Mitochondrial DNA Diversity in South America and the Genetic History of Andean Highlanders

Silvia Fuselli,*† Eduardo Tarazona-Santos,*‡§ Isabelle Dupanloup,† Alonso Soto,‖ Donata Luiselli,* and Davide Pettener*

*Area di Antropologia, Dipartimento di Biologia e. s., Università di Bologna, Bologna, Italy; †Sezione di Biologia Evolutiva, Dipartimento di Biologia, Università di Ferrara, Italy; ‡Department of Biology, University of Maryland; §Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; and ‖Departamento de Medicina, Hospital Hipólito Unanue, Lima, Peru

We analyzed mtDNA sequence variation in 590 individuals from 18 south Amerindian populations. The spatial pattern of mtDNA diversity in these populations fits well the model proposed on the basis of Y-chromosome data. We found evidence of a differential action of genetic drift and gene flow in western and eastern populations, which has led to genetic divergence in the latter but not in the former. Although it is not possible to identify a pattern of genetic variation common to all South America, when western and eastern populations are analyzed separately, the mtDNA diversity in both regions fits the isolation-by-distance model, suggesting independent evolutionary dynamics. Maximum-likelihood estimates of divergence times between central and south Amerindian populations fall between 13,000 and 19,000 years, which is consistent with a Pleistocenic peopling of South America. Moreover, comparison of among-population variability of mtDNA and Y-chromosome DNA seems to indicate that South America is the only continent where the levels of differentiation are similar for maternal and paternal lineages.

Introduction

The first humans to reach the Western Hemisphere arrived in the Pleistocene (i.e., more than 11,000 years ago) from northeastern Asia across the Bering Strait and spread southward, peopling Central America and South America (Dixon 1999). For most of the past century, scholars believed that big-game hunters known as the Clovis, who spread across North America and Central America between 11,200 and 10,500 years ago, were the first inhabitants of the Americas. However, recent archeological discoveries in Monte Verde site (Chile) and Pedra Pintada (Brazil) have shown that generalized hunter-gatherers were already settled in South America 12,000 years ago (Dixon 1999; Dillehay 2000). These findings, demonstrating that early South American pre-history predates Clovis culture, have led to the decline of the Clovis paradigm. Similarly, inferences based on mtDNA diversity have placed the population expansion associated with the first peopling of the Americas from Asia in the Pleistocene (Bonatto and Salzano 1997). The new archeological discoveries, and the fact that South America was one of the last continents to be peopled and one of the first where prehistoric civilizations flourished, have renewed the interest in South American prehistory.

Studies of molecular diversity in south Amerindian populations are still scanty (Ward et al. 1996; Bianchi et al. 1998; Mesa et al. 2000; Moraga et al. 2000; Rodriguez-Delfin, Rubin-de-Celis, and Zago 2001; Tarazona-Santos et al. 2001). The Andean region in particular, ruled by the Inca Empire when Europeans arrived, had reached levels of population density and socioeconomic development unmatched in other regions of South America. Notwithstanding their historical importance, Andean populations have rarely been included in studies of Amerindian genetic diversity.

The Andean region encompasses the western South America from Colombia to northern Chile. The well-known Inca Empire, which dominated the region during the 15th and the beginning of the 16th century, is just the tip of the iceberg of a long-term cultural process that began with the development of the first complex societies more than 4,000 years ago (Stanish 2001), leading to a higher cultural and linguistic homogeneity than in eastern South America. For instance, across the entire Andean region, native people speak languages that belong to a unique linguistic family (Andean), whereas in eastern South America natives speak languages belonging to four different linguistic families (Cavalli-Sforza, Menozzi, and Piazza 1994). One of the open questions in South American anthropology is whether this contrasting picture between the culturally homogeneous west and the heterogeneous east of South America is correlated with the patterns of genetic diversity.

An additional open question in American prehistory is the date of the first peopling of South America and even more important, if people who lived in this area during the Pleistocene were ancestors of current south Amerindians. In particular, craniometric data show that the earliest South American crania (dated > 9,000 years ago) have morphology that differs from the more recent and contemporaneous south Amerindians Mongoloid morphology (Powell and Neves 1999). These differences had lead Neves and Pucciarelli (1991) to propose that current Amerindians are not descendants of the earliest (non-Mongoloid) people that lived in North and South America during the late Pleistocene to early Holocene period. Instead, they proposed that current Native Americans would be descendant of a Holocene migration wave of Mongoloid people that could have replaced previous populations. However, this topic is still subject to controversy, and, in fact, Powell and Neves (1999) have suggested that a model such as that previously proposed by Neves and Pucciarelli (1991) would require some unrealistic assumptions.

In a recent study including Andean samples of Y-chromosome DNA, we proposed a model for the evolution of south Amerindian populations (Tarazona-Santos et al. 2001), which is consistent with the analysis of classical markers (Luiselli et al. 2000). Our model suggests contrasting regional patterns of genetic drift and gene flow. Western populations of South America, associated with the Andean area, show a combination of higher long-term effective sizes and gene flow than eastern populations, settled in Brazil. This has caused homogenization of the gene pool in the former but a divergence of population in the latter. If this model fits the history of south Amerindian populations, we expect that different loci will show a pattern of variation matching that observed for Y-chromosome DNA. At the moment, the hypervariable region I (HVRI) of mtDNA is the only genomic region studied in enough south Amerindian populations to be used to test our model. In this paper, we show that mtDNA diversity in South America not only matches that of Y-chromosome DNA but also presents a clearer spatial organization. We date the split between central and south Amerindian populations in the Pleistocene (likely more than 13,000 years ago) and discuss the implication of this result on our understanding of the tempo and mode of the first colonization of South America. Finally, we show that mtDNA and Y-chromosome DNA present similar among-population differentiation.

Materials and Methods
Samples, Populations, and Sequences

We analyzed the pattern of genetic variation of the HVRI on the basis of 590 sequences from 18 populations. Sequences from 105 individuals are presented for the first time and correspond to samples from three native populations from the Peruvian Andes. The Quechua-speaking samples from Tayacaja ($n = 61$) and Arequipa ($n = 22$), who settled above 2,800 m of altitude, belong to traditional populations of local farmers and have been described elsewhere (Luiselli et al. 2000, Tarazona-Santos et al. 2001). The third sample is from San Martin de Pangoa ($n = 22$), a small town on the eastern slope of the Andes inhabited by Quechua and Nmatsiguenga people. All sampled individuals were informed about our objectives and consented to the anonymous use of DNA for research. For the 105 Peruvian individuals, we amplified the HVRI of mtDNA using the primers described in Alves-Silva et al. (2000). We sequenced both strands of the HVRI between bp 16024 and 16383 (Anderson et al. 1981), using the dideoxy BigDye kit (Applied Biosystems). Sequencing products were separated on the ABI PRISM-310 DNA sequencer (Applied Biosystems). Table A of the Supplementary Material online shows the 64 different haplotypes found (GenBank accession numbers AY304083 to AY304146) and their frequencies. All the different lineages fall into one of the four Native American haplogroups: A, B, C, and D (Merriwether, Rothhammer, and Ferrell 1995). We classified the 18 samples into two categories: (1) traditional, collected from geographically restricted rural areas from native populations that reasonably fit the panmictic model, and (2) urban, HVRI sequences belonging to haplogroups A, B, C, or D, collected in cities from admixed populations. Table 1 shows the available samples and their classification, and figure 1 shows the geographic distribution of traditional samples.

Statistical Analysis

We computed different estimators of the parameter $\theta = 2 N_e \mu$ to compare the amount of within-population variability in different populations: (1) θ_k (Ewens 1972), based on the number of observed alleles (k), (2) θ_S (Watterson 1975), based on the number of observed segregating sites (S), and (3) θ_π (Tajima 1983), based on the mean number of pair-wise differences between sequences (π). After changes in population sizes, k and S change faster (i.e., in a few hundred generations) than π (Helgason et al. 2000). Since we were interested in detecting demographic events that occurred after the first peopling of the continent (i.e., within the past 500 generations), we based our comparisons on θ_k and θ_S. We used the F_s statistic (Fu 1997) to investigate which populations show an excess of rare alleles with respect to the mutation-drift equilibrium expectation (i.e., show negative F_s values) and which populations do not.

Genetic distances and the Spatial Autocorrelation analysis (Autocorrelation Index for DNA Analysis [AIDA] [Bertorelle and Barbujani 1995]) were used to investigate the genetic structure of mtDNA diversity among South Amerindian populations. We calculated Φst genetic distances between populations considering the Tamura and Nei model of nucleotide substitution (Tamura and Nei 1993), with a gamma correction for heterogeneity of mutation rates ($\alpha = 0.26$, Meyer, Weiss, and von Haeseler 1999). We applied a regression analysis to test if the scatter-plot of $\Phi st/(1 - \Phi st)$ versus the logarithm of geographic distances fits a linear function with positive slope, as predicted by the Isolation-by-Distance model (IBD) (Wright 1943; Rousset 1997). Because the statistic $\Phi st/(1 - \Phi st)$ is subject to error and is measured in different units respect to the logarithm of geographic distances, we used a Model II regression, namely the Reduced Major Axis regression (Sokal and Rohlf 1995), implemented in the IBD software (Bohonak 2002). Estimators of the parameter θ, the F_s statistic and genetic distances were calculated using the software Arlequin version 2.0 (Schneider, Roessli, and Excoffier 2000).

AIDA measures similarity among individual haplotypes within classes of geographic distances by a standardized statistic called II. Under the null hypothesis of random distribution of allele frequencies in space, the expected value of II is close to 0 (Bertorelle and Barbujani 1995). If the genetic structure fits the IBD model, correlograms (i.e., the scatterplot of II values against geographic distance classes) show an asymptotically decreasing shape from positive II values at low distance classes to nonsignificant II values at large distances (Sokal 1979). We use the AIDA software to calculate the II statistics (Bertorelle and Barbujani 1995).

We used a method (NW) recently developed by Nielsen and Wakeley (2001) to obtain maximum-likelihood estimations of divergence times between native populations

Table 1
Values of Three Estimators of θ, Their 95% Confidence Intervals (95% CI) or Standard Deviations (SD), and the F_s Neutrality Test for South American Populations on the Basis of HVRI of mtDNA

Population	n	θ_k (95% CI)	θ_s (SD)	θ_π (SD)	F_s
Traditional Samples					
1. Tayacaja	61	63.5 (37.6–109.3)	10.68 (3.16)	7.20 (3.79)	−25.2**
2. Arequipa	22	43.9 (17.7–117.4)	7.13 (2.68)	6.05 (3.34)	−10.4**
3. San Martin	22	19.5 (8.7–45.4)	6.03 (2.32)	5.99 (3.31)	−5.4*
4. Yanomama[a]	129	13.3 (8.7–20.1)	5.71 (1.64)	5.47 (2.94)	−10.6*
5. Pehuenche[b]	24	10.8 (5.0–23.4)	5.89 (2.23)	6.95 (3.77)	−1.9
6. Mapuche 1[c]	39	8.5 (4.4–15.9)	4.73 (1.70)	6.54 (3.51)	−1.8
7. Yaghan[b]	15	4.5 (81.7–11.6)	4.30 (1.89)	5.62 (3.21)	0.6
8. Wai-Wai[d]	26	4.4 (2.0–9.7)	6.29 (2.33)	7.66 (4.11)	1.7
9. Zoro[e]	29	4.0 (1.8–8.5)	4.29 (1.64)	4.62 (2.59)	0.3
10. Mapuche 2[b]	34	3.7 (1.7–7.7)	4.18 (1.56)	6.50 (3.51)	2.0
11. Gaviao[e]	28	2.7 (1.1–6.2)	4.15 (1.63)	4.93 (2.76)	1.9
12. Cayapa[f]	30	2.6 (1.1–5.8)	4.04 (1.56)	5.50 (3.02)	3.4
13. Xavante[e]	25	1.1 (0.4–3.0)	2.64 (2.65)	3.42 (2.02)	3.7
14. Surui[d]	24	1.1 (0.4–3.0)	1.87 (0.90)	2.02 (1.31)	1.8
Urban Samples					
Brazilian Southeast[g]	34	90.4 (40.3–218.2)	9.78 (3.26)	8.33 (3.96)	−21.3**
Brazilian North[g]	26	48.6 (21.0–119.8)	8.39 (2.99)	7.85 (4.21)	−11.4**
Brazilian South[g]	11	48.1 (11.8–208.7)	7.51 (3.29)	8.40 (4.76)	−3.4*
Brazilian Northeast[g]	11	48.1 (11.8–208.7)	7.49 (4.28)	6.49 (2.88)	−3.8*

NOTE.—Populations are sorted as a decreasing function of θ_k. * $P < 0.05$; ** $P < 0.01$.

[a] Merriwether et al. (2000).
[b] Moraga et al. (2000).
[c] Ginther et al. (1993).
[d] Bonatto and Salzano (1997).
[e] Ward et al. (1996).
[f] Rickards et al. (1999).
[g] Alves-Silva et al. (2000).

of South America and two samples settled in the Isthmus of Panama, the Wuonan and Embera, who speak languages of the Chocó linguistic family, which is also spoken in northern South America (Kolman and Bermingham 1997). The NW method uses coalescent processes and assumes a demographic model of two populations of equal effective sizes that split from a panmictic ancestral population at some time (T) in the past. The method estimates the likelihood function of the demographic parameters θ, T, and M (the number of migrants per generation between the two populations) given (1) the demographic model, (2) the entire set of data (i.e., DNA sequences) and the likelihood of their possible underlying genealogies calculated using a Markov chain Monte Carlo approach, and (3) a mutational model. In this case, we assumed the finite-site model of Hasegawa, Kishino, and Yano (1985), which allow for recurrent mutations, differences in nucleotide frequencies and a transition/transversion bias. We calculated the N_e from the estimated θ_{ML} using the formula $N_e = \theta_{ML}/2\mu$, where $\mu = 6.66 \times 10^{-6}$ mutations per site per generation (Sigurgardottir et al. 2000), and then we used this N_e values to convert the T values, scaled by N_e, in number of generations. Although in principle the method allows for a more general demographic model with different effective population sizes and asymmetric gene flow, the available version of the *mdiv* software released by the authors (http://www.biom.cornell.edu/Homepages/Rasmus_Nielsen/files.html) does not incorporate such complications. Anyway, the NW method is currently the only one that considers and distinguishes the effects of genetic drift and gene flow using

the entire set of data in a probabilistic framework. Considering the importance of gene flow in determining the genetic structure of human populations, the NW method represents an important progress toward the construction of population genetics models able to better fit the demographic history of human populations. Our analysis in this paper were limited to a subset of samples because calculations for each pair of populations usually require more than 12 h on a personal computer. We ran the program for each pair of populations twice using different seed numbers to check the convergence of the ergodic averages.

To investigate differences in demographic trends among males and females, we performed a major axis regression (MAR) of Y-chromosome–Φst on mtDNA-Φst for 15 pairs of populations studied for both markers (see Tarazona-Santos et al. [2001] for Y-chromosome data). Under the null hypothesis of equal differentiation of male and female lineages, the slope of the regression for the 15 pairs of populations should be 1. The Tayacaja, Arequipa, Cayapa, Wai-Wai, and Xavante populations were used for this analysis as well as a pooled sample of the Gaviao, Zoro, and Surui individuals since only aggregated data are available for Y-chromosome DNA. We used the MAR because it is more appropriate than least-square regression when measures of both variables are subject to errors (Sokal and Rohlf 1995). Y-chromosome data consist of haplotypes of six Y-linked microsatellites belonging to the Native American–specific haplogroup Q, which frequency in Amerindian populations is higher than 80%. This excludes a possible homogenizing effect introduced by

FIG. 1.—Geographic localization of the traditional south Amerindian populations studied for HVRI diversity. The western and eastern samples are represented respectively by filled squares and open squares. The Cayapa sample, located in the Amazonian region near the Andes, is denoted by a filled circle.

recent European gene flow. We also calculated the φ*st* among the six populations for mtDNA and Y-chromosome DNA. Finally, coalescent simulations were used to test some specific evolutionary scenarios (Hudson 1990; Excoffier, Novembre, and Schneider 2000).

Results

Genetic Structure of South Amerindian Populations

The three Andean populations exhibit the highest values of θ_k and the first, second, and fourth highest values of θ_s among all the traditional samples (table 1). In general, traditional populations in western South America (Andes and Chile, excluding the Amazonian Cayapa sample) show higher values of θ_k and θ_s than do populations of the Amazonian region or eastern South America (Kruskal-Wallis test: H(1,13) = 5.22, P = 0.022 for θ_k and H(1,13) = 4.00, P = 0.045 for θ_s).

Moreover, the three traditional Andean samples also have an excess of rare alleles, leading to negative and significant F_s values (table 1). Among the other 11 traditional populations, only the Yanomama shows a similar pattern.

Table 1 shows that urban collections of Native American HVRI sequences from admixed Brazilian cities (Alves-Silva et al. 2000) exhibit an excess of rare alleles and values of θ_k and θ_s similar to those of Andean traditional populations, in contrasts to the low level of diversity observed in native Brazilian populations. This could be due to recurrent and high levels of gene flow from surrounding native populations during the past 500 years,
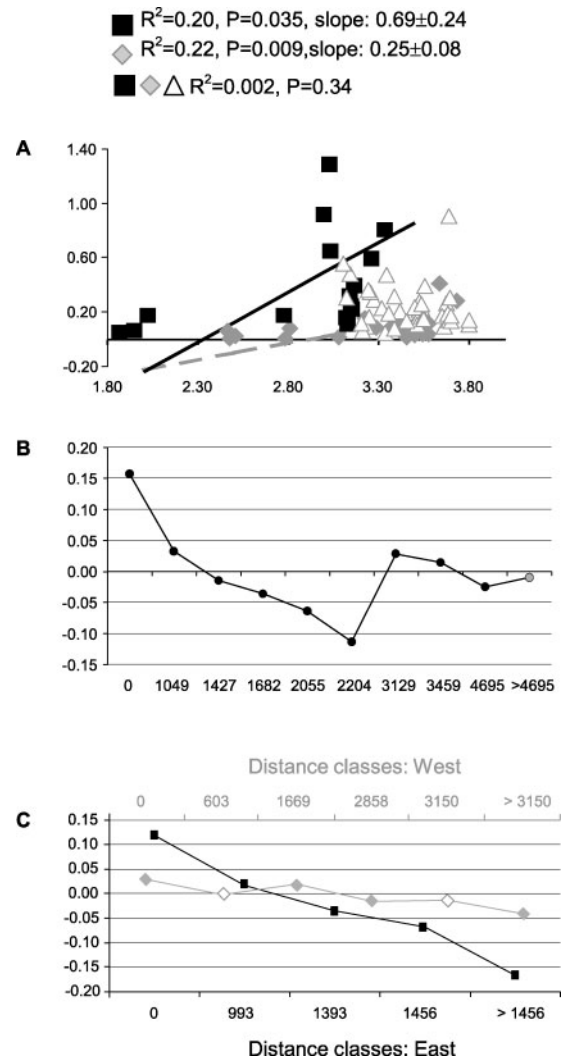

FIG. 2.—Results of the analysis of among-population genetic variability in south Amerindian groups. (A) Reduced major axis regression of the genetic distances ($\Phi st/1 - \Phi st$) on the logarithm of geographic distances, considering 14 traditional populations. The two different regression lines represent comparisons among western populations (diamond) and among eastern populations (square). The symbol Δ represents distances between western and eastern populations. (B and C) Spatial autocorrelation analysis in all of South America (B) and in the eastern (C, in black) and the western regions (C, in gray). The number and limits of geographic distances classes were chosen to have a similar and sufficient number of comparisons within each class. X-axis: higher limit of geographic distance classes (in kilometers) between localities. Y-axis: II index (significance assessed by a randomization test: filled symbols indicate $P < 0.01$; empty symbols indicate nonsignificant at the 1% level).

so that urban populations would be amalgams of the surrounding ones. This explanation is consistent with the results of Chakraborty, Smouse, and Neel (1988), showing that the consequence of population amalgamation is an excess of alleles (particularly rare ones) with respect to the mutation-drift equilibrium expectation.

When all traditional South American samples are considered together, there is no correlation between genetic and geographic distances (fig. 2A). AIDA shows that in the whole South American region, the II values decrease from positive at the first distance class to negative at the sixth

distance class (fig. 2*B*). However, in the seventh distance class, the *II* value increases, and the higher distance classes do not show any clear pattern of autocorrelation. These classes include comparisons between populations from the western region and between these populations and those sampled in the eastern region, but exclude comparisons within the latter area (see figure 1 for definition of the western and eastern samples). These results suggest that it is not possible to invoke a simple population genetic model that reasonably explains the spatial pattern of mtDNA diversity in South America. However, when the samples are divided into two groups, western and eastern populations, a straightforward geographic structure appears: (1) The logarithm of geographic distances explains approximately 20% of the genetic distances in both cases, and the scatter-plots of these variables fit a linear function (fig. 2*A*). Hence, the geographical pattern of mtDNA diversity in the western and eastern regions could be explained in part by the IBD model. (2) The correlogram including only the eastern populations shows an almost linear decrease of *II* values as the geographic distances increase (fig. 2*C*), indicating a clinal pattern. This pattern is typically generated by a population expansion followed by continuous founder effects (Sokal 1979). It should be pointed out that when the Cayapa sample is included in the eastern group, the association between genetic and geographic distances disappears, suggesting a complex genetic history for this population, including interactions with populations from western and eastern South America, as previously proposed (Tarazona-Santos et al. 2001; Rickards et al. 1999). (3) The correlogram of the western populations (fig. 2*C*) also shows decreasing values of *II*, but it has a smoother shape than the eastern populations correlogram and shows two nonsignificant *II* values at classes 2 and 5. Together with the lower slope of the regression line (fig. 2*A*), this can be interpreted as the effect of higher level of gene flow and larger range dispersion in the western region, which would have reduced population differentiation even at large geographic distances. This finding is in part consistent with data of Y-chromosome diversity, which shows no significant differentiation in western South America, but high values of Φst in the east that, in any case, do not fit any simple model of geographic population structure (Tarazona-Santos et al. 2001).

Dating the Split Between Central and South
Amerindian Populations

The maximum-likelihood estimators of the divergence times (*T*), number of migrants (*M*), and θ_{ML} for some Amerindian populations are shown in table 2. The corresponding likelihood surfaces are available in figure A of the Supplementary Material online. Because the genetic structure for mtDNA diversity fits the IBD model in the west and the east of South America, geographically closed populations likely contain redundant information when compared with the Panamanian populations. Therefore, we consider for the maximum-likelihood estimations, which are computationally intensive, four distant populations scattered across South America: San Martin, Mapuche 2, Yanomama, and Gaviao. The Xavante and Wai-Wai

populations were excluded from analysis because preliminary exploration showed that they do not contain enough information to produce reliable estimates of T, showing flat likelihood surfaces (data not shown). Interestingly, while values of *T* between South Amerindian populations range between 5,989 and 12,191 years, considering 27 years per generation (Tremblay and Vezina 2000), divergence times between south Amerindian populations and the Wuonan and Embera, settled in the Isthmus of Panama, are consistently higher, placed in the Pleistocene and, in seven of eight cases, fall between 14,900 and 19,000 years. These estimates already take into account the effect of gene flow and could correspond to the first peopling of South America by the ancestors of current populations.

Comparison Among mtDNA and Y-Chromosome Data

If the Φst values were nearly the same for paternal and maternal lineages, the slope of the regression line of Y-chromosome–φst on mtDNA-Φst should be 1. The calculated slope of the major axis regression line is 0.88 (fig. 3). Moreover, the Φst values for mtDNA and Y-chromosome DNA for the six considered samples are close (0.16 and 0.17 respectively). This suggests that among population divergence is similar for Y-chromosome DNA and mtDNA in South America, a picture that is not observed in other continents, where Φst values are higher for the male lineages (Seielstad, Minch, and Cavalli-Sforza 1998; Perez-Lezaun et al. 1999; Hammer et al. 2001).

Discussion

Estimators of θ based on the number of different alleles (θ_k) and segregating sites (θ_s) show that Andean populations have the highest within-population variability among traditional south Amerindian populations. Moreover, they show an excess of rare alleles with respect to the mutation-drift equilibrium expectation (i.e., negative and significant F_s values), while populations settled in eastern South America (with the only exception of the Yanomama) do not. Different non–mutually exclusive evolutionary factors could have contributed to this pattern of diversity. For instance, a population expansion produces an excess of rare alleles (Slatkin and Hudson 1991). Bonatto and Salzano (1997) have shown that the Pleistocenic peopling of the Americas from the Bering Strait was accompanied by a demographic expansion, whose signal should be present in current Native American populations, unless it had been canceled by more recent (i.e., Holocenic) population instability or bottlenecks typical of small hunter-gatherer populations (Excoffier and Schneider 1999). This could explain at least partially why Andean populations show an excess of rare alleles and traditional groups settled in eastern South America do not. Indeed, archaeological and historical records show that the Andean region hosts the largest South American populations at least since the development of complex societies around 4,000 years B.C. (Stanish 2001), whereas populations settled in the east have been more fragmented, have lower census sizes, and have historically achieved a lower level of socioeconomic development (Salzano and Callegari-Jacques 1988).

Table 2
Maximum-Likelihood Estimates of Demographic Parameters

	San Martin	Yanomama	Gaviao	Mapuche 2	Embera	Wuonan
San Martin	—	11,563	9,993	11,585	18,709	18,793
	—	(0.91)	(0.49)	(0.32)	(0.01)	(0.01)
Yanomama	9.25	—	12,773	6729	14,962	16,509
		—	(0.58)	(0.75)	(0.01)	(0.01)
Gaviao	6.87	5.62	—	5,989	17,018	15,187
			—	(0.26)	(0.01)	(0.01)
Mapuche 2	7.08	5.64	3.66	—	13,331	18,269
				—	(0.11)	(0.01)
Embera	9.8	7.98	5.85	6.11	—	2561
					—	(>3.5)
Wuonan	10.88	9.08	5.94	6.28	6.44	—

NOTE.—Above the diagonal: divergence times (T) in years and gene flow (M/2, in parentheses) among populations. Below the diagonal: θ_{ML}. Assuming 27 years per generation and a mutation rate of $\mu = 0.66 \times 10^{-5}$ per site per generation. Comparisons among Panamanian and south Amerindian populations are the top eight rows of the last two columns.

Ray, Currat, and Excoffier (2003) have recently studied how a spatial range expansion of populations influences the pattern of genetic diversity. If the spatial expansion is accompanied by large level of gene flow between populations (Nm > 20 [number of migrants per generation]), it also generates a pattern of diversity characterized by an excess of rare alleles, but this is not true if the expansion is accompanied by a small number of migrants per generation (Nm < 20). Our results are also compatible with a scenario of range expansion associated with a high number of migrants in the Andean area and a low number of migrants in the eastern South America, which is also consistent with the observed correlograms for the western and eastern regions of the continent (see *Results* and see below). On the other hand, Mishmar et al. (2003) have recently shown that natural selection has shaped regional mtDNA variation in humans. MtDNA contains genes involved in the oxygen metabolism, whose normal function should be important in a hypoxic environment such as the Andean highlands. Because mtDNA does not recombine, natural selection acting on the coding region will also affect the HVRI by a hitchhiking effect or background selection, which determine an allelic spectrum with an excess of rare alleles (Kreitman 2000). Thus, we cannot exclude that natural selection had also contributed to determine the excess of rare alleles observed in Andean populations.

In our previous analysis of classical markers, we found no barriers to gene flow in the Andean region (Luiselli et al. 2000), and Y-chromosome diversity is high in the Andes (Tarazona-Santos et al. 2001), as well as mtDNA diversity (present work). These consistent results, obtained from different loci, suggest that the observed pattern of diversity of mtDNA is at least partially due to different demographic histories of Andean and non-Andean populations, rather than to selective forces. Therefore, it seems that despite the environmental stresses posited by the high altitude environment (hypobaric hypoxia, cold, and a poor nutritional environment [Tarazona-Santos et al. 2000]), the Andean region has maintained a larger effective population sizes and levels of gene flow than other regions of the continent.

To further evaluate which combination of evolutionary factors have generated the observed pattern of within-population variability, we performed computer simulations based on coalescent processes. We generated samples of populations under three scenarios (fig. 4). They roughly match demographic histories proposed for south Amerindian populations. For each scenario we considered three current demes, and we generated 1,000 sets of samples composed of 40 genes with 300 potentially variable DNA sites for each deme. We consider a mutation rate of 6.66×10^{-6} per site per generation (Sigurgardottir et al. 2000). In each of the three demographic scenarios, an ancestral population experiences a large stepwise demographic expansion (with a 100-fold increase of population size) 1,425 generations before present (which could correspond to the expansion associated with the first colonization of the Americas [Excoffier and Schneider 1999; Silva et al. 2002]). Then, 425 generations ago, this population split into three demes, which exchanged migrants at a low rate (Nm = 0.01) for 400 generations. This phase corresponds to the differentiation of Native American populations. Finally, for the past 25 generations, two demes sent migrants to the third at a rate of 1%, representing the unidirectional migration towards urban centers occurring during the past 500 years after the European Conquest. In the first demographic scenario, the sizes of the demes after
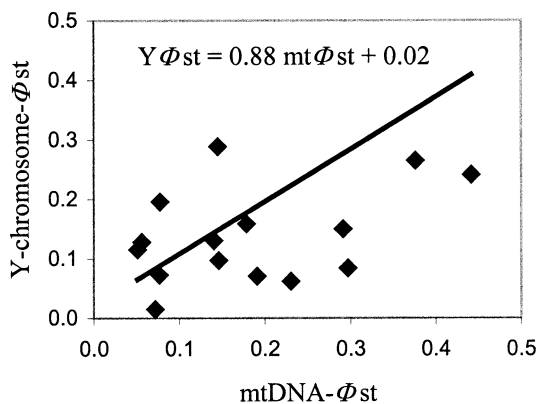

$$Y\Phi st = 0.88\ mt\Phi st + 0.02$$

FIG. 3.—Major axis regression of Y-chromosome–Φst on mtDNA-Φst.

	Scenario 1, N=5000		Scenario 2, N=1000		Scenario 3, N=1000	
	M. S. P.	Urban center	M. S. P.	Urban center	M. S. P.	Urban center
θ_S	8.64 (0.07)	10.98 (1.67)	3.27 (0.05)	5.09 (1.20)	3.98 (0.03)	6.66 (1.45)
θ_π	5.04 (0.00)	5.22 (0.95)	2.97 (0.02)	3.59 (1.14)	3.83 (0.04)	4.62 (1.13)
θ_k	20.84 (0.00)	41.04 (17.68)	4.15 (0.10)	9.15 (3.42)	4.39 (0.01)	10.54 (3.78)
F_S	-9.87 (0.07)	-17.94 (4.73)	-1.17 (0.11)	-4.87 (2.80)	-0.25 (0.04)	-4.14 (2.47)
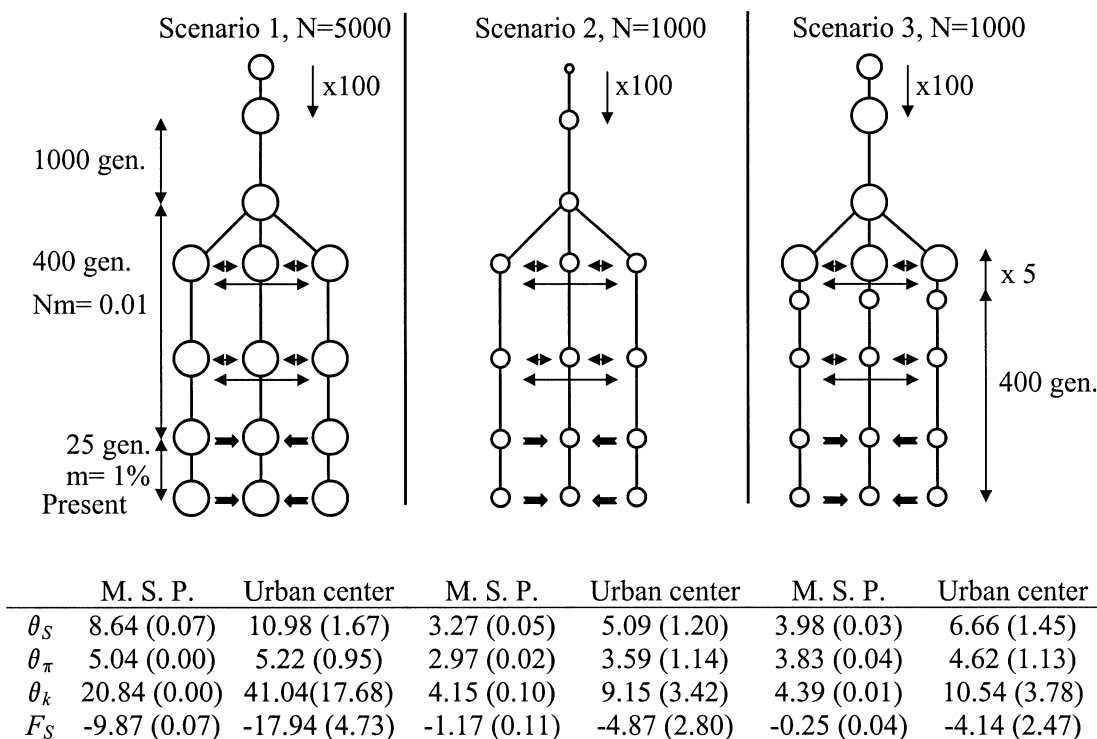
FIG. 4.—Three simulated scenarios representing three possible histories for the south Amerindian populations. For each scenario means and standard deviations (SD) of calculated statistics are presented. See the text for explanations of the evolutionary scenarios. M. S. P. = mean of source populations.

the expansion were 5,000 haploid individuals. In the second scenario, the only difference is a smaller size of the demes after the expansion (1,000 haploid individuals). In the third scenario, the size of the demes after the expansion is also 1,000, but, unlike the second scenario, the three demes experience a bottleneck (with a fivefold decrease of population size) soon after their divergence 400 generations ago, representing the demographic instability of hunter-gatherer populations during the Holocene (Excoffier and Schneider 1999). We propose that the first scenario (higher N_e) best matches the history of the western populations, whereas the second and third scenarios mimic the history of the eastern groups. Figure 4 shows the mean and standard deviation of three estimates of the θ parameter (θ_S, θ_π, and θ_k) and the Fu's F_S statistics for the different sets of simulations. Comparison of the three scenarios reveals a relative excess of low-frequency alleles in the first scenario, as indicated by the high θ_k and the large negative F_S values. This excess seems to be the consequence of the larger size of the demes after the expansion 1,425 generations ago. This differential pattern of genetic diversity mimics the differences among western and eastern south Amerindian populations, supporting our inferences about the higher long-term effective population sizes of the former. On the other hand, the comparison between the simulated urban center and the traditional populations shows an excess of rare alleles across the three scenarios, indicated by negative F_S statistics and the difference between θ_k and the other estimators of θ, in agreement with Chakraborty, Smouse, and Neel (1988).

This suggests that the high diversity and excess of rare alleles observed in Brazilian urban samples are likely due to recent intensive gene flow from surrounding populations and not to a high long-term effective population size. In general, the results for the sampled populations and those for the simulated samples (compare table 1 and figure 4) are similar. Although some of the demographic parameters used in our simulations are arbitrary (as in most simulations performed in population genetics studies), they are realistic. Even if our results do not exclude alternative scenarios, they clearly suggest that our interpretation is compatible with the actual demographic history of south Amerindian populations.

When South Amerindian samples are separated in two groups, western and eastern populations, a straightforward geographic structure appears. Moreover, mtDNA diversity shows a higher degree of spatial organization than male lineages. In fact, no clear spatial pattern of genetic structure was identified for Y-chromosome diversity in the western and eastern regions (Tarazona-Santos et al. 2001). Considering western and eastern populations, regression analysis suggests that mtDNA diversity fits the IBD model, whereas autocorrelation analysis suggests a set of founder effects after a population expansion in both regions. These results are not incompatible. Even under IBD, correlograms tend to show negative autocorrelation at large distances when migration rates are low or short-distance migration predominates (Hardy and Vekemans 1999), which seems to be the case in eastern South America. It is also possible that the

history of eastern and western South Amerindian populations involved a combination of population expansion, founder effects, and subsequent IBD.

The Pleistocenic versus Holocenic settlement of South America is still one of the more controversial points of American archeology. We have used the estimated divergence times between south Amerindian and Panamanian populations to date, for the first time on the basis of genetic data, the divergence of these groups. We included in these comparisons populations settled in the northernmost area of distribution of the Chocó linguistic family, whose languages are spoken in northern South America. In particular, the ancestors of Embera and Wuonan populations were probably settled in South America (Kolman and Bermingham 1997). Three features of the Nielsen-Wakeley (NW) method should be discussed. First, it assumes that the two considered populations do not differ in their effective population sizes. In our calculations, for each pair (A, B) of populations considered, their intervals $\theta_\pi \pm$ SD for A and B overlap as do the intervals $\theta_s \pm$ SD. Moreover, the θ_π statistics for populations A and B fall in the 95% confidence intervals of θ_{ML}. Second, because the NW method uses the entire set of data to calculate the likelihood of the possible underlying genealogies, the effect of a possible demographic expansion (i.e., the concentration of coalescent events in a short time frame and the excess of rare alleles) is incorporated in the model. For this reason, the inclusion of the San Martin and Yanomama populations, which show a significant excess of rare alleles, should not affect a reliable estimation of divergence times. Third, the NW method gives the divergence times in units of N_e. To transform them into number of years or generations (see *Materials and Methods*), we need to assume a mutation rate for the mtDNA. The actual mutation rate of the mtDNA is currently subject to controversy. In this study, we have used a mutation rate that is intermediate among those based on pedigrees (Sigurgardottir et al. 2000), that are usually higher than estimates based on the molecular clock. If future studies show that the HVRI of mtDNA has even higher mutation rate, our divergence times will be reduced. However, because pedigrees that do not show mutations are seldom published, it is possible that mutation rates obtained from pedigrees were biased and overestimate the actual mutation rate. Altogether, we think that the six considered populations from South America and Panama fit reasonably the assumptions of the NW model, and the considered mutation rate is realistic. Although the obtained ML estimates are subject to statistical uncertainty, two important features emerge: (1) Divergence times among South Amerindian populations are lower than those among Panamanian and South Amerindian populations, but gene flow is higher for the former comparisons and lower for the latter. (2) All ML divergence times between Panamanian and South Amerindian populations consistently fall in the Pleistocene. If we assume that ancestors of South Amerindian populations passed across Panama only once (Kolman and Bermingham 1997), following a demographic process that reasonably fits a model of a spatial southward population expansion, and then entered into South America (where they may have followed a different migration pattern), the obtained estimates of divergence times

between south Amerindian and Panamanian populations could correspond to the first peopling of South America by the ancestors of current native populations. Under this model, our results give independent support to archeological evidences claiming an early peopling of South America and suggest that ancestors of the present-day south Amerindians were likely in South America at that time. Interestingly, the likelihood surface is smoother at the right than at the left of ML estimations of T (see Supplementary Material online), suggesting that higher times of divergence are actually more likely than shorter ones.

MtDNA and Y-chromosome diversity show similar φ*st* values. The observed Φst is determined by demographic factors as well as mutational rates and mechanisms. To clarify the roles of demographic factors and mutation in our results, we performed coalescent simulations. We considered two demographic scenarios in which two populations of 5,000 haploid individuals (sample size = 40) diverged, exchanging $N_em = 0.1$ (scenario 1) and $N_em = 0.01$ (scenario 2) migrants per generation. For each scenario, we simulated DNA sequences and haplotypes of six linked microsatellites and then calculated the φ*st* between the simulated populations, repeating the simulations 1,000 times. In the case of DNA sequences, we assumed a simple finite-sites mutation model with a 0.9 transition bias and a mutation rate of 6.6×10^{-6} substitutions per site per generation (Sigurgardottir et al. 2000). For simulations of the six linked microsatellite haplotypes, we assumed a pure stepwise model without size constrains with a mutation rate of 2×10^{-4} mutations per locus per generation (Kayser et al. 2000). The mean and standard deviation (SD) of Φst obtained from simulations for mtDNA were 0.85 ± 0.11 and 0.58 ± 0.20 for N_em values of 0.01 and 0.1, respectively. For simulations of Y chromosomes, we obtained $\Phi st = 0.90 \pm 0.15$ for $N_em = 0.01$ and $\Phi st = 0.61 \pm 0.24$ for $N_em = 0.1$. Therefore, considering the different mutational rates and mechanism but the same demographic history, we expect the ratio Y-chromosome–Φst/mtDNA-Φst to be approximately 1.05. From the reduced major axis regression, we observed a slightly lower ratio: 0.88. These results suggest that the among-population differentiation is similar (or higher) for mtDNA than for Y chromosome in South America. Our results are consistent with those of Mesa et al. (2000) and show for the first time that the observed pattern is probably due to demographic factors and not to the different mutational rates and mechanisms of maternal and paternal lineages.

In conclusion, we propose that ancestors of current south Amerindians would have arrived in South America during the Pleistocene, as suggested by archeology (Dixon 1999; Dillehay 2000). Although the routes of dispersion are not clear, autocorrelation and genetic distance analysis suggest that they expanded and founded new populations and, at certain times, differential patterns of gene flow were established in the western and the eastern parts of the continent. These patterns fit the IBD model in both regions, at least for female lineages. However, the intensity of gene flow, the range dispersion, and the effective population sizes were higher in the western region than in the eastern region, and these differences were likely related

to the higher level of cultural and linguistic homogeneity of the Andean area, which has been involved in a unique cultural process for the past 12,000 years (Tarazona-Santos et al. 2001). Although this depiction is an oversimplification of more complicated phenomena, it is a useful model that can be tested using genetic data. Further analysis of multiple genomic regions would be the best way to test whether the proposed evolutionary scenario resembles important demographic events in the history of these human populations. On the other hand, more populations should be studied to test whether, at a certain point between the continental and microgeographic levels, the proposed model becomes inadequate to explain the pattern of genetic diversity of south Amerindian populations.

Acknowledgments

Literature Cited

Alves-Silva, J., M. da Silva Santos, P. E. Guimaraes, A. C. Ferreira, H. J. Bandelt, S. D. Pena, and V. F. Prado. 2000. The ancestry of Brazilian mtDNA lineages. Am. J. Hum. Genet. **67**:444–461.

Anderson, S., A. T. Bankier, B. G. Barrell et al. (11 co-authors). 1981. Sequence and organization of the human mitochondrial genome. Nature **290**:457–465.

Bertorelle, G., and G. Barbujani. 1995. Analysis of DNA diversity by spatial autocorrelation. Genetics **140**:811–819.

Bianchi, N. O., C. I. Catanesi, G. Bailliet, V. L. Martinez-Marignac, C. M. Bravi, L. B. Vidal-Rioja, R. J. Herrera, and J. S. Lopez-Camelo. 1998. Characterization of ancestral and derived Y-chromosome haplotypes of New World native populations. Am. J. Hum. Genet. **63**:1862–1871.

Bohonak, A. J. 2002. IBD (isolation by distance): a program for analyses of isolation by distance. J. Hered. **93**:153–154.

Bonatto, S.-L., and F. M. Salzano. 1997. Diversity and age of the four major mtDNA haplogroups, and their implications for the peopling of the New World. Am. J. Hum. Genet. **61**:1413–1423

Cavalli-Sforza, L. L., P. Menozzi, and A. Piazza. 1994. The history and geography of human genes. Princeton University Press, Princeton, N.J.

Chakraborty, R., P. E. Smouse, and J. V. Neel. 1988. Population amalgamation and genetic variation: observations on artificially agglomerated tribal populations of Central and South America. Am. J. Hum. Genet. **43**:709–725.

Dillehay, T. D. 2000. The settlement of the Americas: a new prehistory. Perseus Books Group, New York.

Dixon, J. E. 1999. Bones, boats and bison: archeology and the first colonization of western North America. The University of New Mexico Press, Albuquerque.

Ewens, W. J. 1972. The sampling theory of selectively neutral alleles. Theor. Popul. Biol. **3**:87–112.

Excoffier, L., and S. Schneider. 1999. Why hunter-gatherer populations do not show signs of Pleistocene demographic expansions. Proc. Natl. Acad. Sci. USA **96**:10597–10602.

Excoffier, L., J. Novembre, and S. Schneider. 2000. SIMCOAL: a general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. J. Hered. **91**:506–509.

Fu, Y. X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics **147**:915–925.

Ginther, C., D. Corach, G. A. Penacino, J. A. Rey, F. R. Carnese, M. H. Hutz, A. Anderson, J. Just, F. M. Salzano, and M. C. King. 1993. Genetic variation among the Mapuche Indians from the Patagonian region of Argentina: mitochondrial DNA sequence variation and allele frequencies of several nuclear genes. EXS **67**:211–219.

Hammer, M. F., T. M. Karafet, A. J. Redd, H. Jarjanazi, S. Santachiara-Benerecetti, H. Soodyall, and S. L. Zegura. 2001. Hierarchical patterns of global human Y-chromosome diversity. Mol. Biol. Evol. **18**:1189–1203.

Hardy, O. J., and X. Vekemans. 1999. Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. Heredity **83**:145–154.

Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **22**:160–174.

Helgason, A., S. Sigurgardottir, J. R. Gulcher, R. Ward, and K. Stefansson. 2000. mtDNA and the origin of the Icelanders: deciphering signals of recent population history. Am. J. Hum. Genet. **66**:999–1016.

Hudson, R. R. 1990. Gene genealogies and the coalescent process. Pp. 1–44 *in* D. J. Futuyma and J. D. Antonovics, eds. Oxford surveys in evolutionary biology. Oxfords University Press, New York.

Kayser, M., L. Roewer, M. Hedman et al. (11 co-authors). 2000. Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. Am. J. Hum. Genet. **66**:1580–1588.

Kolman, C., and E. Bermingham. 1997. Mitochondrial and nuclear DNA diversity in the Choco and Chibcha Amerinds of Panama. Genetics **147**:1289–1302.

Kreitman, M. 2000. Methods to detect selection in populations with applications to the human. Annu. Rev. Genomics Hum. Genet. **1**:539–559.

Luiselli, D., L. Simoni, E. Tarazona-Santos, S. Pastor, and D. Pettener. 2000. Genetic structure of Quechua-speakers of the Central Andes and geographic patterns of gene frequencies in south Amerindian populations. Am. J. Phys. Anthropol. **113**:5–17.

Merriwether, D. A., B. M. Kemp, D. E. Crews, and J. V. Neel. 2000. Gene flow and genetic variation in the Yanomama as revealed by mitochondrial DNA. Pp. 84–124 *in* C. Renfrew, ed. America past, America present: genes and languages in the Americas and beyond. The McDonalds Institute for Archeological Research, Cambridge.

Merriwether, D. A., F. Rothhammer, R. E. Ferrell. 1995. Distribution of the four founding lineage haplotypes in Native Americans suggests a single wave of migration for the New World. Am. J. Phys. Anthropol. **98**:411–430.

Mesa, N. R., M. C. Mondragon, I. D. Soto et al. (13 co-authors). 2000. Autosomal, mtDNA, and Y-chromosome diversity in Amerinds: pre- and post-Columbian patterns of gene flow in South America. Am. J. Hum. Genet. **67**:1277–1286.

Meyer, S., G. Weiss, and A. von Haeseler. 1999. Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. Genetics **152**: 1103–1110.

Mishmar, D., E. Ruiz-Pesini, P. Golik et al. (13 co-authors). 2003. Natural selection shaped regional mtDNA variation in humans. Proc. Natl. Acad. Sci. USA **100**:171–176.

Moraga, M. L., P. Rocco, J. F. Miquel, F. Nervi, E. Llop, R. Chakraborty, F. Rothhammer, and P. Carvallo. 2000. Mitochondrial DNA polymorphisms in Chilean aboriginal populations: implications for the peopling of the southern cone of the continent. Am. J. Phys. Anthropol. **113**:19–29.

Neves, W. A., and H. M. Pucciarelli. 1991. Morphological affinities of the first Americans: an exploratory analysis based on early South American human remains. J. Hum. Evol. **21**:261–273.

Nielsen, R., and J. Wakeley. 2001. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. Genetics **158**:885–896.

Perez-Lezaun, A., F. Calafell, D. Comas et al. (12 co-authors). 1999. Sex-specific migration patterns in Central Asian populations, revealed by analysis of Y-chromosome short tandem repeats and mtDNA. Am. J. Hum. Genet. **65**:208–219.

Powell, J. F., and W. A. Neves. 1999. Craniofacial morphology of the first Americans: pattern and process in the peopling of the New World. Am. J. Phys. Anthropol. **29**(suppl):153–188.

Ray, N., M. Currat, and L. Excoffier. 2003. Intra-deme molecular diversity in spatially expanding populations. Mol. Biol. Evol. **20**:76–86.

Rickards, O., C. Martinez-Labarga, J. K. Lum, G. F. De Stefano, and R. L. Cann, 1999. mtDNA history of the Cayapa Amerinds of Ecuador: detection of additional founding lineages for the Native American populations. Am. J. Hum. Genet. **65**:519–530.

Rodriguez-Delfin L. A., V. E. Rubin-de-Celis, and M. A. Zago. 2001. Genetic diversity in an Andean population from Peru and regional migration patterns of Amerindians in South America: data from Y chromosome and mitochondrial DNA. Hum. Hered. **51**:97–106.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics **145**:1219–1228.

Salzano, F. M., and S. Callegari-Jacques. 1988. South American Indians: a case study in evolution. Pp. 259 *in* Research monographs on human population biology. Clarendon Press, Oxford.

Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin: a software for population genetics data analysis. Version 2.000. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Seielstad, M. T., E. Minch, and L. L. Cavalli-Sforza. 1998. Genetic evidence for a higher female migration rate in humans. Nat. Genet. **20**:278–280.

Sigurgardottir, S., A. Helgason, J. R. Gulcher, K. Stefansson, and P. Donnelly. 2000. The mutation rate in the human mtDNA control region. Am. J. Hum. Genet. **66**:1599–1609.

Silva, W. A. Jr., S. L. Bonatto, A. J. Holanda et al. (11 co-authors). 2002. Mitochondrial genome diversity of Native Americans supports a single early entry of founder populations into America. Am. J. Hum. Genet. **71**:187–192.

Slatkin, M., and R. R. Hudson. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics. **129**:555–562.

Sokal, R. 1979. Ecological parameters inferred from spatial correlograms. Pp. 167–196 *in* G. Patil and M. Rozenzweig, eds. Contemporary quantitative ecology and related econometrics. International Co-operative Publishing House, Fairland, Md.

Sokal, R. R., and F. J. Rohlf. 1995. Biometry: the principles and practice of statistics in biological research, 3rd edition. Freeman, New York.

Stanish, C. 2001. The origin of state societies in South America. Annu. Rev. Anthropol. **30**:41–64.

Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics **105**:437–460.

Tamura, K., and M. Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. **10**: 512–526.

Tarazona-Santos, E., D. R. Carvalho-Silva, D. Pettener, D. Luiselli, G. F. De Stefano, C. M. Labarga, O. Rickards, C. Tyler-Smith, S. D. Pena, and F. R. Santos. 2001. Genetic differentiation in south Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. Am. J. Hum. Genet. **68**:1485–1496.

Tarazona-Santos, E., M. Lavine, S. Pastor, G. Fiori, and D. Pettener. 2000. Hematological and pulmonary responses to high altitude in Quechuas: a multivariate approach. Am. J. Phys. Anthropol. **111**:165–176.

Tremblay, M., and H. Vezina. 2000. New estimates of intergenerational time intervals for the calculation of age and origins of mutations. Am. J. Hum. Genet. **66**:651–658.

Ward, R. H., F. M. Salzano, S. L. Bonatto, M. H. Hutz, C. E. A. Coimbra Jr., and R. V. Santos. 1996. Mitochondrial DNA polymorphism in three Brasilian Indian tribes. Am. J. Hum. Biol. **8**:317–372.

Watterson, G.-A. 1975. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7**:256–276.

Wright, S. 1943. Isolation by distance. Genetics **28**:114–138.