

mtDNA Diversity of the Peruvian Andean Altiplano

Natalie M. Myres(1), Joel E. Myres(1,2), Mike Malan(2), Jeff Fuller(2), Matt McNairy(2), Mark J. Rowe(2), Ray T. Matheny(2), Heather A. McCammon(1), Robert Hughes(1), Scott R. Woodward(1,2).
1) Sorenson Molecular Genealogy Foundation, Salt Lake City, Utah 2) Department of Microbiology and Molecular Biology, Brigham Young University, Provo, Utah 84115

A more adequate understanding of New World settlements and biologic history is accessible within the context of selected population studies. Examination of regional populations will more completely address the nature of New World settlement within the framework provided by broader studies. Selected regional studies can yield a more refined assessment of settlement scenarios that include intra-continental migration because of the different developmental history of each group. The current study examines the population structure of Andean Altiplano populations from the Lake Titicaca Basin within the context of Peruvian populations. We report the mtDNA haplogroup distributions of 454 samples from island and mainland populations of the Lake Titicaca region and identify control region sequence polymorphisms of 58 individuals. A selectively dispersed mtDNA control region motif is also presented as a diagnostic marker for tracing gene flow and founding events within the Lake Titicaca region and surrounding areas.

Populations

Transacted by the border of Peru and Bolivia, the Lake Titicaca Basin in the Andean Altiplano, sustains the densest rural population in the two countries (Fig. 1). Lake Titicaca is over 161 by 64 kilometers, and is located at 3850 meters above sea level. The modern indigenous residents of the Peruvian Altiplano represent a population that has remained in relative isolation until recent times, providing evidence of original Paleo-Indian founding populations. Eight Peruvian Altiplano locations were selected for this study; mainland settlements: Taraco (n=36), Juliaca (n=108), Puno (n=47), Desaguadero (n=21), and islands: Amantani (n=81), Isla Soto (n=29), Taquile (n=39), and the Uru (n=93). Samples were approximately evenly divided between mainland (47%) and island (53%) locations.

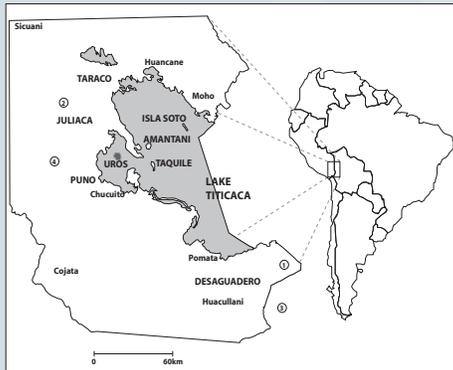


Figure 1 - Map of study area along Peru/Bolivia border.

Haplogroup Distribution

All four major Native American haplogroups were observed in the Peruvian Lake Titicaca Basin samples. Haplogroup B was predominant, with a frequency of 0.782, followed by the less frequent haplogroups C, A, D, and O (non-A, B, C or D), at 0.086, 0.066, 0.057, and 0.009 respectively (Table 1). The majority of total samples were further classified (i.e. A+/A-) based on the Haell1 site at nt 16519. This polymorphic position can discriminate between closely related lineages and revealed significant substructuring within populations, particularly the exclusivity of haplotype A-(17/17) and D+(10/10) within the Uru sample. This finding is significant to understanding the isolated placement of the Uru within the Lake Titicaca basin as a historically marginalized group.

POPULATION (n#)	h	mtDNA Lineage Classification												
		A	A+	A-	B+	B	B-	C+	C	C-	D+	D	O+	O
Taraco (34)	0.315	1 (2.9)	1 (2.9)	28 (82.4)										
Juliaca (78)	0.493	2 (2.5)	3 (3.8)	70 (5.5)	1 (1.2)	2 (2.5)	2 (2.5)	7 (9.2)	1 (1.2)	5 (6.4)				
Puno (36)	0.487			25 (69.4)	1 (2.8)	3 (8.3)	7 (19.4)	1 (2.8)						
Desaguadero (21)	0.496	1 (4.8)	1 (4.8)	15 (71.4)	1 (4.8)	2 (9.5)	1 (4.8)							
Amantani (51)	0.184			46 (90.2)	2 (7.8)	4 (7.8)								
Isla Soto (27)	0.572			17 (63)	5 (18.5)	3 (11.1)	2 (7.4)							
Taquile (35)	0.058			34 (97.1)	1 (2.9)									
Uru (86)	0.514			17 (17.9)	56 (66.7)	1 (1.2)	4 (2.4)							

Table 1 - Haplotype distribution among Peruvian Altiplano populations. Major New World haplogroups are as characterized after (Wallace and Torroni, 1992) with haplotypes based on the Haell1 status at nt 16519. Significant variation between locations is seen in the frequency of haplotypes. This is reflected in the variability of the haplotype diversity (h) for each population.

Genetic distance between islands, and between islands and mainland locations was generally significant (Table 2). Exceptions to significance between Amantani and Taquile may indicate gene flow due to contact and/or shared ancestry influenced by geographic distance. Similarly, significant differences between mainland locations were not detected except for between Taraco and Puno. Furthermore, this lack of population subdivision suggests there is less restriction on gene flow around the circumference of the lake, facilitated by transportation, migration, and urban sprawl, than for locations with limited or restricted access and cultural barriers (i.e. islands).

	Fst Values and Test of Population Differentiation							
	Amantani	Isla Soto	Taquile	Uru	Taraco	Juliaca	Puno	Desaguadero
Amantani	-	0.004*	0.124	0.000*	0.003*	0.129	0.005*	0.042*
Isla Soto	0.119*	-	0.000*	0.000*	0.023*	0.021*	0.005*	0.128
Taquile	0.021	0.213*	-	0.000*	0.039*	0.395	0.015*	0.013*
Uru	0.117*	0.062*	0.146*	-	0.000*	0.000*	0.007*	0.046*
Taraco	0.032	0.056*	0.070*	0.067*	-	0.518	0.000*	0.058
Juliaca	0.029*	0.040*	0.063*	0.042*	0.002	-	0.012*	0.491
Puno	0.099*	0.034	0.150*	0.030*	0.064*	0.031	-	0.192
Desaguadero	0.044	-0.003	0.130*	0.009	0.014	-0.013	-0.007	-

Table 2 - Below diagonal: Pairwise FST values between populations. The p-value of the test is the proportion of 100 permutations leading to a FST value greater than or equal to the observed one. Above diagonal: A test of population differentiation based on a random distribution of haplotypes among the populations. Potential states of the contingency table are considered with a Markov chain to estimate the probability of observing an equal or less likely outcome as seen in a null hypothesis of panmixia. Islands values are boxed. (* = p < 0.05)

Altiplano Sequence Analysis

58 randomly selected samples from each haplogroup were sequenced at HVI (nt positions 16051-16374) and revealed 30 polymorphic sites that defined 25 haplotypes. Nineteen sequences were singletons (76.0%). Nucleotide diversity among total samples was 0.016 (0.0087), compared to 0.007 (0.004) for the group B subset. The nucleotide diversity estimate for the entire Altiplano population is consistent with other South American indigenous populations shown to have low diversity measures. However, the nt diversity estimate for Altiplano haplogroup Bs is significantly lower, suggesting a relatively homogenous founding population followed by limited gene flow with surrounding groups.

Wide-spread in both island and mainland locations was a distinct grouping of polymorphisms found in 67% of group-B sequences and constituting a common motif characterized by nt 16183 (A-C), nt 16188 (C-T), nt 16189 (T-C), and nt 16217 (T-C). This Titicaca sequence motif (Tc1) with variants C-T at nt 16190 (Tc6), C-T at nt 16354 (Tc7), A-G at nt 16170 (Tc8), occur against a backdrop of polymorphisms generally used to define haplogroup B; 9-bp deletion, +Haell1 at nt 16519, and nt 16189 (T-C). The haplotype Tc1 was selectively distributed on the northern side of the lake indicating genetic affinity with island populations and possibly the geographic origin of this type.

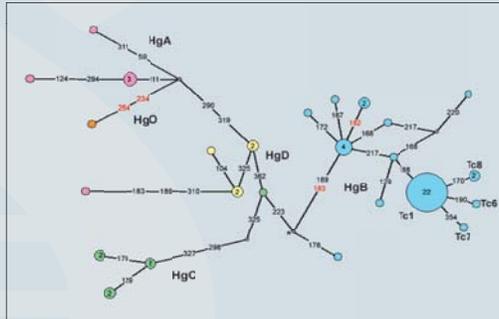


Figure 2 - Median joining network of Altiplano mtDNA sequences. Node sizes are proportional to haplotype frequencies which are indicated when >1. Variable sites are shown along branches with transversions in red. Node colors indicate haplogroup membership as follows: HgA-pink, HgB-blue, HgC-green, HgD-yellow, and Other(HgO)-orange.

Phylogenetic analysis of Altiplano sequences reveals clades representing the major Native American haplogroups with an elevated frequency of haplogroup B (Fig. 2). B-haplotypes harboring the Titicaca sequence motif (Tc1, Tc6, Tc7, Tc8) appear to be derived from closely related founding populations. These findings support the position of an independent founding population in the Altiplano followed by subsequent lineage differentiation.

Peruvian Sequence Analysis

To consider the Altiplano sequence data within a broader geographical context, data from indigenous Peruvian populations was compiled from published sources [2,3]. Samples were collected from the departments of San Martin (n=21), Ancash (n=31), Huancavelica (n=61), and Arequipa (n=23) (Fig.3). Sequences with incomplete data were not used in our analysis.

A median joining network of Altiplano sequence data and published sequence data from regions of Peru show Altiplano mtDNA lineages to cluster tightly with other Peruvian groups (Fig 4). Nodes within each of the

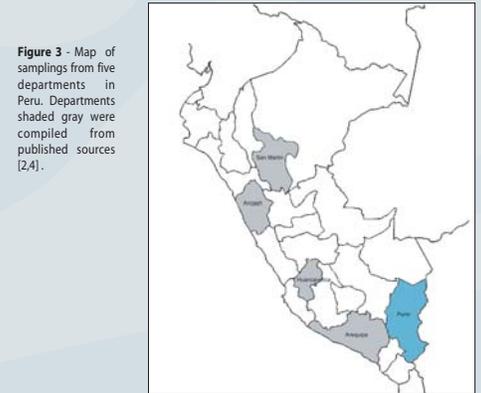


Figure 3 - Map of samplings from five departments in Peru. Departments shaded gray were compiled from published sources [2,4].

clades for Haplogroups A, C, and D contain relatively equal representations of haplotypes from each of the 5 populations. A notable exception is demonstrated in the clade representing Haplogroup B (shaded blue) where haplotypes from the Altiplano are strictly partitioned from those of remaining regions. Haplogroup B nodes within the contoured area are composed exclusively of haplotypes from the Altiplano sampling. Conversely, the remaining Haplogroup B nodes contain haplotypes from all samplings except the Altiplano, lending further support to the hypothesis that Altiplano populations descend from a relatively homogenous founding group that experienced limited gene flow with surrounding areas.

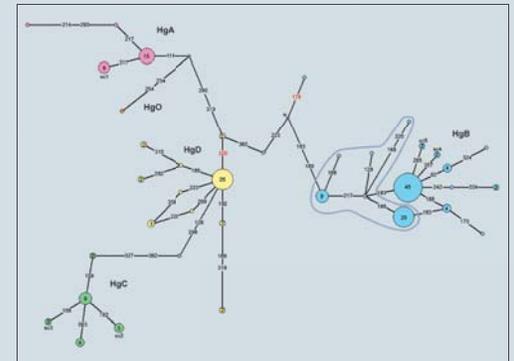


Figure 4 - Median joining network of Peruvian mtDNA sequences. Node sizes are proportional to haplotype frequencies which are indicated when >1. Star contraction clusters are labeled 'scf'. Variable sites are shown along branches with transversions in red. Node colors indicate haplogroup membership as follows: HgA-pink, HgB-blue, HgC-green, HgD-yellow, and Other(HgO)-orange. Nodes within the contoured area contain haplotypes exclusively from the Altiplano sampling.

Conclusion

mtDNA lineage distribution demonstrates the selective dispersal of the Titicaca motif throughout island populations, as well as around the Lake, and provides an informative marker for exploring gene flow between other populations. These findings support evidence that haplogroup B participated in the early settlement of South America. Based on these findings, we propose that variants of haplogroup B constituted a distinct intra-continental migration responsible for settlement of the Andean Altiplano. The geographic and environmental conditions over time have facilitated their isolation and possible adaptation in the region.

Methods and Materials

Hair samples were obtained from volunteers with informed consent. DNA was isolated and haplogroups were defined according to published criteria. Products from PCR were also sequenced and read with Sequencer version 3.1.1 (Gene Code Corp.). Median joining networks were drawn in Network [7]. Estimates of nucleotide diversity were calculated according to Nei, 1987 [8]. Pairwise FST values based on haplotype frequencies were computed in Arlequin version 2.0 [9].

Reference

- [1] Batista O, Kolman J, Bermingham E (1995) Human Molecular Genetics 4:921-929
- [2] Fuselli S, Tarazona-Santos E, Dupanloup I, Soto A, Luiselli D, Pettener D (2003) Mol. Biol. Evol. 20:1682-1691
- [3] Lewis C, Tito R, Lizattaga B, Stone A (2004) AJPA (www.interscience.wiley.com)
- [4] Kolman C, Bermingham E, Cooke R, Ward R, Arias T, Guionneau-Sinclair F (1995) Genetics 140:275-283
- [5] Rickards O, Martinez-Labarga C, Lum J, DeStefano G, Cann R (1999) AJHG 65:519-530
- [6] Ward R, Salzano R, Bonatto S, Hutz M, Coimbra C, Santos R (1996) AJHG 83:17-323
- [7] http://www.fluxus-engineering.com
- [8] Nei M. 1987. Molecular Evolutionary Genetics. New York: Columbia University Press
- [9] http://www.anthropologie.unige.ch.arlequin. Arlequin Ver. 2.000