Int J Legal Med (2007) 121:234–237
DOI 10.1007/s00414-007-0153-y

SHORT COMMUNICATION

Y-chromosome genetic structure in sub-Apennine populations of Central Italy by SNP and STR analysis

Valerio Onofri · Federica Alessandrini · Chiara Turchi · Barbara Fraternale · Loredana Buscemi · Mauro Pesaresi · Adriano Tagliabracci

Received: 18 August 2006 / Accepted: 22 December 2006 / Published online: 8 February 2007

Abstract To define the Y-chromosome genetic structure in Apennine populations, 17 Y-chromosome short tandem repeats (Y-STRs) and 37 Y-single nucleotide polymorphisms (Y-SNPs) were typed in 162 subjects living in the upland area of the Marches (Central Italy). A total number of 155 haplotypes (haplotype diversity was 0.9994) and 14 SNP haplogroups were observed. Testing high-resolution Y-chromosome data sets, e.g. using Yfiler and SNPs, increases the discriminatory capacity in individual identification for forensic purposes. It is also useful in autochthonous population and micro-population studies to highlight the most informative loci for evolutionary aims.

Keywords Y-chromosome · Short tandem repeats · Single nucleotide polymorphisms · Population data · Italy

Introduction

The Y chromosome plays a crucial role in forensic investigations, especially in resolving the male component in a mixture from crime samples. In such cases, knowledge of marker frequencies and their distributions across different populations is a determining factor.

Many data on Y-chromosome short tandem repeats (Y-STRs) data were collected during the last few years, but precious information may also come either from both data from Y-single nucleotide polymorphisms (Y-SNPs) haplogroup distribution or microsatellite variability inside each SNP clade [1]. Like STRs, Y-chromosome SNPs are useful markers in human identification (e.g. mass disasters), paternity tests, interpretation and of mixed male–female samples. However, their non-random distribution among populations makes them a unique tool for evolutionary and population studies. This study therefore defines the genetic structure of the Y chromosome in a sub-Apennine sample of the Marches (Central Italy) using Y microsatellites and SNPs.

The 17-microsatellite panel used here offers very high discriminatory power. In addition, our six Y-SNP multiplex approach allowed very high discrimination between haplogroups and within sub-clades, yielding a more clearly defined picture of binary polymorphism structure in populations.

Materials and methods

Y-chromosome markers were examined in 162 healthy male donors, unrelated and with different surnames, selected from 19 upland areas of the Marches region (Fig. S1), Fabriano (n=44), Urbino (n=40), Ascoli Piceno (n=38), Matelica (n=9), Camerino (n=4), Roccafluvione (n=4), Acquasanta Terme (n=4), Rotella (n=4), Spinetoli (n=3), Pergola (n=2), Castignano (n=2), Amandola (n=1), San Severino (n=1), Urbisaglia (n=1), Montefelcino (n=1), Castelleone di Suasa (n=1), Apiro (n=1), Patrignone (n=1) and Arquata del Tronto (n=1).

Electronic supplementary material The online version of this article (doi:10.1007/s00414-007-0153-y) contains supplementary material, which is available to authorized users.

V. Onofri · F. Alessandrini · C. Turchi · B. Fraternale · L. Buscemi · M. Pesaresi · A. Tagliabracci (✉)
Institute of Legal Medicine, Università Politecnica delle Marche, Via Conca-Torrette,
60020 Ancona, Italy
e-mail: a.tagliabracci@univpm.it

Table 1 Relative frequency and haplotype diversity (HD) within the observed haplogroups of this study; for each haplogroup is also indicated the terminal derived marker of the typed SNP set

Haplogroup	Terminal marker	*n*	Relative frequency	HD
E3b*(xE3b1,E3b2,E-M224)	M35	14	0.0864	0.9890±0.0314
P*(xR1)	M45	1	0.0062	1.0000±0.0000
E3b1*	M78	16	0.0988	1.0000±0.0221
F*(xK,J2,I,H,G)	M89	22	0.1358	0.9947±0.0178
G	M201	12	0.0741	1.0000±0.0340
I*(xI1a2,I1a3,I1b2,I1c)	M170	9	0.0556	1.0000±0.0524
I1b2*	M26	1	0.0062	1.0000±0.0000
I1c	M223	2	0.0123	1.0000±0.5000
J2*(xJ2a1a, J2a1b*, J2a1c, J2a1d, J2a1e, J2b*)	M172	12	0.0741	1.0000±0.0340
J2a1b*	M67	8	0.0494	0.9643±0.0772
K*(xP,N,O)	M9	6	0.0370	1.0000±0.0962
R1b3	M269	57	0.3519	0.9969±0.0042
R1a1	M17	1	0.0062	1.0000±0.0000
R1b*(xR1b,R1b3)	P25	1	0.0062	1.0000±0.0000

n: number of individuals observed for each haplogroup.

All samples were genotyped for 17 Y-chromosome short tandem repeats using the AmpFlSTR® Yfiler™ PCR Amplification Kit (Applied Biosystems, Forest City, CA, USA), which allows co-amplification of the core set of the European Minimal Haplotype (minHt) and eight other loci (DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, GATA H4.1).

A total of 37 Y-SNPs in 6 homemade multiplex polymerase chain reactions (PCR) were typed for all 162 samples to obtain a highly discriminative picture of haplogroups spanning these micro-populations.

DNA was extracted from peripheral blood by the phenol–chloroform protocol [2] or from buccal swabs by DNA IQ™ System (Promega, Madison, WI, USA) [3].

Y-STRs were amplified according to the manufacturer's instructions (Applied Biosystems) modifying the final volume to 6.25 μl.

Y-STR alleles were assigned with the internal allele standard provided in the Applied Biosystems kit. Random samples were re-typed for confirmation by the PowerPlex Y System (Promega). Our laboratory participates in the International Forensic Y-User Group quality control program (http://www.yhrd.org).

Amplification of 37 Y-SNPs in 6 multiplex PCRs and subsequent single base extension reactions were carried out as previously described [4, 5]. Typing was accomplished by ABI PRISM 310 Genetic Analyzer (AB Applied Biosystems). STR and SNP allele calling was performed with GeneScan v3.7 or GeneMapper v3.2 (AB Applied Biosystems). According to the International Society of Forensic Genetics (ISFG) recommendations [6], the name of GATA H4 locus was replaced by GATA H4.1. Also, as the locus allele nomenclature of the Yfiler kit was based on the variable core repeat TAGA and the locus structure consists of a core repeat region now designated as AGAT plus the non-variable repeats $(AGAT)_4CTAT(AGAT)_2(AGGT)_3$ [7], we added nine repeats to the Yfiler allele number.

Statistical analyses were performed by Arlequin software ver. 2.000 [8]; DYS385 was excluded from the calculation of haplotype diversity and analysis of molecular variance (AMOVA) because of the impossibility of assigning each allele to one or the other locus of this ambiguous system.

Results and discussion

Y-STRs

In 162 subjects, 155 haplotypes were observed, 149 of which were unique (Supplementary Table S1). At the

Table 2 Comparison between haplotype diversity (HD) values (of Yfiler kit and within regrouped Y-SNP haplogroups) observed in this study and in an Austrian male sample

	This study		Berger et al. [1]	
	Number of haplotypes	HD	Number of haplotypes	HD
Yfiler	155	0.9994±0.0008	130	0.9994±0.0010
E	29	0.9977±0.0094	18	1.0000±0.0185
F*(xK)	64	0.9991±0.0029	51	0.9985±0.0040
K*(xR1a1)	61	0.9976±0.0034	43	0.9980±0.0052
R1a1	1	1.0000±0.0000	18	0.9942±0.0193

Table 3 Haplogroup frequencies of the Marches population compared to other heterogeneous Italian samplings

Haplogroup	Frequency		
	This study (n=162)	Semino et al. [12] (n=50)	Di Giacomo et al. [13] (n=524)
R1b	0.36	0.62	0.36
DE	0.19	0.02	0.16
J2	0.12	0.14	0.21
I	0.07	0.08	0.07
G	0.07	0.10	0.06
R1a	0.01	0.04	0.03
Other	0.18	–	0.11

DYS458 locus, one duplication (Ht129) and four micro-variants were observed (Ht1, Ht50, Ht92, Ht93, Ht94, Ht99, Ht109, Ht110, Ht129). One duplication was also observed at the DYS389II locus (Ht99). Samples containing duplications or micro-variants were re-amplified and sequence characterization was performed. The discrimination capacity, determined by dividing the number of observed haplotypes by the number of individuals, was 0.9568 (minHt=0.8086 only).

Haplotype diversity calculation resulted in a high level of differentiation (0.9994±0.0008), a value similar to other Yfiler data sets [1, 9], but also slightly lower than a single only 11 Y-STR loci assay used for a large East Asian population sample [10]. Analysis of molecular variance (AMOVA) was performed between the three largest samples: Fabriano (n=44), Urbino (n=40) and Ascoli Piceno (n=38). The variation among the three samples was 0.33% and 99.67% among individuals within the group with a statistically non-significant percentage of variation among the three areas suggesting background homogeneity.

Ninety-one haplotypes, inclusive of the minHt and the other loci, were submitted to the YHRD (release 18); remaining data will be transferred.

In this study, we observed the most frequent minimal haplotype without DYS385 (14–13–29–24–11–13–13) in 12 sub-Apennine males (7.4% of the sample), all of them R1b3. Consistently, this haplotype recurred in about 5.5% of the European meta-population sample included in YHRD (24,497 haplotypes, http://www.yhrd.org) and was indicated as the ancestral Y-STR haplotype of the P(xR1a) haplogroup [11].

The relative frequencies of Yfiler loci are shown in Supplementary Table S2; for DYS385 see Supplementary Table S3. This multi-locus marker obviously showed the highest gene diversity (0.9202±0.0001). However, we should also note that in our population, DYS458 showed the second highest value of gene diversity (0.8277±0.0123), i.e. far higher than in the Northeast Italian (0.725) [9] and Austrian samples (0.7878±0.0134) [1]. This is probably due to the 5.6% of allele micro-variants of this locus in the Marches sample (all belonging to the F*(xK,J2,I,H,G) haplogroup) compared with a lower value (2.2%) observed in Austrians (F*(xK)) and 0% in Northeast Italians (no SNPs typed).

Y-SNPs

A total number of 14 Y-SNP haplogroups was observed with a haplogroup diversity of 0.8224±0.0002 and a discrimination power of 7.29%. As previously observed in European populations [12], the most frequent haplogroup in the Marches population was R1b3, which recurs in 35.19% of the sample.

To define which Y-SNP clade showed the lowest microsatellite variability, haplotype diversity within each haplogroup was estimated (Table 1). A low Y-STR haplotypes variability was observed only within J2a1b* (0.9643±0.0772), whereas the other haplogroups tended to the highest variability value. Our values are similar to those observed in Austrians, after grouping the 14 haplogroups at a lower level of resolution (see Table 2).

Haplogroups frequencies found in our sample were compared with those published for other Italian populations (north–central Italy [12], and central–south Italy predominant [13], Table 3). This analysis showed that the Marches sample lies close to the central–south Italian population (F_{ST}=0.00975).

We also observed four alleles of different STRs recurring in all subjects of a particular haplogroup (Supplementary Table S2, notes). These were allele 11 of DYS392 and allele 12 of DYS393 in J2 males, allele 10 of DYS438 in I males and allele 14 of DYS437 in E3b males. The last allele was also typical in E3b Austrian males.

Consistent with the calculation of discrimination capacity before and after adding SNP-typing to STR-typing (both 0.9568), biallele markers cannot increase the discrimination between male subjects. In our case, in fact, none of the individuals sharing a haplotype could be resolved with SNPs. The main usefulness of Y-SNPs in forensic practice lies in the possibility of using obtained data for inferring information about the geographic origin of male subjects.

Acknowledgements Special thanks go to Dr. M. Piani and to all volunteers. This study was supported by a grant from the Ministero per l'Istruzione, l'Università e la Ricerca (MIUR grant no. 2005064329_002).

References

1. Berger B, Lindinger A, Niederstatter H, Grubwieser P, Parson W (2005) Y-STR typing of an Austrian population sample using a 17-loci multiplex PCR assay. Int J Legal Med 119:241–246

2. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
3. DNA IQ™ System—Small sample casework protocol, Promega Company, http://www.promega.com/profiles/
4. Onofri V, Alessandrini F, Turchi C, Pesaresi M, Buscemi L, Tagliabracci A (2006) Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs. Forensic Sci Int 157:23–35
5. Alessandrini F, Turchi C, Onofri V, Buscemi L, Pesaresi M, Tagliabracci A (2005) Multiplex PCR development of Y-chromosomal biallelic polymorphisms for forensic application. J Forensic Sci 50(3):519–525
6. Gusmao L, Butler JM, Carracedo A et al (2006) DNA Commission of the International Society for Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. Int J Legal Med 120:191–200
7. Mulero JJ, Budowle B, Butler JM, Gusmao L (2006) Letter to the editor—nomenclature and allele repeat structure update for the Y-STR locus GATA H4. J Forensic Sci 51(3):694
8. Schneider S, Roessli D, Excoffier L (2000) Arlequin ver. 2.000: a software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland
9. Turrina S, Atzei R, De Leo D (2006) Y-chromosomal STR haplotypes in Northeast Italian population sample using 17plex loci PCR assay. Int J Legal Med 120:56–59
10. Kwak KD, Jin HJ, Shin DJ et al (2005) Y-chromosomal STR haplotypes and their applications to forensic and population studies in east Asia. Int J Legal Med 119:195–201
11. Roewer L, Croucher PJP, Willuweit S et al (2005) Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. Int J Legal Med 116:279–291
12. Semino O, Passarino G, Oefner PJ et al (2000) The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. Science 290:1155–1159
13. Di Giacomo F, Luca F, Anagnou N et al (2003) Clinal patterns of human Y chromosomal diversity in continental Italy and Greece are dominated by drift and founder effects. Mol Phylogenet Evol 28:387–395