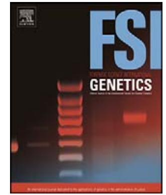
Forensic Science International: Genetics 5 (2011) 100–104

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig



Hierarchical Y-SNP assay to study the hidden diversity and phylogenetic relationship of native populations in South America

Maria Geppert [a,*], Miriam Baeta [b], Carolina Núñez [b], Begoña Martínez-Jarreta [b], Sarah Zweynert [c], Omar Wladimir Vacas Cruz [d], Fabricio González-Andrade [e], Jorge González-Solorzano [f], Marion Nagy [a], Lutz Roewer [a]

[a] Department of Forensic Genetics, Institute of Legal Medicine and Forensic Sciences, Charité-Universitätsmedizin Berlin, Germany
[b] Department of Legal Medicine, University of Zaragoza, Spain
[c] Department Psychiatry and Psychotherapy, Charité-Universitätsmedizin Berlin, Germany
[d] Pontifical Catholic University of Ecuador, Quito, Ecuador
[e] Faculty of Medical Sciences, Central University of Ecuador, Quito, Ecuador
[f] Department of Medicine, Metropolitan Hospital, Quito, Ecuador

ARTICLE INFO

Keywords:
Y-SNPs
SNaPshot
Minisequencing
South America
Ecuador
Native Americans

ABSTRACT

Studying the Y chromosomes of indigenous tribes of Ecuador revealed a lack of strategic SNP assays to examine the substructure of South American native populations. In most studies dealing with South American samples so far only the most common Y-SNP M3 of haplogroup Q was analyzed, because this is known to define a founder group in South America. Studies of SNPs ancestral to Q-M3 (Q1a3a) to confirm the results or the typing of Q subclades have often been neglected. For this reason we developed a SNaPshot assay, which allows first for a hierarchical testing of all main haplogroups occurring in South American populations and second for a detailed analysis of haplogroups Q and C thought having ancient Asian descent. We selected 16 SNPs from the YCC haplogroup tree and established two multiplexes. The first multiplex ("SA Major") includes 12 Y-SNPs defining the most frequent haplogroups occurring in South America (M42, M207, M242, M168, M3, M145, M174, M213, RPS4Y711, M45, P170, and M9). The second multiplex ("SA SpecQ") contains Y-SNPs of haplogroup Q, especially of the subclade Q-M3 (M19, M194, P292, M3, and M199). Within our Ecuadorian sample, haplogroup Q-M3 (xM19, M194, P292, and M199) was predominant, but we also found haplogroup E and R, which can be attributed to recent admixture. Moreover, we found four out of 65 samples, which were tested to be haplogroup C3* (C-M217) the modal haplogroup in Mongolians and widespread in indigenous populations of the Russian Far East as well as in Eastern Asia. This haplogroup is not known to be the result of recent admixture and has been found only one time before in South America. Since haplogroup C occurs in Asia and in North America (C3b or C-P39), we assume that these C-lineages are ancient as well. Therefore, we established a third multiplex ("SA SpecC"), which allows the further subtyping of haplogroup C, mainly of subclade C3 defined by the Y-SNP M217 (M407, M48, P53.1, M217, P62, RPS4Y711, M93, M86, and P39). Altogether, these three multiplexes cover the most frequent haplogroups in South America and allow for a maximal resolution of the Y-chromosomal SNP diversity in Amerindian population samples.

© 2010 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

The analysis of Y-chromosomal haplogroups, defined by binary polymorphisms is a useful tool to study the origin as well as the past demography of human populations. The updated phylogenetic tree of the Y Chromosome Consortium [1] displays the ancestral relationship of all recent Y chromosomes. Using this tree, SNP-typing strategies can be developed to address the issue of population history and movements within geographic territories of interest. The colonization of South America as the last continent settled by modern humans represents one of the more controversial topics in the study of human evolution. While it is generally accepted that the ancestors of Native Americans came from Asia, the number of migratory waves, the places of origin and the timing is a matter of debate between archaeologists, anthropologists and geneticists [2–6]. Also, varying models on the initial genetic differentiation in South Americans across the major geographical regions have been devised [7,8], but due to the paucity of data from Amerindian tribal populations genetic inferences remain still premature. The Spanish and Portuguese conquest around 1500, the

* Corresponding author. Tel.: +49 30450525039; fax: +49 30450525912.
E-mail address: maria.geppert@charite.de (M. Geppert).

1872-4973/$ – see front matter © 2010 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.fsigen.2010.08.016

slave trade and the influx of millions of European settlers in the 19th century result in a sex-biased admixture of contemporary South American populations. The directional mating between European males and Native American as well as African females has clearly had a much more detrimental effect on the Native American paternal lineages than it has had on the frequency of their maternal lineages [9,10].

The study of Y chromosome polymorphisms (STRs and SNPs) assists in inferring the forces that shaped the contemporary Native American gene pool [4] . It has been demonstrated that although the tribal populations of South America diverge remarkably in their Y-STR haplotypes composition [11,12] only one ancient haplogroup is predominant among Amerindians: haplogroup Q1a3a, which has been explained by a major founder event [13]. Other more frequent haplogroups (mainly R and E) are introduced through directional mating from European settlers and African slaves. Therefore, in most studies dealing with South American samples only a few Y-SNPs were analyzed [7,14–18]. The main focus is on haplogroup Q and in particular on the subclade Q1a3a defined by the Y-SNP M3. However, hidden variability of the ancient Amerindian genepool exists as it has been shown by the detection of haplogroup C-M217 in two individuals of the Wayuu inhabiting the La Guajira Peninsula straddling the Venezuela-Colombian border [3,4]. Consequentially, we designed three multiplexes in a hierarchical order for the analysis of all Y-SNPs, which are expected to occur in South American population samples. The hierarchical approach developed earlier for a number of populations including the Europeans [19,20] is advantageous, because it reduces the number of SNPs that have to be typed and thus saves time, financial resources and DNA. Especially the latter point is most important for samples, which were collected during field research, because there is only a restricted amount of genomic material available.

2. Materials and methods

2.1. DNA samples

A total of 65 males of two Waorani communities located in the Ecuadorian Amazon region were analyzed (Supplementary Table 1). All male relatives have been included in the study. The ethnographic description of this tribe is given in [11]. While the majority of villagers belong to the Waorani ethnic group ($n = 42$), individuals of other self-defined ethnic groups were represented—15 Kichwa speakers, 2 Shuar, 4 Mestizos and 2 Afroecuadorians. A written document on informed consent approved by an ethical commission was translated by a native Waorani translator to the members of the communities, some of them voluntarily agreed in providing buccal swabs. Genomic DNA was extracted using the BioRobot EZ1 workstation with the EZI DNA Blood Card (Qiagen). The Y-STR typing was performed with the AmpFlSTR® Yfiler® PCR amplification kit (Applied Biosystems) according to the protocol. Four additional Y-STRs (DYS471, 570, 576, 643) were analyzed as well [21]. YHRD accession numbers were issued after evaluation of data (YA003552 and YA003553), which can be used to search the samples in the YHRD (www.yhrd.org/Search/Contributions). Two in-house control DNAs, whose haplogroups were assigned by two different methods (pyrosequencing and minisequencing), served as positive controls for gel and capillary electrophoresis.

2.2. Multiplex PCRs for Y-SNPs

Three multiplexes "South American (SA) Major", "SA SpecC" and "SA SpecQ" were performed in a final volume of 14.4, 10.4 and 7.8 μl, respectively, with 1× Multiplex Mastermix (Qiagen), 2 ng genomic DNA per reaction and primers at the concentrations indicated in Supplementary Table 2. The cycling conditions were 95 °C for 15min followed by 32 cycles of 94 °C for 30 s, 60 °C for 90 s, 72 °C for 60 s and a final extension of 72 °C for 10 min. All primers, except M9 [19], were in-house designed to restrict the template size to a maximum of 300 bp. Primer Quest (http://eu.idtdna.com/Scitools/Applications/Primerquest/), Primer3 PLUS software (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) and OligoCalulator (http://www.basic.northwestern.edu/biotools/oligocalc.html) were used to design and check the primers for secondary structures. Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastAlnAd), BLAT Search Genome (http://genome.ucsc.edu/cgibin/hgBlat?command=start), and UCSC In-Silico PCR (http://genome.ucsc.edu/cgi-bin/hgPcr?org=Human&db=hg18&hgsid=145163942) were used to check binding sites of the primers in the genome and to assure that all PCR primers are specific for the Y chromosome. Primer sequences, concentrations and further information are given in Supplementary Table 2. Sizes of the amplicons were chosen to be at least 4bp different to enable the checking of amplification results by gel electrophoresis. The amplicons were separated by using a 2.5% agarose gel (2 h, 100 mV).

2.3. Multiplex single base extension (SBE)

Before single base extension, 2 μl PCR-product was cleaned up with 1 μl ExoSAP-IT® (USB) and incubated 60 min at 37 °C followed by 15 min at 75 °C to inactivate the enzymes. Multiplex single base extension reaction was performed in a final volume of 11 μl, consisting of 3 μl *ABI PRISM® SNaPshot™ Multiplex Ready Reaction Mix* (Applied Biosystems), 3 μl cleaned up PCR-product, SBE primers and dH_2O. The cycling conditions were 96 °C for 10 s, 50 °C for 5 s and 60 °C for 30 s, for 25 cycles. Design of SBE primers was conducted with the same software as PCR primers. Additionally, each SBE primer was extended to a certain length with an unspecific non-human sequence [22]. All SBE primers and concentrations are shown in Supplementary Table 3. SBE products were cleaned up with 1 U of shrimp alkaline phosphatase (USB) by incubation at 37 °C for 60 min and the enzyme was inactivated at 85 °C for 15 min.

2.4. Capillary electrophoresis

SBE products were detected on an ABI 3130xl Genetic Analyzer (Applied Biosystems) by running 19.5 μl Hi-Di™ formamide (Applied Biosystems), 0.5 μl size standard GeneScan™ 120 LIZ® and 0.5 μl of the sample with Dye Set E5 and Matrix DS-02. Analysis of the electropherograms was performed with Gene Mapper® ID 3.2.1.

3. Results

The combined Y-STR/Y-SNP data from 65 inhabitants of two Waorani villages are presented in Supplementary Table 1. The assignment of haplogroups presented in this study follows the updated phylogenetic tree of the Y Chromosome Consortium [1]. The selection of Y-SNPs is based on literature [4,23–26] and comprises all haplogroup-specific markers, which have been detected in South American populations to date. Altogether 24 Y-SNPs were selected to be typed within three multiplexes/SBE-reactions, in which the SNPs were arranged in a hierarchical order according to their position in the YCC tree (Fig. 1). First multiplex "SA Major" includes 12 Y-SNPs, which define the most frequent major haplogroups (Fig. 2). The second and third multiplex "SA SpecC" and "SA SpecQ" are applicable to analyze subclades of the autochthonous haplogroup C and haplogroup Q in detail. Multiplex "SA SpecC" comprises nine Y-SNPs, which allow particular for a specific typing of the subclade C-M217 (Fig. 2). For the subclade

Fig. 1. Phylogenetic tree according to [1]. Solid lines indicate haplogroups, which can be typed by the three new Y-SNP assays. Markers that have been typed are indicated upon the lines. Dotted lines indicate haplogroups, which are not included in the multiplexes. The length of each branch has no significance.

C3c, both markers M48 and M86 were analyzed, because they were found to define different branches. Pakendorf et al. [27] showed that individuals, which were typed to be ancestral for C-M48 were derived for C-M86. Multiplex "SA SpecQ" comprises five Y-SNPs with the focus on subclade Q-M3 (Fig. 2). Due to the few data available for the subclades of Q-M3 two (M199 and P292) of the three known markers of subclade Q1a3a3 were analyzed, to increase the informative value of the multiplex and to allow for the detection of unknown diversity within subclade Q-M3. Both specific multiplexes include each a single Y-SNP, which is already included in multiplex "SA Major". This repetition of SNP determination serves as a confirmation of the typing results in the first multiplex and can be helpful to avoid a mix up of samples. The 24 Y-SNPs allow for a determination of 17 haplogroups, although in our sample only 4 haplogroups have been found (Supplementary Table 1). While 55 of the 65 tested individuals belong to the most common haplogroup in South America Q-M3, six individuals belong to non-Amerindian haplogroups R (4) and E (2). Interestingly, four unrelated male individuals (3 Waorani and 1 Kichwa) carried haplogroup C-M217 (C3*), which has been found only once before in South America in two Wayuu males. Though, in contrast to our full-resolution approach only two of six markers downstream to M217 were analyzed (i.e. P39 AND M86) [3,4].

4. Discussion

We present here for the first time hierarchical Y-SNP assays for the native populations of South America which completely resolve autochthonous Q and C lineages. Application to an Ecuadorian native population sample confirmed the robustness of the system as well as the sensitivity to handle low concentrated or degraded DNA. With regard to the technical protocols, all three multiplexes were optimized in the manner that the SBE-reactions show well-balanced heights of all peaks in order to increase the chance of receiving full profile results from all DNA samples. Therefore, concentrations of primers were individually adapted for optimal efficiency. To obtain absolute reliable results, alleles must be determined with certainty and misinterpretation should be excluded. Therefore, we aimed to reduce the background in the electropherograms to a minimum by cleaning up the PCR- and SBE-reaction carefully. Therefore we ensure to carry out the protocol quickly, because the enzymes are sensible to room temperature, and use fresh ExoSAP in the required volume. Additionally, the number of Y-SNPs included in a single multiplex/SBE-reaction was kept to a low number and only those markers were chosen, which are well studied and known to have no paralogous sequence variants. Concerning the SBE-primers the extension with the unspecific non-human sequence resulted in electropherograms with less background than extending the primers with poly (dC) tails, which in our experiments had a tendency to degrade rapidly and to generate incorrect elongations. While the multiplexes represent the most current state of the phylogeny, the full potential of the multiplexes is not utilized and new markers can be implemented without problems.

Besides haplogroups Q1a3a* (55; 85%), R (4; 6%) and E (2; 3%), the haplogroup C3* was determined in 4 out of 65 Waorani and

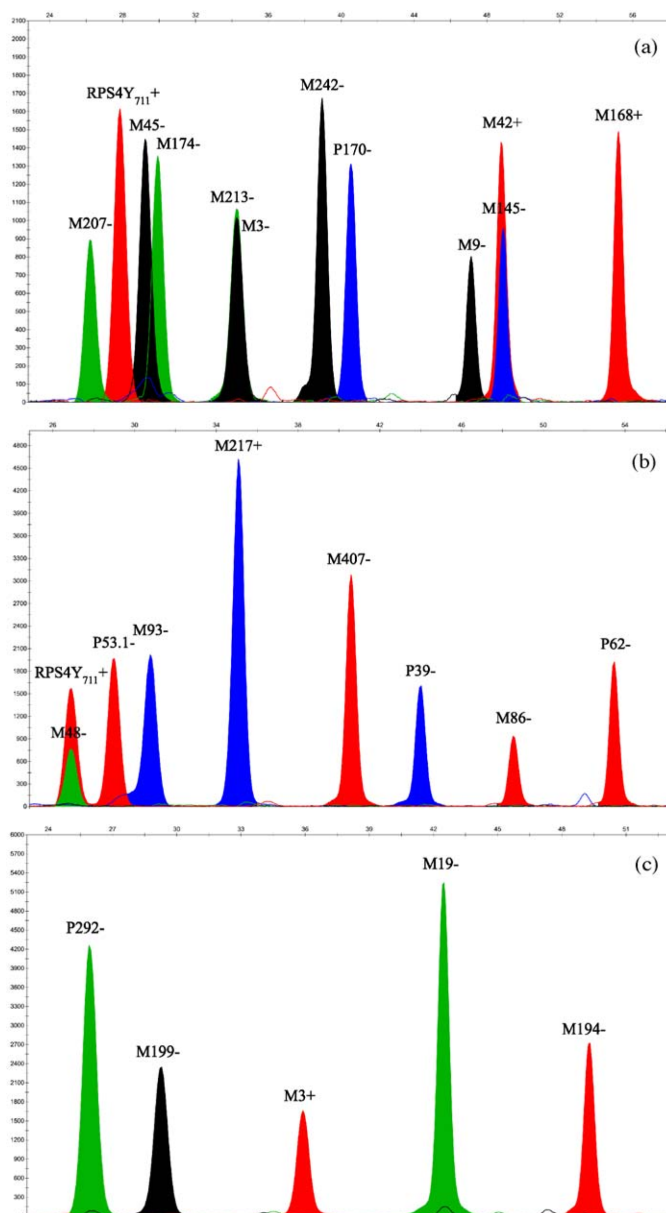


Fig. 2. Electropherograms of the SBE-reactions of two different samples. SNPs are named above the corresponding peaks. A plus indicates derived status of the SNP and a minus indicates ancestral status. (a) Multiplex ''SA Major'' including 12 Y-SNPs from a sample assigned to haplogroup C-RPS4Y_{711}. (b) Multiplex ''SA SpecC'' including 9 Y-SNPs from the same sample as (a) assigned to haplogroup C-M217. (c) Multiplex ''SA SpecQ'' including 5 Y-SNPs from a sample assigned to haplogroup Q-M3.

Kichwa males (6%). In comparison to North American haplogroups, where C3b (defined by P39) is the most frequent in Native Americans, haplogroup C3* is more ancient. This finding is surprising, because in view of the hypothesis of the peopling of South America by Asians migrating by land through North America it was expected to find the same C lineage or a more derived one as in North America. With now six Y chromosomes belonging to the Central and Eastern Asian haplogroup C3* found in three tribes living in the Northwest of South America (Wayuu, Lowland Kichwa and Waorani) existing models of the entry and early distribution of settler in South America could be challenged. Reanalysis of existing poorly resolved samples with the proposed SNaPshot assays may lead to further observations of rare founder lineages in South America.

5. Conclusion

The three multiplexes presented here encompass the most frequent haplogroups in South America and allow for an evaluation of the complete Y-chromosomal diversity in Amerindian populations. Detection of the full spectrum of haplogroups occurring in tribal populations and the maximal resolution of the autochthonous haplogroups Q and C is the most remarkable advantage of the assay. More sampling projects and a more comprehensive analysis (and reanalysis of samples) is necessary to detect hidden traces of ancient migration waves into and within South America.

Acknowledgements

We are very grateful to the indigenous peoples of Ecuador who donated the samples analyzed here. This research is part of a joint initiative called the EQGen project aiming to study the genetics of the Native American populations in Ecuador.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.fsigen.2010.08.016.

References

[1] T.M. Karafet, F.L. Mendez, M.B. Meilerman, P.A. Underhill, S.L. Zegura, M.F. Hammer, New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree, Genome Res. 18 (2008) 830–838.
[2] J.H. Greenberg, C.G.T. Ii, S.L. Zegura, L. Campbell, J.A. Fox, W.S. Laughlin, E.K.J.E. Szathmary, K.M. Weiss, E. Woolford, The settlement of the Americas: a comparison of the linguistic, dental, and genetic evidence [and comments and reply], Curr. Anthropol. 27 (1986) 477–497.
[3] T.M Karafet, S.L. Zegura, O. Posukh, L. Osipova, A. Bergen, J. Long, D. Goldman, W. Klitz, S. Harihara, P. de Knijff, V. Wiebe, R.C. Griffiths, A.R. Templeton, M.F. Hammer, Ancestral Asian source(s) of new world Y-chromosome founder haplotypes, Am. J. Hum. Genet. 64 (1999) 817–831.
[4] S.L. Zegura, T.M. Karafet, L.A. Zhivotovsky, M.F. Hammer, High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas, Mol. Biol. Evol. 21 (2004) 164–175.
[5] D.A Merriwether, F. Rothhammer, R.E. Ferrell, 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 (1995) 411–430.
[6] C.J. Mulligan, A. Kitchen, M.M. Miyamoto, Updated three-stage model for the peopling of the Americas, PLoS One 3 (2008) e3199, doi:10.1371/journal.pone.0003199.
[7] E Tarazona-Santos, D.R. Carvalho-Silva, D. Pettener, D. Luiselli, G.F. De Stefano, C.M. Labarga, O. Rickards, C. Tyler-Smith, S.D.J. Pena, F.R. Santos, Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome, Am. J. Hum. Genet. 68 (2001) 1485–1496.
[8] F.M. Salzano, Molecular variability in Amerindians: widespread but uneven information, An. Acad. Bras. Cienc. 74 (2002) 223–263.
[9] D.A. Bolnick, D.I. Bolnick, D.G. Smith, Asymmetric male and female genetic histories among native Americans from Eastern North America, Mol. Biol. Evol. 23 (2006) 2161–2174.
[10] M. Jobling, M.E. Hurles, C. Tyler-Smith, Human Evolutionary Genetics: Origins, Peoples and Disease, Garland Science Publishing, London/New York, 2004 p. 523.
[11] F. Gonzalez-Andrade, L. Roewer, S. Willuweit, D. Sanchez, B. Martinez-Jarreta, Y-STR variation among ethnic groups from Ecuador: Mestizos, Kichwas, Afro-Ecuadorians and Waoranis, Forensic Sci. Int. Genet. 3 (2009) e83–e91. , doi:10.1016/j.fsigen.2008.08.003.
[12] T.J. Palha, E.M. Rodrigues, S.E. Dos Santos, Y-STR haplotypes of native American populations from the Brazilian Amazon region, Forensic Sci. Int. Genet., in press.
[13] P.A. Underhill, L. Jin, R. Zemans, P.J. Oefner, L.L. Cavalli-Sforza, A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 196–200.
[14] J.T. Lell, M.D. Brown, T.G. Schurr, R.I. Sukernik, Y.B. Starikovskaya, A. Torroni, L.G. Moore, G.M. Troup, D.C. Wallace, Y chromosome polymorphisms in native American and Siberian populations: identification of native American Y chromosome haplotypes, Hum. Genet. 100 (1997) 536–543.
[15] A.R. Marrero, W.A. Silva-Junior, C.M. Bravi, M.H. Hutz, M.L. Petzl-Erler, A. Ruiz-Linares, F.M. Salzano, M.C. Bortolini, Demographic and evolutionary trajectories of the Guarani and Kaingang natives of Brazil, Am. J. Phys. Anthropol. 132 (2007) 301–310.
[16] S. Mazieres, E. Guitard, E. Crubezy, J.M. Dugoujon, M.C. Bortolini, S.L. Bonatto, M.H. Hutz, E. Bois, F. Tiouka, G. Larrouy, F.M. Salzano, Uniparental (mtDNA, Y-chromosome) polymorphisms in French Guiana and two related populations—implications for the region's colonization, Ann. Hum. Genet. 72 (2008) 145–156.

[17] A. Ruiz-Linares, D. Ortiz-Barrientos, M. Figueroa, N. Mesa, J.G. Munera, G. Bedoya, I.D. Velez, L.F. Garcia, A. Perez-Lezaun, J. Bertranpetit, M.W. Feldman, D.B. Goldstein, Microsatellites provide evidence for Y chromosome diversity among the founders of the New World, Proc. Natl. Acad. Sci. U.S.A. 96 (1999) 6312–6317.
[18] N.O. Bianchi, C.I. Catanesi, G. Bailliet, V.L. Martinez-Marignac, C.M. Bravi, L.B. Vidal-Rioja, R.J. Herrera, J.S. Lopez-Camelo, Characterization of ancestral and derived Y-chromosome haplotypes of New World native populations, Am. J. Hum. Genet. 63 (1998) 1862–1871.
[19] M. Brion, B. Sobrino, A. Blanco-Verea, M.V. Lareu, A. Carracedo, Hierarchical analysis of 30 Y-chromosome SNPs in European populations, Int. J. Legal Med. 119 (2005) 10–15.
[20] V. Onofri, F. Alessandrini, C. Turchi, M. Pesaresi, L. Buscemi, A. Tagliabracci, Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs, Forensic Sci. Int. 157 (2006) 23–35.
[21] M. Geppert, J. Edelmann, R. Lessig, The Y-chromosomal STRs DYS481, DYS570, DYS576 and DYS643, Legal Med. 11 (2009) S109–S110.
[22] K. Lindblad-Toh, E. Winchester, M.J. Daly, D.G. Wang, J.N. Hirschhorn, J.P. Laviolette, K. Ardlie, D.E. Reich, E. Robinson, P. Sklar, N. Shah, D. Thomas, J.B. Fan, T. Gingeras, J. Warrington, N. Patil, T.J. Hudson, E.S. Lander, Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse, Nat. Genet. 24 (2000) 381–386.
[23] M.C. Bortolini, F.M. Salzano, M.G. Thomas, S. Stuart, S.P. Nasanen, C.H. Bau, M.H. Hutz, Z. Layrisse, M.L. Petzl-Erler, L.T. Tsuneto, K. Hill, A.M. Hurtado, D. Castro-de-Guerra, M.M. Torres, H. Groot, R. Michalski, P. Nymadawa, G. Bedoya, N. Bradman, D. Labuda, A. Ruiz-Linares, Y-chromosome evidence for differing ancient demographic histories in the Americas, Am. J. Hum. Genet. 73 (2003) 524–539.
[24] A. Blanco-Verea, J.C. Jaime, M. Brión, A. Carracedo, Y-chromosome lineages in native South American population, Forensic Sci. Int. Genet. 4 (2010) 187–193.
[25] G. Bailliet, V. Ramallo, M. Muzzio, A. Garcia, M.R. Santos, E.L. Alfaro, J.E. Dipierri, S. Salceda, F.R. Carnese, C.M. Bravi, N.O. Bianchi, D.A. Demarchi, Brief communication: restricted geographic distribution for Y-Q* paragroup in South America, Am. J. Phys. Anthropol. 140 (2009) 578–582.
[26] J.T. Lell, R.I. Sukernik, Y.B. Starikovskaya, B. Su, L. Jin, T.G. Schurr, P.A. Underhill, D.C. Wallace, The dual origin and Siberian affinities of native American Y chromosomes, Am. J. Hum. Genet. 70 (2002) 192–206.
[27] B. Pakendorf, I.N. Novgorodov, V.L. Osakovskij, A.P. Danilova, A.P. Protod'jakonov, M. Stoneking, Investigating the effects of prehistoric migrations in Siberia: genetic variation and the origins of Yakuts, Hum. Genet. 120 (2006) 334–353.