Human Y- chromosome SNP characterization by multiplex amplified product-length polymorphism analysis.

Journal:	*ELECTROPHORESIS*
Manuscript ID:	elps.201400020.R1
Wiley - Manuscript type:	Short Communication
Date Submitted by the Author:	02-May-2014
Complete List of Authors:	Jurado Medina, Laura; Instituto Multidisciplinario de Biología Celular, Muzzio, Marina; Instituto Multidisciplinario de Biología Celular, ; Facultad de Ciencias Naturales y Museo. Universidad de La Plata, Schwab, Marisol; IMBICE Instituto Multidisciplinario de Biología Celular, Bravi Constantino, María; IMBICE Instituto Multidisciplinario de Biología Celular, Barreto, Guillermo; Universidad del Valle, Departamento de Biología Bailliet, Graciela; Instituto Multidisciplinario de Biología Celular,
Keywords:	APLP, SNP, South America, Y chromosome

Short Communication

Human Y- chromosome SNP characterization by Multiplex amplified product-length polymorphism analysis.

Laura Smeldy Jurado Medina[1], Marina Muzzio[1,2], Marisol Schwab[1], María Leticia Bravi Constantino[1], Guillermo Barreto[3], Graciela Bailliet[1]

1. Laboratorio de Genética Molecular Poblacional, Instituto Multidisciplinario de Biología Celular (IMBICE). CCT- CONICET-La Plata / CICPBA, Argentina.

2. Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata, Argentina

3. Laboratorio de Genética Molecular Humana. Departamento de Biología, Universidad del Valle, Colombia.

Name, address, telephone number, fax number, and email address of the person to whom the proof is to be sent: Graciela Bailliet. IMBICE. 526 e/ 10 y 11. CC 403. 1900 La Plata, Argentina. Phone: 54 221 4210112 Ext.220. Fax.: 54 221 4210112 Ext.222.

E-mail: gbailliet@imbice.gov.ar

Key words: APLP, SNP, South America, Y chromosome.

Word count for the body of the manuscript: 2160

Abstract

We designed an allele-specific amplification protocol to optimize Y-chromosome SNP typing, which is an unavoidable step for defining the phylogenetic status of paternal lineages. It allows the simultaneous highly-specific definition of up to six mutations in a single reaction by amplifying products of different size (APLPs) without the need of specialized equipment, at a considerably lower cost than that based on single-base primer extension (SNaPshot™) technology or PCR-RFPL systems, requiring as little as 0.5 ng DNA and compatible with the small fragments characteristic of low-quality DNA. By designation of 2 primers recognizing the derived and ancestral state for each SNP, which can be differentiated by size by the addition of a non-complementary nucleotide tail, we could define major Y clades E, F, K, R, Q, and subhaplogroups R1, R1a, R1b, R1b1b, R1b1c, J1, J2, G1, G2, I1, Q1a3, and Q1a3a1 through amplification fragments that ranged between 60 and 158bp.

It is possible to study the origin, dispersal and migration routes into the continents of human populations through the analysis of the non-recombining region of the Y chromosome (1,2). In virtue of the uniparental transmission of that region (its mutations are transmitted from father to son) researchers can find out the phylogenetic status of a lineage and assign a possible geographical origin (2, 3) by defining haplogroups from single nucleotide polymorphisms (SNPs), with a specific nomenclature (4,5).

There are several ways to analyze SNPs: restriction length polymorphisms (RFLP) (6), fluorescence techniques such as SNaPshot (7), and Real-Time PCR (8); however these methods are expensive and require specialized equipment. Software predictions from short tandem repeats (STRs) combinations (haplotypes) (9) are an alternative strategy to define haplogroups in the wet lab, though they have been questioned for their imprecision and high probability of wrong assignation (10).

We present here a new allele-specific amplification method for the identification of 17 Y-chromosome haplogroups present in America, by amplifying products of different size (PCR-AFLP) in a multiplex reaction (Table 1, Fig. 1). We have selected these haplogroups following our experience in relation with their frequency in human populations from Argentina (11, 12, 13, 14). This technique of allele-specific SNPs has been previously applied to mitochondrial DNA (15), resulting in a low-cost method with high specificity and sensitivity.

Our method is based on the modification of a primer that preferably recognizes an allelic variant. Under high-specificity conditions, efficient PCR reactions can be achieved when the base at 3'hydroxyl end of the primer is complementary with the corresponding nucleotide, while the absence of the complementary nucleotide results in the absence or replication

deficiency in the other allele. Three primers were designed for each SNP: 2 that specifically recognize the nucleotide abutting sequence for derived and ancestral alleles respectively, one of which was added a tail of non-complementary nucleotides to distinguish such amplification products in the two alleles, and the third oligonucleotide which delimits the size of the fragment.

a) We selected SNPs in order to characterize subhaplogroups from the most frequent haplogroups in human populations from Argentina reported in previous studies from our laboratory (R, F, and Q) (11, 12, 13, 14), particularly those with frequency differences between European and Middle East populations, i.e. subhaplogroups G1, G2 (16) , J1, J2 (17), I1 (18), R1, R1a1 (19) , R1b, R1b1b, R1b1c (20). Once the mutation was located (http://www.ncbi.nlm.nih.gov/ SNP? Term = 20sapiens homo%), an adjacent fragment of 100 bp before and 100 bp after the mutated nucleotide was selected. The sequence was then analyzed, and we determined which of the primers (forward or reverse) included the SNP.

b) We used Primer3 v0.3.0 software (http://frodo.wi.mit.edu) (21) for the design of oligonucleotides. The size was delimited in accordance with the experimental design, taking into account the number of markers included in each Multiplex reaction in order to avoid overlapping between the fragments.

c) The allele-specific primers had complementary bases to the variant allele that we wanted to amplify at the 3'OH end. A tail of the non-complementary nucleotide was added to the flanking sequence of the primer at the 5 'end. The nucleotide tail generally unbalances the melting temperature (Tm) and so, when differences between Tm were high, a mismatch was introduced for stabilization. Once the primer was modified, we tested again with Primer3 software in order

to confirm the stability of the sequence. We followed the same procedure to design the primer that recognizes the other allelic variant, changing the nucleotide at the 3'end and removing the tail nucleotides at 5´ end, so as to observe the difference of amplification fragments for each allele.

d) As a last check, we obtained thermodynamic stability values (free energy; ΔGo) using the OligoCalc program (http://www.basic.northwestern.edu/biotools/oligocalc.html) (22), which allowed us to detect the interactions between primers (formation of hairpins, homodimers, or heterodimers) that could cause errors in the PCR reactions. This software allows recognizing primer complementarities or other forms of secondary structure. After this, the oligonucleotides were ready to be synthesized.

The Multiplex APLP was used to analyze biological samples from voluntary donors from Argentina, who provided informed consent, with the approval of Argentine Ethics Committees. Samples were coded and submitted in anonymity for DNA testing. We also searched for positive controls in CEPH samples.

The amplification was tested with varying DNA concentrations from 0.5 to 30ng (Figure 2) to check the reliability of the method at low sample concentrations. We amplified the positive control of Y23Powerplex Amplification Kit (Promega) at 10ng/uL DNA concentration. We analyzed Multiplex I in some DNA samples from exhumed human bones that were provided by the Forensics Laboratory at our research institute, and from degraded low quality DNA in order to check the efficiency of amplification (Figure 3 and 4). We used NanoDrop (Thermo Scientific) for quantification of DNA samples.

PCR reactions were organized into four Multiplex reactions (Table 1, Figure 1, Multiplex I to IV). Each set of primers was mixed in a single reaction tube, and produced fragments between 60 and 160bp. Concentrations and primer sequences used are shown in Table 1. The final volume of the PCR reactions was 10µL, with the following composition: 1ng genomic DNA, 0.025U Platinum® Taq DNA Polymerase (Invitrogen), 0.75 mM MgCl2, 1X buffer (Buffer) reaction, 100µM dNTP, and ultrapure water added to the final volume of each primer (optimum concentration) (Table 1). The amplification reaction was run in a Biometra T3000 thermocycler. After 4 min denaturation at 94°C, 33 cycles of amplification (1 min 94°C, 1 min 56°C, and 1 min 72°C) were run, followed by a final extension step end of 72°C for 10 min.

All samples were amplified for Multiplex I (Fig. 5). Every time all SNPs from Multiplex I showed ancestral state, P152 was analyzed in order to determine E haplogroup. These SNPs could not be included in Multiplex I due to size incompatibility with SNP fragments, and lack of phylogenetic consistency with any of the other Multiplexes. The finding of the derived state of M242 demonstrated the presence of Q haplogroup, in which case we amplified for Multiplex II in order to determine Q subhaplogroups Q1a3* or Q1a3a (Fig. 6). When the derived state M89, ancestral M9, and M172 indicated the presence of F haplogroup, we amplified Multiplex III in order to identify J1, I1, G1, and G2 subhaplogroups (Fig. 7). When derived state from M207 and M173 was detected, Multiplex IV was analyzed in order to determine the R1 sub-haplogroup status R1, R1a, R1b, R1b1b, or R1b1c (Fig. 8).

We run 5µl PCR product in 10% neutral polyacrylamide electrophoresis gel . The DNA bands were detected by GelRed staining. Each allele was assigned by comparison of the different sizes of PCR products. Figures 5, 6, 7 and 8 show the PCR amplicons obtained (Table 1, Figure 1).

APLP analyses started by the amplification of samples acting as controls that had been previously analyzed by RT-PCR and published, for example those from the CEPH that were YAP+ (12), and Native American origin of haplogroup Q (11). We observed complete concordance of results from both methods. Besides, the different amplification fragments inside each Multiplex reaction acted as control of the phylogenetic status of lineages (Figure 1). It is a two-step method that requires minimum budget and minimum equipment, such as conventional PCR, vertical electrophoresis, and digital photography.

References

1. Karafet, T.M., S.L. Zegura, O. Posukh, L. Osipova, A. Bergen, J. Long, D .Goldman, W. Klitz, et al., 1999. Ancestral Asian Source(s) of New World Y-Chromosome Founder Haplotypes. American Journal of Human Genetics. 64: 817–831.

2. Underhill, P.A., G. Passarino, A.A. Lin, P. Shen, M. Mirazón Lahr, R.A. Foley, P. Oefner, L.L. Cavalli-Sforza. 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Annals of Human Genetics. 65:43-62

3. Hammer, M.F., A.B. Spurdle, T.M. Karafet, M.R. Bonner, E.T. Wood, A. Noveletto , P. Malaspina, R.J. Mitchell, et al. 1997. The Geographic Distribution of Human Y Chromosome Variation. Genetics. 145: 787-805.

4. Karafet, T.M., F.L. Mendez, M.B. Meilerman, P.A. Underhill, S.L. Zegura, M.F. Hammer. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Research. 18(5):830-8.

5. Y Chromosome Consortium. 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Research 12:339–348.

6. Torroni, A., T.G. Schurr, C. Yang, E.J.E. Szathmary, R.C. Williams, M.S. Schanfierld , C. Troup, W.C. Knowler, et al. 1992. Native American Mitochondrial DNA Analysis Indicates That the Amerind and the Nadene Populations Were Founded by Two Independent Migrations. Genetics. 130:153–162.

7. Van Oven, M. &, R. A. Kayser M. 2011. An efficient multiplex genotyping approach for detecting the major worldwide human Y-chromosome haplogroups. International Journal Legal Medicine. 125(6):879-85.

8. Zuccarelli, G., E. Alechine, M. Caputo, C. Bobillo, D. Corach, A. Sala. 2010. Rapid screening for Native American mitochondrial and Y-chromosome haplogroups detection in routine DNA analysis. Forensic Science International: Genetics. 5: 105-108.

9. Athey, T.W. 2005. Haplogroup Prediction from Y-STR Values Using an Allele Frequency Approach. Journal of Genetic Genealogy. 1:1-7.

10. Muzzio, M., V. Ramallo, J.M.B. Motti, M.R. Santos, J.S. López-Camelo, G. Bailliet. 2010. Software for Y-haplogroup predictions: a word of caution. International Journal Legal Medicine. doi: 10.1007/s00414-009-0404-1.

11. Bailliet, G. V. Ramallo, M. Muzzio, A. García, M.R. Santos, E.L. Alfaro, J.E. Dipierri, S. Salceda, F.R. Carnese, N.O. Bianchi, D.A. Demarchi. 2009. Restricted geographic distribution for Y-Q* paragroup in South American. Am. J. Phys. Anthrop 140:578–582.

12. Bravi, C.M., G. Bailliet, V.L. Martínez-Marignac, N.O. Bianchi. 2000 Origin of YAP+ lineages of the human Y-chromosome. Am. J. Phis. Antrop. 112 (2):149-158.

13. Dipierri J.E., E.L. Alfaro, V.L. Martínez-Marignac, G. Bailliet, C.M. Bravi, S. Cejas, N.O. Bianchi.1998. Paternal directional mating in two amerindian subpopulations from the northwest of Argentina. Hum. Biol. 70 (6):1001-1010.

14. Ramallo V, Mucci JM, García A, Muzzio M, Motti JMB, Santos MR, Perez ME, Alfaro EL, Dipierri JE, Demarchi DA, Bravi CM, Bailliet G. 2009. Comparison of Y chromosome haplogroup frequencies in 8 Provinces of Argentina. Forensic Science Int: Genetics Supplement Series 2:431–432.

15. Umetsu, K., M. Tanaka, I. Yuasa, N. Saitou, T. Takeyasu, N. Fuku, E. Naito, K. Ago, et al. 2001. Multiplex amplified product-length polymorphism analysis for rapid detection of human mitochondrial DNA variations. Electrophoresis. 22: 3533–3538.

16. Jobling, M.A. and C. Tyler-Smith. 2003. The human Y chromosome: Anevolutionary marker comes of age. Nat. Rev. Genet. 4: 598–612.

17. Semino, O., C. Magri, G. Benuzzi, A.A. Lin, N. Al-Zahery, L. Battaglia, V.,Maccioni, C. Triantaphyllidis,P. Shen, P.J. Oefner, et al. 2004. Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: Inferences on the neolithization of Europe and later migratory events in the Mediterranean area. Am. J. Hum. Genet. 74: 1023–1034.

18. Rootsi, S., C. Magri, T. Kivisild, G. Benuzzi, H. Help, M. Bermisheva, I. Kutuev, L. Barac, M. Pericic, O. Balanovsky, O., et al. 2004. Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. Am. J. Hum. Genet. 75: 128–137.

19. Underhill, P.A., N.M. Myres, S. Rootsi, M. Metspalu, L.A. Zhivotovsky, R.J. King, A.A. Lin, C-E.T. Chow, O. Semino, V. Battaglia. 2010. Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. Eu J Human Genetics: 18, 479–484.

20. Myres, M.R., S. Rootsi, A.A. Lin, M. Jarve, R.J. King, I. Kutuev, V.M. Cabrera, E.K. Khusnutdinova, A. Pshenichnov, B. Yunusbayev, O. Balanovsky, E. Balanovska, P. Rudan, M.

Baldovic, R.J. Herrera, J. Chiaroni, J. Di Cristofaro, R.Villems, T. Kivisild and P.A. Underhill. 2011. A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe European Journal of Human Genetics (2011) 19, 95–101

21. Rozen, S. & H.J. Skaletsky. J. 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp 365-

22. Kibbe, W.A. 2007. OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Research. 35:43-46.

Figure legends

Table 1. Sequence of oligonucleotide primers used for Multiplex APLP method. The non-complementary nucleotides are shown in lower case letters.

Figure 1. Y chromosome phylogenetic tree; mutation names are indicated on the branches.

Figure 2. Multiplex I, using gradient concentration of Control DNA from Y23Powerplex (Promega) (Table 1): lanes 1 to 6, 0.2ng, 0.3ng, 0.5ng, 1ng, 5ng, 10ng; lane M, 25 bp DNA ladder (Invitrogen)

Figure 3. Genomic DNA run in 1% agarose gel. Lines 1-4 low quality and degraded DNA samples; line 5, DNA from exhumed human bones (Forensic Lab); line M, 100bp ladder (Invitrogen).

Figure 4. Multiplex I from low quality DNA samples (S. Fig. 1). Line 1, good quality DNA; lines 2-4, low quality DNA; line 5, DNA from exhumed human bones; line M, 25 bp DNA ladder (Invitrogen).

Figure 5. APLPs-Multiplex I (Table 1). Lanes 1 and 6, Q haplogroup; lane 2, J2 ; lane 3, DE ; lanes 4 and 7, R1; lane 5, F(XJ2); lane M, 25 bp DNA (Invitrogen).

Figure 6. APLPs-Multiplex II. Q Clade of Y chromosome haplogroups (Table 1), Lanes 1 and 2, derivate state for M242, M346, and M3; lane 3, ancestral state for M242, M346, and M3; lane M, 25 bp DNA (Invitrogen).

Figure 7. APLPs-Multiplex III (Table 1). F Clade of Y chromosome haplogroups. Lane 1 ancestral for all SNP; lane 2, J1; lane 3, G1; lane 4, I1; lane 5, G2; lane M, 25 bp DNA (Invitrogen).

Figure 8. APLPs- Multiplex IV (Table 1). R Clade of Y chromosome haplogroups. Lane 1 and 3, R1b1b; lane 2, R1a; lane 4, Q1a3a; lane M, 25 bp DNA ladder (Invitrogen).

Acknowledgements

Authors thank K Umetsu for technical support. We are also grateful to MC Bortolini, DA Demarchi, and LB Vidal Rioja for the provision of control samples for some SNP derivate state analyses.

Grant sponsorship: CONICET, CICPBA, ANPCyT. G. Bailliet and M. Muzzio are CONICET researchers.

Competing interest statement

The authors declare no competing interests

Table 1. Sequence of oligonucleotide primers used for multiplex APLP method

SNP code.	Haplogroup	Primer	Sequence (5'-3')	Concentration (pmol/10uL)	amplicon size (pb)
Multiplex I					
rs2032624	R1	M173F1C	cCAAGGGCATTgAGAACC	3.75	150 (C)
		M173F2A	taatgtatTCAAGGGCATaTAGAACA	3.75	158 (A)
		M173R	CAGCCTTCAAAGCTTCTCCT	3.75	
rs8179021	Q	M242F1C	AAAGGTGACCAAGGTGCTC	2.81	131 (C)
		M242F2T	AacaatAAAGGTtACCAAGGTGCTT	3.75	137 (T)
		M242R	TCAGCATAATACCTTACCTAGAACAAC	2.81	
rs2032658	R	M207F1G	CgAGTCAAGCAAGcAATTTAG	5	124 (G)
		M207F2A	gttcacTAAGTCAAGCAAGAAATcTAA	10	118 (A)
		M207R	AAAAGCTGAAGGAAAAGTGGA	5	
rs2032604	J2	M172F1G	AAACCCATTTgGATGCTTG	6.25	96 (G)
		M172F2T	ctaatAAACCCATTgTGATGCTTT	7.5	101 (T)
		M172R	TTAAATATCAGCCAGGTACAGAGA	6.25	
rs2032652	F (XK,Q,R)	M89R1	CgCAGGCAAAGTGAGAGATA	2	85 (T)
		M89R2	tttaaTCAGGCAAAGTGAGAGATG	2	89 (C)
		M89F	CACAGAAGGATGCTGCTCA	2	
rs3900	K(x Q,R)	M9F1G	GGCCTAAGATGGTTGgATG	3.75	70 (G)
		M9F2C	tttgGaCCTAAGATGGTTGAATC	3.75	74 (C)
		M9R	TGTAAGACATTGAACGTTTGgA	3.75	
Multiplex II					
rs8179021	Q	M242F1C	AAAGGTGACCAAGGTGCTC	3.75	131 (C)
		M242F2T	aacaatAAAGGTtACCAAGGTGCTT	3.75	137 (T)
		M242R	TCAGCATAATACCTTACCTAGAACAAC	3.75	
rs3894	Q1a3a	M3R1T	cGTACCAGCTCTTCCcAATTA	2.5	115 (T)
		M3R2C	attttaaGTACCAGCTCTTCCTAATTG	2.5	121 (C)
		M3F	AAGGTACATTCGCGGGATA	2.5	
Y-position 2947155	Q1a3	M346R1G	cGCAATTTACAAcGACAAGC	5	160 (G)
		M346R2C	aataatgtGCAATTTACAAAGACAAGG	5	167 (C)
		M346F	AAGGCCTGAAAATGTGGAAA	5	
Multiplex III					
rs4116820	G2	P287F1G	TGACTGACATGGCgTTGG	10	149 (G)
		P287F2T	aaaatatatTGACTGACATtGCCTTGT	2.5	158 (T)
		P287R	AAGGGTAATGAAGCCTATAGTCCA	3.75	
rs9341296	I1	M253R1	TCAATGAAGAACCTGGAGAAA	6.25	132 (A)
		M253R2	ttattaTCAATGAAGAACCTGGAGAAG	2.5	138 (G)
		M253F	GCTGATCTGTTTCTTTTTGGTG	2.5	
rs9341313	J1	M267F1G	ggTGGAAGCATTTTgGTAAATAG	3.75	117 (G)
		M267F2T	taagcaTGGAAGCATTTTTGTAAATAT	3.75	121 (T)
		M267R	CGTAGCTcCAAGCAATCCAC	3.75	
rs13447378	G1	M285F1G	CcCATCATCTACATTTCTCCTG	2.5	80 (G)

		M285F2C	gtacaCTCATCATCTACATTTCTCCTC	2.5	85 (C)
		M285R	ATCGAATCCGCTATCCAGACT	2.5	
Multiplex IV					
rs9786184	R1b	M343F1G	GAGTGCCaTCGTGTTCCAC	2.5	140 (C)
		M343F2A	attaaattAGTGaCCTCGTGTTCCAA	2.5	147 (A)
		M343R	ATATGCAAATGCCAGCGTTA	2.5	
rs9785702	R1b1b	P297R1	GCCAAAGATGTATAAGcGGAAG	2.5	104 (C)
		P297R2	tctaaGCCAAAGATGTATAAGAGGAAC	2.5	109 (G)
		P297F	TGGGAGAGAGAGAGACAGGTG	2.5	
Y-position 21148755	R1a	SRY10831.2F1	AggTgAACCTTgAAAATgCTAC	2.5	76 (C)
		SRY10831.2F2	TCgATggTgAACCTTgTAAATgTTAT	2.5	72 (T)
		SRY10831.2R	CATgggATCATTCAgTATCTgg	2.5	
Y-position 13535789	R1b1c	M335R1	ATGCCCTTGAATTGTAAGcAAA	4.5	66 (A)
		M335R2	gattATGCCCTTGAATTGTAAGAAAT	4.5	62 (T)
		M335F	cCTGTAACTTCCTAGAAAATTGGAAA	4.5	
Single					
rs9786634	E	P152F1G	GCcCATCCGTTTTCTGAAG	3.75	141 (G)
		P152F2C	ataaatGCTCATCCGTTTTCTGAAC	3.75	147 (C)
		P152R	TCTCaGGGTTTGTGCAGATTT	3.75	

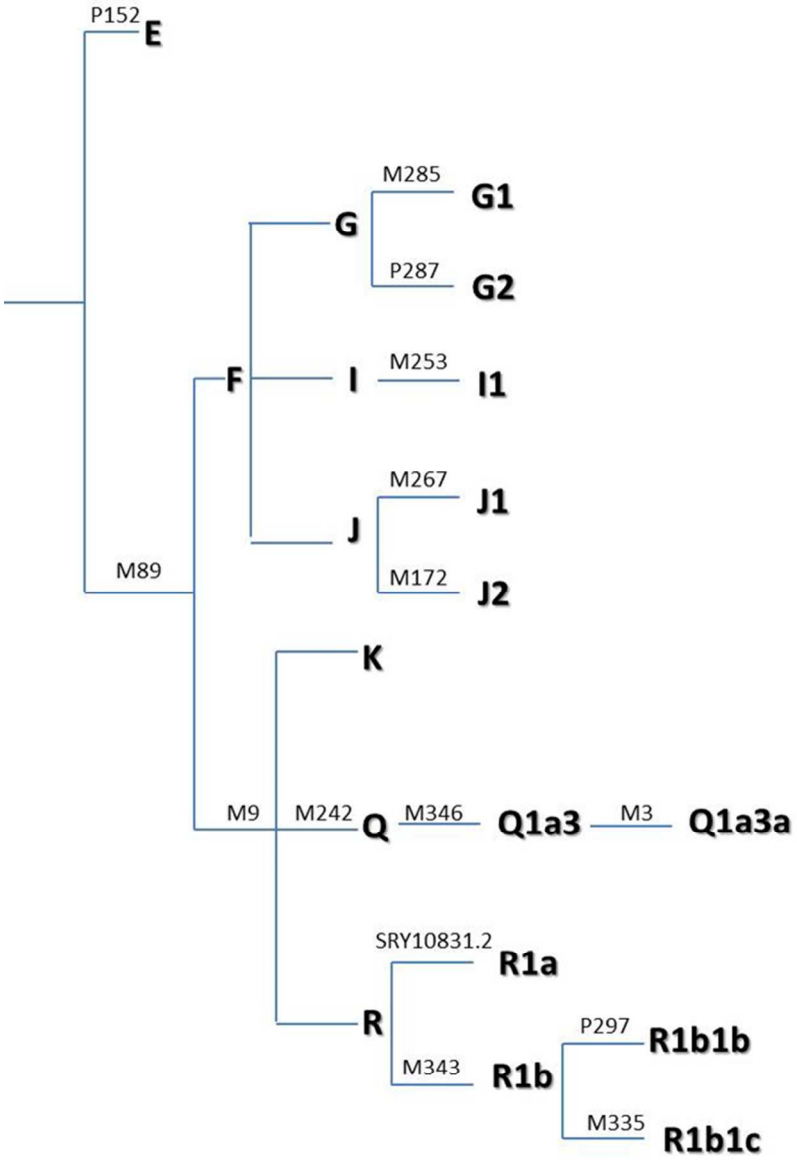

Figure 1. Y chromosome phylogenetic tree; mutation names are indicated on the branches.
119x175mm (150 x 150 DPI)