Ethnic fragmentation and degree of urbanization strongly affect the discrimination power of Y-STR haplotypes in central Sahel.

Chiara Della Rocca [a], Francesco Cannone [b], Eugenia D'Atanasio [c], Maria Bonito [a], Paolo Anagnostou [d,e], Gianluca Russo [f], Filippo Barni [b], Eugenio Alladio [b], Giovanni Destro-Bisol [d,e], Beniamino Trombetta [a], Andrea Berti [b], Fulvio Cruciani [a,c*]

[a] *Dipartimento di Biologia e Biotecnologie "Charles Darwin", Sapienza Università di Roma, P.le Aldo Moro 5, 00185 - Rome, Italy.*
[b] *Reparto Carabinieri Investigazioni Scientifiche di Roma – Sezione di Biologia, Viale Tor di Quinto 119, 00191 - Rome, Italy.*
[c] *Istituto di Biologia e Patologia Molecolari, CNR, Rome, Italy*
[d] *Dipartimento di Biologia Ambientale, Sapienza Università di Roma, P.le Aldo Moro 5, 00185 - Rome, Italy.*
[e] *Istituto Italiano di Antropologia, Rome, Italy*
[f] *Dipartimento di Sanità Pubblica e Malattie Infettive, Sapienza Università di Roma, P.le Aldo Moro 5, 00185 - Rome, Italy.*

[*]correspondence: fulvio.cruciani@uniroma1.it

ABSTRACT

Y chromosome short tandem repeats (Y-STRs) are commonly used to identify male lineages for investigative and judicial purposes and could represent the only source of male-specific genetic information from unbalanced female-male mixtures. The Yfiler Plus multiplex, which includes twenty conventional and seven rapidly-mutating Y-STRs, represents the most discriminating patrilineal system commercially available to date. Over the past five years, this multiplex has been used to analyze several Eurasian populations, with a reported discrimination capacity (DC) approaching or corresponding to the highest possible value. However, despite the inclusion of rapidly mutating Y-STRs, extensive haplotype sharing was still reported for some African populations due to a number of different factors affecting the effective population size.

In the present study, we analyzed 27 Y-STRs included in the Yfiler Plus multiplex and 82 Y-SNPs in central Sahel (northern Cameroon and western Chad), an African region characterized by a strong ethnic fragmentation and linguistic diversity. We evaluated the effects of population sub-structuring on genetic diversity by stratifying a sample composed of 431 males according to their ethnicity (44 different ethnic groups) and urbanization degree (four villages and four towns).

Overall, we observed a low discrimination capacity (DC = 0.90), with 71 subjects (16.5%) sharing 27 Y-STR haplotypes. Haplotype sharing was essentially limited to subjects with the same binary haplogroup, coming from the same location and belonging to the same ethnic group. Haplotype sharing was much higher in rural areas (average DC = 0.83) than urban settlements (average DC = 0.96) with a significant correlation between DC and census size ($r = 0.89$; $p = 0.003$). Notably, we found that genetic differentiation between villages from the same country ($\Phi_{ST} = 0.14$) largely exceeded that found among countries ($\Phi_{ST} = 0.02$). These findings have important implications for the choice of the appropriate reference population database to evaluate the statistical relevance of forensic Y-haplotype matches.

Keywords: Y-STRs, Y-SNPs, Haplotype, Yfiler Plus, Africa, Urbanization

1. Introduction

Y chromosome short tandem repeats (Y-STRs) are commonly used in family testing when the putative father is not available and could represent the only source of available genetic information from mixtures with an excess of female DNA such as those recovered from sexual assaults. In addition, Y-STRs (mainly those in single copy), due to their straightforwardness of interpretation, are often useful in the estimation of the number of male individuals contributing to a mixture of DNA in a stain, thus assisting in the weight of evidence evaluation [Gusmao et al. 2006; Kayser 2017; Andersen and Balding 2019].

For male individualization purposes, a major problem with haplotypes generated from conventional Y-STRs is their inability to differentiate between subjects belonging to the same paternal descent line, due to their relatively low mutation rate (around 2×10^{-3} per locus per generation) [Gusmao et al. 2005]. To overcome this problem, through a Y chromosome-wide scan for new informative microsatellites, a panel of thirteen Y-STRs characterized by mutation rates higher than 10^{-2} per locus per generation was developed [Ballantyne et al. 2010]. Haplotypes generated from this set of rapidly mutating Y-STRs (RM Y-STRs) allowed the individualization of more than 99% of about 12,000 unrelated males [Ballantyne et al. 2014], as well as about half of brother pairs [Ballantyne et al. 2012; Adnan et al. 2016].

Recently, seven RM Y-STRs (five single copy and the two copy system DYF387S1) have been included in a 27 Y-STR multiplex termed Yfiler™ Plus PCR amplification kit (ThermoFisher Scientific), which represents to date the most discriminating Y-STRs multiplex commercially available for capillary electrophoresis [Gopinath et al. 2016]. In the last five years, this multiplex has been used to analyze several populations worldwide. The reported discrimination capacity (DC) was close to (or corresponded to) the highest possible value (DC = 1) in most populations of Eurasian origin analyzed so far [Oloffson et al. 2015; Ottaviani et al. 2015; Pickrahn et al. 2016; Rapone et al. 2016; Wang et al. 2016; He et al. 2019; Lang et al. 2019]; however, despite the contribution of the RM Y-STRs, extensive haplotype sharing was still reported for autochthonous

populations from Africa [Oloffson et al. 2015; Iacovacci et al. 2017; D'Atanasio et al. 2019], Australia [Henry et al. 2019] and the Middle East [Khubrani et al. 2018].

Reasons for the observed lower discrimination power of Yfiler Plus in Africans and/or tribal populations have been attributed to a number of different factors which may reduce the effective population size, such us endogamy, sudden expansion of successful male lineages and patrilocality [Oloffson et al. 2015; Iacovacci et al. 2017; Khubrani et al. 2018; D'Atanasio et al. 2019; Henry et al. 2019].

In this study, we report on the analysis of 82 Y-SNPs and 27 Yfiler Plus Y-STRs in populations from central Sahel (northern Cameroon and western Chad), a region characterized by a strong ethnic fragmentation and linguistic diversity and interested by intense emigration flows towards European countries. We evaluated the effects of population sub-structuring on genetic diversity by stratifying a sample composed of 431 males according to their ethnicity (44 different ethnic groups) and urbanization degree (four villages and four urban settlements).

Overall, we observed a relatively low discrimination capacity, especially in rural areas, as a likely consequence of relative isolation among groups from different locations and belonging to distinct ethnic groups. Notably, we found that inter-population genetic distances between geographically close villages largely exceeded that found among countries (Cameroon and Chad), an evidence which has important implications for the choice of the appropriate reference population database to evaluate the statistical relevance of forensic Y-haplotype matches.

2. Materials and Methods

2.1 DNA samples

A total of 431 males were sampled in eight different locations (four rural villages and four urban areas) from northern Cameroon and western Chad (Table 1, Supplementary Table 1 and Fig. 1). Despite efforts to avoid the inclusion of relatives during the sampling process, the presence of related males in the sample could not be excluded due to the unavailability of genealogical information (see also [Coia et al. 2009]). For each subject, the ethnic identity was assessed by self-identification. In total, 44 different ethnic groups are represented in the sample. Some groups were represented by few subjects ($N < 10$), while ten ethnic groups with $N \geq 10$ accounted for 80% of the sample (Supplementary Table 1). Most of the subjects have been previously studied [Cruciani et al. 2010; Trombetta et al. 2011, 2015; D'Atanasio et al. 2018] for some of the SNPs analyzed in the present study. Appropriate informed consent was obtained from all participants. The research project ethically complies with the ISFG guidelines for the publication of genetic population data [Gusmao et al. 2017] and was formally approved by the "Reparto Carabinieri Investigazioni Scientifiche di Roma" and by the "Sapienza, Università di Roma" Ethical committee (Document number 2755/15).

Table 1: Sample composition

Location	*Country*	*Latitude*	*Longitude*	*Census Size (x 10^3)*[a]	*N*	*Ethnic Diversity*[b]
Towns						
N'Djamena	Chad	12.138	15.054	721.1	93	0.93
Mora	Cameroon	11.046	14.140	55.2	13	0.92
Maroua	Cameroon	10.593	14.321	319.9	28	0.84
Garoua	Cameroon	9.324	13.394	436.9	108	0.77
Villages						
Mémé	Cameroon	10.971	14.233	6.5	73	0.08
Cashiga	Cameroon	9.433	13.367	2.0	40	0.69
Kossoumou	Cameroon	9.271	13.295	0.9	23	0.17
Banaye	Cameroon	9.459	13.506	0.3	53	0.04
					431	0.89

[a] Census sizes is from the The GeoNames geographical database (https://www.geonames.org/) (for towns) and [BUCREP 2005] (for villages)

[b] Ethnic diversity is defined as the probability that two randomly chosen subjects from the same location belong to different ethnic groups

Figure 1: Sampling location in Central Sahel. Towns are indicated in red; villages are indicated in blue.

2.2. DNA extraction and quantification

Genomic DNA was extracted either from blood using a standard phenol-chloroform protocol or from saliva using the BioRobot EZ1® Advanced XL Workstation (Qiagen). DNA quantification was performed using Quantifiler® Trio DNA Quantification Kit (ThermoFisher Scientific) and/or Qubit™ 4 Fluorometer (ThermoFisher Scientific).

2.3. Single nucleotide polymorphism genotyping

Overall, 82 SNPs of the MSY were analyzed in order to assign 431 Y chromosomes to specific binary haplogroups/paragroups. SNPs analyzed are reported in Fig. 2. SNPs were genotyped according to a phylogenetic hierarchical approach based on the minimal reference MSY tree [van Oven et al. 2014], using PCR amplified products and subsequent heteroduplex DHPLC analysis, RFLP analysis or Sanger sequencing. The level of resolution of the resulting phylogeny corresponds to that of the minimal reference phylogeny [van Oven et al. 2014] adopted as a forensic standard by the YHRD database (https://yhrd.org/pages/resources/ysnps).

2.4. Y-STRs multiplex genotyping

Multiplex amplification of 27 Y-STRs was performed on an Applied Biosystem® Veriti™ 96-Well Thermal Cycler (ThermoFisher Scientific) by using the Yfiler™ Plus PCR Amplification Kit (ThermoFisher Scientific) according to the manufacturer's protocol utilizing 1 ng of genomic DNA. Amplified DNAs were then electrophoresed on the 24-capillary Applied Biosystems® 3500xL Genetic Analyzer (ThermoFisher Scientific) and the fragment analysis was performed with GeneMapper® ID-X software v.1.4 (ThermoFisher Scientific). Haplotype data were submitted to the Y-chromosomal haplotype reference database (www.yhrd.org) (YHRD accession numbers pending). The contributors successfully passed the quality control test.

2.5. Y-STR microvariant sequence analysis

Six loci that were found to carry micro-variant alleles (DYF387S1, DYS385, DYS449, DYS458, DYS518, DYS570) were PCR-amplified and sequenced. PCR primers for loci DYF387S1, DYS385, DYS449 and DYS458 were previously described [Iacovacci et al. 2017 or D'Atanasio et. al 2019]. For both multilocus markers DYF387S1 and DYS385, which consist of two paralogous sequences, two pairs of different paralog-specific primers were used as described in [Iacovacci et al. 2017 and D'Atanasio et al. 2019], which made it possible to amplify and sequence the two copy of each locus separately. PCR primers for the loci DYS518 (FOR: 5'-TGCCAACGAAGAAACTACCC and REV: 5'- CAGCCTGGACAACAGACTGA) and DYS570 (FOR: 5'-GAGGAGATTAGGAGCACAGTGA and REV: 5'-TCCTGCACATCTTGGGACTT) were designed on the basis of the MSY sequence reported in the UCSC Genome Browser web site (February 2009 assembly of the human genome; http://genome.ucsc.edu/) using Primer3 software v. 4.1.0 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) [Untergasser et al. 2012] to obtain a PCR product of 926 bp and 406 bp, respectively. Sequencing templates were obtained through PCR in a 50 µl reaction containing 20 ng of genomic DNA, 200 mM of each dNTP, 2.5 mM MgCl2, 1 unit of Taq polymerase, and 10 pmoles of each primer. A touchdown PCR program was used with an annealing temperature decreasing from 63 to 56 °C over 14 cycles, followed by 30 cycles with an annealing temperature of 56 °C. After DNA amplification, the PCR products were purified using the MultiScreen FB plates (MERCK).

Cycle sequencing was performed using the BigDye Terminator Cycle Sequencing Kit with AmpliTaq® DNA Polymerase (ThermoFisher Scientific) and an internal or PCR primer. Cycle sequencing products were purified by ethanol precipitation and run on a 3730xl DNA Analyzer (ThermoFisher Scientific). Chromatograms were aligned and analyzed for microvariant characterization using Sequencher v. 4.8 (Gene Codes Corporation, Ann Arbor, MI, USA).

2.6 Forensic and population genetic parameter estimation

The genetic diversity (GD) for each locus was calculated according to Nei and Tajima [1981].

Haplotype frequencies were determined using the counting method. The haplotype diversity (HD) was calculated as $HD = n(1-\Sigma p_i^2)/(n-1)$, where n is the sample size and p_i the frequency of the i^{th} haplotype. The haplotype match probability (*HMP*) was estimated using the formula $HMP = \Sigma p_i^2$, whereas the discrimination capacity (*DC*) was calculated as the ratio between the number of different haplotypes and the total number of chromosomes in the dataset. For each location analyzed, an "ethnic diversity" parameter (*ED*) was also calculated as $ED = 1-\Sigma e_i^2$, where e_i is the proportion of subjects belonging to the e_i ethnic group in the location.

2.7 Analysis of molecular variance

Analysis of molecular variance (AMOVA, [Excoffier et al. 1992]) was performed using Arlequin, v.3.5.2.2 [Excoffier et al. 2010]. Different grouping schemes were used taking into account: (1) degree of urbanization (individuals grouped into 8 sampling locations classified as villages or towns); (2) ethnic affiliation (only 10 major groups with $N \geq 10$ were considered); (3) country of provenance (2 groups: Cameroon and Chad). The analyses were performed either considering or not the molecular distances between Y-STR haplotypes (Φ-statistics and F-statistics, respectively). As a measure of molecular distance in the Φ-statistics analysis, we used the number of different Y-STR loci between haplotypes, i.e. a number ranging from 0 (identical haplotypes) to 25 (i.e. haplotypes differing at all loci) considering both DYS385 and DYF387S1 multi-copy systems as a single locus. Significance levels of both Φ- and F-statistics were obtained by comparing the actual values with the distribution of 10,000 values obtained by randomization. For all the statistical analyses, the alleles of the DYS389II locus were converted to the DYS389B nomenclature by subtracting the repeat number of the DYS389I locus from that of the DYS389II locus.

2.8 Dendrogram generation

As a simple way to visualize similarities among Y-STR haplotypes, a dendrogram based on

UPGMA cluster analysis was generated through the PAST software v. 3.07, using the total number of differences in repeat number as a distance between haplotype pairs (Manhattan distance). It should be noted that, because of recurrent mutations at microsatellite loci, the dendrogram generated cannot be regarded as a true phylogenetic tree, especially for the deepest branches.

3. Results and discussion

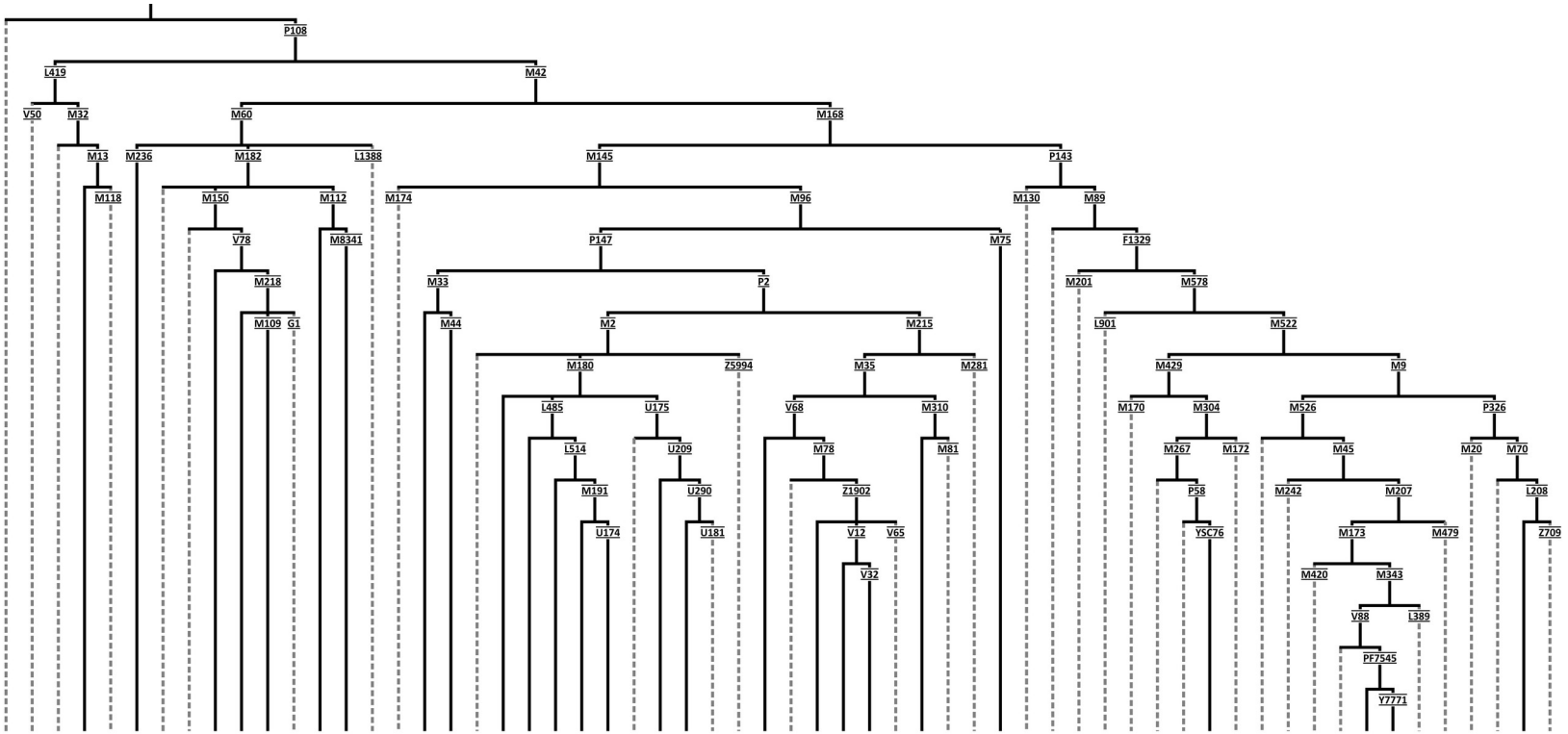
3.1 SNP-defined haplogroup distribution in Central Sahel

We used a hierarchical Y-SNP genotyping approach to assign each chromosome to terminal branches of the minimal reference Y-phylogeny [van Oven et al. 2014] adopted as a standard for the forensic community by the YHRD [Willuweit and Roewer 2015] (Y Chromosome Haplotype Reference Database, https://yhrd.org/pages/resources/ysnps).

Overall, we observed 26 different haplogroups or paragroups, whose frequencies are shown in Fig. 2 (for ethnic groups) and in supplementary Fig. 1 (for sampling locations). The most common haplogroup, R-V88 [Cruciani et al. 2010], accounted for about one third (32.5%) of the Y chromosomes analyzed and was represented (with a single exception) by the African-specific branch R-Y7771. Consistently with some previous studies [Cruciani et al. 2010, Cruciani et al. 2011; Kulichova et al. 2017], the highest frequencies of this haplogroup were observed among Afroasiatic-speaking populations (in particular Daba, Mandara, and Ouldeme from Cameroon).

Eight different branches within the common sub-Saharan E-M2 haplogroup were observed at varying frequencies accounting on the whole for another third of the sample (35.0%) (Fig. 2 and supplementary Fig. 1). The area under study is very close to the supposed cradle of the Bantu civilization, in the southern part of the Cameroon/Nigeria boundary [Newman 1995]. About four to three millennia ago, Bantu-speaking populations rapidly propagated agriculture and iron working from their homeland in central Africa towards eastern and southern Africa [Newman 1995], leading to the concomitant spread at high frequencies of E-M2 chromosomes in those areas [Cruciani et al. 2002; Montano et al. 2011]. The relatively low frequency of the E-M2 haplogroup in central Sahel suggests that the area under study has only been marginally touched by this expansion, despite its proximity to the supposed Bantu homeland. Consistently, the relative proportions of E-M2 sub-lineages in central Sahel are quite different from those reported for Bantu-speaking populations and could have been involved in other human movements such as those fostered by the Sahara desertification [D'Atanasio et al. 2018; Xue 2018]. In particular, one of the observed E-M2

lineages, E-L485*(xL514), was only observed in the Fulbe population (16.9%). The Fulbe population also differentiated from other central Sahelian groups for the high frequencies (36.9%) of the otherwise rare E-M44 haplogroup (Fig. 2). Both these haplogroups could represent a signature of the known recent historical westward migration of the Fulbe pastoralists into central Sahel [Newman 1995].



	Y*	A2-V50	A2-M32*	A2-M13*	A2-M118	B1-M236	B2-M182*	B2-M150*	B2a-V78*	B2a-M218*	B2a-M109	B2a-G1	B2b-M112*	B2b-M8341	B3-L1388	D-M174	E1a-M33*	E1a-M44	E1b1a-M2*	E1b1a-M180*	E1b1a-L485*	E1b1a-L514*	E1b1a-M191*	E1b1a-U174	E1b1a-U175*	E1b1a-U209*	E1b1a-U290*	E1b1a-U181	E1b1a-Z5994	E1b1b-V68*	E1b1b-M78*	E1b1b-Z1902*	E1b1b-V12*	E1b1b-V32	E1b1b-V65	E1b1b-M310*	E1b1b-M81	E1b1b-M281	E2-M75	C-M130	F-M89*	G-M201	H-L901	I-M170	J1-M267*	J1-P58*	J1-YSC76	J2-M172	K-M526	Q-M242	R1a-M420	R1b-V88*	R1b-PF7545*	R1b-Y7771	R1b-L389	R2-M479	L-M20	T-M70*	T-L208*	T-Z709	N
Daba	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	2	3	-	-	-	-	-	-	-	-	-	-	-	4	-	-	-	-	-	-	-	-	-	-	-	-	-	15	-	-	-	-	-	-	**26**
Fali	-	-	-	-	-	-	-	-	-	-	28	-	-	-	-	-	-	-	-	-	14	-	19	-	2	7	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	22	-	-	-	-	-	-	**93**
Fulbe	-	-	-	6	-	-	-	-	-	-	-	-	-	-	-	-	24	-	-	11	3	-	3	-	1	2	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	9	-	-	-	-	4	-	**65**
Mandara	-	-	-	10	-	-	-	-	-	-	3	-	-	-	-	-	-	-	2	1	2	-	3	-	-	1	-	-	-	-	-	1	-	-	1	-	-	16	-	-	-	-	-	-	-	-	-	-	-	-	-	33	-	-	-	-	1	-	**74**
Moundang	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	2	-	-	1	-	-	-	-	-	-	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	5	-	-	-	-	-	-	**12**
Ngambai	-	-	-	1	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	4	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	**10**
Ouldeme	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	20	-	-	-	-	-	-	**21**
Sara Madjingay	-	-	-	1	-	1	-	-	-	-	-	-	1	-	-	-	-	-	-	1	-	3	-	4	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	**13**
Tali	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5	-	-	1	-	4	-	4	-	4	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	**21**
Toupouri	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	7	-	-	-	-	-	-	**14**
Other populations	-	-	-	4	-	1	-	-	1	1	3	-	-	-	-	-	-	-	13	-	8	-	11	-	5	4	-	-	-	-	1	-	5	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	24	-	-	-	-	-	-	**82**
Total	**-**	**-**	**-**	**23**	**-**	**2**	**-**	**-**	**1**	**1**	**39**	**-**	**1**	**1**	**-**	**-**	**5**	**24**	**-**	**18**	**12**	**38**	**2**	**46**	**-**	**15**	**20**	**-**	**-**	**1**	**-**	**1**	**1**	**6**	**-**	**1**	**-**	**-**	**26**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**1**	**-**	**-**	**-**	**-**	**-**	**1**	**140**	**-**	**-**	**-**	**-**	**5**	**-**	**431**

Figure 2: MSY reduced consensus tree [van Oven et al. 2014] and haplogroup distribution in central Sahel. Lineages not observed in our dataset are represented by dashed lines.

3.2 Molecular characterization of intermediate-size alleles

Overall, we observed 54 interalleles at six different Y-STRs (43 interalleles at DYF387S1, 4 at DYS385, 3 at DYS570, two at DYS449, and one at DYS458 and DYS518), four of which were rapidly mutating loci (Table 2).

Sequence analysis was conducted to characterize the molecular features of the interalleles detected (Table 2). Chromosomes with DYF387S1.2 alleles were found to carry a two base indel 5' to the $(AAAG)_n$ repeat, on the DYF387S1a locus located on the left arm of the P1 palindrome. Consistently with previous studies [Iacovacci et al. 2017; D'Atanasio et al. 2019], all the 43 chromosomes with the DYF387S1.2 allele belonged to haplogroup B-M182 and derivatives (Fig. 2 and Supplementary Fig. 2), suggesting a common monophyletic ancestry. We did not observe additional signs of gene conversion between the two copies of the DYF387S1 locus [Iacovacci et al. 2017] since all the B2 chromosomes carried one interallele and one conventional allele, suggesting that this form of recombination could be relatively uncommon in this portion of the P1 palindrome.

The DYS385 locus has been previously reported to be prone to recurrent mutations generating intermediate-size alleles occurring on several different haplogroup backgrounds, and most of the interalleles so far sequenced were in the form of the DYS385.2 variant [Myres et al. 2009; D'Atanasio et al. 2019]. Conversely, the four interalleles observed in the present study were DYS385.1 variants and were due to at least two independent events as suggested by their molecular features (Table 2) and haplogroup affiliation (E1b1a-L514* and R1b-V88). The three subjects belonging to haplogroup E1b1a carried a microsatellite-like variation at a short trinucleotide repeat within the DYS385 amplicon (CAA_3 to CAA_2), while the R1b subject carried a 3 base deletion in the variable (GAAA) repeat. In a previous study [Berniell-Lee et al. 2009], about 45% (21 out of 46) of the R1b chromosomes from Gabon (immediately south to the Cameroon) were found to carry a DYS385.2 interallele. Interestingly, we did not observe this intermediate-sized allele in 140 R1b chromosomes from central Sahel, suggesting that the previously described DYS385.2 allele could

have marked a local southern R-V88 expansion probably not associated with the Chadic-speaking population expansion.

Among the Yfiler Plus loci, DYS449 has been shown to be prone to interallele generating mutations, since interalleles sequenced so far have been assigned to at least ten different binary haplogroup backgrounds [Myres et al 2009; Iacovacci et al. 2017; D'Atanasio et al. 2019]. The two DYS449 interalleles here characterized also occurred on two additional binary haplogroups (Table 2), contributing to the sequence diversity of this complex locus.

Table 2. Sequence structure of microvariant alleles at six Y-STR loci

Y-STR	Sample	Genotype	Haplogroup	Structure	N[a]
DYS458	L1-79	17.2	J1-YSC76	(GAAA)15 **AA** (GAAA)2 GGAGGG	1
DYS458	Reference	19		(GAAA)19 GGAGGG	
DYS518	M6-15	38.1	E1b1a-U290	[AAAG]3 [GAAG]1 [AAAG]15 [GGAG]1 [AAAG]4 N6 [AAAG]10 **A** [AAAG]4	1
DYS518	Reference	38		[AAAG]3 [GAAG]1 [AAAG]16 [GGAG]1 [AAAG]4 N6 [AAAG]13	
DYS570	M4-54, M4-56, M4-71	20.2	E1b1a-U209*	(TTTC)14-**TT**-(TTTC)6	3
DYS570	Reference	17		(TTTC)17	
DYS385b	L1-51	11.1	E1b1a-L514*	**(CAA)2** N21 (AAGG)4 N14(AAAG)3 N12 (AAAG)3 N29 (AAGG)6(GAAA)12	1
DYS385b	M4-62	14.1	E1b1a-L514*	**(CAA)2** N21 (AAGG)4 N14(AAAG)3 N12 (AAAG)3 N29 (AAGG)6(GAAA)15	1
DYS385b	L1-47	14.1	E1b1a-L514*	**(CAA)2** N21 (AAGG)4 N14(AAAG)3 N12 (AAAG)3 N29 (AAGG)6(GAAA)15	1
DYS385a	M5-45	14.1	R1b-Y7771	(CAA)3 N21 (AAGG)4 N14(AAAG)3 N12 (AAAG)3 N29 (AAGG)6 (GAAA)2 **A** (GAAA)12	1
DYS385a	Reference	15		(CAA)3 N21 (AAGG)4 N14(AAAG)3 N12 (AAAG)3 N29 (AAGG)6(GAAA)15	
DYS385b	Reference	16		(CAA)3 N21 (AAGG)4 N14(AAAG)3 N12 (AAAG)3 N29 (AAGG)6(GAAA)16	
DYS449	L1-65	28.2	E1b1a-L514*	(TTTC)1 **TT** (TTTC)12 N14 (TTTC)2 CTTC (TTTC)2 N16(TTTC)15	1
DYS449	M6-16	33.3	B2b-M8341	(TTTC)9 **TTC** (TTTC)8 N14 (TTTC)2 CTTC (TTTC)2 N16 (TTTC)15	1
DYS449	Reference	30		(TTTC)16 N14 (TTTC)2 CTTC (TTTC)2 N16 (TTTC)15	
DYF387S1a	L1-63	41.2	B2b-M112*	(AAAG)3 (GTAG)1 (GAAG)4 N20 (GAAG)10 **AA** (AAAG)18	1
DYF387S1a	M6-14	39.2	B2a-M109	(AAAG)3 (GTAG)1 (GAAG)4 N20 (GAAG)9 **AA** (AAAG)17	1
DYF387S1a	L1-31	40.2	B2a-M109	(AAAG)3 (GTAG)1 (GAAG)4 N20 (GAAG)10 **AA** (AAAG)17	4
DYF387S1a	L1-15	41.2	B2a-M109	(AAAG)3 (GTAG)1 (GAAG)4 N20 (GAAG)9 **AA** (AAAG)19	26
DYF387S1a	L1-54	42.2	B2a-M109	(AAAG)3 (GTAG)1 (GAAG)4 N20 (GAAG)9 **AA** (AAAG)20	6
DYF387S1a	M5-6	43.2	B2a-M109	(AAAG)3 (GTAG)1 (GAAG)4 N20 (GAAG)9 **AA** (AAAG)21	1
DYF387S1a	M5-9	44.2	B2a-M109	(AAAG)3 (GTAG)1 (GAAG)4 N20 (GAAG)11 **AA** (AAAG)20	1
DYF387S1a	M3-11	45.2	B2a-M218*	(AAAG)3 (GTAG)1 (GAAG)4 N20 (GAAG)12 **AA** (AAAG)20	1
DYF387S1a	L1-77	44.2	B2a-V78*	(AAAG)3 (GTAG)1 (GAAG)4 N20 (GAAG)12 **AA** (AAAG)19	1
DYF387S1a	M6-16	41.2	B2b-M8341	(AAAG)3 (GTAG)1 (GAAG)4 N20 (GAAG)10 **AA** (AAAG)18	1
DYF387S1a	Reference	35		(AAAG)3 (GTAG)1 (GAAG)4 N20 (GAAG)9 (AAAG)13	

[a]Number of samples sharing the intermediate allele. For the DYF387S1a locus only 10 out of 43 microvariant alleles were sequenced

3.3 Multiallelic patterns and null alleles

We observed 30 extra-alleles at two loci (24 diallelic patterns at DYS448 and 6 triallelic patterns at DYF389S1) and 4 null alleles at the locus DYS488.

The diallelic pattern at the DYS448 showed little variation and, in the present study, was always associated with Fulbe subjects carrying the E1a-M44 haplogroup (Fig. 2, Supplementary Fig. 2, Supplementary Table 2 and 3), while five phylogenetically related chromosomes belonging to E1a-M33* paragroup (Fig. 2) showed a conventional DYS448 single allele. The DYS448 duplication (without haplogroup information) has been recently reported to be quite common in Portuguese subjects of African descent (19 out of 400 males) [Dente et al. 2019], and has also been observed in two western African Mandenka (Haplogroup E1a-M33) [Balaresque et al. 2008] and two E1a-M33*(xM44) northern Africans [D'Atanasio et al. 2019]. Taken together, these data suggest that a single DYS448 duplication occurred in a E-M33 chromosome before the mutation that generated the E1a-M44 haplogroup and could now represent a relatively common pattern in African subjects.

Triallelic patterns at the usually diallelic locus DYF387S1 have been observed in six subjects belonging to three different haplogroups (Supplementary Table 2). Triallelic and tetrallelic pattern at the DYF387S1 locus are relatively common and have been previously reported (YHRD, https://yhrd.org) [Willuweit and Roewer 2015]. Three of the subjects described in this study (M4-85, M4-98 and M4-100) share the same DYF387S1 alleles (36,38,39) and the same R1b-Y7771 haplogroup, and are clearly related based on their complete haplotype profile (Supplementary Table 3 and Supplementary Fig. 2). The unbalanced RFU values for these subjects (Supplementary Table 2) also suggest that the apparent triallelic pattern is indeed a masked quadriallelic pattern, i.e. a duplication of the entire diallelic DYF387S1 locus. Interestingly, these three chromosomes also carry a concomitant null allele at DYS448, also found in a fourth phylogenetically related R1b-Y7771 subject (subject M4-106, Supplementary Table 3 and Supplementary Fig. 2), which carry a diallelic DYF387S1 pattern characterized by unusually high RFUs as compared to other loci (data

not shown), compatible with a full DYF387S1 duplication. Several cases of null alleles at the DYS448 locus have been previously reported [Budowle et al. 2008; Balaresque et al. 2008; Westen et al. 2015; Shonhai et al. 2020] which can result from point mutations at primer binding sites or locus deletions. In our case, the absence of a detectable DYS448 PCR product using an alternative external primer pair (see materials and methods) suggests a locus deletion. Both DYS448 and DYF387S1 are located in a region of the Y chromosome characterized by a high rate of instability due to non-allelic recombination between duplicated portions. We suggest that a single recombination event could have generated at the same time the DYF387S1 duplication and the DYS448 deletion, although further molecular studies are requested to confirm this hypothesis.

3.4 Intra-population haplotype diversity

The complete list of 431 Y-STR haplotypes and corresponding Y-SNP haplogroups is reported in Supplementary Table 3 and the resulting forensic parameters are reported in Table 3.

In total, 387 haplotypes have been observed, among which 360 were unique and 27 were shared by 71 subjects (Table 3), with two haplotypes shared by as much as 6 and 8 males, respectively (Supplementary Table 4). Thus, despite the inclusion in the Yfiler Plus multiplex of seven RM Y-STRs, the diversity in the global sample remains relatively low, with a DC = 0.898. This figure is similar to that previously reported for two other broad African regions, eastern Africa (DC = 0.887) [Iacovacci et al. 2017] and northern Africa (DC = 0.908) [D'Atanasio et al. 2019], suggesting that DC values could be generally low across the African continent.

Haplotype sharing was always limited to subjects carrying the same binary haplogroup and coming from the same country (supplementary Table 3 and Supplementary Fig. 2). Also, similarly to previous studies concerning African populations [Iacovacci et al. 2017; D'Atanasio et al. 2019], males sharing haplotypes, with a single exception, were from the same ethnic group. Finally, haplotype sharing was limited to males from the same sampling location (with only two exceptions, both concerning geographically close sampling areas), likely as a consequence of patrilocality.

Haplotype sharing was not observed for individuals who at the same time came from different localities and belonged to different ethnicities. Overall, these results suggest that haplotype sharing for the 27 Y-STRs is not incidental, but likely due to some degree to paternal relatedness. Although the sampling approach adopted for the subjects under study lead to the exclusion of known related individuals, the lack of precise genealogical information during the sampling phase could not prevent the inclusion of hidden related individuals. Further studies, based on autosomal STRs, could help to assess the proportion of related individuals and their degree of relatedness, among those found to share the same Y-STR haplotype.

In any case, despite the relatively low DC observed, the additional Y-STRs in the Yfiler Plus multiplex resulted in a significant increase in the discrimination capacity compared to the 17 conventional YSTRs included in the Yfiler multiplex (43 haplotypes shared by 122 subjects, DC = 0.817) (Supplementary Tables 4-6 and Figure 3).

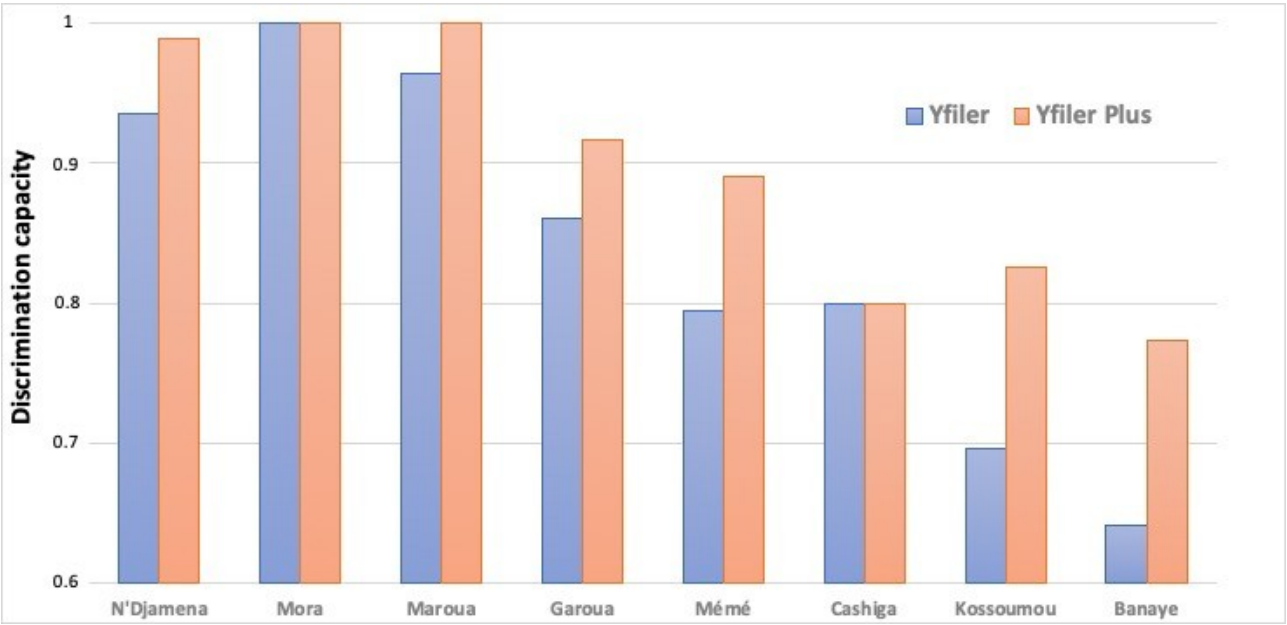


Figure 3: Comparison of the discrimination capacity for the Yfiler and the Yfiler Plus haplotypes

3.5 Comparison between urban settlements and villages

In order to evaluate the possible effect of the degree of urbanization on the discrimination power of the Y-STRs here analyzed, we subdivided our global sample in eight sampling locations classified as towns (census size > 50,000 inhabitants) and villages (census size < 10,000 inhabitants) (Table 1).

The DC was much lower in villages (average DC = 0.831, range 0.774 - 0.890) where 27.5% of the males carried a shared haplotype, than in towns (average DC = 0.959, range 0.916 - 1.0) where haplotype sharing was limited to 7.9% of the subjects (Table 3 and Fig. 3). This fact was also reflected in a significant positive correlation between DC and census size (DC *vs.* Log census size, r = 0.890, p = 0.003) for the eight locations analyzed (Fig. 4). Reduced haplotype diversity and DC have also been previously observed for rural villages in the Tyrolean Alps (Austria) [Niederstatter et al 2016]. These findings could be the consequence of reduced male-specific gene flow (patrilocality), higher level of endogamy and lower effective population size in rural areas with respect to large urban settlements

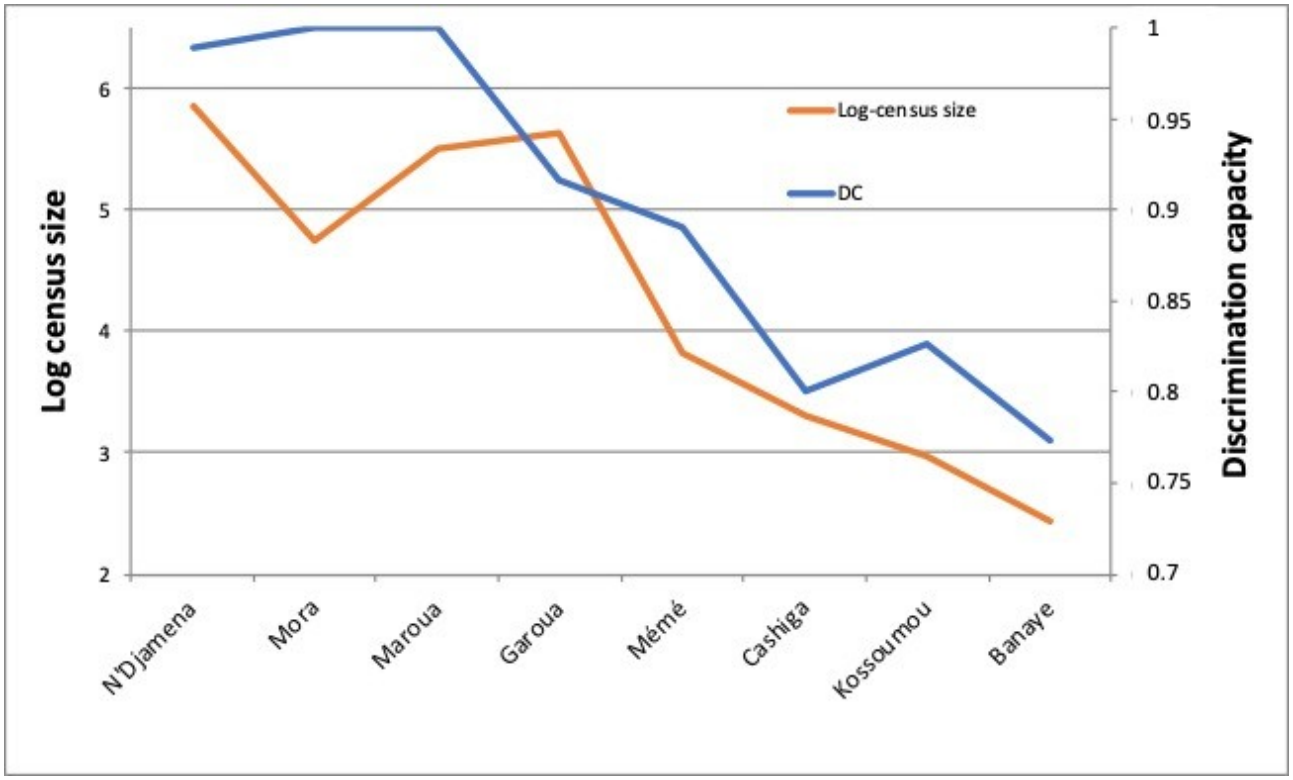


Figure 4: Correlation between discrimination capacity and census size for 8 sampling locations.

Table 3. Forensic parameters for 8 locations from Cameroon and Chad analyzed in this study by using Yfiler Plus PCR Amplification kit

	Country	*Sample size*	*Number of haplotypes*	*Haplotype match probability*	*Discrimination capacity*	*Mean number of pairwise differences*	*Average gene diversity over loci*	Haplotype diversity	Matching haplotypes	# subjects matching haplotypes	% subjects matching haplotypes
Towns											
N'Djamena	Chad	93	92	0.011	0.989	17.154	0.686	0.999	1	2	2.2
Mora	Cameroon	13	13	0.077	1	16.295	0.652	1	0	0	0
Maroua	Cameroon	28	28	0.036	1	17.979	0.719	1	0	0	0
Garoua	Cameroon	108	99	0.011	0.916	17.595	0.704	0.998	8	17	15.7
All towns[a]		242	232	0.034	0.959	17.256	0.690	0.999	9	19	7.9
Villages											
Mémé	Cameroon	73	65	0.018	0.890	16.796	0.672	0.996	6	14	19.2
Cashiga	Cameroon	40	32	0.048	0.800	17.547	0.702	0.977	4	12	30
Kossoumou	Cameroon	23	19	0.062	0.826	9.794	0.392	0.980	3	7	30.4
Banaye	Cameroon	53	41	0.037	0.774	14.537	0.581	0.981	5	17	32.1
All villages[a]		189	157	0.041	0.831	14.669	0.587	0.984	19	52	27.5
Total		431	387	0.003	0.898	17.641	0.706	0.999	27	71	16.5

[a] Forensic indexes reported as arithmetic means across four locations

3.6 Pattern of inter-population diversity

To evaluate the degree of differentiation between Sahelian populations, we performed an AMOVA analysis at different level of classifications, either considering or not molecular differences between haplotypes (Table 4). We observed a relatively low level (although significant) of molecular differentiation between countries (Cameroon and Chad, $\Phi_{ST} = 0.016$) and between four urban settlements ($\Phi_{ST} = 0.016$). Conversely, we found a marked level of differentiation between ten major ethnic groups ($\Phi_{ST} = 0.120$) but also between the four villages ($\Phi_{ST} = 0.144$). The same pattern was observed when molecular distances between haplotypes were not taken into account (conventional F_{ST} values, Table 4). In particular, the average level of differentiation among three villages near the town of Garoua was extremely high ($\Phi_{ST} = 0.209$), even if the three locations were only a few dozens of kilometers apart (Fig. 1). These features are at least in part consequence of the fact that villages are characterized by a much lower ethnic diversity, being in some cases inhabited by a single ethnic group (Supplementary Table 1). However, the fact that some ethnic groups have been sampled in multiple locations (Supplementary table 1), and that haplotype sharing remains substantially limited to males from the same location, suggests that both patrilocality and ethnicity contribute to the observed pattern of intra- and inter-population diversity.

3.7 Concluding remarks

Our findings of reduced intra-population diversity and high inter-population differentiation in central Sahel have some consequences for forensic applications.

The level of differentiation between villages is so high to make non-trivial (if not impossible) to define an adequate reference population database to be used for statistical evaluation of Y-haplotype matches for crimes committed in rural areas, especially where ethnic groups are fragmented or subject to other factors (e.g. endogamy due to social structure) leading to reduced gene flow among sub-populations. The identification of new RM Y-STRs and their inclusion in validated multiplex seem necessary to reach an acceptable discrimination power among males in

this area as well as other regions characterized by genetic isolation and ethnic fragmentation. As a positive counterpart, given the strong population structuring, a haplotype match could provide relevant investigative leads about both the ethnicity and the provenance of an unknown person of interest.

Table 4. Analysis of molecular variance for Y-STR haplotypes in central Sahel

	Number of sub-groups	*Number of groups*	*ϕ-statistics*[a]			*F-statistics*[a]		
			ϕ_{ST}	ϕ_{CT}	ϕ_{SC}	F_{ST}	F_{CT}	F_{SC}
Town vs. Village	*8*	2	0.07006**	- 0.00622	0.07581**	0.00664**	- 0.00073	0.00736**
Town level	*4*	1	0.01564**			0.00077*		
Village level	*4*	1	0.14396**			0.01494**		
Country level	*2*	1	0.01617**			0.00085**		
Ethnic level[b]	*10*	1	0.12007**			0.00823**		

[a]P-values obtained as the fraction of cases in which a ϕ-value greater than the quoted value is obtained in a permutation test of samples across populations (ϕ_{ST}/F_{ST} and ϕ_{CT}/F_{CT}) and populations across groups (ϕ_{CT}/F_{CT}). Significant p-values are indicated with ** ($p < 10^{-4}$) or * ($p = 0.031$).
[b]Only ethnic groups with $N \geq 10$ were considered

References

A. Adnan, A. Ralf, A. Rakha, N. Kousouri, M. Kayser, Improving empirical evidence on differentiating closely related men with RM Y-STRs: A comprehensive pedigree study from Pakistan, Forensic Sci. Int. Genet. 25 (2016) 45–51. https://doi.org/https://doi.org/10.1016/j.fsigen.2016.07.005.

M.M. Andersen, D.J. Balding, Y-profile evidence: Close paternal relatives and mixtures, Forensic Sci. Int. Genet. 38 (2019) 48–53. https://doi.org/https://doi.org/10.1016/j.fsigen.2018.10.004.

P. Balaresque, G.R. Bowden, E.J. Parkin, G.A. Omran, E. Heyer, L. Quintana-Murci, L. Roewer, M. Stoneking, I. Nasidze, D.R. Carvalho-Silva, C. Tyler-Smith, P. de Knijff, M.A. Jobling, Dynamic nature of the proximal AZFc region of the human Y chromosome: multiple independent deletion and duplication events revealed by microsatellite analysis, Hum. Mutat. 29 (2008) 1171–1180. https://doi.org/10.1002/humu.20757.

K.N. Ballantyne, M. Goedbloed, R. Fang, O. Schaap, O. Lao, A. Wollstein, Y. Choi, K. van Duijn, M. Vermeulen, S. Brauer, R. Decorte, M. Poetsch, N. von Wurmb-Schwark, P. de Knijff, D. Labuda, H. Vézina, H. Knoblauch, R. Lessig, L. Roewer, R. Ploski, T. Dobosz, L. Henke, J. Henke, M.R. Furtado, M. Kayser, Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications, Am. J. Hum. Genet. 87 (2010) 341–353. https://doi.org/10.1016/j.ajhg.2010.08.006.

K.N. Ballantyne, V. Keerl, A. Wollstein, Y. Choi, S.B. Zuniga, A. Ralf, M. Vermeulen, P. de Knijff, M. Kayser, A new future of forensic Y-chromosome analysis: Rapidly mutating Y-STRs for differentiating male relatives and paternal lineages, Forensic Sci. Int. Genet. 6 (2012) 208–218. https://doi.org/10.1016/j.fsigen.2011.04.017

K.N. Ballantyne, A. Ralf, R. Aboukhalid, N.M. Achakzai, M.J. Anjos, Q. Ayub, J. Balažic, J. Ballantyne, D.J. Ballard, B. Berger, C. Bobillo, M. Bouabdellah, H. Burri, T. Capal, S. Caratti, J. Cárdenas, F. Cartault, E.F. Carvalho, M. Carvalho, B. Cheng, M.D. Coble, D. Comas, D. Corach, M.E. D'Amato, S. Davison, P. de Knijff, M.C.A. De Ungria, R. Decorte, T. Dobosz, B.M. Dupuy, S. Elmrghni, M. Gliwiński, S.C. Gomes, L. Grol, C. Haas, E. Hanson, J. Henke, L. Henke, F. Herrera-Rodríguez, C.R. Hill, G. Holmlund, K. Honda, U.-D. Immel, S. Inokuchi, M.A. Jobling, M. Kaddura, J.S. Kim, S.H. Kim, W. Kim, T.E. King, E. Klausriegler, D. Kling, L. Kovačević, L. Kovatsi, P. Krajewski, S. Kravchenko, M.H.D. Larmuseau, E.Y. Lee, R. Lessig, L.A. Livshits, D. Marjanović, M. Minarik, N. Mizuno, H. Moreira, N. Morling, M. Mukherjee, P. Munier, J. Nagaraju, F. Neuhuber, S. Nie, P. Nilasitsataporn, T. Nishi, H.H. Oh, J. Olofsson, V. Onofri, J.U. Palo, H. Pamjav, W. Parson, M. Petlach, C. Phillips, R. Ploski, S.P.R. Prasad, D. Primorac, G.A. Purnomo, J. Purps, H. Rangel-Villalobos, K. Rębała, B. Rerkamnuaychoke, D.R. Gonzalez, C. Robino, L. Roewer, A. Rosa, A. Sajantila, A. Sala, J.M. Salvador, P. Sanz, C. Schmitt, A.K. Sharma, D.A. Silva, K.-J. Shin, T. Sijen, M. Sirker, D. Siváková, V. Skaro, C. Solano-Matamoros, L. Souto, V. Stenzl, H. Sudoyo, D. Syndercombe-Court, A. Tagliabracci, D. Taylor, A. Tillmar, I.S. Tsybovsky, C. Tyler-Smith, K.J. van der Gaag, D. Vanek, A. Völgyi, D. Ward, P. Willemse, E.P.H. Yap, R.Y.Y. Yong, I.Z. Pajnič, M. Kayser, Toward male individualization with rapidly mutating y-chromosomal short tandem repeats, Hum. Mutat. 35 (2014) 1021–1032. https://doi.org/10.1002/humu.22599.

G. Berniell-Lee, F. Calafell, E. Bosch, E. Heyer, L. Sica, P. Mouguiama-Daouda, L.

van der Veen, J.-M. Hombert, L. Quintana-Murci, D. Comas, Genetic and Demographic Implications of the Bantu Expansion: Insights from Human Paternal Lineages, Mol. Biol. Evol. 26 (2009) 1581–1589. https://doi.org/10.1093/molbev/msp069.

BUCREP (Bureau central des recensements et des etudes de population), Répertoire actualisé des villages du Cameroun. Troisième Recensement Général de la Population et de l'Habitat du Cameroun. Volume IV - Tome 07 (2005).

B. Budowle, X.G. Aranda, R.E. Lagace, L.K. Hennessy, J.V. Planz, M. Rodriguez, A.J. Eisenberg, Null allele sequence structure at the DYS448 locus and implications for profile interpretation, Int. J. Legal Med. 122 (2008) 421–427. https://doi.org/10.1007/s00414-008-0258-y.

V. Coia, F. Brisighelli, F. Donati, V. Pascali, I. Boschi, D. Luiselli, C. Battaggia, C. Batini, L. Taglioli, F. Cruciani, G. Paoli, C. Capelli, G. Spedini, G. Destro-Bisol, A multi-perspective view of genetic variation in Cameroon, Am. J. Phys. Anthropol. 140 (2009) 454-64. https://doi.org/10.1002/ajpa.21088

F. Cruciani, P. Santolamazza, P. Shen, V. Macaulay, P. Moral, A. Olckers, D. Modiano, S. Holmes, G. Destro-Bisol, V. Coia, D.C. Wallace, P.J. Oefner, A. Torroni, L.L. Cavalli-Sforza, R. Scozzari, P.A. Underhill, 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 (2002) 1197–1214. https://doi.org/https://doi.org/10.1086/340257.

F. Cruciani, B. Trombetta, D. Sellitto, A. Massaia, G. Destro-Bisol, E. Watson, E. Beraud Colomb, J.M. Dugoujon, P. Moral, R. Scozzari, Human Y chromosome haplogroup R-V88: a paternal genetic record of early mid Holocene trans-Saharan connections and the spread of Chadic languages, Eur. J. Hum. Genet. 18 (2010) 800–807. https://doi.org/10.1038/ejhg.2009.231.

F. Cruciani, B. Trombetta, A. Massaia, G. Destro-Bisol, D. Sellitto, R. Scozzari, A revised root for the human Y chromosomal phylogenetic tree: the origin of patrilineal diversity in Africa, Am. J. Hum. Genet. 88 (2011) 814–818. https://doi.org/10.1016/j.ajhg.2011.05.002.

E. D'Atanasio, B. Trombetta, M. Bonito, A. Finocchio, G. Di Vito, M. Seghizzi, R. Romano, G. Russo, G.M. Paganotti, E. Watson, A. Coppa, P. Anagnostou, J.M. Dugoujon, P. Moral, D. Sellitto, A. Novelletto, F. Cruciani, The peopling of the last Green Sahara revealed by high-coverage resequencing of trans-Saharan patrilineages, Genome Biol. 19 (2018) 20. https://doi.org/10.1186/s13059-018-1393-5.

E. D'Atanasio, G. Iacovacci, R. Pistillo, M. Bonito, J.M. Dugoujon, P. Moral, F. El-Chennawi, M. Melhaoui, A. Baali, M. Cherkaoui, D. Sellitto, B. Trombetta, A. Berti, F. Cruciani, Rapidly mutating Y-STRs in rapidly expanding populations: Discrimination power of the Yfiler Plus multiplex in northern Africa, Forensic Sci. Int. Genet. 38 (2019) 185–194. https://doi.org/https://doi.org/10.1016/j.fsigen.2018.11.002.

Â. Dente, A. Amorima, H. Afonso Costa, M.J Porto, E. Cunha, F. Corte Real, C. Vieira, C. Silva, Study of Y chromosome markers with forensic relevance in Lisbon immigrants from African countries – Allelic variants study, Forensic Sci. Int. Genet. Supp. Series 7 (2019) 906-907. https://doi.org/10.1016/j.fsigss.2019.11.016.

L. Excoffier, P.E. Smouse, J.M. Quattro, Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data, Genetics. 131 (1992) 479–491. https://pubmed.ncbi.nlm.nih.gov/1644282.

L. Excoffier, H.E.L. LischerI, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, Mol. Ecol. Resour. 10 (2010) 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x.

S. Gopinath, C. Zhong, V. Nguyen, J. Ge, R.E. Lagacé, M.L. Short, J.J. Mulero, Developmental validation of the Yfiler® Plus PCR Amplification Kit: An enhanced Y-STR multiplex for casework and database applications, Forensic Sci. Int. Genet. 24 (2016) 164–175. https://doi.org/https://doi.org/10.1016/j.fsigen.2016.07.006.

L. Gusmão, P. Sánchez-Diz, F. Calafell, P. Martín, C.A. Alonso, F. Álvarez-Fernández, C. Alves, L. Borjas-Fajardo, W.R. Bozzo, M.L. Bravo, J.J. Builes, J. Capilla, M. Carvalho, C. Castillo, C.I. Catanesi, D. Corach, A.M. Di Lonardo, R. Espinheira, E. Fagundes de Carvalho, M.J. Farfán, H.P. Figueiredo, I. Gomes, M.M. Lojo, M. Marino, M.F. Pinheiro, M.L. Pontes, V. Prieto, E. Ramos-Luis, J.A. Riancho, A.C. Souza Góes, O.A. Santapa, D.R. Sumita, G. Vallejo, L. Vidal Rioja, M.C. Vide, C.I. Vieira da Silva, M.R. Whittle, W. Zabala, M.T. Zarrabeitia, A. Alonso, A. Carracedo, A. Amorim, Mutation rates at Y chromosome specific microsatellites, Hum. Mutat. 26 (2005) 520–528. https://doi.org/10.1002/humu.20254.

L. Gusmão, J.M. Butler, A. Carracedo, P. Gill, M. Kayser, W.R. Mayr, N. Morling, M. Prinz, L. Roewer, C. Tyler-Smith, P.M. Schneider, DNA Commission of the International Society of Forensic Genetics (ISFG): An update of the recommendations on the use of YSTRs in forensic analysis, Forensic Sci. Int. 157 (2006) 187–197. https://doi.org/https://doi.org/10.1016/j.forsciint.2005.04.002.

L. Gusmão, J.M. Butler, A. Linacre, W. Parson, L. Roewer, P.M. Schneider, A. Carracedo, Revised guidelines for the publication of genetic population data, Forensic Sci. Int. Genet. 30 (2017) 160–163. https://doi.org/10.1016/j.fsigen.2017.06.007.

G. He, Z. Wang, Y. Su, X. Zou, M. Wang, X. Chen, B. Gao, J. Liu , S. Wang, Y. Hou, Genetic structure and forensic characteristics of Tibeto-Burman-speaking Ü-Tsang and Kham Tibetan Highlanders revealed by 27 Y-chromosomal STRs, Sci. Rep. 9 (2019) 7739. https://doi.org/10.1038/s41598-019-44230-2.

J. Henry, H. Dao, L. Scandrett, D. Taylor, Population genetic analysis of Yfiler® Plus haplotype data for three South Australian populations, Forensic Sci. Int. Genet. 41 (2019) e23–e25. https://doi.org/10.1016/j.fsigen.2019.03.021.

G. Iacovacci, E. D'Atanasio, O. Marini, A. Coppa, D. Sellitto, B. Trombetta, A. Berti, F. Cruciani, Forensic data and microvariant sequence characterization of 27 Y-STR loci analyzed in four Eastern African countries, Forensic Sci. Int. Genet. 27 (2017) 123–131. https://doi.org/https://doi.org/10.1016/j.fsigen.2016.12.015.

M. Kayser, Forensic use of Y-chromosome DNA: a general overview, Hum. Genet. 136 (2017) 621–635. https://doi.org/10.1007/s00439-017-1776-9.

Y.M. Khubrani, J.H. Wetton, M.A. Jobling, Extensive geographical and social

structure in the paternal lineages of Saudi Arabia revealed by analysis of 27 Y-STRs, Forensic Sci. Int. Genet. 33 (2018) 98–105. https://doi.org/https://doi.org/10.1016/j.fsigen.2017.11.015.

I. Kulichová, V. Fernandes, A. Deme, J. Nováčková, V. Stenzl, A. Novelletto, L. Pereira, V. Černý, Internal diversification of non-Sub-Saharan haplogroups in Sahelian populations and the spread of pastoralism beyond the Sahara, Am. J. Phys. Anthropol. 164 (2017) 424–434. https://doi.org/10.1002/ajpa.23285.

M. Lang, H. Liu, F. Song, X. Qiao, Y. Ye, H. Ren, J. Li, J. Huang, M. Xie, S. Chen, M. Song, Y. Zhang, X. Qian, T. Yuan, Z. Wang, Y. Liu, M. Wang, Y. Liu, J. Liu, Y. Hou, Forensic characteristics and genetic analysis of both 27 Y-STRs and 143 Y-SNPs in Eastern Han Chinese population, Forensic Sci. Int. Genet. 42 (2019) e13–e20. https://doi.org/https://doi.org/10.1016/j.fsigen.2019.07.011.

V. Montano, G. Ferri, V. Marcari, C. Batini, O. Anyaele, G. Destro-Bisol, D. Comas, The Bantu expansion revisited: a new analysis of Y chromosome variation in Central Western Africa, Mol. Ecol. 20 (2011) 2693–2708. https://doi.org/10.1111/j.1365-294X.2011.05130.x.

N.M. Myres, K.H. Ritchie, A.A. Lin, R.H. Hughes, S.R. Woodward, P.A. Underhill, Y-chromosome short tandem repeat intermediate variant alleles DYS392.2, DYS449.2, and DYS385.2 delineate new phylogenetic substructure in human Y-chromosome haplogroup tree, Croat. Med. J. 50 (2009) 239–249. https://doi.org/10.3325/cmj.2009.50.239.

M. Nei, F. Tajima, DNA Polymorphism Detectable by Restriction Endonucleases, Genetics 97 (1981) 145–163. http://www.genetics.org/content/97/1/145.

J.L. Newman, The peopling of Africa, a geographic interpretation. Yale University Press (1995).

H. Niederstätter, B. Berger, M. Kayser, W. Parson, Differences in urbanization degree and consequences on the diversity of conventional vs. rapidly mutating Y-STRs in five municipalities from a small region of the Tyrolean Alps in Austria, Forensic Sci. Int. Genet. 24 (2016) 180–193. https://doi.org/https://doi.org/10.1016/j.fsigen.2016.07.009.

J.K. Olofsson, H.S. Mogensen, A. Buchard, C. Børsting, N. Morling, Forensic and population genetic analyses of Danes, Greenlanders and Somalis typed with the Yfiler® Plus PCR amplification kit, Forensic Sci. Int. Genet. 16 (2015) 232–236. https://doi.org/https://doi.org/10.1016/j.fsigen.2015.02.006.

E. Ottaviani, S. Vernarecci, P. Asili, A. Agostino, P. Montagna, Preliminary assessment of the prototype Yfiler® Plus kit in a population study of Northern Italian males, Int. J. Legal Med. 129 (2015) 729–730. https://doi.org/10.1007/s00414-014-1078-x.

I. Pickrahn, E. Müller, W. Zahrer, B. Dunkelmann, J. Cemper-Kiesslich, G. Kreindl, F. Neuhuber, Yfiler® Plus amplification kit validation and calculation of forensic parameters for two Austrian populations, Forensic Sci. Int. Genet. 21 (2016) 90–94. https://doi.org/https://doi.org/10.1016/j.fsigen.2015.12.014.

C. Rapone, E. D'Atanasio, A. Agostino, M. Mariano, M.T. Papaluca, F. Cruciani, A. Berti, Forensic genetic value of a 27 Y-STR loci multiplex (Yfiler® Plus kit) in an Italian

population sample, Forensic Sci. Int. Genet. 21 (2016) e1–e5. https://doi.org/10.1016/j.fsigen.2015.11.006.

M. Shonhai, T. Nhiwatiwa, T. Nangammbi, S. Mazando, Genetic analysis of 27 Ychromosomal STR loci in a Zimbabwean Shona ethnic group, Leg. Med. 43 (2020) 101660. https://doi.org/10.1016/j.legalmed.2019.101660.

B. Trombetta, F. Cruciani, D. Sellitto, R. Scozzari, A new topology of the human Y chromosome haplogroup E1b1 (E-P2) revealed through the use of newly characterized binary polymorphisms, PLoS One. 6 (2011) e16073–e16073. https://doi.org/10.1371/journal.pone.0016073.

B. Trombetta, E. D'Atanasio, A. Massaia, M. Ippoliti, A. Coppa, F. Candilio, V. Coia, G. Russo, J.M. Dugoujon, P. Moral, N. Akar, D. Sellitto D, G. Valesini, A. Novelletto, R. Scozzari, F. Cruciani, Phylogeographic Refinement and Large Scale Genotyping of Human Y Chromosome Haplogroup E Provide New Insights into the Dispersal of Early Pastoralists in the African Continent, Genome Biol. Evol. 7 (2015) 1940–1950. https://doi.org/10.1093/gbe/evv118.

A. Untergasser, I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm, S.G. Rozen, Primer3—new capabilities and interfaces, Nucleic Acids Res. 40 (2012) e115–e115. https://doi.org/10.1093/nar/gks596.

M. van Oven, A. Van Geystelen, M. Kayser, R. Decorte, M.H.D. Larmuseau, Seeing the Wood for the Trees: A Minimal Reference Phylogeny for the Human Y Chromosome, Hum. Mutat. 35 (2014) 187–191. https://doi.org/10.1002/humu.22468.

Y. Wang, Y. Zhang, C. Zhang, R. Li, Y. Yang, X. Ou, D. Tong, H. Sun, Genetic polymorphisms and mutation rates of 27 Y-chromosomal STRs in a Han population from Guangdong Province, Southern China, Forensic Sci. Int. Genet. 21 (2016) 5–9. https://doi.org/https://doi.org/10.1016/j.fsigen.2015.09.013.

A.A. Westen, T. Kraaijenbrink, L. Clarisse, L.J.W. Grol, P. Willemse, S.B. Zuniga, E.A. Robles de Medina, R. Schouten, K.J. van der Gaag, N.E.C. Weiler, A.J. Kal, M. Kayser, T. Sijen, P. de Knijff, Analysis of 36 Y-STR marker units including a concordance study among 2085 Dutch males, Forensic Sci. Int. Genet. 14 (2015) 174–181. https://doi.org/https://doi.org/10.1016/j.fsigen.2014.10.012.

S. Willuweit, L. Roewer, The new Y Chromosome Haplotype Reference Database, Forensic Sci. Int. Genet. 15 (2015) 43–48. https://doi.org/https://doi.org/10.1016/j.fsigen.2014.11.024.

Y. Xue, A history of male migration in and out of the Green Sahara, Genome Biol. 19 (2018) 30. https://doi.org/10.1186/s13059-018-1410-8.

Conflicts of interest

The authors declare that there are no conflicts of interest

Acknowledgements: Samples from Cameroon were collected by Prof. Gabriella Spedini, who is gratefully acknowledged. This work was supported by Sapienza University of Rome (grant number RM11916B7A0F76B1 and RM118164337D9395 to FC).