

Molecular Variability of the 16p13.3 Region in Amerindians and its Anthropological Significance

J. Battilana¹, L. Cardoso-Silva¹, R. Barrantes², K. Hill³, A. M. Hurtado³, F. M. Salzano¹ and S. L. Bonatto^{4,*}

¹Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil

²Sección de Genética Humana, Instituto Nacional de Investigaciones en Salud (INISA), Universidad de Costa Rica, San José, Costa Rica

³Department of Anthropology, University of New México, Albuquerque, New Mexico 87131, USA

⁴Centro de Biologia Genômica e Molecular, Faculdade de Biociências, Pontifícia Universidade Católica do Rio Grande do Sul, 90619-900 Porto Alegre, RS, Brazil

Summary

A total of 1558 base pairs in the 16p13.3 region were investigated in 98 individuals of Mongolian, Northern Arctic and Amerindian affiliation, and the results compared with those obtained in a previous worldwide study of the same genomic region. Fifty-five polymorphic sites could be classified into thirty-five haplotypes from the total data. A median joining network based on the haplotypes revealed two distinct clusters: one with low diversity, with haplotypes found in all five geographic-ethnic categories; while the other, with the most divergent haplotypes, was composed mainly of Africans and a few Amerindians. Almost all neutrality parameters yielded significantly negative values. Demographic simulations with the exclusively Amerindian dataset rejected all scenarios, including a bottleneck beginning more than 12,000 years ago. The demographic scenarios tested considering population growth were similar among the Amerindian and worldwide or Eurasian data sets. The results suggest that Amerindians are a representative sample of Eurasian populations, preserving the signal of demographic growth from the out of Africa exodus and, together with data from uniparental markers, support a scenario of a bottleneck of moderate intensity during the peopling of the New World.

Keywords: 16p13.3, Amerindians, molecular variability

Introduction

Studies of human genetic variation provide a powerful tool for elucidating the genetic, evolutionary, and demographic factors shaping the human genome. Considerable work has accumulated over the last decades documenting DNA sequence variation in humans. Early studies focused primarily on mitochondrial DNA

(mtDNA) (e.g. Vigilant *et al.* 1991; Ingman *et al.* 2000) and the Y chromosome (e.g. Hammer, 1995; Whitfield *et al.* 1995; Underhill *et al.* 2000). More recent single-locus studies have focused on the X chromosome (e.g. Nachman *et al.* 1998; Harris & Hey, 1999; Kaessmann *et al.* 1999; Nachman & Crowell, 2000; Gilad *et al.* 2002; Hammer *et al.* 2004; Garrigan *et al.* 2005) and autosomes (reviewed in Przeworski *et al.* 2000; Excoffier, 2002). Particularly for autosomal noncoding regions the number of surveys is still low, but they are beginning to reveal more complex demographic events than those revealed by uniparental markers (Yu *et al.* 2002). In relation to Amerindians the number of surveys and diversity of markers has been growing (reviewed in Salzano, 2002;

*Correspondence to: Sandro L. Bonatto, Centro de Biologia Genômica e Molecular, Faculdade de Biociências, PUCRS, 90619-900 Porto Alegre, RS, Brazil. Phone +55 51 3320-3500 Ext. 4717, Fax+55-51-3320-3612. E-mail address: slbonatto@puers.br

Bortolini *et al.* 2003; Dornelles *et al.* 2004; Schmitt *et al.* 2004; Kohlrausch *et al.* 2005; Mateus-Pereira *et al.* 2005). However, there are very few studies describing sequence variability in nuclear loci in populations from this continent, and most only screened a few individuals from a couple of populations. Two exceptions are Heller *et al.* (2004) and Fagundes *et al.* (2005) who studied the genetic variability in the 3'-UTR of the *LDLR* gene from many individuals of African, Asian, European and especially Amerindian ancestry. They found a lower value of genetic diversity in Amerindians as compared to other populations, suggesting a single origin and a bottleneck during the peopling of the American continent. However, neutrality statistic tests point to the possible action of worldwide balancing selection at this locus.

Alonso & Armour (2001) studied a region localized between the minisatellite loci D16S309 (MS205) and D16S83 (EKMDA2), immediately flanking the 5' end of minisatellite MS205 at 16p13.3. This region contains the first nine exons of the α 1H subunit of a human voltage-dependent T-type Ca^{2+} calcium channel (*CACNA1H*) (Cribbs *et al.* 1998). The sequence analysed (about 1.5 kb) maps inside intron 2 (which is approximately 41 kb long) of this gene, about 35 kb downstream of exon 2 and 6 kb upstream of exon 3, and is G+C rich (65%). The estimated average substitution rate per site per year found by Alonso & Armour (2001) (2.19×10^{-9}), obtained from 100 chromosomes sampled from different African and Eurasian populations, was higher than most other estimates in the literature (e.g. Harding *et al.* 1997; Harris & Hey, 1999; Jaruzelska *et al.* 1999; Kaessmann *et al.* 1999; Nachman & Crowell, 2000). Alonso & Armour (2001) found that this region is highly polymorphic and could constitute a very informative source for investigating recent human evolution. In addition, it maps to a region of high recombination which may help to shield it from the distorting effects of genetic hitchhiking or background selection. Badge *et al.* (2000) had previously studied this region and described a recombinational hot spot in the 85 kb separating the 3' end of minisatellite MS205 and the 5' end of minisatellite EKMDA2. Recombination, however, does not seem to disrupt the reconstruction of the evolutionary history of this region, as Alonso & Armour (2001) estimated the recombination rate to be 0

in this specific sequence. As the region analyzed is short (1.56 kb) this also reduces the chance of it containing a recombinational hot spot.

We have sequenced 1558 base pairs of this region in a set of different ethnic Amerindian and Asian groups and integrated these data with those of Alonso & Armour (2001). Our objectives were to investigate the molecular variability of this region, to make inferences about the early peopling of the Americas, and also to evaluate if the extended dataset would shed new light on the process of human continental diversification.

Subjects and Methods

Samples

Information about the studied populations is presented in Table 1. A total of 98 individuals have been tested: 4 Mongolians (2 Khalkh, 2 Khoton), 4 Northern Arctic (2 Chukchi, 2 Eskimo), 12 Central Americans (from six tribes) and 78 South Americans (affiliated with 17 tribes or ethnic groups). As indicated by the geographical coordinates, they are spread all over Central and South America, while non-Americans are located in places from which the prehistoric colonization of the New World has presumably taken place. Wide linguistic representation is also found among the Central and South Americans.

DNA Extraction, PCR Amplification, and DNA Sequencing

The region analyzed includes the first 1558 base pairs (bp) of the 1742bp sequence studied by Alonso & Armour (2001). The genomic DNA samples were extracted from plasma and glycerolized red blood cells stored in Porto Alegre, using the QIAamp DNA Mini Kit (Qiagen). Six primer pairs were designed with overlapping regions to amplify the 1558 base pairs (one of them designed by Alonso & Armour, 2001; primer sequences are available on request). Amplification was performed using 10–50 ng of genomic DNA, 1.5 mM of MgCl_2 , 0.2 mM of each dNTP, 0.2 μM of each primer, and 0.5 U of *Taq* DNA polymerase (Ampli-Taq Gold [Applied Biosystems] or Platinum [Invitrogen Life Technologies]). Cycle conditions were 94°C for

Table 1 Sample size, geographic location, and linguistic information for the studied populations

Population	No. of individuals	Geographic location	Linguistic family ^a	References for further information
South Americans				
Aché	5	23° 30'–24° 10' S; 55° 50'–56° 30' W	Guayaki	Brown <i>et al.</i> (1974); Hill & Hurtado (1996)
Ayoreo	5	16–22° S; 58–63° W	Ayore	Dornelles <i>et al.</i> (2004)
Arara	3	3° 30' S; 53° W	Arara	Salzano <i>et al.</i> (1991)
Cinta Larga	4	9° 50'–12° 30' S; 59° 10'–60° 50' W	Tupi Mondé	Callegari-Jacques <i>et al.</i> (1994)
Foz do Içana	3	1° N; 67° 30' W	Cubeo, Tucano, Baniwa, Tariana	Salzano <i>et al.</i> (1986)
Gavião	4	10° 10' S; 61° 08' W	Tupi Mondé	Hutz <i>et al.</i> (1997); Andrade <i>et al.</i> (2000)
Guarani	3	23° 6' S, 55° 12' W; 23° 12' S, 55° 6' W 23° 48' S, 54° 30' W	Guarani	Salzano <i>et al.</i> (1997)
Kaingang	4	27° 20' S; 52° 45' W	Ge-Kaingan	Salzano <i>et al.</i> (1980)
Lengua	3	23° S; 56° W	Macro-Panoan	Brown <i>et al.</i> (1974); Goicoechea <i>et al.</i> (2001)
Pacaás Novos	7	11° 8' S; 65° 5' W	Chapacura	Salzano <i>et al.</i> (1985)
Parakanã	14	5° S, 50° 10' W; 4° 30' S, 50° W 5° 55' S; 52° 42' W	Tupi	Black <i>et al.</i> (1988)
Quechua	5	12° 33' S, 75° 83' W; 16° 38' S, 71° 52' W	Quechua	Tarazona-Santos <i>et al.</i> (2001)
Suruí	4	10° 50' S; 61° 10' W	Tupi Mondé	Hutz <i>et al.</i> (1997); Andrade <i>et al.</i> (2000)
Waiãpi	3	1° N; 53° W	Oyampi	Mestriner & Salzano (1998)
Wai Wai	3	0° 40' S; 58° W	Carib	Callegari-Jacques <i>et al.</i> (1996)
Xavante	4	13° 20' S; 51° 40' W	Gê	Coimbra <i>et al.</i> (2002)
Zoró	4	10° 20' S; 60° 20' W	Tupi Mondé	Heidrich <i>et al.</i> (1995); Andrade <i>et al.</i> (2000)
Central Americans				
Bribri	2	9° 38' N; 82° 50' W	Chibchan	Barrantes (1993)
Cabecar	2	9° 26' N; 83° 09' W	Chibchan	Barrantes (1993)
Guatuso	2	10° 40' N; 84° 49' W	Chibchan	Barrantes (1993)
Guaymi	2	8° 13' N; 82° 57' W	Chibchan	Barrantes (1993)
Huetar	2	9° 53' N; 84° 14' W	Chibchan	Barrantes (1993)
Teribe	2	9° 20' N; 82° 35' W	Chibchan	Barrantes (1993)
Northern Arctic				
Chukchi	2	64° N; 175° W	Chukot	Erdesz <i>et al.</i> (1994); Krylov <i>et al.</i> (1995)
Eskimo	2	64° N; 175° W	Yupic	Erdesz <i>et al.</i> (1994); Krylov <i>et al.</i> (1995)
Mongolians				
Khalkh	2	46° N; 106° E	Altaic	Kolman <i>et al.</i> (1996)
Khoton	2	45° N; 94° E	Qotong	Munkhbat <i>et al.</i> (1997)

^aAccording to Campbell (1997).

1 min, 62°C for 2 min, and 72°C for 2 min, with an initial denaturing step of 94°C for 1 min and a final extension step of 72°C for 10 min. The PCR products were purified with exonuclease I and alkaline phosphatase (Amersham Biosciences) and sequenced on both strands using the amplification primers. Sequencing was performed with the DYEnamic ET Dye

Terminator Kit (MegaBACE, Amersham Biosciences) as instructed and read in a MegaBace1000 (Amersham Biosciences) automated system. All determinations which indicated possible variants were confirmed by exhaustive re-sequencing, including for some samples sequencing in an ABI Prism 3100 (Applied Biosystems) machine.

Haplotype Assignment

Samples with confirmed multiple variants had their haplotypes determined experimentally by cloning the PCR products with the Topo TA Cloning Kit for Sequencing (Invitrogen Life Technologies). Plasmid DNA extraction was carried out according to Sambrook & Russell (2001). Sequencing of the cloned fragments was performed as indicated above.

Data Analysis

All chromatograms were manually checked using the CHROMAS 1.45 (www.technelysium.com.au/index.html) program, and all sequences were manually aligned using BIOEDIT 6.0.7 (www.mbio.ncsu.edu/BioEdit/bioedit.html). To test departures from the neutral model, such as selection or population change, Tajima's D (Tajima, 1989), Fu & Li's D^* and F^* (Fu & Li, 1993) and Fu's F_s (Fu, 1997) statistics were calculated; additionally, diversity parameters such as haplotype (H_d) and nucleotide (π) diversity (Nei, 1987), as well as theta (θ) (Watterson, 1975), were obtained, all employing the DnaSP 4.0 (www.ub.es/DNASP) software. The relationships among haplotypes were obtained by the median joining method using NETWORK 4.1.0.8 (www.fluxus-engineering.com). The ARLEQUIN 2.0 package (Schneider *et al.* 2000) was used to perform the Analysis of Molecular Variance (AMOVA).

Mean divergence (K) between humans and one chimpanzee (GenBank accession number AJ252012) was calculated using the PAUP* 4.0 program (Swofford, 2002), with a Tamura-Nei with invariants distance selected by MODELTEST (Posada & Crandall, 1998). The nucleotide substitution rate was inferred from the divergence value using the formula $\mu = K/(2t)$, considering a divergence time (t) of 6 million years.

To additionally test if there are signals of past demographic fluctuations in the samples of all individuals (worldwide sample), and Amerindians alone, coalescence simulations were performed using Rogers' algorithm (Rogers, 1995) as implemented by Wooding *et al.* (2004) in the DFSC 1.0 program (www.xmission.com/~wooding/DFSC/). These sim-

ulations were also performed with the original data of Alonso & Armour (2001) only and with a Eurasian subset. This algorithm assumes that, t generations ago, a population increased suddenly from an ancient population size (N_0) to a larger population size (N_1), with an infinite-sites mutation rate (μ). The use of simulations based on coalescent theory to test complex demographic scenarios from genetic polymorphisms is considered one of the most powerful methods to study human demographic evolution (Rosenberg & Nordborg, 2002). This is achieved due to the efficiency of the simulations, which can simulate a wide range of scenarios with a reduced number of assumptions. The simulations employed here considered a combination of demographic scenarios, from growth factors of 1-fold (stationary) to 500-fold, beginning 3,000 to 150,000 years ago for all individuals; and for the Amerindian dataset, between 2,500 and 50,000 years ago, respectively. For Amerindians assumptions about population decline (bottleneck) were performed, considering reduction factors of 1-fold to 50-fold beginning 12,000 to 50,000 years ago. Different values of theta (θ) were used in the simulations (0.01, 0.1, 1.0, 10) to verify how the effective population size (N_e) influences the proposed scenarios.

Results

Polymorphic Sites and Haplotype Determination

A total of 1558 base pairs of 196 chromosomes were sequenced. Of the 98 individuals tested 95 had their haplotypes directly determined, since they were either homozygous or heterozygous for a single site. The remaining three had their phase assigned by cloning (seven clones of each individual were sequenced): they comprise one Aché, one Kaingang, and one Parakanã. In the Aché and the Kaingang we found some recombinant clones between the two alleles that may be explained as an artifact of the PCR process, perhaps originating as described by Ennis *et al.* (1990).

Combining our data with those obtained by Alonso & Armour (2001), 55 polymorphic sites (including a 5 bp indel in two haplotypes) were observed in this seg-

ment. Of these 12 were observed in this study for the first time. These new substitutions can be characterized as follows: nine transitions at sites 143, 735, 1535 (C/T); 264, 468, 491, 535, 722 (G/A); 1510 (T/C); and three transversions at sites 1132 (A/C); and 1141, 1251 (A/T).

The haplotypes observed by Alonso & Armour (2001) and those inferred in this study were combined, resulting in a total of 35 haplotypes (Table 2). By far the most common was HP1, occurring in 62% of the 296 chromosomes studied. HP5 has a frequency of 11%, while all the others have a prevalence of 5% or less. Twelve occur in Amerindians and Northern Arctic only, and were newly identified in the present investigation. Nineteen others were found by Alonso & Armour (2001) only, in Asian, European, and African subjects, while four were observed by both studies. HP1 occurs in the five ethnic-geographic categories established; HP13 in four, HP5 in three, and HP10 in two of them. In Amerindians, among the 180 chromosomes tested, the three most common haplotypes are (in percentages) HP1 (68), HP5 (14), and HP6 (8).

Of the 27 local ethnic groups studied here (listed in Table 1), seven comprise individuals who are all homozygous for the most common allele (HP1): South Americans; Arara, Lengua, Pacaás Novos, Wai Wai; Central Americans; Huetar, Teribe; and Mongolians; Khalkh. The following haplotypes are found in only one ethnic group: HP3 (Quechua), HP4 (Waiãpi), HP11 (Gavião), HP12 (Foz do Içana), HP14 (Cinta Larga), HP15 (Chukchi), and HP16 (Aché). Haplotype six occurs in 11 individuals belonging to the Aché, Parakanã, Suruí, and Xavante tribes. Parakanã is the tribe with the highest diversity, presenting HP6, HP1, HP2, HP5, HP7, HP8, and HP9 haplotypes.

Median Joining Network

A Median Joining Network was constructed with all haplotypes (those detected here and those listed by Alonso & Armour, 2001), and is pictured in Figure 1. Two clusters are apparent, one with the most divergent haplotypes and the other with a starlike structure, where haplotype 1 is the central and most numerous one. This second cluster is more compact, with most haplotypes differing from HP1 by just one mutation.

Included in this are haplotypes found in the five major geographic-ethnic categories established in the present work. Of the haplotypes found exclusively in Amerindians the most frequent is HP6, which differs from HP1 by a mutation at site 491. Four other haplotypes, HP7, HP8, HP9, HP11, also differ from HP1 by just one mutation. The most derived Amerindian haplotype in this cluster is HP2, found in the Parakanã, which differs from HP1 at three sites, and from HP8 at two sites.

The second cluster originates from HP27, the haplotype identified by Alonso & Armour (2001) as the root of their tree. Mostly exclusively African haplotypes are represented in this cluster, but several exceptions occur. HP28 was found in one Basque individual: HP10, found by Alonso & Armour (2001) in Kenya, was also observed among the Kaingang. Three other haplotypes in this clade are found exclusively in Amerindians: HP3 in the Quechua, HP16 in one Aché and HP14 in one Cinta Larga. HP15, which is very derived, was found in the Chukchi. The reticulations observed derive from haplotypes confirmed by cloning.

Nucleotide Substitution Rates

Mean divergence (K) between humans and one chimpanzee was 0.026 ± 0.00082 . The average substitution rate, estimated from this divergence, was 2.16×10^{-9} per site per year. The substitution rate per sequence (1558 sites) per generation (20 years) was estimated at 6.75×10^{-5} .

Genetic Structure

To investigate the partition of variance within and among populations, an AMOVA analysis was performed comparing five geographic arrangements. The results are shown in Table 3. In the first comparison, involving all the geographic-ethnic categories considered, $\sim 87\%$ of the variance occurs at the intrapopulation level, $\sim 10\%$ among populations within these categories, and $\sim 3\%$ among these categories. There is no clear differentiation between Central and South Amerindians (negative value), but moderate variation occurs among the different tribes or populations

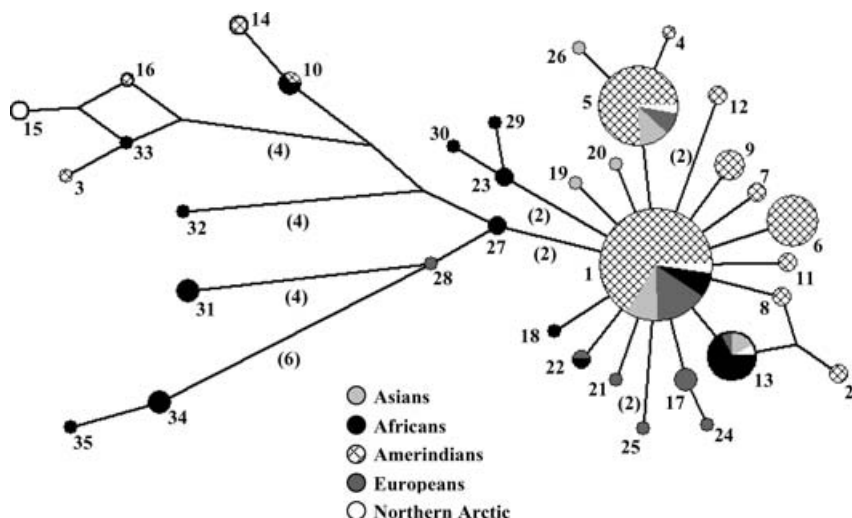


Figure 1 Median-joining network showing the relationships among haplotypes. Each circle represents a different haplotype, and its size is proportional to its relative frequency. For each haplotype, the different shading and the diagonal cross pattern indicate the fraction of observations in Asians, Africans, Amerindians, Europeans and the Northern Arctic. The numbers in parentheses under the branches correspond to the amount of nucleotide substitutions in excess of one, and the number beside each circle is the haplotype assignment. Reticulations indicate ambiguous relationships. HP27 is the root of the tree.

Table 3 AMOVA results: intra- and interpopulation variability considering different hierarchical arrangements

Population Structure	Among Groups (%)	Among Populations Within Groups (%)	Within Populations (%)
Geographic-Ethnic Categories ^a	3.32 ^{ns}	9.70	86.98
Central vs South Amerindians	-0.34 ^{ns}	15.12	85.22
Amerindians vs Asians	-3.33 ^{ns}	14.00	89.33
Amerindians vs Northern Arctic	31.52	6.45	62.03
Africans vs non-Africans ^b	11.44	8.05	80.51

^aGroups: Africans, Amerindians (South and Central), Asians, Europeans and Northern Arctic.

^bnon-Africans: Amerindians, Asians, Europeans and Northern Arctic.

^{ns}: p > 0.1; all the other values are significant at the 0.1% or lower level.

Genetic Diversity

Results of haplotype and nucleotide diversity, as well as the neutrality tests, are presented in Table 4. Haplotype and nucleotide diversity indices are generally similar among the major ethnic groups, with the exception of Africans who present with significantly higher values. Northern Arctic populations also have high values, but these are probably due to one Chukchi that is homozygous for a very divergent haplotype (HP15). Our results for the world sample are similar to those found by Alonso & Armour (2001), although diver-

sity values from our data are smaller, as expected, since the low-diversity Amerindian group comprises a high proportion of the sample. Almost all neutrality statistics (Tajima's D, Fu and Li's D* and F*, Fu's Fs) gave negative values, except those for Northern Arctic populations that were positive, but not significantly; Fu and Li's D* for Amerindians was also non-significantly positive. Europeans and the world sample gave significant values for all neutrality tests, and Amerindians for Tajima's D and Fu's Fs. Asian estimates were significant for Fu's Fs only.

Table 4 16p haplotype and nucleotide diversity, as well as neutrality test results, by major geographic-ethnic categories

Characteristics	Amerindians		Northern		Europeans	Africans	World	World A & A ^a
	Central	South	Arctic	Asian				
No. of chromosomes	24	156	8	28	40	40	296	100
No. of haplotypes		14	4	6	9	14	35	28
No. of polymorphic sites		23	13	5	11	28	49	42
Haplotype diversity (Hd)		0.519	0.750	0.598	0.508	0.853	0.598	0.765
Nucleotide diversity (π) (SE) ^b %		0.064 (0.011)	0.335 (0.116)	0.048 (0.010)	0.050 (0.013)	0.243 (0.037)	0.098 (0.012)	0.147 (0.020)
Tajima's D		-2.08*	0.20	-1.18	-2.12*	-1.46	-2.32**	-2.14*
Fu and Li's F*		-0.66	0.90	-1.58	-3.55*	-2.07	-2.88*	-3.76*
Fu and Li's D*		0.46	1.00	-1.44	-3.45*	-1.90	-2.46*	-3.80*
Fu's Fs		-7.60**	2.30	-2.68*	-5.97**	-2.72	-32.70**	-18.42**

^aAlonso & Armour (2001); ^bSE; standard error, *: $p < 0.05$; **: $p < 0.01$

Demographic Simulations

As most of the values of Tajima's D statistic were significantly more negative than expected under the neutral model (assuming constant population size and no selection, see discussion), they are consistent with the hypothesis of a population bottleneck followed by growth or a selective sweep. Considering the vast evidence for population growth in the history of modern humans (reviewed in Harpending & Rogers, 2000) and the evidence presented by Alonso & Armour (2001) that the 16p13.3 region is neutral, we only considered the hypothesis of demographic change. To better investigate this possibility with our datasets a wide variety of demographic scenarios (see Subjects and Methods) were considered in the coalescent simulations.

Two historical events were tested with the exclusively Amerindian dataset: population decline (bottleneck) and growth (Figure 2I). All scenarios of population decline (>1-fold to 50-fold) were rejected (data not shown). Simulations considering population increase, with thetas of 0.01 and 10, rejected almost all scenarios, while those with theta of 0.1 and 1.0 rejected only scenarios with more recent expansions (Figure 2a-d).

Our and Alonso & Armour's (2001) samples were assembled into a worldwide dataset. For this dataset all scenarios with $\theta = 0.01$ were rejected (Figure 2e). For other theta values the results were not very different from those of the Amerindians. The implication is that the Amerindian genetic pool may be considered to be

a representative sample from the world population. To test if the results with the world dataset could have been influenced by the large Amerindian sample size, simulations were also run exclusively with Alonso & Armour's (2001) dataset (Figure 2i-l). The results were similar, suggesting that the total set was not appreciably affected by the inclusion of the Amerindians.

As there is a clear difference in the demographic history of Africans vs non-Africans (e.g. Harpending & Rogers, 2000), and it is widely accepted that the latter are Amerindians' ancestors, it was interesting to compare the Amerindian with the Eurasian scenarios (Figure 2m-p); the results were also similar.

Discussion

We have expanded the information on the 16p13.3 region for Amerindian and Asian groups, and integrated these data with those from African, Europe and Asian populations, to investigate this region's molecular variability and make inferences about the early peopling of the Americas, as well as the subsequent process of Amerindian genetic diversification. Our results have revealed new and divergent haplotypes, the latter previously found in Africans only. The five recognized geographic-ethnic categories were differentially sampled: by far the largest was the Amerindian, with 180 chromosomes, which may partially explain why they contained the largest number (11) of specific haplotypes. In relation to haplotype frequencies, in both Alonso & Armour (2001) and the present investigation HP1 (or

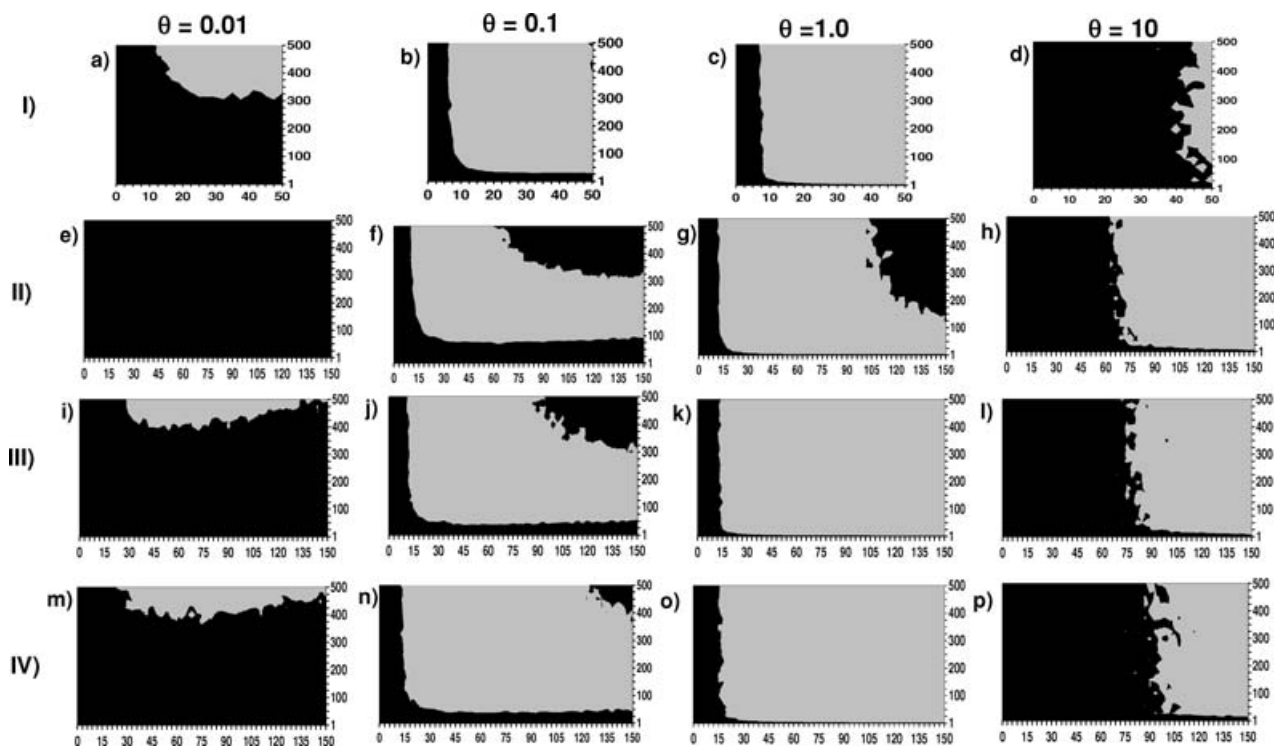


Figure 2 Simulation tests based on coalescence in four datasets: I) Amerindians; II) worldwide; III) data from Alonso & Armour (2001) only; and IV) Eurasians. Black shading indicates rejected scenarios ($P < 0.025$) and light shading scenarios that could not be rejected; on the axes are the magnitude of growth (1 to 500-fold) and time of growth (thousands of years ago).

Hpa in the former nomenclature) was the most frequent (respectively 62% and 52%).

Considering the substitution rate estimated for all populations together, this differed little from that calculated by Alonso & Armour (2001). Divergence (K) between human and chimpanzee, as calculated here and by the previous study, was very similar (0.029 and 0.026 respectively), the same being true for the substitution rates (2.18×10^{-9} and 2.19×10^{-9} respectively). For the whole dataset nucleotide diversities were recalculated to include the same number of base pairs. As expected, Africans gave the highest value (0.243%), with Amerindians presenting a value of the same order of magnitude (0.064%) as Europeans (0.050%) and Asians (0.048%).

When the data were hierarchically analyzed, as is usually found in human populations (e.g. Wilder *et al.* 2004) most of the variation was found within populations. Central and South Amerindians showed no significant differences. But in Central America only the two most common haplotypes (HP5 and HP1) were found, and this low genetic variability may have occurred be-

cause of isolation followed by a population reduction (Barrantes *et al.* 1990; Azofeifa *et al.* 2001). The difference observed among populations within groups ($\sim 15\%$) was probably due to genetic structure within South America. The very high between-group diversity value found between Amerindians and the Northern Arctic was unexpected, and could be explained by the presence of a highly divergent Chukchi sample. Consistent with this hypothesis, when we performed the AMOVA without this individual (data not shown) the result was reduced to a negative value (-6.74% , $P > 0.1$), in agreement with the other results. As expected, after this correction the highest between-group value was that for the Africans vs non-Africans comparison.

The statistical neutrality tests calculated for this and the Alonso & Armour (2001) worldwide datasets resulted in significant negative values (Table 4); this is a signal of population growth but may also be due to genetic hitchhiking, as this region is located near an exon. However, we agree with Alonso & Armour (2001) who suggested that genetic hitchhiking is not the likely explanation, as there are recombination hotspots near this

region. Additionally, the pattern of distribution of the significant values is totally concordant with the most widely accepted human demographic scenario. When separated by the main continental groups, Africans did not show the significant values found in Eurasians, corroborating the widely supported hypothesis that the strong past population growth occurred after the exodus from Africa. This scenario is further corroborated by the median-joining network with two clusters, one with very divergent haplotypes, mainly found in Africans, and another with a starlike shape mainly due to Eurasians. This last mainly non-African cluster suggests a strong population expansion from a restricted source, as extensively found by mitochondrial (Ingman *et al.* 2000) and some nuclear data (Williamson *et al.* 2005). However, we also found some very divergent haplotypes in Amerindians and Northern Arctic populations (grouped into the African cluster). As these divergent haplotypes were found at very low frequencies they could have been missed if our sample sizes were small and/or the ethnic diversity low. In relation to the out-of-Africa exodus this result supports a scenario of a larger and more diverse founding population than that initially suggested by Alonso & Armour (2001). These results also recommend being cautious about conclusions concerning genetic diversity in human populations that are supported by samples from very limited sources.

The signal of population growth observed in Amerindians is most likely the same as that of the out-of-Africa exodus, and would not have been erased during and after the demographic events that led to the colonization of the New World, since the statistics (Table 4) and scenarios (Figure 2) from this continent are very similar to the worldwide sample. These results concur with the simulations that also rejected all scenarios of a past population decline (bottleneck) in the Amerindian sample. Therefore, the hypothesis of a strong bottleneck during the entrance of prehistoric humans into the Americas is not supported by the 16p13.3 locus, which on the contrary indicates that the genetic variability of the source population has been maintained to a certain degree in Amerindians. The question of a population decline associated with the prehistoric human arrival in the Americas is controversial. A recent study of a noncoding region in Amerindian and Asian pop-

ulations (Heller *et al.* 2004) did not find evidence of a significant bottleneck, as was true for the investigation of Hutz *et al.* (2002) with five Brazilian Indian populations and fifteen short tandem repeat polymorphisms (STRs). On the other hand, Fagundes *et al.* (2005), studying the low density lipoprotein receptor gene (*LDLR*) in 222 chromosomes from individuals of African, Asian, Caucasian and Amerind ancestry, found a signal of a moderate population bottleneck. This last result agreed with polymorphisms in the Y chromosome (Bortolini *et al.* 2003) and mtDNA (Bonatto & Salzano, 1997) that also suggested a bottleneck during the peopling of the Americas. However, mtDNA and the non-recombining region of the Y chromosome are uniparentally inherited, and have effective population sizes that are one fourth of the autosomal locus considered here; their diversity would therefore be much more strongly affected by a moderate population size reduction than the latter (Avice, 2000; Birky, 2001). Also, independent autosomal loci need not show similar population histories in stationary scenarios due to the vagaries of the coalescent process (Rosenberg & Nordborg, 2002), and some loci may be subjected to the effects of different natural selection regimes which should affect only the local genomic region containing the target of selection (Stajich & Hahn, 2005). On the other hand, a strong bottleneck should affect all loci equally, and its effects should be directly proportional to the intensity of the event (Galtier *et al.* 2000). Considering all these diverse situations, a scenario suggesting a bottleneck of moderate intensity in the peopling of the New World may be most compatible with all the data presented so far.

The debate about the colonization of South America is strongly linked to the theories on the peopling of North America, as most scholars agree that people arrived in South America after traversing Central America (see Salzano & Callegari-Jacques, 1988). Although there have been early suggestions that the entry into South America was followed by a population bottleneck (see Salzano & Callegari-Jacques, 1988), recent studies have not found statistically significant differences for mtDNA and nuclear DNA markers (reviewed in Salzano, 2002 and e.g. Heller *et al.* 2004; Mateus-Pereira *et al.* 2005) between South and North American Amerindians. Considering the results presented here, as

the majority of our data were from South American native populations the conclusions above concerning the absence of a strong bottleneck in the colonization of the New World may also be valid for the settlement of South America.

Acknowledgments

This work was supported by grants from Programa de Apoio a Núcleos de Excelência (PRONEX), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS). The Amerindian field collections were approved by the Brazilian Ethics National Committee (CONEP, resolution 123/98). We would also like to thank Jomar Laurino for assistance and support in the cloning process; Nelson J. R. Fagundes and Felipe Gobbi Graziotin for their contribution to some of the results, and M. L. Petzl-Erler, L. T. Tsuneto, Eduardo Tarazona-Santos, Mikhail Krylov, L. Benevolenskaya, and B. Munkhbat for providing some DNA samples.

References

Alonso, S. & Armour, J. A. (2001) A highly variable segment of human subterminal 16p reveals a history of population growth for modern humans outside Africa. *Proc Natl Acad Sci U S A* **98**, 864–869.

Andrade, F. M., Coimbra, C. E. A., Jr., Santos, R. V., Goicoechea, A., Carnese, F. R. & Salzano, F. M. et al. (2000) High heterogeneity of apolipoprotein E gene frequencies in South American Indians. *Ann Hum Biol* **27**, 29–34.

Awise, J. C. (2000) *Phylogeography: the history and formation of species*, Harvard University Press, Cambridge.

Azofeifa, J., Ruiz, E. & Barrantes, R. (2001) Blood group, red cell, and serum protein variation in the Cabecar and Huetar, two Chibchan Amerindian tribes of Costa Rica. *Am J Hum Biol* **13**, 57–64.

Badge, R. M., Yardley, J., Jeffreys, A. J. & Armour, J. A. (2000) Crossover breakpoint mapping identifies a subtelomeric hotspot for male meiotic recombination. *Hum Mol Genet* **9**, 1239–1244.

Barrantes, R. (1993) *Evolución en el trópico: los Amerindios de Costa Rica e Panamá*, Editorial de la Universidad de Costa Rica, San José.

Barrantes, R., Smouse, P. E., Mohrenweiser, H. W., Gershowitz, H., Azofeifa, J., Arias, T. D. et al. (1990) Microevolution in Lower Central America - genetic-characterization of the Chibcha-speaking groups of Costa Rica and Panama, and a consensus taxonomy based on genetic and linguistic affinity. *Am J Hum Genet* **46**, 63–84.

Birky, C. W., Jr. (2001) The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annu Rev Genet* **35**, 125–148.

Black, F. L., Santos, S. E., Salzano, F. M., Callegari-Jacques, S. M., Weimer, T. A. & Franco, M. H. et al. (1988) Genetic variation within the Tupi linguistic group: new data on three Amazonian tribes. *Ann Hum Biol* **15**, 337–351.

Bonato, S. L. & Salzano, F. M. (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.

Bortolini, M. C., Salzano, F. M., Thomas, M. G., Stuart, S., Nasanen, S. P. & Bau, C. H. et al. (2003) Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am J Hum Genet* **73**, 524–539.

Brown, S. M., Gajdusek, D. C., Leyshon, W. C., Steinberg, A. G., Brown, K. S. & Curtain, C. C. (1974) Genetic studies in Paraguay: blood group, red cell, and serum genetic patterns of the Guayaki and Ayore Indians, Mennonite settlers, and seven other Indian tribes of the Paraguayan Chaco. *Am J Phys Anthropol* **41**, 317–343.

Callegari-Jacques, S. M., Salzano, F. M., Weimer, T. A., Franco, M. H. L. P., Mestriner, M. A. & Hutz, M. H. et al. (1996) The Wai Wai Indians of South America: History and genetics. *Ann Hum Biol* **23**, 189–201.

Callegari-Jacques, S. M., Salzano, F. M., Weimer, T. A., Hutz, M. H., Black, F. L. & Santos, S. E. et al. (1994) Further blood genetic studies on Amazonian diversity-data from four Indian groups. *Ann Hum Biol* **21**, 465–481.

Campbell, L. (1997) *American Indian languages: the historical linguistics of Native America*, Oxford University Press, New York.

Coimbra, C. E. A., Jr., Flowers, N. M., Salzano, F. M. & Santos, R. V. (2002) *The Xavante in transition: health, ecology, and bioanthropology in central Brazil.*, University of Michigan Press, Ann Arbor.

Cribbs, L. L., Lee, J. H., Yang, J., Satin, J., Zhang, Y. & Daud, A. et al. (1998) Cloning and characterization of alpha 1H from human heart, a member of the T-type Ca²⁺ channel gene family. *Circ. Res.* **83**, 103–109.

Dornelles, C. L., Battilana, J., Fagundes, N. J., Freitas, L. B., Bonatto, S. L. & Salzano, F. M. (2004) Mitochondrial DNA and *Alu* insertions in a genetically peculiar population: the Ayoreo Indians of Bolivia and Paraguay. *Am J Hum Biol* **16**, 479–488.

Ennis, P. D., Zemmour, J., Salter, R. D. & Parham, P. (1990) Rapid cloning of HLA-A,B cDNA by using the polymerase chain reaction: frequency and nature of errors produced in amplification. *Proc Natl Acad Sci USA* **87**, 2833–2837.

Erdesz, S., Shubin, S. V., Shoch, B. P., Krylov, M., Mylov, N. M. & Chekalina, N. A. et al. (1994) Spondyloarthropathies in circumpolar populations of Chukotka

- (Eskimos and Chukchi): epidemiology and clinical characteristics. *J Rheumatol* **21**, 1101–1104.
- Excoffier, L. (2002) Human demographic history: refining the recent African origin model. *Curr Opin Genet Dev* **12**, 675–682.
- Fagundes, N. J., Salzano, F. M., Batzer, M. A., Deininger, P. L. & Bonatto, S. L. (2005) Worldwide genetic variation at the 3'-UTR region of the *LDLR* gene: possible influence of natural selection. *Ann Hum Genet* **69**, 389–400.
- Fu, Y. X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915–925.
- Fu, Y.-X. & Li, W.-H. (1993) Statistical tests of neutrality of mutations. *Genetics* **133**, 693–709.
- Galtier, N., Depaulis, F. & Barton, N. H. (2000) Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. *Genetics* **155**, 981–987.
- Garrigan, D., Mobasher, Z., Severson, T., Wilder, J. A. & Hammer, M. F. (2005) Evidence for Archaic Asian Ancestry on the Human X Chromosome. *Mol Biol Evol* **22**, 189–192.
- Gilad, Y., Rosenberg, S., Przeworski, M., Lancet, D. & Skorecki, K. (2002) Evidence for positive selection and population structure at the human MAO-A gene. *Proc Natl Acad Sci USA* **99**, 862–867.
- Goicoechea, A. S., Carnese, F. R., Dejean, C., Avena, S. A., Weimer, T. A. & Estalote, A. C. *et al.* (2001) New genetic data on Amerindians from the Paraguayan Chaco. *Am J Hum Biol* **13**, 660–667.
- Hammer, M. F. (1995) A recent common ancestry for human Y-chromosomes. *Nature* **378**, 376–378.
- Hammer, M. F., Garrigan, D., Wood, E., Wilder, J. A., Mobasher, Z., Bigham, A., Krenz, J. G. & Nachman, M. W. (2004) Heterogeneous patterns of variation among multiple human X-linked loci: the possible role of diversity-reducing selection in non-Africans. *Genetics* **167**, 1841–1853.
- Harding, R. M., Fullerton, S. M., Griffiths, R. C., Bond, J., Cox, M. J. & Schneider, J. A. *et al.* (1997) Archaic African and Asian lineages in the genetic ancestry of modern humans. *Am J Hum Genet* **60**, 772–789.
- Harpending, H. & Rogers, A. (2000) Genetic perspectives on human origins and differentiation. *Annu Rev Genomics Hum Genet* **1**, 361–385.
- Harris, E. E. & Hey, J. (1999) X chromosome evidence for ancient human histories. *Proc Natl Acad Sci USA* **96**, 3320–3324.
- Heidrich, E. M., Hutz, M. H., Salzano, F. M., Coimbra, C. E. A., Jr. & Santos, R. V. (1995) D1S80 locus variability in three Brazilian ethnic-groups. *Hum Biol* **67**, 311–319.
- Heller, A. H., Salzano, F. M., Barrantes, R., Krylov, M., Benevolenskaia, L. & Arnett, F. C. *et al.* (2004) Intra and intercontinental molecular variability of an *Alu* insertion in the 3' untranslated region of the *LDLR* gene. *Hum Biol* **76**, 591–604.
- Hill, K. & Hurtado, A. M. (1996) *Ache life history : the ecology and demography of a foraging people*, Aldine de Gruyter, New York.
- Hutz, M. H., Callegari-Jacques, S. M., Almeida, S. E., Armbrorst, T. & Salzano F. M. (2002) Low levels of STRP variability are not universal in American Indians. *Hum Biol* **74**, 791–806.
- Hutz, M. H., Mattevi, V. S., Callegari-Jacques, S. M., Salzano, F. M., Coimbra, C. E. A., Jr. & Santos, R. V. *et al.* (1997) D1S80 locus variability in South American Indians. *Ann Hum Biol* **24**, 249–255.
- Ingman, M., Kaessmann, H., Paabo, S. & Gyllensten, U. (2000) Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708–713.
- Jaruzelska, J., Zietkiewicz, E., Batzer, M., Cole, D. E., Moisan, J. P. & Scozzari, R. *et al.* (1999) Spatial and temporal distribution of the neutral polymorphisms in the last ZFX intron: analysis of the haplotype structure and genealogy. *Genetics* **152**, 1091–1101.
- Kaessmann, H., Heissig, F., von Haeseler, A. & Paabo, S. (1999) DNA sequence variation in a non-coding region of low recombination on the human X chromosome. *Nat Genet* **22**, 78–81.
- Kohlrausch, F. B., Callegari-Jacques, S. M., Tsuneto, L. T., Petzl-Erler, M. L., Hill, K., Hurtado, A. M., Salzano, F. M. & Hutz, M. H. (2005) Geography influences microsatellite polymorphism diversity in Amerindians. *Am J Phys Anthropol* **126**, 463–470.
- Kolman, C. J., Sambuughin, N. & Bermingham, E. (1996) Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* **142**, 1321–1334.
- Krylov, M., Erdesz, S., Alexeeva, L., Benevolenskaya, L., Arnett, F. C. & Reveille, J. D. (1995) HLA class II and HLA-B27 oligotyping in two Siberian native population groups. *Tiss Antig* **46**, 382–386.
- Mateus-Pereira, L. H., Socorro, A., Fernandez, I., Masleh, M., Vidal, D., Batzer, M. A., Bonatto, S. L., Salzano, F. M. & Herrera, R. J. (2005) Phylogenetic information in polymorphic L1 and Alu insertions from East Asians and Native American populations. *Am J Phys Anthropol* **128**, 171–184.
- Mestriner, M. A. & Salzano, F. M. (1998) Monomorphic and polymorphic enzyme genetic markers of the Waiãpi Indians of Amapá and of inhabitants of Manaus, Amazonas. *Gen Mol Biol* **21**, 311–314.
- Munkhbat, B., Sato, T., Hagihara, M., Sato, K., Kimura, A. & Munkhtuvshin, N. *et al.* (1997) Molecular analysis of HLA polymorphism in Khoton-Mongolians. *Tiss Antig* **50**, 124–134.
- Nachman, M. W. & Crowell, S. L. (2000) Contrasting evolutionary histories of two introns of the Duchenne muscular dystrophy gene, *Dmd*, in humans. *Genetics* **155**, 1855–1864.

- Nachman, M. W., Bauer, V. L., Crowell, S. L. & Aquadro, C. F. (1998) DNA variability and recombination rates at X-linked loci in humans. *Genetics* **150**, 1133–1141.
- Nei, M. (1987) *Molecular evolutionary genetics*. New York: Columbia University Press.
- Posada, D. & Crandall, K. A. (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818.
- Przeworski, M., Hudson, R. R. & Di Rienzo, A. (2000) Adjusting the focus on human variation. *Trends Genet* **16**, 296–302.
- Rogers, A. R. (1995) Genetic evidence for a Pleistocene population explosion. *Evolution* **49**, 608–615.
- Rosenberg, N. A. & Nordborg, M. (2002) Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nat Rev Genet* **3**, 380–390.
- Salzano, F. M. (2002) Molecular variability in Amerindians: widespread but uneven information. *An Acad Bras Cienc* **74**, 223–263.
- Salzano, F. M., Black, F. L., Callegari-Jacques, S. M., Santos, S. E., Weimer, T. A. & Mestriner, M. A. et al. (1991) Blood genetic systems in four Amazonian tribes. *Am J Phys Anthropol* **85**, 51–60.
- Salzano, F. M., Callegari-Jacques, S. M., Franco, M. H., Hutz, M. H., Weimer, T. A. & Silva, R. S. et al. (1980) The Caingang revisited: blood genetics and anthropometry. *Am J Phys Anthropol* **53**, 513–524.
- Salzano, F. M., Callegari-Jacques, S. M., Weimer, T. A., Franco, M. H. L. P., Hutz, M. H. & PetzlErler, M. L. (1997) Electrophoretic protein polymorphisms in Kaingang and Guarani Indians of southern Brazil. *Am J Hum Biol* **9**, 505–512.
- Salzano, F. M., Gershowitz, H., Mohrenweiser, H., Neel, J. V., Smouse, P. E. & Mestriner, M. A. et al. (1986) Gene flow across tribal barriers and its effect among the Amazonian Içana river Indians. *Am J Phys Anthropol* **69**, 3–14.
- Salzano, F. M., Weimer, T. A., Franco, M. H. L. P., Mestriner, M. A., Simões, A. L. & Constans, J. et al. (1985) Population-structure and blood genetics of the Pacaás Novos Indians of Brazil. *Ann Hum Biol* **12**, 241–249.
- Salzano, F. M. & Callegari-Jacques, S. M. (1988) South American Indians. A Case Study in Evolution. Clarendon Press, Oxford.
- Sambrook, J. & Russell, D. W. (2001) *Molecular cloning : a laboratory manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- Saunders, M. A., Hammer, M. F. & Nachman, M. W. (2002) Nucleotide variability at G6pd and the signature of malarial selection in humans. *Genetics* **162**, 1849–1861.
- Schmitt, R., Bonatto, S. L., Freitas, L. B., Muschner, V. C., Hill, K., Hurtado, A. M. & Salzano, F. M. (2004) Extremely limited mitochondrial DNA variability among the Ache Natives of Paraguay. *Ann Hum Biol* **31**, 87–94.
- Schneider, S., Roessli, D. & Excoffier, L. (2000) *Arlequin ver. 2.000: a software for population genetics*, Genetics and Biometry Laboratory, University of Geneva, Geneva.
- Stajich, J. E., & Hahn, M. W. (2005) Disentangling the effects of demography and selection in human history. *Mol Biol Evol* **22**, 63–73.
- Swofford, D. L. (2002) *PAUP*: Phylogenetic Analysis Using Parsimony, Version 4.0b10 For Macintosh*. Sinauer Associates Inc., Sunderland MA.
- Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.
- Tarazona-Santos, E., Carvalho-Silva, D. R., Pettener, D., Luiselli, D., De Stefano, G. F. & Labarga, C. M. et al. (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.
- Underhill, P. A., Shen, P. D., Lin, A. A., Jin, L., Passarino, G. & Yang, W. H. et al. (2000) Y chromosome sequence variation and the history of human populations. *Nat Genet* **26**, 358–361.
- Verrelli, B. C., McDonald, J. H., Argyropoulos, G., Destro-Bisol, G., Froment, A. & Drousiotou, A. et al. (2002) Evidence for balancing selection from nucleotide sequence analyses of human G6PD. *Am J Hum Genet* **71**, 1112–1128.
- Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. & Wilson, A. C. (1991) African populations and the evolution of human mitochondrial DNA. *Science* **253**, 1503–1507.
- Watterson, G. A. (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* **7**, 256–276.
- Whitfield, L. S., Sulston, J. E. & Goodfellow, P. N. (1995) Sequence variation of the human Y-chromosome. *Nature* **378**, 379–380.
- Wilder, J. A., Kingan, S. B., Mobasher, Z., Pilkington, M. M. & Hammer, M. F. (2004) Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration rates of females versus males. *Nat Genet* **36**, 1122–1125.
- Williamson, S. H., Hernandez, R., Fledel-Alon, A., Zhu, L., Nielsen, R. & Bustamante, C. D. (2005) Simultaneous inference of selection and population growth from patterns of variation in the human genome. *Proc Natl Acad Sci USA* **102**, 7882–7887.
- Wooding, S., Kim, U. K., Bamshad, M. J., Larsen, J., Jorde, L. B. & Drayna, D. (2004) Natural selection and molecular evolution in *PTC*, a bitter-taste receptor gene. *Am J Hum Genet* **74**, 637–646.
- Yu, N., Chen, F. C., Ota, S., Jorde, L. B., Pamilo, P. & Patthy, L. et al. (2002) Larger genetic differences within Africans than between Africans and Eurasians. *Genetics* **161**, 269–74.

Received: 22 November 2005

Accepted: 29 March 2006