

# Supporting Information

Fehren-Schmitz et al. 10.1073/pnas.1403466111

## SI Materials and Methods

**Bayesian Coalescent Simulations and Description of Models.** We used the software program BayeSSC (Bayesian serial SimCoal) to determine the demographic scenario that would most likely support the observed population genetic diversity and differentiation observed among the ancient South American populations. Under the null hypothesis (H0), we considered one continuous, panmictic population for the whole Rio Grande de Nasca drainage (RGND), with an effective population size ranging from 1,000 to 1 million females and an exponential growth starting from a small deme of 5–1,000 females 448 generations ago. The maximum value of the prior distribution of the South American Late Intermediate Period (LIP) deme effective population size ( $N_e$ ) represents ~3% (1/28th) of the present-day census size of ~27.9 million people in Peru based on the July 29, 2007 estimate by the United Nations (1). Growth rates were drawn from the uniform prior distributions based on the natural logarithm function of the respective deme sizes.

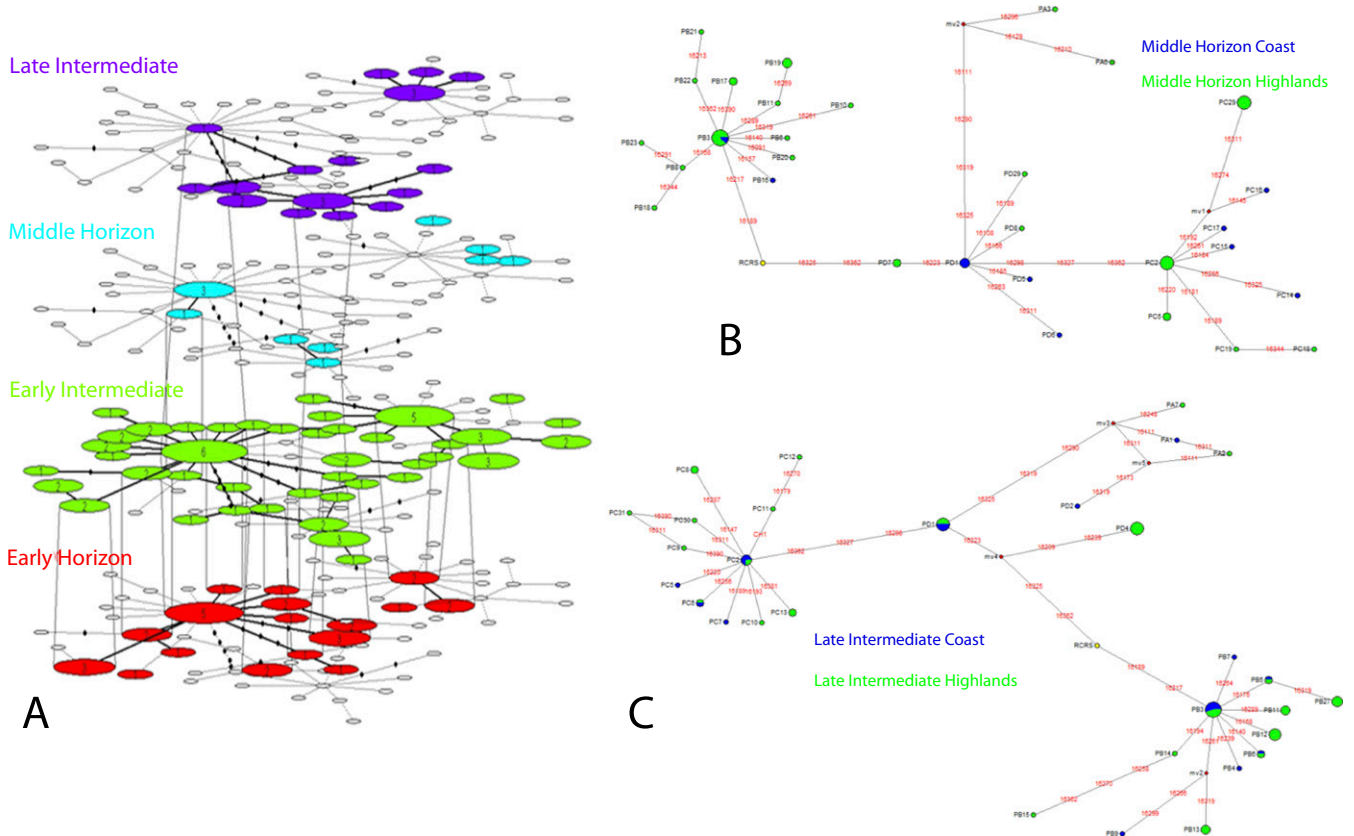
Model H1 assumes two exponentially growing populations, a coastal deme and a highland deme, of equal size (1,000–1 million) with no gene flow between the demes, and that these coalesce sometime between 81 and 500 generations ago (allowing for a population divergence at any time since the peopling of the continent until the Early Horizon, ~600 BC). Model H2 is identical in structure but allows for constant bidirectional gene flow between the two demes at a rate of  $m = 0.05$  (5% per generation). Model H3 differs from the previous scenario in allowing only unidirectional gene flow from the highlands to the coast at a rate of 2% per generation. Models H4–H9 use similar scenarios with two demes but, instead of a constant gene flow, we included distinct migration events with various proportions of migrants and both directions, exploring the timing of the events across the temporal range between our observed ancient groups. Model H4 simulates a single substantial migration event of 25% migrants from the highlands to the coast occurring 1–15 generations (25–375 y) before our start date 0 (~1400 AD). This model assumes that no emigration to the highlands occurred at the end of the Nasca Period (Early Intermediate Period–Middle Horizon transition) but that subsequent climatic and social conditions pushed people to migrate into the lower valley in the early to middle LIP (exploring the time from 1000 to 1375 AD). The three variants (a, b, c) under model H5 extend the previous scenario by exploring the contribution of migrants from the highlands to the coast using three different percentages (a, 25%; b, 50%; c, 75%) and adding a single reverse migration (10%) from the coast to the highlands 17–55 generations before 1400 AD. This model considers the proposed exodus from the lower valleys in the Late Nasca Period by exploring the timing of events

between 25 and 975 AD, which includes the start of the LIP (~600 AD). Variants of model H6 are identical to model H5 but assume a larger number of migrants (20%) to the highlands during the earlier event. Model H7 is similar in setup to H5 but with the direction of migration reversed to provide a comparable “null hypothesis” to our migration event models. Model H8a assumes a balanced proportion of migrants during both migration events (50% migrants from the coast to the highlands 17–55 generations ago, followed by a back-migration to the coast 1–15 generations before 1400 AD), whereas model H8b explores the same proportions of migrants but in the reverse direction (50% migrants from the highlands to the coast 17–55 generations ago, followed by a back-migration to the highland 1–15 generations before 1400 AD). Model H9a is again identical in setup but assumes a small number of migrants (5%) during the early migration to the highlands and a rather extreme number of recent back-migrants (80%) from the highlands to the coast 1–15 generations ago. H9b explores the reverse direction of migrants. We considered these models in comparison with H5–H8 to test whether the actual number of migrants per event or the ratio of migrants in different directions is driving our models. Schematic illustrations of the 10 basic models (H0–H9) are given in Fig. S2.

All models were simulated using 1 million genealogies in BayeSSC ([www.stanford.edu/group/hadlylab/ssc/index.html](http://www.stanford.edu/group/hadlylab/ssc/index.html)). Summary statistics were calculated on the observed and simulated data using a customized executable script to overcome the differences in underlying parameter calculations (SCStat.exe). To condition the runs and compare simulated and observed data with summary statistics, two within-population parameters (Tajima's D and haplotype diversity) and one between-population parameter ( $F_{ST}$ ; average pairwise distances) were chosen, which best reflected population differentiation between the South American cultures (Fig. 2 and Fig. S2). We compared simulated and observed values by applying an approximate Bayesian computation (ABC) framework (2) in R 2.14.1 ([www.r-project.org](http://www.r-project.org)) using available scripts ([www.stanford.edu/group/hadlylab/ssc/index.html](http://www.stanford.edu/group/hadlylab/ssc/index.html)). The fraction of simulations with the smallest Euclidian distance to observed population statistics was retained (1%) to construct posterior distributions of population parameters using the *mle* function to select the “maximum credible” version of each model. The prior distribution of these model versions was replaced with maximum-likelihood estimation values from the posterior distributions, and each model was run for 1,000 genealogies. Goodness of fit for the different models tested was compared using Akaike information criteria (3) and Akaike weights  $\omega$  (4). Results of the simulations and ABC are given in summarized form in Table S2 and Dataset S2.

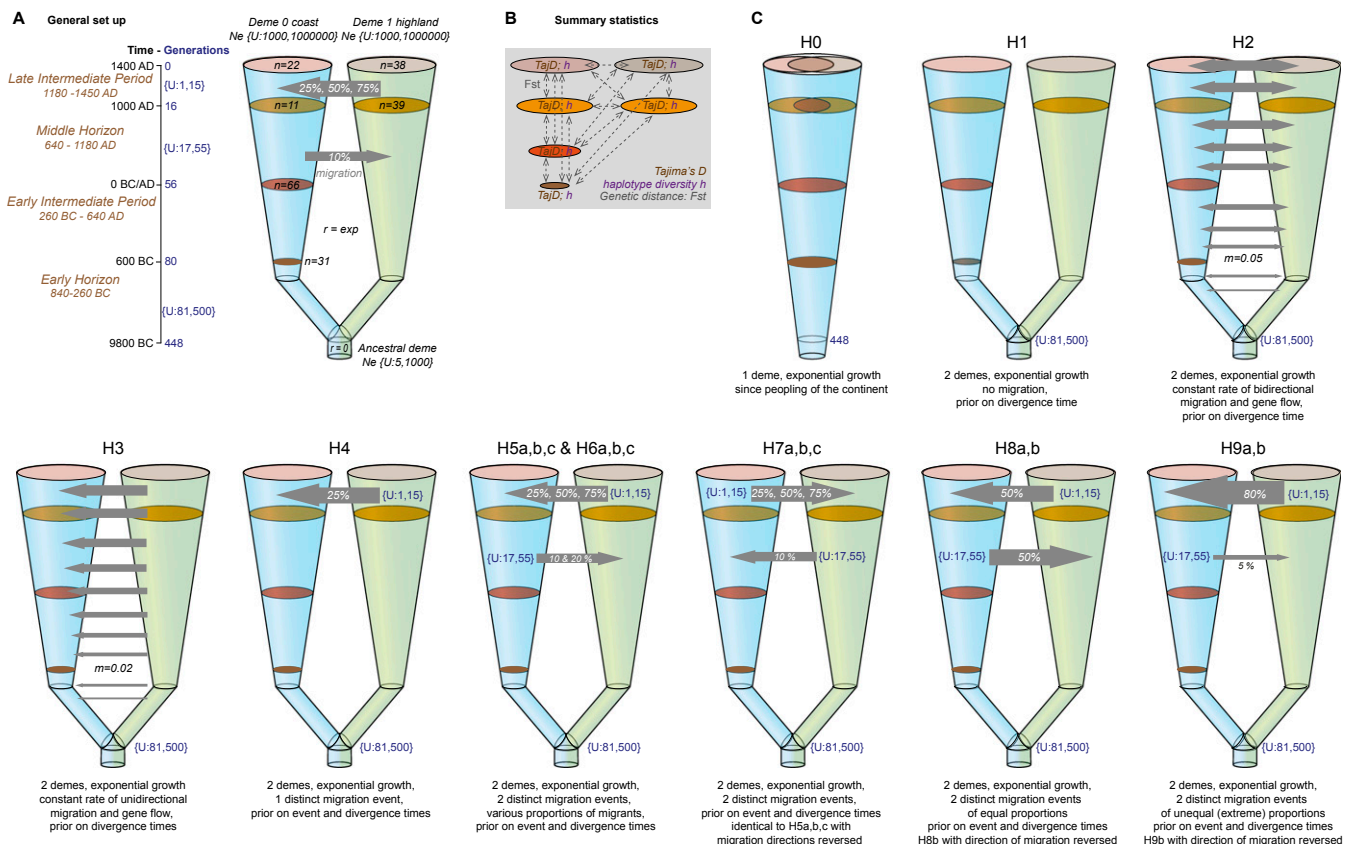
1. United Nations (2007) *World Population Prospects: The 2006 Revision, Highlights* (Department of Economic and Social Affairs, Population Division, United Nations, New York), Working Paper No ESA/P/WP.202.
2. Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics* 162(4):2025–2035.

3. Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Automat Contr* 19(6):716–723.
4. Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: Advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Syst Biol* 53(5):793–808.



**Fig. S1.** Temporal network and median-joining network analysis of all ancient mitochondrial haplotypes applied in this study. The left temporal network (A) combines all sequences diachronically, whereas the right median-joining networks compare the synchronous LIP populations of the coast and the highlands (C) or the Middle Horizon populations, respectively (B). The latter networks have been calculated using the recommended weighting of nucleotide positions following Bandelt et al. (1) and by excluding substitutions at nucleotide position (np)16182 and np16283.

1. Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16(1):37–48.



**Fig. S2.** General setup and schematic illustration of the main demographic scenarios. (A) The general setup, timeline, priors, and parameter settings used in our scenarios. (B) The relationship of summary statistics between the observed and simulated values between our six groups. (C) A number of schematic scenarios exploring population splits, independent growth, gene flow, and/or migration events to explain the observed genetic differentiation between our ancient groups, as shown in Fig. 2.

**Table S1. Mitochondrial haplotypes shared between the six analyzed populations**

Population	C_EH	C_EIP	C_MH	C_LIP	H_MH
C_EIP	9 (7)				
C_MH	1 (0)	3 (1)			
C_LIP	3 (0)	5 (0)	2 (0)		
H_MH	1 (0)	3 (1)	1 (0)	6 (4)	
H_LIP	2 (0)	4 (1)	2 (0)	7 (4)	6 (3)

Numbers in parentheses denote diagnostic haplotypes, whereas the other number includes the founding haplotypes of A, B, C, and D commonly shared between all Native American populations (Table S2). Population: ecogeographical origin: C, coast/foothills; H, highlands/upper valleys; archaeological period: EH, Early Horizon; EIP, Early Intermediate Period; LIP, Late Intermediate Period; MH, Middle Horizon.

**Table S2. Ancient and modern indigenous South American populations used for comparison in this study**

Population*	In MDS <sup>†</sup>	n	Location	Source
<b>Laramate (MH)</b>	HMH	39	Central Andes (highlands)	This study and (1)
<b>Laramate (LIP)</b>	HLIP	38	Central Andes (highlands)	(1)
<b>Palpa (LIP)</b>	CLIP	22	Central Andes (coast)	This study
<b>Palpa (EH)</b>	CEH	31	Central Andes (coast)	(2)
<b>Palpa (EIP)</b>	CEIP	66	Central Andes (coast)	This study and (2)
<b>Palpa (MH)</b>	CMH	11	Central Andes (coast)	This study and (2)
Arequipa	ARE	22	Central Andes	(3)
San Martin	SAN	21	Central Andes	(3)
Ancash	ANC	33	Central Andes	(4)
Mapuche 1	MAP	34	South Andes	(5)
Pehuenche1	PEHU	24	South Andes	(5)
Yaghan 1	YAG	15	Tierra del Fuego	(5)
Puno Quechua	QUP	30	Central Andes	(6)
Yungay	YUN	36	Central Andes	(6)
Tupe	TUP	16	Central Andes	(6)
Puno Aymara	AYP	14	Central Andes	(6)
<b>Conchapata (MH)</b>	COMH	10	Central Andes (highlands)	(7)
<b>Huari (LIP)</b>	HUA	17	Central Andes (highlands)	(7)
<b>Pampa Grande (EIP-MH)</b>	PAM	19	Central Andes (highlands)	(8)
Titicaca Quechua	QUT	37	Central Andes	(9)
Titicaca Aymara	AYT	20	Central Andes	(9)
Titicaca Uros	URT	7	Central Andes	(9)
Bolivia Quechua	QUB	93	Central Andes	(10)
Bolivia Aymara	AYB	97	Central Andes	(10)
<b>Tompullo (LH)</b>	TOM	24	Central Andes (highlands)	(11)
Kawésqar	KAW	13	Tierra del Fuego	(12)
Atacameño	ATA	28	Central Andes (highlands)	(12)
Chile Aymara	AYC	38	Central Andes	(12)
Huilliche	HUI	47	South Andes	(12)
Mapuche2	MAP2	19	South Andes	(12)
Pehuenche2	PEHU2	42	South Andes	(12)
Telhueche	TEL	23	South Andes	(12)
Yamana	YAM	21	Tierra del Fuego	(12)

\*Ancient population is set in bold; archaeological period is set in parentheses.

<sup>†</sup>Fig. 2.

- Fehren-Schmitz L, et al. (2011) Diachronic investigations of mitochondrial and Y-chromosomal genetic markers in pre-Columbian Andean highlanders from south Peru. *Ann Hum Genet* 75(2):266–283.
- Fehren-Schmitz L, Reindel M, Cagigao ET, Hummel S, Herrmann B (2010) Pre-Columbian population dynamics in coastal southern Peru: A diachronic investigation of mtDNA patterns in the Palpa region by ancient DNA analysis. *Am J Phys Anthropol* 141(2):208–221.
- Fuselli S, et al. (2003) Mitochondrial DNA diversity in South America and the genetic history of Andean highlanders. *Mol Biol Evol* 20(10):1682–1691.
- Lewis CM, Jr., Tito RY, Lizárraga B, Stone AC (2005) Land, language, and loci: mtDNA in Native Americans and the genetic history of Peru. *Am J Phys Anthropol* 127(3):351–360.
- Moraga ML, et al. (2000) Mitochondrial DNA polymorphisms in Chilean aboriginal populations: Implications for the peopling of the southern cone of the continent. *Am J Phys Anthropol* 113(1):19–29.
- Lewis CM, Jr., et al. (2007) Mitochondrial DNA and the peopling of South America. *Hum Biol* 79(2):159–178.
- Kemp BM, Tung TA, Summar ML (2009) Genetic continuity after the collapse of the Wari empire: Mitochondrial DNA profiles from Wari and post-Wari populations in the ancient Andes. *Am J Phys Anthropol* 140(1):80–91.
- Carnese FR, et al. (2010) Paleogenetic study of pre-Columbian samples from Pampa Grande (Salta, Argentina). *Am J Phys Anthropol* 141(3):452–462.
- Barbieri C, Heggarty P, Castri L, Luiselli D, Pettener D (2011) Mitochondrial DNA variability in the Titicaca basin: Matches and mismatches with linguistics and ethnohistory. *Am J Hum Biol* 23(1):89–99.
- Gayà-Vidal M, et al. (2011) mtDNA and Y-chromosome diversity in Aymaras and Quechuas from Bolivia: Different stories and special genetic traits of the Andean Altiplano populations. *Am J Phys Anthropol* 145(2):215–230.
- Baca M, Doan K, Sobczyk M, Stankovic A, Węgleński P (2012) Ancient DNA reveals kinship burial patterns of a pre-Columbian Andean community. *BMC Genet* 13:30.
- de Saint Pierre M, et al. (2012) An alternative model for the early peopling of southern South America revealed by analyses of three mitochondrial DNA haplogroups. *PLoS ONE* 7(9):e43486.

## Other Supporting Information Files

[Datasets S1 and S2 \(XLSX\)](#)