. Doyle, G. D. Poznik, V. Gudmundsdottir, R. Yadav, A.-S. Malaspina, S. S. White V, M. E. Allentoft, O. E. Cornejo, K. Tambets, A. Eriksson, P. D. Heintzman, M. Karmin, T. S. Korneliussen, D. J. Meltzer, T. L. Pierre, J. Stenderup, L. Saag, V. M. Warmuth, M. C. Lopes, R. S. Malhi, S. Brunak, T. Sicheritz-Ponten, I. Barnes, M. Collins, L. Orlando, F. Balloux, A. Manica, R. Gupta, M. Metspalu, C. D. Bustamante, M. Jakobsson, R. Nielsen, E. Willerslev, The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* **506**, 225–229 (2014).
62. M. Rasmussen, Y. Li, S. Lindgreen, J. S. Pedersen, A. Albrechtsen, I. Moltke, M. Metspalu, E. Metspalu, T. Kivisild, R. Gupta, M. Bertalan, K. Nielsen, M. Z. P. Gilbert, Y. Wang, M. Raghavan, P. F. Campos, H. M. Kamp, A. S. Wilson, A. Gledhill, S. Tridico, M. Bunce, E. D. Lorenzen, J. Binladen, X. Guo, J. Zhao, X. Zhang, H. Zhang, Z. Li, M. Chen, L. Orlando, K. Kristiansen, M. Bak, N. Tommerup, C. Bendixen, T. L. Pierre, B. Grønnow, M. Meldgaard, C. Andreasen, S. A. Fedorova, L. P. Osipova, T. F. G. Higham, C. B. Ramsey, T. v. O. Hansen, F. C. Nielsen, M. H. Crawford, S. Brunak, T. Sicheritz-Pontén, R. Vilems, R. Nielsen, A. Krogh, J. Wang, E. Willerslev, Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762 (2010).
63. M. Rasmussen, M. Sikora, A. Albrechtsen, T. S. Korneliussen, J. V. Moreno-Mayar, G. D. Poznik, C. P. E. Zollikofer, M. S. P. de León, M. E. Allentoft, I. Moltke, H. Jónsson, C. Valdiosera,

- R. S. Malhi, L. Orlando, C. D. Bustamante, T. W. Stafford Jr., D. J. Meltzer, R. Nielsen, E. Willerslev, The ancestry and affiliations of Kennewick Man. *Nature* **523**, 455–458 (2015).
64. J. H. Marcus, J. Novembre, Visualizing the geography of genetic variants. *Bioinformatics* **33**, 594–595 (2017).
65. A. G. Hogg, Q. Hua, P. G. Blackwell, M. Niu, C. E. Buck, T. P. Guilderson, T. J. Heaton, J. G. Palmer, P. J. Reimer, R. W. Reimer, C. S. M. Turney, S. R. H. Zimmerman, SHCal13 Southern Hemisphere calibration, 0–50,000 years cal BP. *Radiocarbon* **55**, 1889–1903 (2016).
66. A. Parnell, Bchron: Radiocarbon dating, age-depth modelling, relative sea level rate estimation, and non-parametric phase modelling (2016); <https://CRAN.R-project.org/package=Bchron>.
67. J. V. Moreno-Mayar, L. Vinner, P. de Barros Damgaard, C. de la Fuente, J. Chan, J. P. Spence, M. E. Allentoft, T. Vimala, F. Racimo, T. Pinotti, S. Rasmussen, M. W. Pedersen, A. Margaryan, M. I. Orbeago, D. Milopotamitaki, M. Wooller, C. Bataille, L. Becerra-Valdivia, D. Chivall, D. Comeskey, T. Devièse, D. K. Grayson, L. George, H. Harry, V. Alexandersen, C. Primeau, T. J. Gergersen, P. Endzweig, J. Flansdon, D. Jenkins, T. Stafford Jr., J. Cybulski, C. Vullo, F. Morello, M. Vilar, S. Wells, K. Gregersen, K. Lykke Hansen, N. Lynnerup, M. M. Lahr, K. Kjær, A. Strauss, M. Alfonso-Durruty, A. Salas, H. Schroeder, T. Higham, R. S. Malhi, J. T. Rasic, L. Souza, F. R. Santos, A.-S. Malaspina, M. Sikora, R. Nielsen, Y. S. Song, D. J. Meltzer, E. Willerslev, Early human dispersals within the Americas. *Science* **10.1126/science.aav2621** (2018).
68. C. Posth, N. Nakatsuka, I. Lazaridis, P. Skoglund, S. Mallick, T. C. Lamnidis, N. Rohland, K. Nägele, N. Adamski, E. Bertolini, N. Broomandkhoshbacht, A. Cooper, B. J. Culleton, T. Ferraz, M. Ferry, A. Furtwängler, W. Haak, K. Harkins, T. K. Harper, T. Hünemeier, A. M. Lawson, B. Llamas, M. Michel, E. Nelson, J. Oppenheimer, N. Patterson, S. Schiffels, J. Sedig, K. Stewardson, S. Talamo, C.-C. Wang, J.-J. Hublin, M. Hubbe, K. Harvati, A. N. Delaunay, J. Beier, M. Francken, P. Kaulicke, H. Reyes-Centeno, K. Rademacher, W. R. Trask, M. Robinson, S. M. Gutierrez, K. M. Prüfer, D. C. Salazar-García, E. Nunes Chim, L. Müller Plumm, M. Lopes Alves, A. Liryo, M. Ingles, R. E. Oliveira, D. V. Bernardo, A. Barioni, V. Wesolowski, N. A. Scheifler, M. A. Rivera, C. R. Plens, P. G. Messineo, L. Figuti, D. Corach, C. Scabuzzo, S. Eggers, P. DeBlasis, M. Reindel, C. Méndez, G. Politis, E. Tomasto-Cagigao, D. J. Kennett, A. Strauss, L. Fehren-Schmitz, J. Krause, D. Reich, Reconstructing the Deep Population History of Central and South America. *Cell* **10.1016/j.cell.2018.10.027** (2018).
69. C. O. Lovejoy, Dental wear in the Libben population: Its functional pattern and role in the determination of adult skeletal age at death. *Am. J. Phys. Anthropol.* **68**, 47–56 (1985).
70. M. Stuiver, H. A. Polach, Reporting of  $^{14}\text{C}$  data. *Radiocarbon* **19**, 355–363 (1977).
71. M. Schubert, S. Lindgreen, L. Orlando, AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC. Res. Notes* **9**, 88 (2016).
72. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
73. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin; 1000 Genome Project Data Processing Subgroup, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
74. 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).
75. Q. Fu, H. Li, P. Moorjani, F. Jay, S. M. Slepchenko, A. A. Bondarev, P. L. F. Johnson, A. Aximu-Petri, K. Prüfer, C. de Filippo, M. Meyer, N. Zwyns, D. C. Salazar-García, Y. V. Kuzmin, S. G. Keates, P. A. Kosintsev, D. I. Razhev, M. P. Richards, N. V. Peristov, M. Lachmann, K. Douka, T. F. G. Higham, M. Slatkin, J.-J. Hublin, D. Reich, J. Kelso, T. B. Viola, S. Pääbo, Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445–449 (2014).
76. K. Prüfer, F. Racimo, N. Patterson, F. Jay, S. Sankararaman, S. Sawyer, A. Heinze, G. Renaud, P. H. Sudmant, C. de Filippo, H. Li, S. Mallick, M. Dannemann, Q. Fu, M. Kircher, M. Kuhlwillm, M. Lachmann, M. Meyer, M. Ongyerth, M. Siebauer, C. Theunert, A. Tandon, P. Moorjani, J. Pickrell, J. C. Mullikin, S. H. Vohr, R. E. Green, I. Hellmann, P. L. F. Johnson, H. Blanche, H. Cann, J. O. Kitzman, J. Shendure, E. E. Eichler, E. S. Lein, T. E. Bakken, L. V. Golovanova, V. B. Doronichev, M. V. Shunkov, A. P. Derevianko, B. Viola, M. Slatkin, D. Reich, J. Kelso, S. Pääbo, The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
77. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytzky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
78. P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin; 1000 Genomes Project Analysis Group, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
79. 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).
80. 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* **18**, 373 (2017).
81. E. Tamm, T. Kivisild, M. Reidla, M. Metspalu, D. G. Smith, C. J. Mulligan, C. M. Bravi, O. Rickards, C. Martinez-Labarga, E. K. Khushnudinova, S. A. Fedorova, M. V. Golubenko, V. A. Stepanov, M. A. Gubina, S. I. Zhadanov, L. P. Ossipova, L. Damba, M. I. Voevoda, J. E. Dipierri, R. Villemes, R. S. Malhi, Beringian standstill and spread of Native American founders. *PLOS ONE* **2**, e829 (2007).
82. G. Renaud, V. Slon, A. T. Duggan, J. Kelso, Schmutzi: Estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 224 (2015).
83. A. L. Price, N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, D. Reich, Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
84. M. Sikora, M. L. Carpenter, A. Moreno-Estrada, B. M. Henn, P. A. Underhill, F. Sánchez-Quinto, I. Zara, M. Pitalis, C. Sidore, F. Busonero, A. Maschio, A. Angius, C. Jones, J. Mendoza-Revilla, G. Nekhrizov, D. Dimitrova, N. Theodosios, T. T. Harkins, A. Keller, F. Maixner, A. Zink, G. Abecasis, S. Sanna, F. Cucca, C. D. Bustamante, 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).
85. 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. *Bioinformatics* **32**, 2817–2823 (2016).
86. B. S. Weir, C. C. Cockerham, Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
87. S. Gravel, F. Zakharia, A. Moreno-Estrada, J. K. Byrnes, M. Muzzio, J. L. Rodriguez-Flores, E. E. Kenny, C. R. Gignoux, B. K. Maples, W. Guiblet, J. Dutil, M. Via, K. Sandoval, G. Bedoya; 1000 Genomes Project, T. K. Oleksyk, A. Ruiz-Linares, E. G. Burchard, J. C. Martinez-Cruzado, C. D. Bustamante, Reconstructing native american migrations from whole-genome and whole-exome data. *PLOS Genet.* **9**, e1004023 (2013).
88. M. Livi-Bacci, The depopulation of Hispanic America after the conquest. *Pop. Dev. Rev.* **32**, 199–232 (2006).

**Acknowledgments:** We would like to thank J. Blangero and S. Blangero for contribution with the Aymara samples. We would like to thank C. Jeong and S. Nakagome for helpful discussions throughout the project. We would like to thank M. DeGiorgio for helping with the  $f_3$  statistic analysis. Field support for R.H. was provided by the Collasuyo Archaeological Research Institute, C. Justo Chavez, V. Incacona Huaraya, M. Incacona Huaraya, N. Condori Flores, A. Pilco Quispe, D. Pilco Incacona, D. Pilco Incacona, L. M. Pilco Incacona, L. Hayes, and the community of Mulla Fasiri, Peru. Archeological data recovery at SMP and international export of artifacts, including those from Jiskairumoko, Kaillachuro, and the Rio Uncallane sites, were carried out under Peruvian Ministry of Culture permit nos. 064-2013-DGPA-VMPCIC/MC and 138-2015-VMPCIC/MC. Multiple permits were issued by the Peruvian Instituto Nacional de Cultura for research at Jiskairumoko, Kaillachuro, and the Rio Uncallane sites from 1994 to 2002. See Moreno-Mayar *et al.* (67) and Posth *et al.* (68) for related analyses of ancient DNA samples from the Americas. **Funding:** This work was supported in part by NIH grant R01HL119577 and the National Science Foundation grants BCS-1528698 (awarded to M.A.I., C.W., and A.D.R.) and BSC-9221724 (awarded to C.B.). J.L. was funded by a University of Chicago Provost's Postdoctoral Scholarship. Support for archeological excavation and artifact analysis was provided to R.H. by the National Science Foundation (BCS-1311626), the American Philosophical Society, and the University of Arizona. Survey and data recovery at the Rio Uncallane sites was supported by grants to M.A.I. from the National Geographic Society (5245-94) and the H. John Heinz III Charitable Trust. Excavations at Jiskairumoko and Kaillachuro were supported by grants to M.A.I. from the National Science Foundation (SBR-9816313 and SBR-9978006). The Huilliche-Pehuenche datasets were funded by the Chilean National Council of Science and Technology (CONICYT) grants USA2013-0015 and FONDEF D1011007. **Author contributions:** M.A.I., R.H., C.B., J.T.W., and C.V.L. provided samples for the study. J.L., A.D.R., C.W., and C.H. contributed to the experimental design. M.A.P., J.L., A.D.R., J.N., and D.W. analyzed data. M.M. and R.A.V. generated the Huilliche-Pehuenche datasets. J.L., A.D.R., and M.A.I. wrote the initial draft of the manuscript. C.W. and R.H. contributed to the writing of the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors. 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 R.A.V. at the Universidad de Chile.

Submitted 15 June 2018  
Accepted 12 October 2018  
Published 8 November 2018  
10.1126/sciadv.aau4921

**Citation:** Lindo, R. Haas, C. Hofman, M. Apatá, M. Moraga, R. A. Verdugo, J. T. Watson, C. Viviano Llave, D. Witonsky, C. Beall, C. Warinner, J. Novembre, M. Aldenderfer, A. Di Rienzo, The genetic prehistory of the Andean highlands 7000 years BP through European contact. *Sci. Adv.* **4**, eau4921 (2018).

## The genetic prehistory of the Andean highlands 7000 years BP though European contact

John Lindo, Randall Haas, Courtney Hofman, Mario Apata, Mauricio Moraga, Ricardo A. Verdugo, James T. Watson, Carlos Viviano Llave, David Witonsky, Cynthia Beall, Christina Warinner, John Novembre, Mark Aldenderfer and Anna Di Rienzo

*Sci Adv* 4 (11), eaau4921.  
DOI: 10.1126/sciadv.aau4921

### ARTICLE TOOLS

<http://advances.sciencemag.org/content/4/11/eaau4921>

### SUPPLEMENTARY MATERIALS

<http://advances.sciencemag.org/content/suppl/2018/11/05/4.11.eaau4921.DC1>

### RELATED CONTENT

<http://science.sciencemag.org/content/sci/362/6415/627.full>

### REFERENCES

This article cites 78 articles, 19 of which you can access for free  
<http://advances.sciencemag.org/content/4/11/eaau4921#BIBL>

### PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)