MUTATION IN BRIEF

Sub-Populations Within the Major European and African Derived Haplogroups R1b3 and E3a Are Differentiated by Previously Phylogenetically Undefined Y-SNPs

Lynn M. Sims [1], Dennis Garvey [4], and Jack Ballantyne [1-3*]

[1] *Biomolecular Sciences, Graduate Program in Chemistry, University of Central Florida, Orlando, Florida;* [2]*Department of Chemistry, University of Central Florida, Orlando, Florida;* [3]*National Center for Forensic Science, Orlando, Florida;* [4] *Department of Physics, Gonzaga University, Spokane, Washington*

*Correspondence to: Jack Ballantyne, Ph.D., Department of Chemistry, University of Central Florida, Bldg #5, 4000 Central Boulevard, Orlando, FL 32816-2366; Tel.: 407-823-4440; Fax: 407-823-2252; E-mail: jballant@mail.ucf.edu

Grant sponsor: National Institute of Justice (NIJ), Department of Justice; grant numbers: 1998-IJ-CX-K003 and 2005-Mu-MU-K044.

Communicated by Michael Dean

Single nucleotide polymorphisms on the Y chromosome (Y-SNPs) have been widely used in the study of human migration patterns and evolution. Potential forensic applications of Y-SNPs include their use in predicting the ethnogeographic origin of the donor of a crime scene sample, or exclusion of suspects of sexual assaults (the evidence of which often comprises male/female mixtures and may involve multiple perpetrators), paternity testing, and identification of non- and half-siblings. In this study, we used a population of 118 African- and 125 European-Americans to evaluate 12 previously phylogenetically undefined Y-SNPs for their ability to further differentiate individuals who belong to the major African (E3a)- and European (R1b3, I)-derived haplogroups. Ten of these markers define seven new sub-clades (equivalent to E3a7a, E3a8, E3a8a, E3a8a1, R1b3h, R1b3i, and R1b3i1 using the Y Chromosome Consortium nomenclature) within haplogroups E and R. Interestingly, during the course of this study we evaluated M222, a sub-R1b3 marker rarely used, and found that this sub-haplogroup in effect defines the Y-STR Irish Modal Haplotype (IMH). The new bi-allelic markers described here are expected to find application in human evolutionary studies and forensic genetics. © 2006 Wiley-Liss, Inc.

KEY WORDS: Y-SNPs; Y-chromosome; E3a; R1b3; M222; IMH; forensic; ethnogeographic origins

INTRODUCTION

Single nucleotide polymorphisms (SNPs) are the smallest and most abundant type of human DNA polymorphisms (Brookes, 1999). SNPs have been extensively used in the study of human evolutionary and migratory patterns (Shastry, 2002) and are increasingly being used in genome-wide association studies (Syvanen, 2005). It is unclear the extent to which SNPs will augment STRs as the primary method of genotyping in forensic

Received 15 May 2006; accepted revised manuscript 16 August 2006.

© 2006 WILEY-LISS, INC.
DOI: 10.1002/humu.9469

science but their potential use in determining population origin specific Y chromosome and mtDNA haplogroups is growing (Sobrino and Carracedo, 2005). Y-SNPs, in particular, are of interest due to their paternal inheritance, lack of recombination, abundance, and low mutation rate and are currently being investigated for characterizing male population structure and individualization in forensic science (Brion, et al., 2005; Hammer, et al., 2005; Jobling, 2001; Kidd, et al., 2005; Sanchez, et al., 2003; Vallone and Butler, 2004). Unique mutations within the non-recombining region (NRY) of the Y-chromosome (mainly SNPs) have created population specific paternal haplogroups that have persisted throughout human history. Potential forensic applications of Y-SNPs include their use in predicting the ethnogeographic origin of the donor of a crime scene sample, inclusion or exclusion of suspects of sexual assaults (the evidence of which often comprises male/female mixtures and may involve multiple perpetrators), paternity testing, and identification of non- and half-siblings.

A large scale parsimonious phylogenetic tree representing world wide Y chromosomal variation has been constructed and comprises the major haplogroups A-R (Jobling and Tyler-Smith, 2003; YCC, 2002). Many of these polymorphisms have proven highly informative in tracing human prehistoric migrations and generating new hypotheses on human colonizations and migrations (Rosser, et al., 2000). It is suspected that migration restrictions and population expansions following the Last Glacial Maximum (LGM) have resulted in the survival of a limited number of particular European haplogroups (Semino, et al., 2000). In the US, for example, most European Americans belong to haplogroups I and R (Hammer, et al., 2005; Vallone and Butler, 2004). African Americans are descendents of forced migration from certain western and western central African populations (Quintana-Murci, et al., 1999). Recent studies, including the data presented in this paper, indicate that most African Americans and Caucasians in the United States belong to one of only two major sub-haplogroups, E3a (58-62%) and R1b (47-58.3%) respectively (Hammer, et al., 2005; Vallone and Butler, 2004).

Additional markers that provide higher resolution differentiation of sub-populations within the E3a and R1b major haplogroups would be useful in forensic genetics and in evolutionary studies, including admixture analysis. Here, we describe 10 previously phylogenetically undefined Y-SNPs that define four new E3a sub-haplogroups and three new R1b3 sub-haplogroups. Two additional markers are reported that also have the potential to differentiate populations within haplogroup I, but for which further population studies are required.

MATERIALS AND METHODS

Study Subjects

A total of 243 unrelated individuals including 118 African Americans (AA) and 125 European Americans (EA) whose major Y-SNP haplogroups (hgs) were determined with 56 well-defined Y-SNPs using a hierarchical typing strategy with the pyrosequencing technology described below, for genotyping. All DNA samples were obtained with the individual's informed consent in accordance with the University of Central Florida's Institutional Review Board.

Candidate Marker Selection

A recent genome-wide SNP survey (Hinds et al, 2005) genotyped 334 Y-SNPs in 33 chromosomes. Several of these SNPs had been phylogenetically characterized in earlier studies (Underhill 2001, YCC 2002). These characterized SNPs allow the 33 chromosomes examined in the survey to be assigned to YCC haplogroups. For example, eight of the 33 chromosomes can be assigned to haplogroup E3a (Jobling, 2003) on the basis of M180 (rs2032598: T>C). It was noted that a number of the uncharacterized SNPs showed variation among these eight E3a chromosomes, and these SNPs were therefore selected as candidates for additional study. A candidate list was prepared of SNPs that were polymorphic inside E3a, R, or I, and a total of 12 SNPs were chosen from the candidate list for further study.

Genomic DNA Isolation, PCR, and Genotyping

Genomic DNA was extracted from whole blood or buccal swabs using standard organic extraction protocols. PCR and extension primers were designed using a combination of Primer3 (Skaletsky, 2000) and SNP Primer Design Pyrosequencing AB v.1.0.1.software (http://primerdesign.pyrosequencing.com/jsp/TemplateInput.jsp) to specifically amplify regions flanking and including the SNPs. In all instances, female controls were genotyped to ensure candidate markers were confined to the Y-chromosome. In all assays, sequences were detected in only male individuals. The ancestral and derived allelic states were ascertained by genotyping a male chimpanzee for

all candidate markers as well as by typing samples with the candidate markers from individuals belonging to the more ancient haplogroups (e.g. hgA, hgB). The allelic states of the candidate SNP and the corresponding primer information are listed in Table 1. The 50 µL PCR single-plex reaction contained: 0.5 ng DNA, 0.08-0.2 µM each primer (Forward and Reverse), 125 µM dNTPs, 1X PCR Buffer II (10mM Tris-HCl, pH 8.3, 50 mM KCl), 2.0 mM MgCl2, 10µg non-acetylated BSA (Sigma, St. Louis, MO, USA, http://www.sigmaaldrich.com). and 1.5 units of AmpliTaq® Gold Polymerase (Applied Biosystems, Foster City, CA, USA, http://www.appliedbiosystems.com). Cycling conditions were: (1) 95ºC for 10 min, (2) 45 cycles: 95ºC for 15 s, 50ºC for 30s, 72ºC for 15s, and (3) final extension at 72ºC for 5 min. Genotyping was performed by pyrosequencing on a PSQ™ 96 MA instrument according to the manufacturer's recommendations (Biotage, Uppsala, Sweden, http://www.biotage.com).

Phylogenetic and Statistical Analysis

Genotype data were collected and the phylogenetic relationships were depicted in a phylogenetic tree showing the number of individuals and the corresponding frequencies for each haplogroup observed. For each population, the probability of discrimination (DP) (Jones, 1972) was calculated as: $DP=1-\sum p_i^2$, where p_i is the frequency of the derived allele at each of the i haplogroups.

Table 1. Detailed List of Y-SNP Markers, Primers and Corresponding Positions

Marker	Rs #	Forward PCR primer (5'-3')	Reverse PCR primer (5'-3')	Product Size (bp)	SNP position (bp)
U106	rs16981293: C>T	gctctggtgcatagggattc	agtctgaactcttgggagatgg	255	46
U152	rs1236440: G>A	cttagctatacagcctctttttgg	aacattccacgcttgaggataa	172	127
U174	rs16980586: G>A	tccctgcagtgaaatagttttg	ctcagactttaggtgagatttgc	107	46
U175	rs16980588: G>A	ctggtcacactaaggcacca	tctaatgaccaggagaagtcaaga	72	48
U179	rs2319818: G>A	aaggggatatgacgactgatt	cagctcctcttttcaactctca	275	220
U181	rs16980589: C>T	cactgacacatggaactgagtg	gtttaccaggaacccccatc	90	41
U186	rs16980370: A>G	gcctaagcccttctcgtaag	atctgctagatttctccttattgg	106	29
U198	rs17222279: G>A	tcatcattgcatgggtataactg	ttaggttctatggtgatttgaactt	77	24
U209	rs16980502: C>T	cccacaggaatgcaaaagat	cacctgcagcattaaatgga	72	33
U247	rs2068150: C>T	gaggggaggttcttctcaat	tcggaggcaatacgctgtaa	78	45
U250	rs17315723: G>C	ccctcatgaaataacagtttgc	caattcactgaccttttttgcat	99	70
U290	rs16980406: T>A	atcgcctggaaagccacta	tgtgcagacaaaagcgtacc	140	95

Note: The marker consists of the letter U (for Unique Event Polymorphism, UEP) followed by an arbitrary number assigned to the markers in the order in which they were listed in the original data-mining set, and the corresponding SNP (ancestral>derived) is listed in the forward direction. The location of the SNP is in relation to the beginning of the forward primer.

RESULTS

In our population, 60% of African Americans belong to hg E3a (Fig. 1). Of the 12 Y-SNPs investigated, seven (U175, U209, U181, U290, U174, U186, and U247) of them were identified only in African Americans belonging to haplogroup E3a. The U175 and U209 polymorphisms create a new monophyletic clade (equivalent to E3a8 using the Y Chromosome Consortium nomenclature) derived from M2 comprising 22.9% of the African Americans belonging to haplogroup E3a (Fig. 1). U181 and U290 were found only in individuals derived at U175/U209, dividing this new sub-M2 clade into three sub-clades (E3a8*, E3a8a*, and E3a8a1), with frequencies of 7.6, 11.0, and 3.4, respectively. The U186 and U247 polymorphisms were found only in the individuals derived at marker M191; thus these markers are phylogenetically equivalent to M191 in our population. U174 was also found in all (+) M191 individuals except one, which may signify a rare back-mutation which occurred only in this individual. This latter marker creates a new sub-clade from M191, called E3a7a.

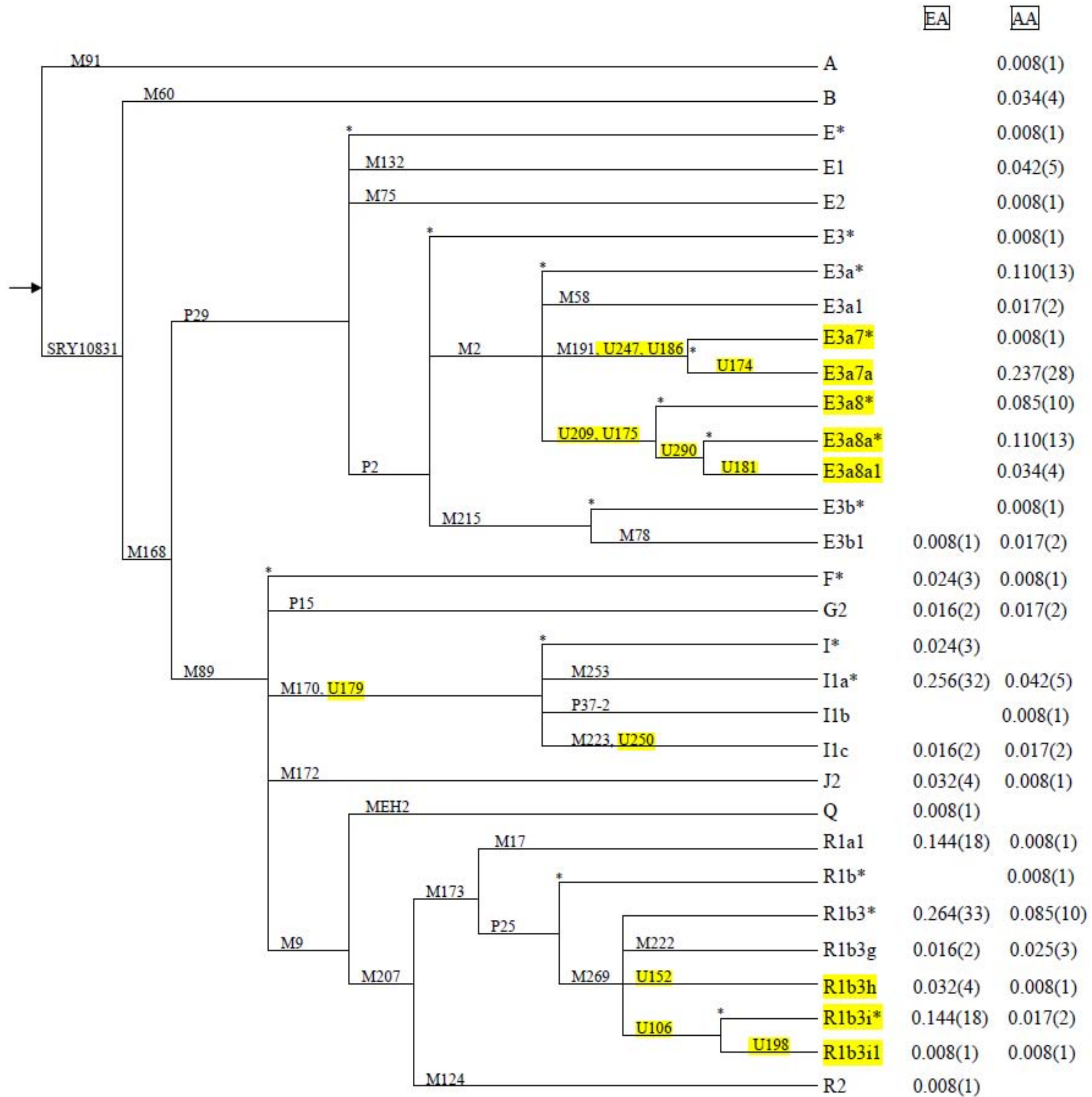

Figure 1. Y-Chromosome Phylogeny. Markers listed on each branch represent the unique event polymorphisms used to investigate our population of 243 individuals. At the ends of the branches are the names of the lineages according to the YCC nomenclature. New markers are highlighted as well as their corresponding proposed new clade, if any. The frequencies of individuals belonging to each haplogroup in the European American (EA) and African American (AA) are listed to the right of the haplogroup names as well as the number of individuals observed, in parentheses.

Forty-six percent of European Americans and 14% African Americans belong to hg R1b3, as defined by M269 (Cruciani, et al., 2002; Moore, et al., 2006). Three of the 12 Y-SNPs (U106, U152, and U198) were found only in individuals belonging to hgR1b3 (Fig. 1). Twenty-six percent (18 EA and 2 AA) of the individuals within hgR1b3 possessed SNP U106 and 6% (4 EA and 1 AA) possessed U152. SNP U198 was only found in 2 (1 EA and 1 AA) individuals who also possessed the polymorphism U106, creating a sub-clade of U106 (Fig. 1). We also included M222, a sub-R1b3 marker seemingly rarely used, in our study and found 5 (2 EA and 3AA) individuals with this polymorphism with frequencies of 1.6% and 2.5% in Europeans Americans and African Americans, respectively.

Additionally, U179 and U250 were found to be phylogenetically equivalent to markers previously described in haplogroup I (Fig. 1).

To ascertain the extent to which the new markers are useful for differentiating individuals within populations, the probability of discrimination (DP) obtained by typing individuals with the 56 well-defined Y-SNP markers was calculated with and without the inclusion of the new markers for each population. The DP was increased from 0.71 to 0.82 (15.5%) for European Americans and from 0.80 to 0.90 (12.5%) for African Americans.

DISCUSSION

Among ancestral populations, haplogroup E3a is restricted to sub-Saharan Africa although it is the major haplogroup in contemporary African Americans, with frequencies of 58-60% (Hammer, et al., 2005; Vallone and Butler, 2004). Y-chromosomal markers representing haplogroup E3a have been used to aid in the study of early western Bantu dispersals (Beleza, et al., 2005; Plaza, et al., 2004). The seven Y-SNPs described here, that divide the African haplogroup E3a into five new haplogroups can provide useful tools for the investigation of human migrations within and out of Africa.

Haplogroup R1b3 increases in frequency from the Middle East to Northwestern Ireland (Moore, et al., 2006) and ranges from 58-62% in European Americans (Hammer, et al., 2005; Vallone and Butler, 2004). All three of the new polymorphisms (U152, U106, and U198) create new haplogroups and could be named, according to the Y Chromosome Consortium nomenclature (YCC, 2002), as R1b3h, R1b3i, and R1b3j1, respectively.

In our population, we found no individuals with the sub-M269 markers typically used (R1b3a-f; M37, M65, M126, M153, M160, SRY2627) so we investigated M222, a marker seemingly rarely used (Sun, et al., 1999). We found 5 individuals (3 AA, and 2 EA) that possess this polymorphism with a frequency of 2% in our total population (Fig. 1). Interestingly, we also discovered these individuals possess Y-STR haplotypes identical or derived from the 17 marker Irish Modal Haplotype (IMH) (Moore, et al., 2006) which is found at high frequencies in NW Ireland (data not shown). To further investigate this latter observation, we searched our locally maintained Y-STR database for samples that possessed the 17-locus IMH and also for those that differed from the IMH by 1 and 2 mutational steps (designated IMH-1 and IMH-2 respectively). A total of 7 individual samples were found to possess the IMH whereas additional samples were one (3 samples) and two (14 samples) mutational steps removed. These 24 IMH-related samples were typed at the M222 locus. Remarkably all 7 of the samples that possessed the IMH and all 3 IMH-1 samples were also positive for the derived M222 G>A substitution. Moreover 4 out of the 14 IMH-2 samples possessed the derived M222 allele. Although it is likely that the majority of individuals more than two steps from the IMH do not belong to the M222 haplogroup we do recognize that, until more comprehensive studies are undertaken, it is possible that there exist M222 derived chromosomes with haplotypes that are more divergent from the IMH. For those interested in improving R1b3 population differentiation in previously reported studies, it may be an option to reanalyze samples using M222 in combination with the three new markers.

Haplogroup I has been shown to account for over 30% of paternal haplogroups in Scandinavian populations and in the northwestern Balkans. Furthermore, sub-haplogroup I1c has been found all over Europe, with the highest frequencies in northwestern Europe (Rootsi, et al., 2004). In this study, we show that marker U179 is phylogenetically equivalent to M170, which defines haplogroup I (Underhill, et al., 2001) and marker U250 is equivalent to M223, which defines haplogroup I1c (Cinnioglu, et al., 2004). Even though these new markers are equivalent to other well-characterized markers in our population, it is possible they could differentiate hg I sub-populations from larger or more diverse populations than the ones we employed.

Previous studies have used a set of common Y-SNP markers which distinguish between major haplogroups (Hammer, et al., 2005; Vallone and Butler, 2004). Many individuals of European and African ancestry belong to a sub-set of the major haplogroups, namely E3a and R1b3, with a significant amount of admixture being irresolvable with the battery of Y chromosomal bi-allelic markers currently available. The 12.5-15.5 % increase in the probability of discrimination with the addition of these now- phylogenetically-defined markers demonstrates their potential value in differentiating populations within haplogroups R1b3 and E3a. The new markers described herein could help differentiate these major populations for use in human history migration investigations and ethno-geographic prediction in forensic genetics.

ACKNOWLEDGMENTS

This work was supported under Award Numbers 1998-IJ-CX-K003 and 2005-MU-MU-K044 from the Office of Justice Programs, National Institute of Justice, Department of Justice. Points of view in this manuscript are those of the authors and do not necessarily represent the official position of the US Department of Justice. The authors would like to acknowledge Mr. Jeremy Fletcher and Ms. Kathleen Mayntz-Press for their assistance in the initial phase of this work.

REFERENCES

Beleza S, Gusmao L, Amorim A, Carracedo A, Salas A. 2005. The genetic legacy of western Bantu migrations. Hum Genet 117(4):366-375.

Brion M, Sanchez JJ, Balogh K, Thacker C, Blanco-Verea A, Borsting C, Stradmann-Bellinghausen B, Bogus M, Syndercombe-Court D, Schneider PM, Carracedo A, Morling N. 2005. Introduction of an single nucleodite polymorphism-based "Major Y-chromosome haplogroup typing kit" suitable for predicting the geographical origin of male lineages. Electrophoresis 26(23):4411-4420.

Brookes AJ. 1999. The essence of SNPs. Gene 234(2):177-186.

Cinnioglu C, King R, Kivisild T, Kalfoglu E, Atasoy S, Cavalleri GL, Lillie AS, Roseman CC, Lin AA, Prince K, Oefner PJ, Shen P, Semino O, Cavalli-Sforza LL, Underhill PA. 2004. Excavating Y-chromosome haplotype strata in Anatolia. Hum Genet 114(2):127-148.

Cruciani F, Santolamazza P, Shen P, Macaulay V, Moral P, Olckers A, Modiano D, Holmes S, Destro-Bisol G, Coia V, Wallace DC, Oefner PJ, Torroni A, Cavalli-Sforza LL, Scozzari R, Underhill PA. 2002. A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. Am J Hum Genet 70(5):1197-1214.

Hammer MF, Chamberlain VF, Kearney VF, Stover D, Zhang G, Karafet T, Walsh B, Redd AJ. 2005. Population structure of Y chromosome SNP haplogroups in the United States and forensic implications for constructing Y chromosome STR databases. Forensic Sci Int. (in press)

Jobling MA. 2001. Y-chromosomal SNP haplotype diversity in forensic analysis. Forensic Sci Int 118(2-3):158-162.

Jobling MA, Tyler-Smith C. 2003. The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4(8):598-612.

Jones DA. 1972. Blood samples: probability of discrimination. J Forensic Sci Soc 12(2):355-359.

Kidd KK, Pakstis AJ, Speed WC, Grigorenko EL, Kajuna SL, Karoma NJ, Kungulilo S, Kim JJ, Lu RB, Odunsi A, Okonofua F, Parnas J, Schulz LO, Zhukova OV, Kidd JR. 2005. Developing a SNP panel for forensic identification of individuals. Forensic Sci Int. (in press)

Moore LT, McEvoy B, Cape E, Simms K, Bradley DG. 2006. A y-chromosome signature of hegemony in gaelic ireland. Am J Hum Genet 78(2):334-338.

Plaza S, Salas A, Calafell F, Corte-Real F, Bertranpetit J, Carracedo A, Comas D. 2004. Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola. Hum Genet 115(5):439-447.

Quintana-Murci L, Semino O, Bandelt HJ, Passarino G, McElreavey K, Santachiara-Benerecetti AS. 1999. Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. Nat Genet 23(4):437-441.

Rootsi S, Magri C, Kivisild T, Benuzzi G, Help H, Bermisheva M, Kutuev I, Barac L, Pericic M, Balanovsky O, Pshenichnov A, Dion D, Grobei M, Zhivotovsky LA, Battaglia V, Achilli A, Al-Zahery N, Parik J, King R, Cinnioglu C, Khusnutdinova E, Rudan P, Balanovska E, Scheffrahn W, Simonescu M, Brehm A, Goncalves R, Rosa A, Moisan JP, Chaventre A, Ferak V, Furedi S, Oefner PJ, Shen P, Beckman L, Mikerezi I, Terzic R, Primorac D, Cambon-Thomsen A, Krumina A, Torroni A, Underhill PA, Santachiara-Benerecetti AS, Villems R, Semino O. 2004. Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in europe. Am J Hum Genet 75(1):128-137.

Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, Amos W, Armenteros M, Arroyo E, Barbujani G, Beckman G, Beckman L, Bertranpetit J, Bosch E, Bradley DG, Brede G, Cooper G, Corte-Real HB, de Knijff P, Decorte R, Dubrova YE, Evgrafov O, Gilissen A, Glisic S, Golge M, Hill EW, Jeziorowska A, Kalaydjieva L, Kayser M, Kivisild T,

Kravchenko SA, Krumina A, Kucinskas V, Lavinha J, Livshits LA, Malaspina P, Maria S, McElreavey K, Meitinger TA, Mikelsaar AV, Mitchell RJ, Nafa K, Nicholson J, Norby S, Pandya A, Parik J, Patsalis PC, Pereira L, Peterlin B, Pielberg G, Prata MJ, Previdere C, Roewer L, Rootsi S, Rubinsztein DC, Saillard J, Santos FR, Stefanescu G, Sykes BC, Tolun A, Villems R, Tyler-Smith C, Jobling MA. 2000. Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. Am J Hum Genet 67(6):1526-1543.

Sanchez JJ, Borsting C, Hallenberg C, Buchard A, Hernandez A, Morling N. 2003. Multiplex PCR and minisequencing of SNPs--a model with 35 Y chromosome SNPs. Forensic Sci Int 137(1):74-84.

Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA. 2000. The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. Science 290(5494):1155-1159.

Shastry BS. 2002. SNP alleles in human disease and evolution. J Hum Genet 47(11):561-566.

Skaletsky SRaHJ. 2000. Primer3 on the WWW for general users and for biologist programmers. Bioinformatics Methods and Protocols: Methods in Molecular Biology:365-386.

Sobrino B, Carracedo A. 2005. SNP typing in forensic genetics: a review. Methods Mol Biol 297:107-126.

Sun C, Skaletsky H, Birren B, Devon K, Tang Z, Silber S, Oates R, Page DC. 1999. An azoospermic man with a de novo point mutation in the Y-chromosomal gene USP9Y. Nat Genet 23(4):429-432.

Syvanen AC. 2005. Toward genome-wide SNP genotyping. Nat Genet 37 Suppl:S5-10.

Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, Oefner PJ, Cavalli-Sforza LL. 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet 65(Pt 1):43-62.

Vallone PM, Butler JM. 2004. Y-SNP typing of U.S. African American and Caucasian samples using allele-specific hybridization and primer extension. J Forensic Sci 49(4):723-732.

YCC. 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Res 12(2):339-348.