Published in final edited form as:

Am J Hum Biol. 2014 ; 26(3): 321–330. doi:10.1002/ajhb.22507.

Mitochondrial Variation among the Aymara and the Signatures of Population Expansion in the Central Andes

KEN BATAI^{1,2,*} and SLOAN R. WILLIAMS³

¹Cancer Education and Career Development Program, Institute for Health Research and Policy, University of Illinois at Chicago (M/C 275), Chicago, Illinois 60608

²Institute of Human Genetics, College of Medicine, University of Illinois at Chicago, Chicago, Illinois 60607

³Department of Anthropology, University of Illinois at Chicago, Chicago, Illinois 60607

Abstract

Objectives—The exploitation of marine resources and intensive agriculture led to a marked population increase early in central Andean prehistory. Constant historic and prehistoric population movements also characterize this region. These features undoubtedly affected regional genetic variation, but the exact nature of these effects remains uncertain.

Methods—Mitochondrial DNA (mtDNA) hypervariable region I sequence variation in 61 Aymara individuals from La Paz, Bolivia, was analyzed and compared to sequences from 47 other South American populations to test hypotheses of whether increased female effective population size and gene flow influenced the mtDNA variation among central Andean populations.

Results—The Aymara and Quechua were genetically diverse showing evidence of population expansion and large effective population size, and a demographic expansion model fits the mtDNA variation found among central Andean populations well. Estimated migration rates and the results of AMOVA and multidimensional scaling analysis suggest that female gene flow was also an important factor, influencing genetic variation among the central Andeans as well as lowland populations from western South America. mtDNA variation in south central Andes correlated better with geographic proximity than with language, and fit a population continuity model.

Conclusion—The mtDNA data suggests that the central Andeans experienced population expansion, most likely because of rapid demographic expansion after introduction of intensive agriculture, but roles of female gene flow need to be further explored.

Domesticated plant cultivation may have started as early as 10,000 years ago in the Peruvian highlands (Lynch, 1980), while both rich marine resources and domesticated plants provided abundant food sources for coastal populations (Moseley, 1992). These resources supported

rapid population growth in both environments. As a consequence, complex societies emerged very early both on the coast and in the highlands (Burger, 1995; Shimada, 1994). Around Lake Titicaca, a number of cultures, including Chiripa, Pukara, and Tiwanaku, flourished on a subsistence base of tubers and camelids. The people who live in the *altiplano* (the Andean plateau around the Lake Titicaca) today are mostly rural Aymara, traditional agropastoralists who herd sheep, cows, llamas and alpacas, and grow a wide variety of altitude-tolerant tubers and grains (Vitzthum et al., 2009).

The *altiplano* is a very linguistically diverse region and many of its inhabitants were bi- or trilingual (Mannheim, 1985; Stanish, 2003). The dominant indigenous languages spoken today are Quechua and Aymara. Quechua is spoken by the largest number of people (~7–11 million speakers) with a wide geographic distribution ranging from Columbia to Argentina and Chile (Stark, 1985a,b). Aymara is spoken by over two million people who live around Lake Titicaca, in highland Bolivia and Peru, the mid-altitude valleys of Chile and in lowland Bolivia (Briggs, 1985b). Linguists divide Quechua into two dialect groups, either Quechua I and II (Torero, 1964) or Quechua A and B (Parker, 1963). They hypothesize that the older group, Quechua I (B) spread through the northern and central Andes in 7th–10th centuries. The Inca later re-introduced their dialect (Quechua II or A) into the north and central Andes (Stark, 1985a,b). The Spanish expanded its distribution even further when they adopted Quechua II as a lingua franca during the Colonial era (Hardman, 1985; Mannheim, 1991). Researchers initially believed that Quechua and Aymara were closely related to each other, but have now determined that the languages belong to different language families and any similarities are the result of borrowing (Mannheim, 1985, 1991). Aymara is still widely spoken in the central Andes and is a member of the Jaqi language family, which also includes the nearly extinct Kawki and Jaqaru (Briggs, 1985a). At least two other now extinct languages, Pukina and Uru, were spoken in the *altiplano* at the time of Spanish contact (Browman, 1994; Murra, 1968).

Many archaeologists (Browman, 1994; Janusek, 2004, 2008; Kolata, 1993; Stanish, 2003) have argued that the Aymara, who live in the core area of the prehistoric Tiwanaku state, are likely Tiwanaku descendants. Archaeologists base their arguments largely on the distribution of languages at the time of Spanish conquest and the continuity of ceramics and other archaeological data. Stanish (2003) argues that the change in subsistence practice from agriculture to pastoralism in the period after the Tiwanaku state collapse is more likely linked to a severe drought at that time than to a population replacement. Linguists and anthropologists, on the other hand, have generally favored a recent Aymara immigration after the Tiwanaku state collapse and considered Pukina the language of the ancient Tiwanakans (Albó, 1987; Bouysse-Cassagne, 1987; Shady Solís, 1987; Torero, 1987; Wachtel, 1987). They propose different hypotheses about the original Aymara homeland. Torero (1987) argued that the Aymara originally lived on the Peruvian coast in the Department of Lima where the other Jaqi language speakers are found today. He characterized the Aymara entrance into the Lake Titicaca area as a violent military conquest, using glottochronology to place their arrival in the thirteenth century. Other linguists, most prominently Bouysse-Cassagne (1987, but also Albo, 1987; Wachtel, 1987), used origin

myths and linguistic evidence (Briggs, 1985) to support a southern homeland on the north central Chilean coast around Coquimbo.

Various population movements, from small scale and temporary movements of individuals and families tending fields at different elevations to larger movements of highland colonists to lower elevations in the Middle Horizon or recent migrants to urban centers, have been important factors in Andean settlement history (D'Altroy, 2002; Murra, 1968; Skeldon, 1977). Murra (1968, 1985) has written about what he terms “vertical archipelago” systems that colonial Aymara kingdoms utilized to take advantage of the diverse Andean ecological zones. They established permanent colonies arranged vertically on both sides of the Andean slopes. Maize and other crops grown at lower elevations held nutritional and ritual importance for the people who lived on the *altiplano*. This vertical control of resources is ancient in the Andes, and the Tiwanaku state maintained colonies at lower elevations, for example in the Moquegua Valley (Goldstein, 1993, 2005). Although ancient DNA analysis suggested genetic differences between the Moquegua Middle Horizon colonists and the *altiplano* residents (Lewis et al., 2007a), bone chemistry and cranial data indicated the presence of highland migrants in the area (Blom et al., 1998; Knudson, 2008). The Quechua-speaking Inca also sent *mitmaq* colonists into many parts of the Andes, including the Lake Titicaca region, either to fulfill short-term labor tax obligations or to pacify the people they conquered and disrupt any brewing rebellions.

These population growth and movements undoubtedly influenced regional patterns of genetic variation. Initial mitochondrial DNA (mtDNA) studies (Bert et al., 2001, 2004; Corella et al., 2007; Fuselli et al., 2003; Lewis et al., 2005, 2007b) included many Quechua speakers, but few Aymara speakers. Research focused mainly on the differences between Andean and Amazonian populations to contrast the micro-evolutionary processes that each region experienced and to test whether multiple migrations into South America could account for the differences. Studies of mtDNA and other genetic markers have demonstrated that Andean populations show high within-population genetic diversity with small population differentiation, while Amazonian populations exhibit low within-group variability with high population differentiation (Fuselli et al., 2003; Lewis and Long, 2008; Lewis et al., 2005, 2007b; Luiselli et al., 2000; Tarazona-Santos et al., 2001). Recent studies (Barbieri et al., 2011; de Saint Pierre et al., 2012; Gaya-Vidal et al., 2011) have added more Aymara population samples, which allow us to examine the relationship between Quechua and Aymara speakers more carefully and to test the competing histories of Aymara origins.

Few studies have, however, attempted to examine different demographic expansion processes in central Andes. Fuselli et al. (2003) noted that demographic expansion, spatial (also called a range) expansion with high gene flow among groups, or selection could explain the Andean mtDNA pattern of high within-population diversity with an excess of rare alleles observed by researchers. Human populations that expanded spatially have large migration rates and exhibit large within-population genetic diversity values that are similar to populations that experienced pure demographic expansions (Excoffier, 2004; Ray et al., 2003).

In this study, we added a new Aymara population sample from La Paz, Bolivia, and analyzed the Hypervariable Region I (HVRI) in the control region of the mtDNA in 61 Aymara individuals. We also took advantage of the growing number of population samples to characterize the nature of the central Andean population expansion and to examine how demographic expansion through increasing female effective population size and spatial expansion through female gene flow influenced the mtDNA variation in central Andean and other South American populations. If the Aymara experienced a pure demographic expansion driven by agricultural intensification, but with limited female gene flow, we expected to observe high within-population genetic diversity values, low migration rates, and large genetic differences with other populations. On the other hand, if Aymara culture and language spread spatially, expanding geographically by incorporating other cultural and language groups, we expected to observe high within-population genetic diversity, high migration rates, and little genetic differentiation from other populations. If the Aymara had migrated to the *altiplano* from the north central Andes as Torero and others have proposed, we would expect to find closer affinities between the Aymara and the north central Andean populations, while a southern homeland hypothesis would be supported by evidence of genetic similarity with southern Andean populations and greater differences between the Aymara and Quechua populations.

SAMPLES AND METHODS

Samples

The genetic material used in the study was taken from blood spot samples collected from 61 healthy unrelated Aymara women indigenous to the Bolivian *altiplano*. The deidentified samples were originally collected in 1995 in La Paz as a part of a study of exercise performance (Brutsaert et al., 2002). Informed consent was obtained from all participants according to the protocol approved by the Human Subjects Review Committee, University of California, Riverside, and by the Comité de Ética of the Instituto Boliviano de Biología de Altura, La Paz. The permission for DNA analysis was granted by the University of Illinois at Chicago Institutional Review Board.

Laboratory methods

The Puregene® DNA purification kit was used to extract DNA from blood spot samples following the protocol provided. The Hypervariable Region I (HVRI) of the mtDNA control region was amplified using a touchdown PCR protocol and either of the primer sets (L15985 5'-GCACCCAAAGCTAAGATTCTAA-3' and H404 5'-AAAGTGCATACCGCCAAAAG-3' or L15926 5'-TCAAAGCTTACACCAGTCTTGTAAC-3' and H16498 5'-CCTGAAGTAG GAACCAGATG-3'). The amplified product was sequenced in both directions using the BigDye Terminator Cycle Sequencing Kit, version 3.1 (Applied Biosystems) and analyzed on an ABI 3730 DNA sequencer. DNA extraction, PCR amplification, and sequencing were performed in the University of Illinois at Chicago and the Field Museum's Pritzker Molecular Biology Laboratory.

Sequences were edited and aligned in Sequencher 4.1.4 (GeneCodes). The nucleotide positions (nps) between 16024 and 16383 relative to the Cambridge Reference Sequence

(Anderson et al., 1981) were edited, but subsequent analyses were confined to nps 16056–16383 to facilitate comparisons with previous work. When a T→C transition at np 16189 was present, nps 16182 and 16183 were excluded from analysis because of heteroplasmy (Bendall and Sykes, 1995; Pfeiffer et al., 1999). Haplogroups A, C, and D were assigned based on diagnostic mutations observed in the HVRI sequence. Additionally, the 9 bp deletion was screened, and the haplogroup B2 was assigned based on diagnostic mutations in HVR1 as well as the 9 bp deletion.

Analytical methods

The La Paz Aymara mtDNA HVRI sequence variation was compared to 47 published population samples from South America (Fig. 1, Table 1, and Supporting Information Table 1). Two additional population samples where mtDNA haplogroup frequencies, but not sequences, were reported (Quechua from Department of Pasco and Lima, Peru, and a population sample from Quebrada de Huamahuaca, Argentina), were included in haplogroup frequency comparisons, but not in other analyses (Table 1). The haplogroup frequencies from the lowland Bolivian Aymara samples are shown in Table 1, but were excluded in our analyses due to very small sample size.

The Arlequin population genetics software program (Excoffier et al., 2005; Schneider et al., 2000) was used to estimate within-population genetic diversity, haplotype diversity (h), θ_k , θ_S , and θ_p , to conduct two tests of molecular neutrality (Tajima's D and Fu's F_s), and to analyze mismatch distributions. θ_k , θ_S , and θ_p are three different estimators of $\theta = 2N_{fe}\mu$, where N_{fe} is the female effective population size and μ is the mutation rate. The estimated θ should reflect the differences in female effective population size, because human populations have the same mutation rate. The value of θ_k is based on the relationship

between θ the number of sequences (θ) and the sample size, $E(k) = \theta \sum_{i=0}^{n-1} \frac{1}{\theta+i}$ (Ewens, 1972), while θ_S is based on the relationship between the number of segregating sites (S) and

the sample size, $\theta_S = S / \sum_{i=1}^{n-1} \frac{1}{i}$ (Watterson, 1975). The θ_π is a measure of mean pairwise

differences (π), $\theta_\pi = \sum_{i=1}^k \sum_{j < i} p_i p_j d_{ij}$, where d_{ij} is the number of mutational differences between sequence i and j (Tajima, 1983). While ancient demographic history affects the mean pairwise nucleotide differences (π) and θ_p , the two other θ estimators (θ_k and θ_S) are more sensitive to recent demographic events (Helgason et al., 2000, 2003; Rogers, 1995; Tajima, 1989). Fu's F_s statistic calculates the probability of observing a number of haplotypes (k) less than or equal to the observed k given an observed θ_π . Excoffier and Schneider (1999) have noted that populations with significantly negative F_s values often exhibit other genetic evidence of population expansion, while Tajima's D is a more conservative test for detecting evidence of recent population expansion (Aris-Brosou and Excoffier, 1996).

Mismatch distributions were used to evaluate whether observed within-population variation could better be explained by demographic or spatial expansion models. Under a sudden demographic expansion model, a unimodal mismatch distribution is interpreted as evidence of demographic expansion assuming that the populations are panmictic (Rogers and

Harpending, 1992; Slatkin and Hudson, 1991). Many large agricultural populations are genetically diverse and have unimodal mismatch distributions, while foragers' population sizes have stayed consistently small, so they are less genetically diverse and more likely to have multimodal mismatch distributions (Excoffier and Schneider, 1999; Rogers, 1995; Watson et al., 1996). However, expanding populations that are genetically diverse because of high levels of gene flow may exhibit the same unimodal distributions as populations that experienced a demographic expansion. Ray et al. (2003) analyzed the effects of gene flow between subdivided populations on within-population genetic diversity and demonstrated that when the migration rate is large, a population will have also have a unimodal mismatch distribution, a genetic signature of expansion similar to demographic expansions.

A sum of square deviation (SSD) between the observed and expected (simulated) mismatch distributions was used to assess whether a demographic expansion or spatial expansion model best explained the observed mismatch distribution of populations. The expected mismatch distribution for each demographic model was generated through coalescent simulation. The parameters estimated from the empirical data were used to test the hypothesis that the parameters generated under the model were correct.

To examine patterns of subdivision and interactions among the Aymara and other Latin American groups, first, pairwise population genetic distances (Φ_{ST}) and analysis of molecular variance (AMOVA) were undertaken in Arlequin. Next, the population pairwise genetic distances were visualized using Multidimensional scaling (MDS) analysis with SPSS statistical software (IBM). We also used two different methods to estimate female migration rates. Arlequin was used to calculate the migration rate, $M = 2N_{fe}m$, under a spatial expansion model (Excoffier, 2004), and the migration rate, $M = 2N_{fe}m$, between each pair of populations was estimated using MIGRATE, a coalescent-based maximum likelihood method (Beerli and Felsenstein, 1999, 2001).

Western South American populations were grouped by ecological, cultural and linguistic similarity and placed into six regional/linguistic groups: Aymara, Quechua, southern Andes, Lowland Bolivians, Gran Chaco, and northwestern lowland South Americans. MIGRATE uses an n-Island model to estimate the migration rate, M , Θ , $\Theta = 2N_{fe}\mu$, another estimator of effective population size, for three or more sampled populations simultaneously by exploring genealogical trees, including the topology of branch lengths and various migration scenarios. The parameters, M and Θ , were estimated using the averages from three independent runs in each regional set through 10 short chains (10,000 genealogies per chain) and three long chains (100,000 genealogies per chain) with increments of 20 and 200 steps, respectively. Rather than exploring all possible genealogical trees, parameters were calculated using a Markov chain Monte Carlo approach focusing on the trees with the highest likelihoods. After each chain, the parameters were recalculated and the likelihood was reevaluated. The first 100,000 trees in each chain were discarded, and the long chains were replicated. Metropolis coupled Markov Chain Monte Carlo, or "heating" was used, setting four temperatures (1, 1.5, 3, and 6). Heating allows the running of four chains and swaps between them to explore more genealogical spaces. The "hotter" chains search more widely in the more distant regions of genealogical space than do the "colder" chains.

RESULTS

General patterns of south central Andean and Aymara mtDNA variation

As expected from previous reports (Bert et al., 2001; Merriwether et al., 1995), haplogroup B2 was the most common haplogroup in the La Paz Aymara sample (62.3%; Table 1 and Supporting Information Table 2) and haplogroup D1 was rare (1.6%). Haplogroup A2 (11.5%) and C1 (23.0%) were relatively common. One sample could not be assigned to a haplogroup from the HVRI sequence data. The most common haplotype (B05) had a 16188–16189–16217 motif that was shared with another haplogroup B2 haplotype (B10). This motif was found among the Aymara and Quechua in the south central Andes in intermediate to high frequencies, suggesting a south central Andean origin for this haplotype (Álvarez-Iglesias et al., 2007; Barbieri et al., 2011; Corella et al., 2007; de Saint Pierre et al., 2012; Fuselli et al., 2003; Gayá-Vidal et al., 2011). This motif is also found in two of the Gran Chaco populations (the Pilaga, and Wichi) (Cabana et al., 2006). Although the 16,189T→C mutation is a part of the haplogroup B2 motif, this hypervariable mutation was also observed in the other three haplogroups in the La Paz Aymara sample: haplogroups A2 (A03), C1 (C04 and C05), and D1 (D01).

The La Paz Aymara haplogroup frequencies were similar to other central Andean populations, especially other south central Andeans (Table 1). However, the La Paz Aymara had the lowest haplogroup B2 frequency among the Aymara and the third lowest in the south central Andes after the Quechua Puno and Coya. The Aymara and Quechua of the lowland Bolivian Department of Beni, and the Quebrada de Humahuca and Atacameno populations living in the border area of northern Argentina, southern Bolivia, and Chile also had similar haplogroup frequencies. The north central Andeans had lower frequencies of haplogroup B2 and higher frequencies of various other haplogroups when compared with the south central Andeans. The southern Andean populations had very different haplogroup frequencies; haplogroup C1 and D (D1 and D4h3a) were common there (de Saint Pierre et al., 2012; Moraga et al., 2000).

Aymara and South American mtDNA genetic diversity

Supporting Information Table 3 shows the summary statistics for each of the population samples, organized by region and arranged in order from high to low h within each group. All four measurements of within-population genetic diversity were significantly correlated ($P < 0.05$). The most genetically diverse populations were the central Andeans who were agriculturalists from areas where prehistoric states once flourished. A large number of low frequency haplotypes (k) and polymorphisms (S) were present in the sampled central Andean populations, resulting in large θ_k and θ_S values which are reflective of recent expansions (Helgason et al., 2003; Tajima, 1989).

Among the central Andeans, the Aymara were genetically less diverse than the Quechua. The values of θ_k greatly varied from the highest observed in the Ancash Quechua to the lowest in the Jaqaru speaking Tupe. The Ancash Quechua also had the highest h , while the Tayacaja Quechua had the highest θ_S . All central Andeans, but especially the Aymara, had high frequencies of haplogroup B2, so their haplogroup diversities were low. Consequently,

their h_p values, which generally reflect more ancient demographic events, were small (Helgason et al., 2000, 2003; Rogers, 1995). Very low within-population genetic diversity was observed among the Tupe.

Fourteen genetically diverse South American populations, including central Andeans (Aymara and Quechua), Pilaga, and Moxo showed evidence of population expansion in significantly negative F_s values, and two populations (the Aymara from Lake Titicaca and the Quechua from Potosi) had statistically significant negative values of Tajima's D . Many low frequency or unique haplotypes were observed in the Aymara and Quechua, so they had large θ_k values relative to the θ_p values and thus significantly negative F_s values.

We used mismatch distribution analysis to test whether individual population samples best fit the demographic expansion model or the spatial expansion model (which assumes that gene flow significantly influenced within-population genetic diversity). A few clear trends were observed (Table 2 and Supporting Information Table 4). First, the demographic expansion model was rejected more often than the spatial expansion model. This pattern was the most prominent among the least genetically diverse South American populations; eight out of 13 forager populations (61.5%) rejected the demographic expansion model, but none of them rejected the spatial expansion model. Second, a contrasting pattern was observed among the populations engaged in intensive agricultures. Five out of 16 (31.3%) intensive agriculturalists rejected the spatial expansion model, but only two (11.8%), the Aymara Puno 2 and lowland Bolivia Quechua samples, rejected the demographic expansion model. Today, the Aymara Puno 2 are Aymara speakers, but were speakers of the now extinct Uro language and are a genetically distinct group from surrounding Aymara and Quechua speakers (Barbieri et al., 2011; Sandoval et al., 2013).

Identifying the pattern of interactions in western South America

Population pairwise genetic distances (U_{ST}) were calculated and the genetic relationships among the South American populations are illustrated on a MDS plot in Figure 2. The stress value (0.118) was within an acceptable range (Sturrock and Rocha, 2000). The plot illustrates how closely the Aymara and other central Andean groups are related to each other and to western lowland South American populations, while the Chibchan and eastern lowland South American populations scatter widely. Three forager populations, the Ach e, Ayoreo, and Xavante, were excluded from the final plot because their wide dispersion and extremely low within-population genetic diversity suggested that strong genetic drift resulted in their dramatic differentiation from their neighbors.

Although their haplogroup frequencies greatly varied, the western South American populations, including the central Andean, southern Andean, and lowland populations, clustered near each other on the right side of the plot. The south central Andeans with very high B2 haplogroup frequencies (the Aymara from Lake Titicaca and Puno and the Quechua from lowland Bolivia) were found in the top right quarter of the plot. The other south central Andeans were located in the upper part of western South American cluster. The La Paz Aymara clustered with other south central Andean groups and with the north central Andean Tupe, who speak a language related to Aymara. The Chibchan and Amazonian lowland populations were scattered widely across the left and lower half of the plot.

The MDS plot of western South American populations (Supporting Information Fig. 1) showed little overlap between the south central and north central Andean populations. The south central Aymara and Quechua clustered together on the right side, while the north central Quechua populations were grouped with some of the populations from the southern Andes, lowland Bolivia, western lowland South America, and the Gran Chaco area.

The AMOVA results supported the pattern observed on the MDS plot and showed similarly close relationships among the central Andeans (Supporting Information Table 5). The small Θ_{ST} values among the western South American groups suggested that these populations were not highly differentiated within each group. Lewis et al. (2007b) found smaller Θ_{ST} values among central Andean, southern Andean, and lowland Bolivian groups than among Amazonian populations. This pattern remained the same when more populations were added. The Chibchan group shows great population differentiation with the largest Θ_{ST} . The lowland Bolivians had the highest Θ_{ST} value among the western South Americans, indicating greater genetic differentiation and explaining their dispersion on the MDS plot. The lowest Θ_{ST} was observed among south central Andeans. All Θ_{ST} values were significant at $P < 0.05$.

The most genetically diverse populations had high migration rates ($M = 2N_{fe}m$) estimated using the mismatch distribution spatial expansion model (Supporting Information Table 3). M was significantly correlated with h_k ($P = 0.001$) and θ_S ($P = 0.003$), but not with θ_π ($P = 0.165$). The migration rate could not be calculated for seven south central Andean populations, perhaps because of their poor fit to the spatial expansion model, and the central Andean populations tend to reject the spatial expansion model.

mtDNA variation of Aymara vs. Quechua

We examined the sequence diversity of each haplogroup found in the central Andes (Supporting Information Table 6). Haplogroup B2 was more diverse than haplogroups A2, C1, and D, but the significantly large negative Tajima's D and Fu's F_S indicated that all haplogroups showed evidence of expansion. The mismatch distributions generally conformed to the sequence diversity estimates (Fig. 3). All of the haplogroups had unimodal distributions showing evidence of expansion. Haplogroup D (D1 and D4h3a) had the lowest θ estimates with a peak at one nucleotide difference, while haplogroup A showed the largest θ_π , with a peak at three nucleotide differences.

The haplogroup B2 sequences were analyzed separately for the Aymara and Quechua and then compared to assess similarities in genetic variation and demographic history. Haplogroup B2 showed evidence of expansion in both groups, and their haplogroup B2 sequence diversity was similar. The Aymara and Quechua mismatch distributions each had a peak at two nucleotide differences at a frequency of about 0.3.

The $\Theta = 2N_{fe}\mu$ of the regional/linguistic groups and the migration rates ($2N_{fe}m$) between each group were estimated using MIGRATE. The Quechua had a larger Θ than the other linguistic/regional groups analyzed, about six times larger than the Aymara who had the second largest Θ estimate (Table 3). Although a number of factors likely biased the estimation of migration rates upward, the migration rates estimated between each pair of

population groups were high. The migration rate between the Aymara and Quechua groups was the highest, but the Quechua population showed high migration rates with all groups. The Aymara migration rates with other groups were lower than those of the Quechua.

DISCUSSION

The mtDNA variation strongly supports the ethnohistorical evidence for an Aymara expansion. The sequence diversity comparisons, the F_s , the mtDNA haplogroup sequence diversity, and the mismatch distributions all clearly indicate that the Aymara and Quechua populations are highly diverse and have recently experienced population expansions. Here, we evaluate how well demographic and spatial expansion models explain the observed genetic signatures of expansion and examine the differences between Aymara and Quechua evolutionary history.

Evidence for demographic and spatial expansion

If the Aymara experienced a pure demographic expansion without extensive female gene flow, we expected to observe high within-population genetic diversity, low migration rates, and large genetic differences with other populations. First, it should be noted that the Quechua and Aymara have large θ_k , θ_S , and Θ (estimated with MIGRATE), but small θ_p values. The large θ_k and θ_S relative to the small θ_p values indicate that the populations accumulated rare or low frequency variants relatively recently as their population sizes increased. Second, the Aymara and Quechua mismatch distribution analyses rejected the spatial expansion model more often than the demographic expansion model. The accumulations of rare and low frequency variants through recent demographic expansion can produce significant F_s , star-like networks, and unimodal mismatch distributions of mtDNA haplogroups like those observed in this study as well as in previous studies (Cabana et al., 2006; Fuselli et al., 2003; Lewis et al., 2005). Gene flow can affect the θ estimates (Ray et al., 2003) that form the basis of these tests, however, so the Θ values which were estimated simultaneously with migration rates should reflect female effective population size better. The large Aymara and Quechua Θ estimates indicate that female effective population size influenced within-population genetic diversity more than gene flow, providing important support for a demographic expansion as the better fit for the observed patterns.

The exploitation of marine resources, cultivation of domesticated plants, and intensification of agricultural production started in central Andes very early in its pre-history causing population size to grow dramatically (Lynch, 1980; Moseley, 1992). On the Bolivian altiplano around the Lake Titicaca, domesticated tubers and camelids supported a number of prehistoric cultures (Kolata, 1993). de Saint Pierre et al.'s (2012) Bayesian analyses of mtDNA HVRI and HVRII supported the archaeological evidence, indicating that the Aymara experienced a drastic increase in female effective population size that began around 8,000 years ago and reached its peak by 5,000 years ago. They attributed this expansion to plant and animal domestication.

These analyses suggest that the observed large within-population genetic diversity can be explained by a demographic expansion, but other analyses show that female gene flow

among the Aymara and Quechua was an important factor as well. Although the central Andean populations rejected the spatial expansion model more than the demographic expansion model, all of the Aymara and Quechua population samples had high migration rates estimated based on the mismatch distribution spatial expansion model and the migration rates estimated between them with MIGRATE were very high. The vertical use of ecozones, a series of prehistoric state expansions, forced migration by Inca and colonial governments, and urban immigrations after European contact (D'Altroy, 2002; Murra, 1968; Skeldon, 1977) all likely increased the rate of female gene flow, but teasing apart the effects of each these factors is difficult. The results of simulation analyses conducted by Fuselli et al. (2003) showed that large female population size in the Andes with high migration rates either before or after European contact can explain the observed mtDNA variation. MIGRATE and MDS analyses further suggest that the network of interaction or gene flow extends beyond central Andes to the transition area between the Andes and the Amazon along the western side of South America (Corella et al., 2007).

The migration rates between the Quechua and Aymara and between the central Andean and lowland populations are high, but large effective population sizes, shared founding haplotypes, unsampled populations, and small divergence time can inflate migration rates (Beerli, 2004). More reliable reconstructions of demographic history require the use of model-based methods using longer sequences or more genetic markers, and more population samples. Human populations often violate the important built-in assumptions for estimation of demographic parameters and the population genetics models are more simplistic than actual evolutionary history. More complex population genetics models, that incorporate demographic expansion as well as gene flow and that enable us to simultaneously and more accurately estimate multiple demographic parameters including effective population size, migration rates, and timings of demographic expansion and other events, are also necessary.

Evolutionary history of the Aymara vs. Quechua

No significant genetic differences were found between the south central Aymara and Quechua groups. These genetically diverse populations had high haplogroup B2 frequencies and small AMOVA Θ_{ST} , and they clustered together on the MDS plot. They also had similar haplogroup B2 sequence diversity and mismatch distributions. Despite their different languages and histories, the mtDNA data suggests that linguistic differences were not an important barrier to female gene flow.

The Aymara and north central Quechua also exhibit some genetic similarities. The north central Quechua were slightly more diverse and their haplogroup B2 sequences showed slightly more diversity than those of the Aymara, which may reflect the differences in subsistence strategies between two areas. Nonetheless, overall they were genetically more diverse than other South American populations, and their haplogroup B2 mismatch distributions were very similar. If we assume that natural selection had a minor effect on mtDNA variation, then these characteristics suggest that they experienced similar demographic expansions. Alternatively, Excoffier (2004) suggested that two populations from the same spatial expansion wave will show similar genetic diversity, so these

similarities may instead have been the result of interactions. In our analyses, we observed large migration rates among the central Andean populations.

Geography appears to have influenced mtDNA variation more than language. The south central Aymara and Quechua populations differ from the north central Quechua populations. The north central Andeans tend to be genetically more diverse than the south central Andeans (including the Aymara and Quechua). The Atacameno from northern Chile were plotted closely with the south central Aymara and Quechua on the MDS plots, but south central Andean populations overlap little with north central Andean and southern Andean populations. These genetic similarities and shared demographic history among the south central Andean groups, but genetic differences from other Andean populations, best support a model of population continuity through time in south central Andes and possibly archaeologists' contention that the Aymara were the descendants of the ancient Tiwanaku populations. Small prehistoric genetic samples from Tiwanaku heartland ($n = 9$) and Tiwanaku colony in Moquegua ($n = 23$) do differ from modern highland south central Andeans, however (Lewis et al., 2007a; Rotthammer et al., 2003). It should be noted that language shift or male biased migration could also explain the observed linguistic distributions in the area.

The Quechua language today includes many different dialects and has been spread over a wide area by both the Inca and the Spanish, likely increasing the diversity of the people who have adopted the language. Lowland populations may have interacted more intensively with the north central Andeans than they did with south central Andeans. Migration rates estimated using MIGRATE between the Quechua and other linguistic/regional groups were very high. The north central Andeans had higher frequencies of haplogroup A2, C1, and D than the south central populations, possibly due to gene flow from lowland populations. North central highland and lowland interactions are evident in the archaeological record by the Early Horizon, when the Chavín culture flourished (900 BC-200 BC) (Burger, 1995), and exotic trade goods have been found in many coastal and highland archaeological sites (Alva, 2001). In previous studies, western lowland South American populations exhibited intermediate mtDNA variation between central Andean and eastern lowland South American populations (Cabana et al., 2006; Corella et al., 2007). In their phylogeographic analyses, however, mtDNA control region sequences provided limited capability to assess the intensity and timing of gene flow between highland Andean and lowland populations. Whole mitochondrial genome sequence analysis will likely allow us to further examine migration history (Bodner et al., 2012).

CONCLUSIONS

The mtDNA data suggests that both the Aymara and the Quechua experienced population expansion, most likely as a result of a shared rapid demographic expansion, after the introduction of intensive agriculture. Female gene flow was an important homogenizing factor as well. In the central Andes, geography correlates better with mtDNA variation than language does and the south central Aymara and Quechua are virtually indistinguishable from each other. Small genetic differences among the altiplano population samples suggest that any abrupt, recent population replacements are unlikely. People also moved within

central Andes and into neighboring transitional zones. The extent to which each prehistoric migration event impacted genetic variation, however, needs further examination.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The authors thank the Aymara people who provided the samples. They thank Virginia Vitzthum for sample collection and for reviewing and commenting on this manuscript; Hilde Spielvogel and Carol Worthman for sample collection; Elizabeth Abrams, Mary Ashley, Geoffrey Hayes, and Crystal Patil for helpful discussion and comments on the manuscript; and anonymous reviewers for helpful comments.

Contract grant sponsor: Sigma Xi Grants In Aid of Research; Contract grant number: G2007228164486015; Contract grant sponsor: NCI Training Grants (Cancer Education and Career Development Program); Contract grant number: 5R25 CA057699; Contract grant sponsors: University of Illinois at Chicago Provost Research Award; University of Illinois at Chicago Department of Anthropology Charles Reed Fund.

LITERATURE CITED

- Albó X. Comments on “Lenguas y pueblos altioplánicos en torno al siglo XVI” by A. Torero. *Revista Andina*. 1987; 5:375–376.
- Alva, W. The royal tombs of Sipán: art and power in Moche society.. In: Pillsbury, J., editor. *Moche art and archaeology in ancient Peru*. Yale University Press; New Haven: 2001. p. 223–245.
- Álvarez-Iglesias V, Jaime JC, Carracedo Á, Salas A. Coding region mitochondrial DNA SNPs: targeting East Asian and Native American haplogroups. *Forensic Sci Int Genet*. 2007; 1:44–55. [PubMed: 19083727]
- Anderson A, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG. Sequence and organization of the human mitochondrial genome. *Nature*. 1981; 290:457–470. [PubMed: 7219534]
- Aris-Brosou S, Excoffier L. The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. *Mol Biol Evol*. 1996; 13:494–504. [PubMed: 8742638]
- Barbieri C, Heggarty P, Castri L, Luiselli D, Pettener D. Mitochondrial DNA variability in the Titicaca basin: matches and mismatches with linguistics and ethnohistory. *Am J Hum Biol*. 2011; 23:89–99. [PubMed: 21082684]
- Beerli P. Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. *Mol Ecol*. 2004; 13:827–836. [PubMed: 15012758]
- Beerli P, Felsenstein J. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics*. 1999; 152:763–773. [PubMed: 10353916]
- Beerli P, Felsenstein J. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proc Natl Acad Sci USA*. 2001; 98:4563–4568. [PubMed: 11287657]
- Bendall KE, Sykes BC. Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. *Am J Hum Genet*. 1995; 57:248–256. [PubMed: 7668250]
- Bert F, Corella A, Gene M, Perez-Perez A, Turbon D. Mitochondrial DNA diversity in the Llanos de Moxos: Moxo, Movima and Yuracare Amerindian populations from Bolivia lowland. *Ann Hum Biol*. 2004; 31:9–28. [PubMed: 14742162]
- Bert F, Corella A, Manel G, Pérez-Pérez A, Turbón D. Major mitochondrial DNA haplotype heterogeneity in highland and lowland Amerindian populations from Bolivia. *Hum Biol*. 2001; 73:1–16. [PubMed: 11332638]

- Blom DE, Hallgrímsson B, Keng L, Lozada Cerna MC, Buikstra JE. Tiwanaku “colonization”: bioarchaeological implications for migration in the Moquegua Valley, Peru. *World Archaeol.* 1998; 30:238–261. [PubMed: 16475288]
- Bodner M, Perego UA, Huber G, Fendt L, Röck AW, Zimmermann B, Olivieri A, Gómez-Carballa A, Lancioni H, Angerhofer N, et al. Rapid coastal spread of First Americans: novel insights from South America’s Southern Cone mitochondrial genomes. *Genome Res.* 2012; 22:811–820. [PubMed: 22333566]
- Bouysse-Cassagne T. Comments on “Lenguas y pueblos altiplánicos en torno al siglo XVI” by A. Torero. *Revista Andina.* 1987; 5:377–379.
- Briggs, LT. A critical survey of the literature on the Aymara language.. In: Manelis Klein, HE.; Stark, LR., editors. *South American Indian languages: retrospect and prospect.* University of Texas Press; Austin: 1985a. p. 546-594.
- Briggs, LT. Dialectical variation in Aymara.. In: Manelis Klein, HE.; Stark, LR., editors. *South American Indian languages: retrospect and prospect.* University of Texas Press; Austin: 1985b. p. 595-616.
- Browman DL. Titicaca Basin archaeolinguistics: Uru, Pukina and Ayrara AD 750–1450. *World Archaeol.* 1994; 26:235–251.
- Brutsaert TD, Spielvogel H, Caceres E, Araoz M, Chatterton RT, Vitzthum VJ. Effect of menstrual cycle phase on exercise performance of high-altitude native women at 3600 m. *J Exp Biol.* 2002; 205:233–239. [PubMed: 11821489]
- Burger, RL. *Chavín and the origins of Andean civilization.* Thames and Hudson; London: 1995.
- Cabana GS, Merriwether DA, Hunley K, Demarchi DA. Is the genetic structure of Gran Chaco populations unique? Interregional perspectives on native South American mitochondrial DNA variation. *Am J Phys Anthropol.* 2006; 131:108–119. [PubMed: 16485304]
- Corella A, Bert F, Pérez-Pérez A, Gen M, Turbón D. Mitochondrial DNA diversity of the Amerindian populations living in the Andean Piedmont of Bolivia: Chimane, Moseten, Aymara and Quechua. *Ann Hum Biol.* 2007; 34:34–55. [PubMed: 17536754]
- D’Altroy, TN. *The Incas.* Blackwell Publishing; Cambridge, MA: 2002.
- de Saint Pierre M, Bravi CM, Motti JMB, Fuku N, Tanaka M, Llop E, Bonatto SL, Moraga M. An alternative model for the early peopling of Southern South America revealed by analyses of three mitochondrial DNA haplogroups. *PLoS One.* 2012; 7:e43486. [PubMed: 22970129]
- Ewens WJ. The sampling theory of selectively neutral alleles. *Theor Popul Biol.* 1972; 3:87–112. [PubMed: 4667078]
- Excoffier L. Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Mol Ecol.* 2004; 13:853–864. [PubMed: 15012760]
- Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online.* 2005; 1:47–50. [PubMed: 19325852]
- Excoffier L, Schneider S. Why hunter-gatherer populations do not show signs of Pleistocene demographic expansions. *Proc Natl Acad Sci USA.* 1999; 96:10597–10602. [PubMed: 10485871]
- Fuselli S, Tarazona-Santos E, Dupanloup I, Soto A, Luiselli D, Pettener D. Mitochondrial DNA diversity in South America and the genetic history of Andean highlanders. *Mol Biol Evol.* 2003; 20:1682–1691. [PubMed: 12832631]
- Gayà-Vidal M, Moral P, Saenz-Ruales N, Gerbault P, Tonasso L, Villena M, Vasquez R, Bravi CM, Dugoujon J-M. 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.* 2011; 145:215–230. [PubMed: 21469069]
- Goldstein PS. Tiwanaku temples and state expansion: a Tiwanaku sunken court temple in Moquegua, Peru. *Lat Am Antiq.* 1993; 4:22–47.
- Goldstein, PS. *Andean diaspora: the Tiwanaku colonies and the origins of South American empire.* University of Florida Press; Gainesville, FL: 2005.
- Hardman, MJ. Aymara and Quechua: languages in contact.. In: Manelis Klein, HE.; Stark, LR., editors. *South American Indian languages: retrospect and prospect.* University of Texas Press; Austin: 1985. p. 617-643.

- Helgason A, Nicholson GJ, Stefánsson K, Donnelly P. A reassessment of genetic diversity in Icelanders: strong evidence from multiple loci for relative homogeneity caused by genetic drift. *Ann Hum Genet.* 2003; 67:281–297. [PubMed: 12914564]
- Helgason A, Siguroardottir S, Gulcher JR, Ward R, Stefansson K. mtDNA and the origin of the Icelanders: deciphering signals of recent population history. *Am J Hum Genet.* 2000; 66:999–1016. [PubMed: 10712214]
- Janusek JW. Tiwanaku and its precursors: recent research and emerging perspectives. *J Archaeol Res.* 2004; 12:121–183.
- Janusek, JW. *Ancient Tiwanaku.* Cambridge University Press; Cambridge: 2008.
- Knudson KJ. Tiwanaku influence in the South Central Andes: strontium isotope analysis and Middle Horizon migration. *Lat Am Antiq.* 2008; 19:3–23.
- Kolata, AL. *The Tiwanaku: portrait of an Andean civilization.* Blackwell; Cambridge, MA: 1993.
- Lewis CM, Buikstra JE, Stone AC. Ancient DNA and genetic continuity in the south central Andes. *Lat Am Antiq.* 2007a; 18:145–160.
- Lewis CM Jr, Long JC. Native South American genetic structure and prehistory inferred from hierarchical modeling of mtDNA. *Mol Biol Evol.* 2008; 25:478–486. [PubMed: 18222947]
- Lewis CMJ, Lizárraga B, Tito RY, López PW, Iannaccone GC, Medina A, Martínez R, Polo SI, De La Cruz AF, Cáceres AM, et al. Mitochondrial DNA and the peopling of South America. *Hum Biol.* 2007b; 79:159–178. [PubMed: 18027812]
- Lewis CMJ, Tito RY, Lizarraga B, Stone AC. Land, language, and loci: mtDNA in Native Americans and the genetic history of Peru. *Am J Phys Anthropol.* 2005; 127:351–360. [PubMed: 15584069]
- Luiselli D, Simoni L, Tarazona-Santos E, Pastor S, Pettener D. Genetic structure of Quechua-speakers of the central Andes and geographic patterns of gene frequencies in South American populations. *Am J Phys Anthropol.* 2000; 113:5–17. [PubMed: 10954616]
- Lynch, T., editor. *Guitarrero cave: early man in the high Andes.* Academic Press; New York: 1980.
- Mannheim, B. Southern Peruvian Quechua.. In: Manelis Klein, HE.; Stark, LR., editors. *South American Indian languages: retrospect and prospect.* University of Texas Press; Austin: 1985. p. 644–688.
- Mannheim, B. *The language of the Inka since the European invasion.* University of Texas Press; Austin: 1991.
- Merriwether DA, Rothhammer F, Ferrell RE. Distribution of the four founding lineage haplotypes in Native Americans suggests a single wave of migration for the New World. *Am J Phys Anthropol.* 1995; 98:411–430. [PubMed: 8599378]
- Moraga ML, Rocco P, Miquel JF, Nervi F, Llop E, Chakraborty R, Rothhammer F, Carvallo P. Mitochondrial DNA polymorphisms in Chilean aboriginal populations: implications for the peopling of the southern cone of the continent. *Am J Phys Anthropol.* 2000; 113:19–29. [PubMed: 10954617]
- Moseley, ME. *The Incas and their ancestors: the archaeology of Peru.* Thames and Hudson; London: 1992.
- Murra JV. An Aymara kingdom in 1567. *Ethnohistory.* 1968; 15:115–151.
- Murra, JV. “El archipiélago vertical” revisited.. In: Shozo, M.; Shimade, I.; Morris, C., editors. *Andean ecology and civilization.* University of Tokyo Press; Tokyo: 1985. p. 3–13.
- Parker GJ. La Clasificación genética de los dialectos Quechuas. *Revista del Museo Nacional.* 1963; 32:241–252.
- Pfeiffer H, Brinkmann B, Hühne J, Rolf B, Morris AA, Steighner R, Holland MM, Forster P. Expanding the forensic German mitochondrial DNA control region database: genetic diversity as a function of sample size and microgeography. *Int J Legal Med.* 1999; 112:291–298. [PubMed: 10460419]
- Ray N, Currat M, Excoffier L. Intra-deme molecular diversity in spatially expanding populations. *Mol Biol Evol.* 2003; 20:76–86. [PubMed: 12519909]
- Rogers A, Harpending H. Population growth makes waves in the distribution of pairwise differences. *Mol Biol Evol.* 1992; 9:552–569. [PubMed: 1316531]
- Rogers AR. Genetic evidence for a Pleistocene population expansion. *Evolution.* 1995; 49:608–615.

- Rothhammer F, Moraga M, Rivera M, Santoro C, Standen VG, Garcia F, Carvallo P. Análisis de ADNmt de restos esqueléticos del sitio arqueológico de Tiwanaku y su relación con el origen de sus construcciones. *Chungara*. 2003; 35:269–274.
- Sandoval JR, Lacerda DR, Jota MSA, Salazar-Granara A, Vieira PPR, Acosta O, Cuellar C, Revollo S, Fujita R, Santos FR, et al. The genetic history of indigenous populations of the Peruvian and Bolivian Altiplano: the legacy of the Uros. *PLoS One*. 2013; 8:e73006. [PubMed: 24039843]
- Schneider, S.; Roessli, D.; Excoffier, L. Arlequin 2.0: A software for population genetic data analysis. University of Geneva Genetics and Biometry Laboratory; Geneva, Switzerland: 2000.
- Shady Solís R. Comments on “Lenguas y pueblos altiplánicos en torno al siglo XVI” by A. Torero. *Revista Andina*. 1987; 5:387–392.
- Shimada, I. Pampa Grande and the Mochica culture. University of Texas Press; Austin, TX: 1994.
- Skeldon R. The evolution of migration patterns during urbanization in Peru. *Geogr Rev*. 1977; 67:394–411.
- Slatkin M, Hudson RR. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*. 1991; 129:555–562. [PubMed: 1743491]
- Stanish, C. Ancient Titicaca: the evolution of complex society in southern Peru and northern. University of California Press; Bolivia. Berkeley: 2003.
- Stark, LR. Ecuadorian highland Quechua: history and current status.. In: Manelis Klein, HE.; Stark, LR., editors. South American Indian languages: retrospect and prospect. University of Texas Press; Austin: 1985a. p. 443–479.
- Stark, LR. The Quechua language in Bolivia.. In: Manelis Klein, HE.; Stark, LR., editors. South American Indian languages: retrospect and prospect. University of Texas Press; Austin: 1985b. p. 516–545.
- Sturrock K, Rocha J. A multidimensional scaling stress evaluation table. *Field Methods*. 2000; 12:49–60.
- Tajima F. Evolutionary relationship of DNA sequences in finite populations. *Genetics*. 1983; 105:437–460. [PubMed: 6628982]
- Tajima F. The effect of change in population size on DNA polymorphism. *Genetics*. 1989; 123:597–601. [PubMed: 2599369]
- Tarazona-Santos E, Carvalho-Silva DR, Pettener D, Luiselli D, De Stefano GF, Martinez Labarga C, Rickards O, Tyler-Smith C, Pena SDJ, Santos FR. Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. *Am J Hum Genet*. 2001; 68:1485–1496. [PubMed: 11353402]
- Torero A. Los dialectos quechuas. *Anales Científico de la Universidad Agraria*. 1964; 2:446–478.
- Torero A. Lenguas y pueblos altiplánicos en Torno al Siglo XIV. *Revista Andina*. 1987; 5:329–405.
- Vitzthum VJ, Thornburg J, Spielvogel H. Seasonal modulation of reproductive effort during early pregnancy in humans. *Am J Hum Biol*. 2009; 21:548–558. [PubMed: 19402035]
- Wachtel N. Comments on “Lenguas y pueblos altiplánicos en torno al siglo XVI” by A. Torero. *Revista Andina*. 1987; 5:392–394.
- Watson E, Bauer K, Aman R, Weiss G, von Haeseler A, Pääbo S. mtDNA sequence diversity in Africa. *Am J Hum Genet*. 1996; 59:437–444. [PubMed: 8755932]
- Watterson G-A. On the number of segregation sites in genetical models without recombination. *Theor Popul Biol*. 1975; 7:256–276. [PubMed: 1145509]

Fig. 1.
Map of South America showing the locations of sampled populations.

Fig. 2.
Multidimensional scaling (MDS) plot of South American populations.

Fig. 3.

Mismatch distributions of mtDNA haplogroups. (a) It compares the mismatch distribution of four major mtDNA haplogroups among the central Andeans. (b) It compares the mismatch distribution of haplogroup B2 in Aymara, north central Quechua, and south central Quechua populations.

TABLE 1

Haplogroup frequencies (%) of the Andean populations arranged geographically from north to south

Populations	n	A	B	C	D	Other	References
Yungay (Quechua)	36	2.8	47.2	36.1	13.9	0.0	(Lewis et al., 2007)
Ancash (Quechua)	33	9.1	51.5	18.2	21.2	0.0	(Lewis et al., 2007)
Tupe (Aymaran)	16	0.0	68.8	16.1	0.0	0.0	(Lewis et al., 2007)
Tayacaja (Quechua)	61	21.3	32.8	13.3	29.5	3.3	(Fuselli et al., 2003)
Quechua (Pasco and Lima) ^a	52	3.9	53.8	17.3	19.2	5.8	(Rodriguez-Delfin et al., 2001)
Arequipa (Quechua)	22	9.1	68.2	13.6	9.1	0.0	(Fuselli et al., 2003)
Quechua Puno 1	30	6.7	60.0	23.3	10.0	0.0	(Lewis et al., 2007)
Quechua Puno 2	37	16.2	56.8	13.5	13.5	0.0	(Barbieri et al., 2011)
Quechua (Beni, Bolivia)	32	15.6	75.0	9.4	0.0	0.0	(Bert et al., 2001)
Aymara Puno 1	14	0.0	71.4	14.3	14.3	0.0	(Lewis et al., 2007)
Aymara Puno 2	20	0.0	90.0	5.0	5.0	0.0	(Barbieri et al., 2011)
Aymara La Paz	61	11.5	62.3	23.0	1.6	1.6	This study
Aymara Lake Titicaca (Bolivia)	96	7.3	81.3	6.3	5.1	0.0	(Gayà-Vidalet al., 2011)
Aymara (Beni, Bolivia)	33	0.0	93.3	3.0	3.0	0.0	(Bert et al., 2001)
Aymara (Chile)	172	6.4	67.4	12.2	14.0	0.0	(Merriwether et al., 1995)
Quechua (Potosi)	93	15.1	61.3	19.4	4.2	0.0	(Gayà-Vidal et al., 2011)
Atacamenos	50	12.0	72.0	10.0	6.0	0.0	(Merriwether et al., 1995)
Coya (Quechua)	60	13.3	58.4	23.3	5.0	0.0	(Álvarez-Iglesias et al., 2007)
Quebrada de Humahuaca (Argentina)	46	10.9	67.4	17.4	4.3	0.0	(Dipierri et al., 1998)
Mapuche (Argentina)	136	10.3	35.3	21.3	28.7	5.1	(Ginther et al., 1993; Goicoechea et al., 2001)
Mapuche (Chile) ^a	111	0.0	7.2	44.1	48.7	0.0	(Moraga et al., 2000)
Pehuenche	100	2.0	9.0	37.0	52.0	0.0	(Merriwether et al., 1995)
Huilliche	80	3.8	28.8	18.7	48.7	0.0	(Merriwether et al., 1995)
Tehuelche	29	0.0	20.7	24.1	55.2	0.0	(Goicoechea et al., 2001)
Yaghan	21	1.3	8.0	43.0	47.7	0.0	(Moraga et al., 2000)

^aThese samples were included in haplogroup frequency analyses only.

TABLE 2

Percentage (and number) of populations that rejected the demographic and spatial expansion model

	Total (N)	Demographic expansion model	Spatial expansion model
All Latin Americans	48	37.5% (18)	22.9% (11)
Intensive Agriculturalists	16	12.5% (2)	31.3% (5)
Horticulturalists	19	42.1% (8)	26.3% (5)
Foragers	13	61.5% (8)	0.0 (0)

TABLE 3

Θ and migration rates ($2N_{fe}m$) between linguistic/regional groups obtained using MIGRATE

	<i>n</i>	Θ	Aymara	Quechua	Southern Andes	Lowland Bolivians	Gran Chaco
Aymara ^a	75	0.0740					
Quechua ^b	182	0.4798	195.9752				
Southern Andes ^c	97	0.0106	13.4695	42.4734			
Lowland Bolivians ^d	74	0.0577	36.7992	63.9783	15.4234		
Gran Chaco ^e	204	0.0086	25.3572	41.7774	7.7331	10.9738	
Northwestern Lowland South Americans ^f	104	0.0137	20.4791	72.6017	7.8300	15.9892	6.3225

The estimates are average of five separate runs.

^a Aymara (Aymara La Paz and Aymara Puno).

^b Quechua (Ancash, Arequipa, Quechua Puno, Tayacaja, and Yungay).

^c Southern Andes (Argentina Mapuche, and Mapuche/Pehuenche).

^d Lowland Bolivians, Department of Beni (Chimane, Moseten, Movima, Moxo, and Yuracare).

^e Gran Chaco (Pilaga, Whichi, and Toba).

^f Northwestern lowland South Americans (Cayapa, Embera, and Wounan).