Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/4/11/eaau4921/DC1

Supplementary Materials for

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*, Anna Di Rienzo*

*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.)

Published 8 November 2018, *Sci. Adv.* **4**, eaau4921 (2018)
DOI: 10.1126/sciadv.aau4921

This PDF file includes:

Supplementary Materials and Methods

Archaeological Context. The archaeological samples examined in this study include skeletal remains from seven individuals from five archaeological sites in the Rio Ilave Basin, Peru—16.2º south latitude, 69.7º west longitude, 3800m above sea level. The samples come from three distinct contexts—the Rio Uncallane cemeteries, the site of Kaillachuro, and the site of Soro Mik'aya Patjxa. We describe the contexts here along with radiocarbon determinations.

Rio Uncallane tombs. These sites are found in a mesa-like geological formation along the south side of the Rio Uncallane, an east-west flowing tributary of the Rio Ilave to the west of Kaillachuro and Soro Mik'aya Patjxa. A total of eight tombs, each containing large numbers of human remains, were placed within these crevices. Most of the cemeteries were looted at some time in the past, and human remains were found piled and scattered within the crevices. None of the remains have been examined by an osteologist. Although ceramic sherds were found in many of the tombs, none were diagnostic to a cultural period. The skeletal samples examined in this study are from three tombs: 95-096, 95-107, and 95-116. Individuals 16 and 17 from 95-096 are included in this study, but only Individual 17 was directly dated by radiocarbon and produced a calibrated date range of 1783-1616 B.P. (SHCAL13; Table S1). However, the sample had marginal collagen yield (0.9%) and gave a very high atomic C:N ratio (>4) making the age determination unreliable. Individuals 224 and 229 from 95-107 were also dated. Radiocarbon assays produced calibrated date ranges of 1529-1426 and 1722-1611 B.P., respectively. Individual 80 from 95-116 produced a calibrated date range of 1825-1730 B.P.

Kaillachuro. The site of Kaillachuro (Ilave 95-203) is a large open-air site located on an alluvial terrace overlooking the Rio Grande, a tributary of the Rio Ilave. The site covers approximately 27,000 m^2 and consists of a dense scatter of lithics and nine small, ovoid burial mounds. Numerous secondary burials were discovered during exploratory trenching of two of the mounds. Individual 1, excavated from Feature 7, Trench 2, was examined for this study. Burial 1 consists of the partially complete skeleton of an adult (18-20) female. Age at death is estimated between 18 and 20 years based on mild dental wear (*69*) and limited suture closure (Buikstra and Ubelaker, 1994). Primary sex characteristics such as a wide sciatic notch and secondary sex characteristics of gracile glabella, mastoid process, and nuchal crest are indicative of a female

individual. Preservation is poor. All regions are represented but many elements are missing and damaged with long bones exfoliating. The radiocarbon sample gave zero yield of ultra-filtered collagen and could not be measured.

Soro Mik'aya Patjxa. The site of Soro Mik'aya Patjxa (Ilave 95-259) is an open-air site located on an alluvial terrace near the confluence of the Ilave and Huenque Rivers (*14*). The site consists primarily of a lithic scatter covering approximately 2800 m^2. Excavations resulted in the discovery of 16 distinct cultural pit features and as many human burials at the site. Nineteen radiocarbon assays on associated charcoal and human bone bracket the primary site occupation between 8000 and 6500 cal. B.P. (*5*). Individual 5, excavated from feature 3 located in area 1, excavation unit 25, was examined for this study. The remains consist of a largely complete older adult male. Age-at-death is estimated at approximately 50-55 years based on extremely heavy dental wear (*65*) a and several closed sutures. Secondary sex characteristics of the cranium including a robust nuchal crest and mastoid process and prominent glabella and mental eminence are indicative of a male individual. A radiocarbon assay on collagen from the individual's right rib produced a calibrated date range of 6883-6762 B.P. (SHCAL13).

Radiocarbon analysis. Bone samples from six of the seven individuals were submitted to the University of California, Irvine Keck Carbon Cycle Accelerator Mass Spectrometry Lab (UCI). The results are presented in Table S1. All except the Kaillachuro sample yielded sufficient collagen for a radiometric determination.

UCI reports the following methodological details:

Radiocarbon concentrations are given as fractions of the Modern standard, D14C, and conventional radiocarbon age, following the conventions of Stuiver and Polach (*70*). *Sample preparation backgrounds have been subtracted, based on measurements of 14C-free mammoth and whale bone. All results have been corrected for isotopic fractionation according to the conventions of Stuiver and Polach* (*70*), *with d13C values measured on prepared graphite using the AMS spectrometer. These can differ from d13C of the original material, if fractionation occurred during sample graphitization or the AMS measurement, and are not shown. These*

samples were sonicated in 2:1 chloroform/methanol, methanol and MQ water to remove residual lipids. They were then decalcified in 0.5N HCl, gelatinized at 60°C and pH 2, and ultrafiltered to select a high molecular wt fraction (>30kDa).

Read processing. Adapter sequences and leading/trailing stretches of Ns were trimmed using AdapterRemoval-2 (*71*). Trimmed reads were mapped to the human reference genome build hg19 utilizing bwa-0.6.2 (*72*) and filtered for mapping quality 30 and sorted using samtools (*73*). Data was merged to library level and duplicates removed using Picard MarkDuplicates (http://picard.sourceforge.net) for modern samples and adna-ldup (https://github.com/DReichLab/adna) for ancient samples. The bam files from the enrichment and shotgun sequencing for SMP5 were merged to attain better coverage.

DNA damage patterns. DNA damage (type I and type II) was assessed by comparing T→C/G→A and C→T/A →G transitions, respectively using MapDamage 2.0 (*74*). A specific pattern of DNA damage has been identified in other ancient DNA studies (*75*, *76*). These studies show a pattern of increased type II DNA damage at the beginning and end of degraded DNA fragments. The MapDamage results show signatures of DNA damage, which suggests that the ancient sequences originate from ancient DNA templates and not modern contaminants (Fig. S2).

Alignment and Genotyping. Genotyping of the modern Aymara and Rio Uncallane were conducted via GATK's UnifiedGenotyper tool (*77*), using only sequence data with a base quality greater than 20 and minimum depth of 4. For frequency-based analyses, the ancient samples were further filtered for potentially false calls due to DNA damage, by requiring a minimum of 4 *alleles* to call C -> T or G -> A SNPs. All other analyses removed these transitions.

For the low coverage samples, K1 and SMP5, genotypes were obtained using Samtools pileup, with a minimum base quality of 20. For sites with greater than 1 read, a base call was randomly selected and duplicated to form a homozygous diploid genotype, which were then turned into PED format files for merging. Transitions, which could be due to DNA damage, were removed.

For analyses requiring frequency estimation (PBS and Fastsimcaol2), ANGSD (*57*) was used, using a minimum base quality of 20. For the ancient Rio Uncallane, a minimum of 5 individuals

were used and for the modern Aymara a minimum of 10 individuals were used. None of the Rio Uncallane were determined to be related, using the Ajk statistic, as implemented in vcftools (*78*).

Imputation. Genotype imputation was performed for the set of 39 Huilliche-Pehuenche samples using IMPUTE2 v2.3.2(*79*). We used the 1KGP phase 3 reference data set, downloadable from https://mathgen.stats.ox.ac.uk/impute/1000GP_Phase3.html, merged with the phased genotype calls for the 24 sequenced Aymara as an imputation reference panel by merging them with "-merge_ref_panels" flag in IMPUTE2. For other parameters, we used default values set by the program. Following imputation, genotypes with posterior probability ≥ 0.9 were accepted. Genotypes were assumed to be missing if none of three possible genotypes reached posterior probability threshold of 0.9. Then, we conducted an additional round of quality control by removing SNPs with missing rate higher than 0.05 or HWE *p*-value smaller than 10^{-6}.

MtDNA (Ancient Samples). The samples were aligned to the revised Cambridge Reference Sequence (rCRS), using BWA (bwa aln -l 16500 -t 32 -n 0.01). The program SG-Adviser (*80*) was used to call the haplogroup directly from the sequence alignment file, reporting the highest ranking haplogroup in Table 1. Only reads with mapping quality above 30 and sites with base quality above 30 were used to call a consensus and identify the mtDNA haplogroup. All haplogroups identified in the ancient samples are commonly found in modern day indigenous groups from the Americas (*81*).

Contamination Estimates. Contamination estimates using the mitochondrial genome were run on all seven ancient samples using the Schmutzi program described by Renaud et al. (*82*). The method jointly estimates present day human contamination and reconstructs the endogenous mitochondrial genome by considering both deamination patterns and fragment length distributions. Estimates are shown on Table S2.

Principal component analysis. Principal component analysis was performed on a subset of individuals from the Raghavan et al. (*16*) data set, which excluded African populations, using EIGENSOFT v. 7.2.1 (*83*), and included 199,285 sites, with transitions removed. Each ancient individual had the following overlap with the modern dataset: IL2 (119,882); IL3 (119,709); IL4 (92,969); IL5 (113,246), IL7 (109,375), K1 (26,732), SMP5 (15,377) Native American populations in the data set were masked for non-native ancestry (*16*). Anzick-1(*61*), Kennewick

Man (*63*), Il2, IL3, IL4, IL5, IL7, K1, SMP5, and Saqqaq (*62*) were projected onto the components inferred from these sets of contemporary individuals by using the 'lsqproject' option of smartpca. Heterozygote genotypes were converted to homozygous prior to the analysis, by randomly sampling reads if positions were covered by multiple reads in each ancient individual. To prevent DNA damage from skewing the results, C→T and G→A transitions were removed from both the contemporary and ancient individuals.

***ADMIXTURE* analysis.** We performed model-based clustering analysis using the maximum-likelihood approach implemented in *ADMIXTURE* (*15*). We ran *ADMIXTURE* on a reference panel of 261 individuals from the Raghavan et al. (*16*), only including populations from North America, Central America, South America, and Siberia, and once again masked for non-native ancestry. The analysis included 199,285 sites. The overlap for each ancient individual was the same as in the PCA analysis. Native American populations in the data set were masked for non-native ancestry.

Since the ancient individuals had varying overlap with the contemporary reference panel, we projected the ancient individuals onto the ancestral cluster allele frequencies inferred from the contemporary individuals with the same method described in (*84*) and implemented in Admixture 1.3(*15*). This method also prevents the ancient genotypes from affecting the clustering solutions of the contemporary individuals. The low coverage ancient genotypes were called in the same manner as in the principal coordinate analysis. C→T and G→A transitions were removed from both the contemporary and ancient individuals.

We assumed K=3 to K=15 ancestral clusters. For each value of K, ten replicate runs were performed, and the run with the greatest likelihood was selected for further analysis. Clusters K=2 through K=8 are presented in Fig. S6. The admixture graphs were generated with PONG (*85*).

***TreeMix* analysis.** We started with the identical filtered dataset of called genotypes described in the *SI Appendix*. *TreeMix* (*18*) was applied to the dataset to generate maximum likelihood trees and admixture graphs from allele frequency data. The Yoruban (YRI) from the Simons Genome Project (*46*) was used to root the tree (with the *–root* option). We accounted for linkage disequilibrium by grouping M adjacent sites (with the *–k* option), and we chose M such that a

dataset with L sites will have approximately $L/M \approx 20{,}000$ independent sites. At the end of the analysis (i.e., number of migrations) we performed a global rearrangement (with the *–global* option). We considered admixture scenarios with $m = 0$ and $m = 1$ migration events. Each migration scenario was run with 100 replicates, and the replicate with the highest likelihood was chosen to represent the maximum likelihood tree or graph for the given migration scenario. Figures 3B and 3C display the results for the maximum likelihood tree with no admixture ($m = 0$) events. Figure S5 shows the maximum likelihood tree with one migration event ($m = 1$). Overlapping sites: Rio Uncallane, 1,812,145; K1, 45,455; SMP5, 27,751.

***f*$_3$ Statistics.** To test the genetic affinity of the ancient individuals with global populations, we preformed outgroup f_3 statistic, using the method outlined by Patterson et al. (2012). For each ancient sample, reads with mapping quality below 30 and nucleotides with quality below 20 were not considered. A single allele was sampled from each ancient sample and intersected with the Raghavan et al. (*16*) SNP chip data set that included 2,081 contemporary individuals sampled from across the world. Sites in Native Americans were masked for European ancestry using the exact masking of Raghavan et al. (*16*). We computed outgroup f_3 statistics of the form f_3(A, M; Yoruba), where A is one of the four ancient individuals, M is a contemporary non-African population from the Raghavan et al. (*16*) SNP chip data set, and Yoruba an African population representing the outgroup.

Selection scans. To detect regions under positive selection in both the ancient Rio Uncallane and the modern Aymara, the population branch statistics (PBS) (*24*) was utilized. It employs a set of three populations (call them *X*, *Y*, and *Z*), and assumes that they have the rooted relationship ((*X*,*Y*),*Z*).

In this study, we are concerned with two scenarios. The first is setup where X is the ancient Rio Uncallane, Y is the lowland Huilliche-Pehuenche, and Z is the Han Chinese (CHB from the 1,000 Genomes Project). We therefore are interested in computing

$$PBS_{PRH} = \frac{T_{RU,HP} + T_{RU,CHB} - T_{HP,CHB}}{2}$$

For the second scenario, X is the modern Aymara, Y is the ancient Rio Uncallane, and Z is the Han Chinese (CHB from the 1,000 Genomes Project). We therefore are interested in computing

$$PBS_{PRH} = \frac{T_{Aymara,RU} + T_{Aymara,CHB} - T_{RU,CHB}}{2}$$

Fst that was estimated on a per SNP basis using the Weir and Cockerham (*86*) calculation, before proceeding with the population branch statistic. The scan utilized 103 Han Chinese from the 1000 genomes project (*14*). The top signals for each scan are shown on Table S5.

Demographic history model

The commands used in FastSimCoal2 (v. 2.6.0.3) (*20*) to infer the model were as follows: fsc2603 –t template file –e estimation file -n100000 -N100000 -d –l 10 -L 40 –m –M 0.001 –c0. The inferred parameters and confidence intervals from 100 bootstrap replicates are listed on Table S6. We set value for N_A/N_0=100 to facilitate optimization as in (*87*). The parameter file for the best likelihood after 100 optimizations is as follows:

```
//Parameters for the coalescence simulation program
4 samples to simulate: 200 to 100,000
//Population effective sizes: AYM, ILA, HUI, MIX
1196
1519
249
210
//Samples sizes and samples age
25
5 60
39
3
//Growth rates: negative growth implies population expansion
0
```

```
0
0
0
//Number of migration matrices: 0 implies no migration between demes
0
//historical event: time, source, sink, migrants, new deme size, new growth rate, migration matrix index
8 historical events
17 0 0 0 1.27006889e+00 0 0
13 2 2 0 2.40943971e+01 0 0
14 3 3 0 1.91003706e+01 0 0
60 0 1 1 1 0 0
350 1 2 1 4.46224900e+01 0 0
591 2 3 1 1.04041040e-01 0 0
800 3 3 0 4.93858131e+00 0 0
830 3 3 0 2.02487301e-01 0 0
//Number of independent loci [chromosome]
1 0
//Per chromosome: Number of contiguous linkage Block: a block is a set of contiguous loci
1
//per Block:data type, number of loci, per generation recombination and mutation rates and optional parameters
FREQ 1 0 1.25e-8
```

We also ran two additional models to assess if the highlander collapse could still be inferred without the ancient highland population. In the first scenario, the ancient samples (Rio Uncallane) were removed from the model, allowing the highland collapse to be informed by the modern high-altitude population (the Aymara). We find that the severity of the collapse does increase, however, it does not approach the levels seen in the lowland populations. In this model, the Aymara experience a 37.4% decrease in population size {95% CI(0.44-0.30)}, which is an increase from the original model's inference of 27%. For the next model, we did not merge the Rio Uncallane into the modern Aymara and let them split off into their own modern population. This increased the Aymara collapse to 74% {95% CI(0.71-0.77)}, however, this may also be a consequence of splitting the high-altitude populations. We also find that the difference between the low and highland collpase remains significant (P=0.0163, Chi-squared test).

We were also concerned that an imputation bias may have been introduced in the samples from Chile. We therefore checked the results against a summary statistic, average heterozygosity, to see if model agreed with the result. We first filtered the data sets to only include SNPs found in the Human Origins Array for the SAN (Affymetrix) so that we would have a constantly ascertained set of SNPs. We then calculated the heterozygosity rate for each population using Plink and the formula

$$H_r = \frac{Number\ of\ non-missing\ genotypes\ -\ Observed\ number\ of\ homozygotes}{Number\ of\ non-missing\ genotypes}$$

The results, shown on Figure S3, show a similar pattern of reduced effective population size (with heterozygosity as a proxy) as seen with the demographic model, which utilized both sequence and imputed data.

Finally, we calculated the reduction of average heterozygosity between the Rio Uncallane and Aymara. In order to allow DNA to be simulated right before the bottleneck, we subtracted 17 generations (the time at which the bottleneck is inferred by the model) from all historical events, thereby simulating data right before the population collapse. We then calculated average heterozygosity on the simulated DNA sequence and compared it to the modern Aymara population. We found that the simulated data right before the bottleneck to have an average Heterozygosity of 0.4282, while the Aymara show an average heterozygosity of 0.3287.

Figures and Tables

Fig. S1. Location of ancient and modern samples.

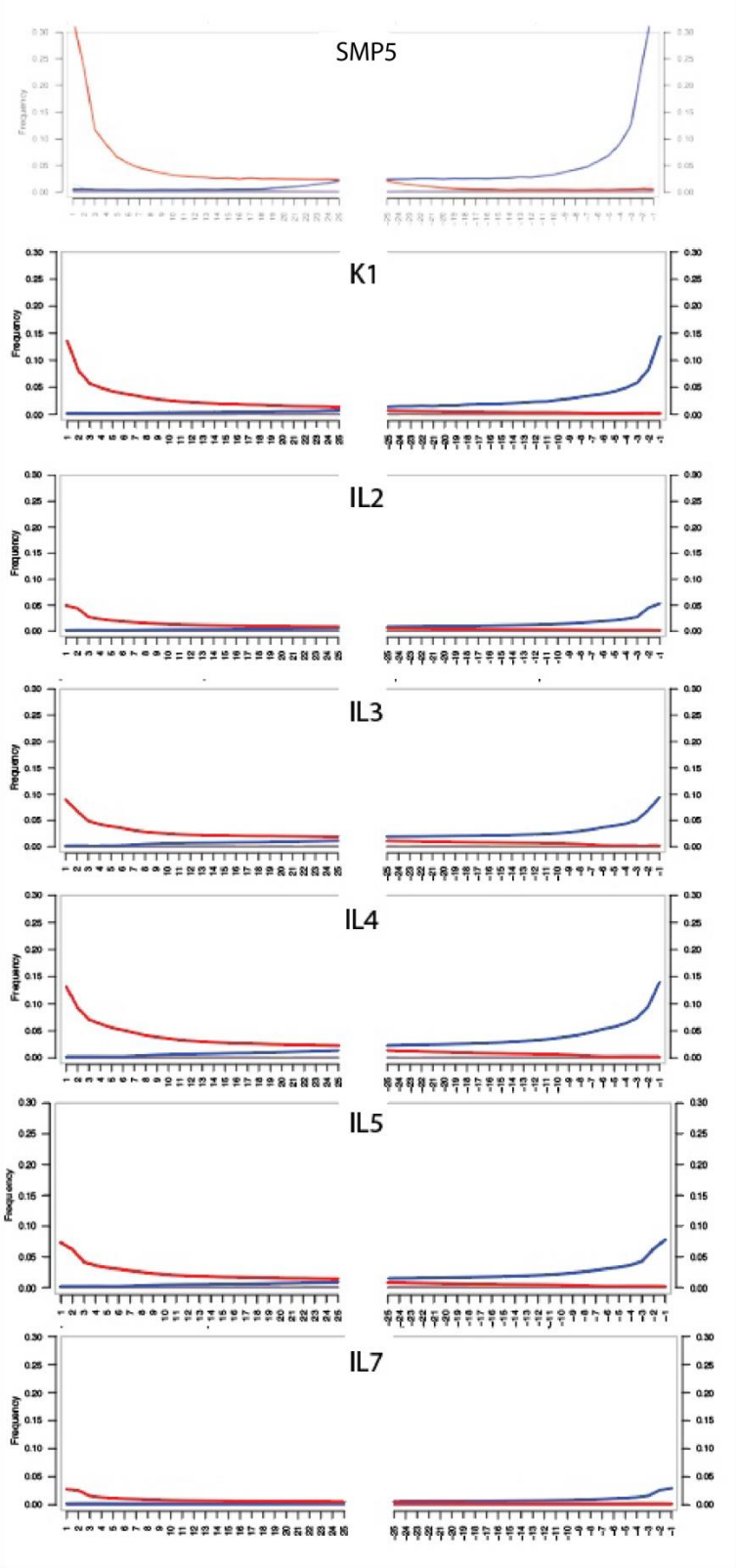

Fig. S2. DNA damage patterns for the ancient individuals. Random subset of all mapped reads for each ancient individual. The mismatch frequency is relative to the reference as function of read position, C to T in red and G to A in blue.

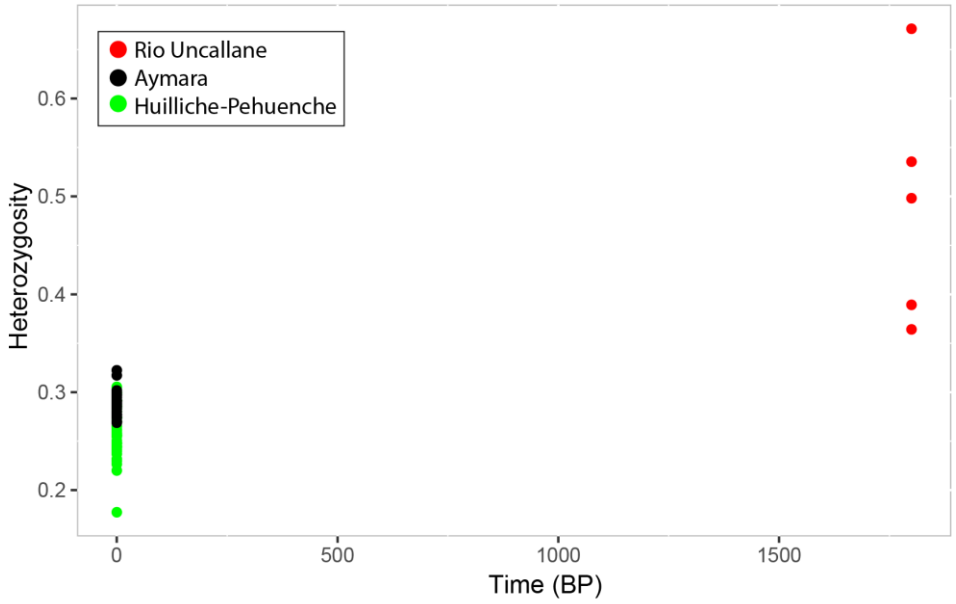

Fig. S3. Heterozygosity rate calculated with overlapping ascertained SNPs from the SAN panel of the Affymetrix Human Origin Array.

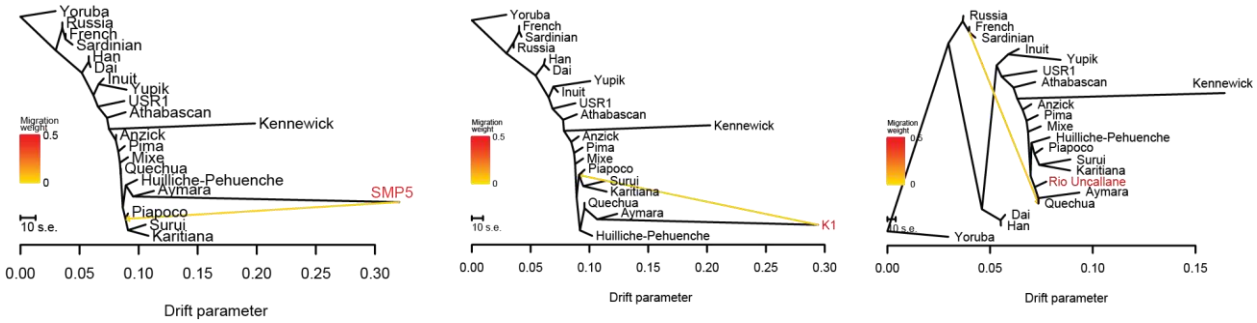

Fig. S4. Maximum likelihood trees with $m = 1$ migration events. Trees were generated by *TreeMix*(*18*) using whole-genome sequencing data from the Simons Genome Project. Trees for individuals K1 (B) and SMP5 (C) were calculated separately due to low overlapping sites.

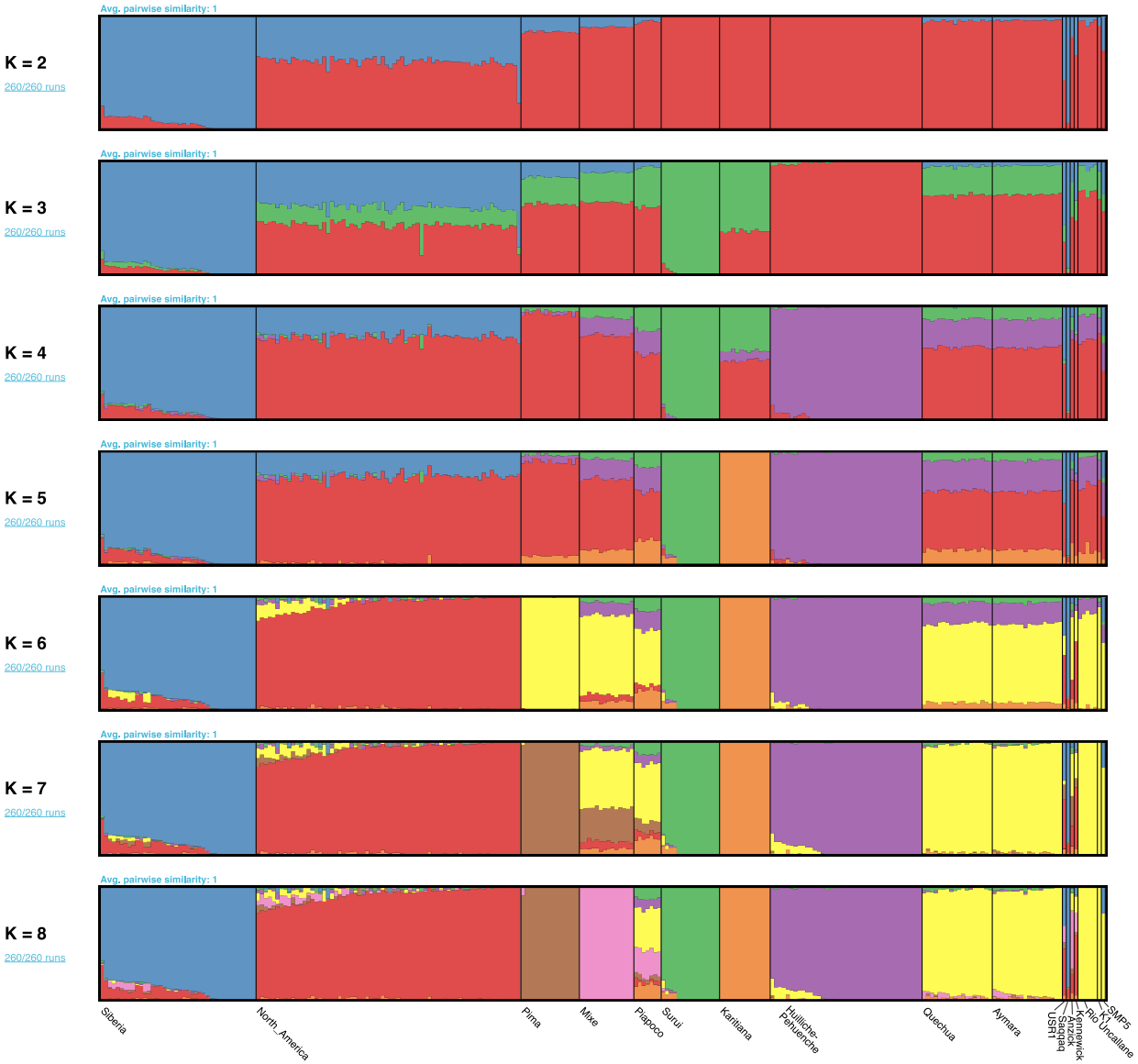

Fig. S5. Cluster analysis generated by ADMIXTURE. Set includes indigenous populations from the Americas, Siberia, and, USR1, Anzick-1, Kennewick, Saqqaq, *Shuká Káa*, Rio Uncallane, K1, and SMP5 samples. The number of displayed clusters is *K*=2 through *K*=8.

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

UCIAMS ID	site	burial	element	$d^{13}C$ (‰)	Fraction modern	$d^{14}C$ (‰)	^{14}C age	>30kDa collagen	C:N (wt%)	C:N (atomic)
191886	SMP	5	right rib	-18.7±0.1	0.4720±0.0012	-528±1.2	6030±20	1.9	2.81	3.27
191891	Ilave 95-096	17	LM^3	-18.8±0.1	0.7961±0.0017	-203.9±1.7	1830±20	1.9	2.78	3.25
191892	Ilave 95-107	249	LM^2	-18.9±0.1	0.7987±0.0019	-201.3±1.9	1805±20	4.4	2.74	3.20
191893	Ilave 95-116	80	RM^3	-18.3±0.1	0.7910±0.0020	-209±2.0	1885±20	3.1	2.74	3.20
191894	Ilave 95-107	224	LM^3	-19.8±0.1	0.8161±0.0018	-183.9±1.8	1630±20	1.5	2.75	3.21

Table S2. Contamination estimates.

Sample	Contamination Estimate	95% Confidence Interval
IL1	0.01	0.00-0.02
IL2	0.02	0.01-0.03
IL4	0.02	0.01-0.03
IL5	0.01	0.00-0.02
IL7	0.04	0.03-0.04
K1	0.01	0.00-0.02
SMP5	0.01	0.00-0.02

Table S3. Modern Aymara sequencing results.

ID	Project_ID	Bases	Coverage
ADR94072	D4-1	12921458861	4.51
ADR94211	D4-2	15226613680	5.31
ADR94238	D4-3	15574140335	5.43
ADR94339	D4-4	15391105405	5.37
ADR94362	D4-5	13747444716	4.79
ADR94215	D4-6	11199349685	3.91
ADR94047	D4-7	14055942150	4.9
ADR94342	D4-8	14419140958	5.03
ADR94318	D4-9	11383750776	3.97
ADR94337	D4-10	13414056652	4.68
ADR94095	D4-11	15563727958	5.43

ADR94351	D4-12	16230507592	5.66
ADR94254	ADR-AY2-254	4231020849	1.48
ADR94364	ADR-AY2-364	7781638308	2.71
ADR94324	ADR-AY2-324	6199879420	2.16
ADR94398	ADR-AY2-398	9622807235	3.36
ADR94325	ADR-AY2-325	6213819407	2.17
ADR94396	ADR-AY2-396	13230773826	4.61
ADR94368	ADR-AY2-368	19009035699	6.63
ADR94071	ADR-AY2-72	8769043847	3.06
ADR94463	ADR-AY2-463	22947491326	8
ADR94573	ADR-AY2-573	13711804508	4.78
ADR94512	ADR-AY2-512	10218751363	3.56
ADR94075	ADR-AY2-75	8985314264	3.13
ADR94233	ADR94233	29831494317	31.2

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

PBS: Ancient Rio Uncallane (High Altitude vs. Low Altitude)								
rs ID	PBS Rank	Gene	Rio Uncallane *f* (High Alt)	Huilliche-Pehuenche *f* (Low Alt)	Aymara (High Alt)	Histone mark: Promotor	Histone mark: Enhancer	Regulatory motifs altered
rs149112613	3	*DST*	G: 0.9	G: 0.0	0.68	Spleen	Brain	4
rs77768615	Top 1%	*MGAM (Promoter Region)*	C: 0.9	C: 0.5	0.71	Duodenum	Intestine, Stomach	14
PBS: Modern Aymara (Post-contact vs. Pre-contact)								
rs ID	PBS Rank	Gene	Aymara *f* (Post-contact)	Rio Uncallane *f* (Pre-contact)	eQTL	Histone mark: Promotor	Histone mark: Enhancer	Regu motifs altered
rs1547616	1	*CD83**	A: 0.97	A: 0.2			Monocytes, Primary B Cells, NK Cells	GR, RFX5, Zbtb12
rs1547617	Top 1%	*CD83**	T:0.97	T:0.5		Blood	Blood, Liver, GI	HEY1,TATA
rs1117797	32	*IL-36R*	C:0.83	C:0.1	*IL18RAP*		Liver	Eomes,Evi-1,TBX5

*Indicates nearest genes. Encode data retrieved from HaploReg v. 4.1 (*38*).

Table S5. Selection scans top signals.

Rank	SNP	PBS	Gene	Function
PBS: Ancient Rio Uncallane (High Altitude vs. Low Altitude)				
1	rs62650368	3.326506021	*MGAM*	Enzyme that plays a role in the final steps of digestion of starch.
2	rs149112613	3.230001259	*DST*	Cytoskeletal linker protein.
3	rs2486896	3.164232142	*NFIB**	Nuclear Factor I B.
4	rs2336845	3.14271949	*MGAM*	Enzyme that plays a role in the final steps of digestion of starch.
5	rs35894927	3.058287389	*MIR548I4*	Non-coding RNAs that are involved in post-transcriptional regulation of gene expression in multicellular organisms by affecting both the stability and translation of mRNAs.
6	rs7577238	3.035163049	*EXOC6B*	Evolutionarily conserved exocyst.
7	rs2336845	2.996506849	*LOC102723330**	Affiliated with the ncRNA class.
8	rs72847967	2.994856295	*KRTAP5-5 **	Related to the keratinization pathway.
9	rs115240547	2.967045613	*TEAD4*	Transcription factor which plays a key role in the Hippo signaling pathway, a pathway involved in organ size control and tumor suppression by restricting proliferation and promoting apoptosis.
10	rs6601815	2.95765172	*LOC101927964**	Affiliated with the ncRNA class.
PBS: Modern Aymara (Post-contact vs. Pre-contact)				
1	rs1547616	1.908567097	CD83*	Member of the immunoglobulin superfamily of receptors. May be involved in the regulation of antigen presentation. A soluble form of this protein can bind to dendritic cells and inhibit their maturation.
2	*rs13093086*	1.667211553	FOXP1	Forkhead box transcription factor that play important roles in the regulation of tissue and cell type-specific gene transcription during both development and adulthood.
3	rs8018314	1.645954736	RPS29*	Ribosomal protein that is a component of the 40S subunit and a member of the S14P family of ribosomal proteins.
4	*rs935200*	1.611793928	TBC1D16	May act as a GTPase-activating protein for Rab family protein(s).
5	*rs4891582*	1.515723505	MIR5011*	Affiliated with the miRNA class.
6	*rs1981427*	1.487360321	CNTN5*	Part of immunoglobulin superfamily, and contactin family, which mediate cell surface interactions during nervous system development.
7	rs4900907	1.478907507	RPS29*	-
8	*rs2059556*	1.387833636	FOXB1*	*Transcription factor activity*.
9	*rs7514930*	1.375128222	RALGPS2	*Guanyl-nucleotide exchange factor activity.*
10	*rs17815229*	1.365853089	CCDC178*	Unknown function.
PBS: Modern Aymara (High Altitude vs. Low Altitude, Modern Only)				
1	*rs4673323*	3.7749766	PARD3B	Putative adapter protein involved in asymmetrical cell division and cell polarization processes.

2	*rs35431164*	3.74286121	ALDH3B2	This gene encodes a member of the aldehyde dehydrogenase family.
3	*rs71458805*	3.61747614	RILPL1	Plays a role in the regulation of cell shape and polarity.
4	*rs9365594*	3.60328207	QKI *	The protein encoded by this gene is an RNA-binding protein.
5	*rs568902253*	3.44302068	LINC00410	RNA Gene, and is affiliated with the non-coding RNA class.
6	*rs529265305*	3.40751427	SPATA31A5 *	May play a role in spermatogenesis.
7	*rs77791520*	3.36011812	HAVCR1 *	May play a role in T-helper cell development and the regulation of asthma and allergic diseases.
8	*rs6903014*	3.35745714	EXOC2 *	Component of the exocyst complex involved in the docking of exocytic vesicles.
9	*N/A, chr2: 25282447*	3.29714407	EFR3B	Required to localize phosphatidylinositol 4-kinase (PI4K) to the plasma membrane.
10	*rs12427299*	3.24272966	CAND1*	Encodes an essential regulator of Cullin-RING ubiquitin ligases.
237**	*rs62650368*	3.326506021	*MGAM*	Enzyme that plays a role in the final steps of digestion of starch.
279**	*rs149112613*	3.230001259	*DST*	Cytoskeletal linker protein.
*Closest gene. **Rank compared to top hits in Ancient Rio Uncallane vs. lowlanders.				

Table S6. Demographical model parameters.

Parameter		Source	95% CI
N_A	873	Inferred	532-2146
N_{ILA}	1519	Inferred	1430-2185
N_{AYM}	1196	Inferred	924-1774
N_{HUI}	249	Inferred	220-274
N_{MIXE}	210	Inferred	205-225
T_{Bot_Aymara}	425 years	Inferred	400-450
T_{Bot_HUI}	325 years	Inferred	300-350
T_{Bot_MIXE}	350 years	Inferred	325-400
$T_{Split\_High\_Low\_Altitude}$	8750 years	Inferred	9250-8200
$T_{Split\_North\_South\_America}$	14,775 years	Inferred	14225-15775
$T_{Entry\_Into\_Americas}$	20,000 years	Fixed (*16*)	-
$T_{European\_Contact}$	485 years	Fixed (*88*)	-