Short Report

Distribution of Y-Chromosome Q Lineages in Native Americans

RAFAEL BISSO-MACHADO,[1] MARILZA S. JOTA,[2] VIRGINIA RAMALLO,[1] VANESSA R. PAIXÃO-CÔRTES,[1] DANIELA R. LACERDA,[2] FRANCISCO M. SALZANO,[1] SANDRO L. BONATTO,[3] FABRÍCIO R. SANTOS,[2] AND MARIA CÁTIRA BORTOLINI[1]*
[1]*Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil*
[2]*Departamento de Biologia Geral, ICB, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil*
[3]*Centro de Biologia Genômica e Celular, Pontifícia Universidade Católica do Rio Grande do Sul, Av. Ipiranga 6681, 90610-001 Porto Alegre, RS, Brazil*

Objectives: This investigation was performed to identify and evaluate the distribution of all 15 Y-chromosome lineages belonging to the Q clade in a sample of natives from South America.

Methods: One hundred and forty-eight individuals from 20 Native American populations, as well as 24 Asian samples including Eskimos, were tested with 18 biallelic loci that can identify all currently known lineages of the Y-Chromosome Q clade. Sequencing was performed in part of the sample (~180,000 nucleotides, which detected, for instance, several downstream markers related to the Q1a3a lineage).

Results: No new mutation was found and Q1a3a was consistently found in high frequencies in all populations, followed at a much lower frequency by Q1a3*, while Q1a3a derived-lineages are probably population/tribe/region-specific.

Conclusion: The number of basal Y chromosome lineages in North America is apparently higher than in South America due probably to a bottleneck during the South American colonization and/or more recent Circum-Arctic gene flow. Am. J. Hum. Biol. 23:563–566, 2011.

The presence of autochthonous Native American Y-chromosomes is well-known since the initial studies in the 1990s. The Q1a3a haplogroup, identified by a mutation at the M3 locus in the nonrecombining region of this chromosome, is present in high frequencies in Native Americans only and in very few Siberian populations, probably due to reverse gene flow from Alaska to western Siberia. Several Q1a3a sublineages, however, are restricted to specific South American regions (Karafet et al., 2008 and references therein; Jota et al., 2011). All these lineages are derived from haplogroup Q chromosomes bearing a C → T mutation at locus M242, today observed in natives from America and Asia (Bortolini et al., 2003; Karafet et al., 2008; Seielstad et al. 2003). Paragroup Q*, a Q chromosome lineage with no derived alleles, has been also observed at low frequencies in Turkey, India, Pakistan, Korea, Japan, and Oceania (revision in Zhong et al., 2011).

Besides Q lineages, Native Americans also present a low frequency of C3b (defined by a mutation at locus P39) found only in North America (Karafet et al., 2008; Zegura et al., 2004), while the more ancient C3* was detected in northwestern South America (Wayuu, Waorani and Kichwa populations; Geppert et al., 2011; Zegura et al., 2004). The presence of other lineages has been attributed to recent admixture with non-Amerindians.

The Y-Chromosome Consortium published the evolutionary relationships among 311 distinct paragroups/haplogroups defined by approximately 600 markers. Within clade Q, besides Q*, 13 paragroups/haplogroups (Q1*, Q1a*, Q1a1, Q1a2, Q1a3*, Q1a3a*, Q1a3a1, Q1a3a2, Q1a3a3, Q1a4, Q1a5, Q1a6, and Q1b) identified by 17 SNPs are recognized. Most of them present geographic distributions restricted to the Americas and/or Asia (Karafet et al., 2008). Another Q1a3a sublineage, Q1a3a4 has been recently described (Jota et al., 2011).

This investigation was performed to identify and evaluate the distribution of all 15 Y-chromosome lineages belonging to the Q clade in a sample of natives from South America with distinct linguistic affiliations and demographic histories. We have also analyzed male samples from Eastern Russia and Siberia.

SUBJECTS AND METHODS

Populations

One hundred and forty-eight Native American individuals from 20 populations widely spread all over South America were tested. Additionally, 24 Asians, including Eskimos, were also investigated. Figure 1 shows the geographical locations and linguistic affiliations of these populations.

Ethical approval for this study was provided by the Brazilian National Ethics Commission (no. 123/1998), as well as by ethics committees in the countries where the non-Brazilian samples were collected.

Laboratory procedures

DNA was extracted from plasma, glycerolized red blood cells, as well as total blood stored in our laboratories as a result of previous studies (review in Salzano, 2002) using the QIAamp DNA MiniKit. Eighteen biallelic loci were

Additional Supporting Information may be found in the online version of this article.

Grant sponsors: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS), Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG), The logistic support of Fundação Nacional do Índio (FUNAI; Brazil).

*Correspondence to: Maria Cátira Bortolini, Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil.
E-mail: maria.bortolini@ufrgs.br

Received 6 January 2011; Revision received 16 February 2011; Accepted 8 March 2011

DOI 10.1002/ajhb.21173

Published online 4 May 2011 in Wiley Online Library (wileyonlinelibrary.com).

POPULATION	N	LANGUAGE (Language family)*	LOCATION
Jamamadi	3	Jamamadi (Arawa)	BRAZIIL
Arara	8	Arara (Carib)	
Wai-Wai	9	Wai-Wai (Carib)	
Araweté	4	Araweté (Tupi)	
Gavião	7	Gavião (dialect), Aruá (Tupi)	
Karipuna	1	Karipuna (Dialect), Wayampí (Tupi)	
Karitiana	9	Karitiana (Tupi)	
Munduruku	1	Munduruku (Tupi)	
Parakanã	17	Akwawa (Tupi)	
Suruí	7	Suruí (Tupi)	
Tenharim	1	Tenharim (dialect) Parintintín (Tupi)	
Zoró	6	Zoró (dialect), Aruá (Tupi)	
Gorotire	14	Gorotire (dialect), Kayapó (Ge)	
Kaingang	1	Kaingang (Ge)	
Mekranoti	9	Mekranoti (dialect), Kayapó (Ge)	
Kuben-Kran-Kegn	9	Kuben-Kran-Kegn (dialect), Kayapó (Ge)	
Xavante	15	Xavante (Ge)	
Xikrin	6	Xikrin (dialect), Kayapó (Ge)	
Pacaás Novo	9	Orowari (Chapacuran)	
Lengua	12	Lengua (Mascoy)	PARAGUAY
Altai	1	Turkic (Altaic)	RUSSIA
Buriat	1	Buriat (Altaic)	
Kalmyc	1	Kalmyc-Oirat (Altaic)	
Tuvan	1	Tuva (Altaic)	
Eskimo	19	Yupic (Eskimo-Aleut)	
Chukchi	1	Chukchi (Chukotko-Kamchatkan)	

*South Americans and Asians according to Campbell (1997) and Lewis (2009) respectively.

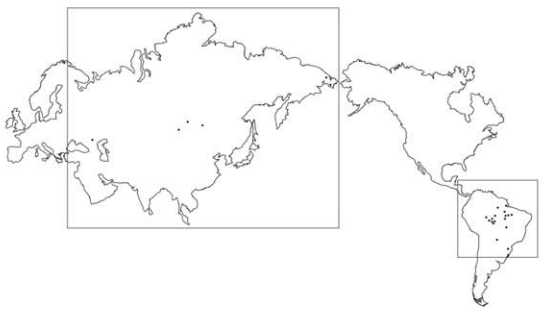

Fig. 1. Geographical distribution and language affiliation of the individuals studies in the present investigation.

analyzed to identify all lineages within the Q clade, whose primers and methods were described elsewhere (Jota et al., 2011; Karafet et al., 2008 and references therein). The allele state at the M3, M346, and M242 loci were typed first using allele competition, PCR-RFLP, TaqMan methods, or sequencing. A hierarchical strategy was used to optimize the genotyping of alleles that characterize the Q subhaplogroups. When a Y-chromosome bears a derived allele for the Q lineage (e.g., M242) or other derived one (e.g., M3 at Q1a3a), the PCR fragments for the likely downstream markers were genotyped using PCR-RFLP, TaqMan assays, or sequencing. For instance, fragments at the M19, M194, M199, P106, and P292 loci (downstream markers related to M3), which detect known SNPs that identify chromosome sublineages within the Q1a3a lineage were sequenced. This procedure led us to sequence more than 50% of the Q1a3a individuals of our sample. The sequencing approach was also adopted as a strategy to identify unknown variants in samples from populations that have never been studied by sequencing. Both DNA strands were sequenced using ABI Prism BigDye and an ABI 310 Genetic Analyzer. Whenever a putative mutation was identified, the sample was re-amplified and resequenced for confirmation.

Data analysis

Sequences were aligned and their qualities, as well as the assessment of the accuracy of the resulting data, were obtained using the Phred, Phrap, Consed (http://

TABLE 1. Y-chromosome haplogroup frequency (%) in 26 Native American and Asian populations investigated in this study

Population	Haplogroups (number of chromosomes)			
	Q1a3a	Q1a3*	Q1a*	Other non-Q lineages[a]
South America				
Arara	100 (8)			
Araweté	75 (3)	25 (1)		
Gavião	100 (7)			
Gorotire	89 (13)	11 (1)		
Jamamadi		100 (3)		
Kaingang				100 (1)
Karipuna				100 (1)
Karitiana	100 (9)			
Kuben-Kran-Kegn	89 (8)	11 (1)		
Lengua	67 (8)	33 (4)		
Mekranoti	67 (6)	11 (1)		22 (2)
Munduruku	100 (1)			
Pacaás Novos	100 (9)			
Parakanã	100 (17)			
Suruí	100 (7)			
Tenharim	100 (1)			
Wai Wai	100 (9)			
Xavante	100 (15)			
Xikrin	100 (6)			
Zoró	67 (4)	33 (2)		
Siberia				
Eskimo	5 (1)	21 (4)	5 (1)	69 (13)
Altai				100 (1)
Buriat				100 (1)
Chukchi				100 (1)
Kalmyk				100 (1)
Tuvan				100 (1)

[a]Among the South Amerindians all non Q chromosomes have a probable origin in the admixture process with non-Indians.

www.phrap.org/phredphrapconsed.html) softwares and the Clustal W algorithm included in the BioEdit 7.0.9 software (http://www.mbio.ncsu.edu/bioedit/bioedit.html).

RESULTS AND DISCUSSION

This is the first study that investigated all known Q sublineages in Amerindian samples composed by speakers of the major South American Native linguistic branches (Tupi, Ge, Carib, and Arawak). Our approach allowed us to sequence ~180,000 nucleotides to check for known SNPs and to screen for new ones in part of our sample, but no new mutation was detected. In South America, some individuals bear chromosomes with the derived allele at M346 and the ancestral allele at M3, thus belonging to paragroup Q1a3*. The presence of Q1a3* was also detected in four Siberian Eskimos but not in other Asians, including the Central (Tuvan, Altai, and Buriat) and extreme East (Chukchi) Siberians, investigated here (Table 1). However, these results should be considered with caution due to the size of the Asian sample analyzed in this study. Bailliet et al. (2009) verified the presence of Q1a3* in populations from Argentina, Chile, and Bolivia, as well as in individuals from North America (two Sioux, one Navajo, and two Zuni; Supporting Information Table S1). Based on these results it is possible to assume that most chromosomes that belong to paragroup Q* in South America are in fact Q1a3*. Supporting Information Table S1 also shows that Q1a3* is present (generally in relatively low frequencies) in West Eurasians, Central Asians, Central-South Asians, and East Asians, as well as in Siberians. Recently, Zhong et al. (2011) postulated that Q1a3 chromosomes probably entered East Asia via a northern route rather than a southern route, as occurred with other common Y chromosomes present in East Asia (for instance, O-M175, D-M174, N-M231, and C-M130/$RPS4Y_{711}$, the latter the C3* ancestor). This agrees with previous postulations for the entry of Q lineage chromosomes in America coming from northeast Asia (Santos et al. 1999).

The wide distribution of Q1a3a in South America is confirmed, while its sublineages show a very restricted distribution (Q1a3a1, Q1a3a3, and Q1a3a4 are found only among the Ticuna/Wayuu, Suruí, and Andean men, respectively; Supporting Information Table S1). The population from which Q1a3a2 was identified remains unknown (Supporting Information Table S1). We did not identify Q1a3a derived-lineages in our sample, confirming the idea that they are probably population/tribe/region-specific.

Our data also revealed the presence of Q1a* in our Siberian Eskimo sample. It is worth mentioning that a ~4,000-year-old permafrost preserved Paleo-Eskimo from Greenland had the Q1a* chromosome (Rasmussen et al., 2010). This result reinforces the view of an identity, by common origin and/or recurrent gene flow, of the members of the two major Eskimo language stocks (Yupik, spoken in Siberia and West Alaska, and Ynyupik spoken in northern Alaska, Canada, and Greenland; Laughlin, 1963). This connection also suggests an extraordinary temporal genetic continuity, since Q1a* is present in both Paleo-Eskimo and in present-day Eskimo individuals.

The other Q clade chromosomes (Q1*, Q1a1, Q1a2, Q1a4, Q1a5, Q1a6, and Q1b) were not identified in our samples.

The results presented here indicate a higher number of distinct Asian and/or Beringian lineages in North America than in South America. Besides the most frequent Q1a3a and Q1a3* chromosomes, Q1a* and C3b are also found in North America, but not in South America. This difference could be due to a founder effect at the beginning of the South America colonization and/or recurrent gene flow between North America and extreme East Asia, especially in the Arctic Circle, after Beringia's disappearance (González-José et al., 2008). The presence of the more ancient C3* in Northwest South America can, however, indicates its existence in present-day (undetected up to now) North America and/or in earlier times, since the peopling of South America by Beringian migrants occurred by land through North America (Geppert et al., 2011; González-José et al., 2008).

ACKNOWLEDGMENTS

We are very grateful to the individuals who donated the samples analyzed here.

LITERATURE CITED

Bailliet G, Ramallo V, Muzzio M, García A, Santos MR, Alfaro EL, Dipierri JE, Salceda S, Carnese FR, Bravi CM, Bianchi NO, Demarchi DA. 2009. Restricted geographic distribution for Y-Q* paragroup in South America. Am J Phys Anthropol 140:578–582.

Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SPK, Bau CHD, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-de-Guerra D, Torres MM, Groot H, Michalski R, Nymadawa P, Bedoya G, Bradman N, Labuda D, Ruiz-Linares A. 2003. Y-chromosome evidence for differing ancient demographic histories in the Americas. Am J Hum Genet 73:524–539.

Campbell L. 1997. American Indian languages: the historical linguistics of native America. New York: Oxford University Press.

Geppert M, Baeta M, Núñez C, Martínezx-Jarreta B, Zweynert S, Wladimir O, Cruz V, González-Andrade F, González-Solorzano J, Nagy M, Roewer L. 2011. Hierarchical Y-SNP assay to study the hidden diversity and phylogenetic relationship of native populations in South America. Forensic Sci Int Genet 5:100–104.

González-José R, Bortolini MC, Santos FR, Bonatto SL. 2008. The peopling of America: craniofacial shape variation on acontinental scale and its interpretation from an interdisciplinary view. Am J Phys Anthropol 137:175–187.

Jota MS, Bisso-Machado R, Lacerda DR, Sandoval JR, Vieira PPR, Santos-Lopes SS, Revollo S, Paz-y-Miño C, Fujita R, Bonatto SL, Bortolini MC, Santos FR, Genographic Consortium. 2011. A new subhaplogroup of Native American Y chromosomes from the Andes. Am J Phys Anthropol (DOI 10.1002/ajpa.21519).

Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res 18:830–838.

Laughlin WS. 1963. Eskimo and Aleuts. Their origins and evolution. Science 142:633–645.

Lewis MP, editor. 2009. Ethnologue: languages of the World,16th ed. Dallas: SIL International. Online version:http://www.ethnologue.com/.

Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, Metspalu M, Metspalu E, Kivisild T, Gupta R, Bertalan M, Nielsen K, Gilbert MT, Wang Y, Raghavan M, Campos PF, Kamp HM, Wilson AS, Gledhill A, Tridico S, Bunce M, Lorenzen ED, Binladen J, Guo X, Zhao J, Zhang X, Zhang H, Li Z, Chen M, Orlando L, Kristiansen K, Bak M, Tommerup N, Bendixen C, Pierre TL, Grønnow B, Meldgaard M, Andreasen C, Fedorova SA, Osipova LP, Higham TF, Ramsey CB, Hansen TV, Nielsen FC, Crawford MH, Brunak S, Sicheritz-Pontén T, Villems R, Nielsen R, Krogh A, Wang J, Willerslev E. 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature 463:757–762.

Salzano FM. 2002. Molecular variability in Amerindians: widespread but uneven information. An Acad Bras Ciênc 74:223–263.

Santos FR, Pandya A, Tyler-Smith C, Pena SD, Schanfield M, Leonard WR, Osipova L, Crawford MH, Mitchell RJ. 1999. The Central Siberian origin for native American Y chromosomes. Am J Hum Genet 64:619–628.

Seielstad M, Yuldasheva N, Singh N, Underhill P, Oefner P, Shen P, Wells RS. 2003. A novel Y-chromosome variant puts an upper limit on the timing of first entry into the Americas. Am J Hum Genet 73:700–705.

Zegura SL, Karafet TM, Zhivotovsky LA, Hammer MF. 2004. High-Resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. Mol Biol Evol 21:164–175.

Zhong H, Shi H, Qi XB, Duan ZY, Tan PP, Jin L, Su B, Ma RZ. 2011. Extended Y-chromosome investigation suggests post-Glacial migrations of modern humans into East Asia via the northern route. Mol Biol Evol 28:717–727.