Int J Legal Med (2009) 123:143–149
DOI 10.1007/s00414-008-0303-x

ORIGINAL ARTICLE

X-chromosome STR sequence variation, repeat structure, and nomenclature in humans and chimpanzees

Iva Gomes · Mechthild Prinz · Rui Pereira · Erik Bieschke · Wolfgang R. Mayr · António Amorim · Angel Carracedo · Leonor Gusmão

Received: 7 July 2008 / Accepted: 14 November 2008 / Published online: 12 December 2008

Abstract Qualitative information on the sequence composition of the allele and locus structure of the X-STRs DXS8378, DXS9898, DXS6789, GATA31E08, and GATA172D05 was generated in this study. Sequence data were obtained from chimpanzees (*Pan troglodytes*) and diverse human population groups including Africans, Caucasians, Asians, African–Americans, and Hispanics. Results revealed DXS8378 as the most stable locus. On the other hand, DXS9898 and GATA172D05 showed unstable regions identified through chimpanzee–human sequence comparison. At DXS6789, intra-allelic variation was found in all human populations, i.e., alleles with same fragment sizes showed structural differences only detected by sequencing. At the GATA31E08 locus, a previously unreported variation between humans and chimpanzees was identified in an adjacent region upstream from the repeat. This resulted in the addition of two repeat units and the proposal of a new allele nomenclature at this locus. Also, the sequence analyses did not detect ethnic differences between the studied population samples that would justify the use of these markers to help identify ethnic origin in an anthropological context.

Keywords X-STRs · Sequence structure · Nomenclature · Chimpanzee

Electronic supplementary material The online version of this article (doi:10.1007/s00414-008-0303-x) contains supplementary material, which is available to authorized users.

I. Gomes (✉) · R. Pereira · A. Amorim · L. Gusmão
IPATIMUP, Institute of Molecular Pathology
and Immunology of the University of Porto,
Rua Dr. Roberto Frias, s/n,
4200-465 Porto, Portugal
e-mail: igomes@ipatimup.pt

I. Gomes · R. Pereira · A. Carracedo
Institute of Legal Medicine,
University of Santiago de Compostela,
15782 Santiago de Compostela, Spain

M. Prinz · E. Bieschke
Department of Forensic Biology,
Office of the Chief Medical Examiner,
New York, NY, USA

W. R. Mayr
Division of Blood Group Serology, Medical University of Vienna,
1090 Vienna, Austria

A. Amorim
Faculty of Sciences, University of Porto,
4050 Porto, Portugal

Introduction

X-chromosome genetic markers have been revealed to be quite useful, particularly in specific cases of kinship analysis. In males, one single copy which is entirely transmitted to female descendants and the absence of recombination allows for direct haplotyping. This characteristic itself can lead to paternity exclusions if the alleged father is unavailable and two sisters or half-sisters are under investigation since they will share the same paternal alleles. This is also true in cases of kinship analysis when the assessment of an alleged paternal grandmother/granddaughter relationship is being examined. In population and forensic genetics, many studies have also focused on the analysis of X-chromosome short tandem repeats (X-STRs) and on the collection of frequency and forensic efficiency information (e.g., [1, 3–5, 7, 8, 12–14, 17–21]). The evident increase of studies in the literature justifies the necessity of a common nomenclature allowing for communication, data exchange,

and data comparison among laboratories. The importance of establishing a common and accurate nomenclature has long been emphasized by several international DNA working groups (e.g., [2, 6, 11]), as well as by many other studies that use STRs located throughout the genome (e.g., [9, 10, 15, 16]). Although many population-data-based studies on X-STRs are available in the literature (e.g., [1, 3–5, 7, 8, 12, 18–20]), few have focused on the analysis of allele and locus sequence structure (e.g., [4, 13, 14, 20]).

In this work, we aimed to study the sequence structure variation at the DXS8378, DXS9898, GATA172D05, DXS6789, and GATA31E08 loci in different human population groups, as well as in chimpanzees. Comparison of both human and chimpanzee sequence composition permitted a more accurate analysis of the X-STRs studied in this work at the sequence structural level, as well as inference of their mutation evolving mechanism.

Materials and methods

Human population groups

A total of 318 male samples from five different population groups were sequenced for the five X-chromosomal STRs. The groups studied in this work included samples from the three major human population groups, namely Africans (Uganda, Angola, and Mozambique), Caucasians (northern Portugal), and Asians (Macau and US Asians), as well as samples from two other US admixed groups, African–Americans and Hispanics. Samples from ten male and four female chimpanzees (*Pan troglodytes*) were also sequenced for the X-linked markers in the present work.

Allele selection

Alleles selected for sequencing of the tetranucleotides DXS8378, DXS9898, GATA172D05, and DXS6789 were based on previous studies [7, 8, 18], except for the Asian samples from Macau (L. Gusmão personal unpublished data).

For GATA31E08, Caucasian data (northern Portugal) were selected from a GEP-ISFG collaborative exercise on X-STRs [12]; for the African–American, Asian, and African data, allele selection was based on L. Gusmão personal unpublished data.

All observed allele classes were sequenced. In addition, the most common alleles were sequenced several times in each population group. The reference DNA samples, 9947A (female) and 9948 (male; taken from Promega commercial kits, Madison, WI, USA) were also sequenced for the five loci.

DNA extraction

The US admixed population groups were extracted from postmortem blood stains available for research purposes using the silica-coated magnetic bead purification technology with the automated M-48 bio-Robot (Qiagen, Hilden, Germany) and following the manufacturer's instructions. The Caucasian, African, and Asian samples were extracted either using a standard phenol-chloroform or chelex procedures.

Primers and PCR

Primer sequences used for amplification and sequencing of DXS8378, DXS9898, GATA172D05, and DXS6789 are described in Gomes et al. [7]. For GATA31E08, primers used were according to Gusmão et al. [12]. Polymerase chain reaction (PCR) singleplex amplification and thermocycling conditions were as follows: 7.5 μl of a 2× QIAGEN Multiplex PCR master mix (Qiagen), 0.2 μM of each primer, and 0.5–5 ng of genomic DNA in a 15-μl final volume. Pre-incubation for 15 min at 95°C, followed by ten cycles of 30 s at 94°C, 90 s at 60°C, 60 s at 72°C; and 20 cycles of 30 s at 94°C, 90 s at 58°C, 60 s at 72°C with a final extension at 72°C for 60 min.

Chimpanzee samples were amplified with the same primers used for PCR reaction of the human samples. To ensure primer binding due to possible mismatches with the human DNA sequence, PCR amplifications were performed by lowering annealing temperature to 54°C for all loci. The observation of unspecific products was minimal or absent when compared with the target region and allowed sequencing of chimpanzee samples using the same primers with no difficulties. All other PCR reaction conditions were the same as described above.

Sequencing conditions

Prior to sequencing, heterozygote female alleles were separated by acrylamide gel electrophoresis and individual alleles were eluted and reamplified as in Gusmão et al. [10]. Amplified fragments were purified with the PCR product cleanup ExoSAP-IT (USB Corporation, Cleveland, USA) following the manufacturer's protocol. Sequencing reaction was performed using the BigDye Terminator v3.1 (Applied Biosystems, Foster City, CA, USA). Thermocyling conditions used were: 96°C pre-incubation for 1 min, followed by 35 cycles of 96°C denaturation for 10 s, annealing at 55°C for 5 s, and extension at 60°C for 2 min. Final sequenced products were purified either with Centri-sep 8 Spin columns (Princeton Separations, Adelphia, NJ, USA) or 5% Sephadex in-house filtration columns. Final products were visualized in an ABI PRISM 3100 Genetic Analyzer electrophoresis

system and analyzed with sequencing analysis 3.7 software (Applied Biosystems).

Alignment of human and chimpanzee sequences was performed either using the sequence similarity search BLAT Genome Browser (www.genome.ucsc.edu) or the EMBL-EBI ClustalW2 sequence alignment tool (www.ebi.ac.uk).

Results

Analysis was performed by comparing the sequence structure of X-STR alleles in different human populations (Caucasians, Africans, Asians, African–Americans, and Hispanics) and between humans and chimpanzees. Sequencing results for the five X-STRs in the human and chimpanzee samples as well as the chimpanzee and human genome sequence alignments are presented in Tables S1–S5. The allele sequence obtained for the reference DNA samples 9947A and 9948 is shown in Table 1. The variation found within human populations and between humans and chimpanzees was restricted to the main repetitive region and to the flanking regions of the varying stretch. Results obtained in this study are discussed individually for each locus.

DXS8378

DXS8378 locus has been described as having a simple uninterrupted repeat motif CTAT, with alleles varying between seven and 14 copies (e.g., [3, 4, 8, 20]). For the human samples sequenced in this work, no sequence structure variations, inside the repeat or flanking regions, were observed in or among the five groups (Table S1). Among individuals, differences for the polymorphic STR were only encountered in the number of repetitions of the CTAT unit.

DXS8378 locus also revealed to be polymorphic among chimpanzees: sequencing of CTAT alleles ranged from eight to 12 repeats, with a total of eight sequenced alleles (in four female chimpanzees). As in humans, no structural variations were found among these primates inside the repeat or in the adjacent regions, except for one of the chimpanzees. In one sample, results showed a C→G transversion at the 18 bp upstream of the CTAT repeat motif.

When comparing both human and chimpanzee sequences, no differences were found in the upstream flanking region and alleles only differed by the number of the repetitive CTAT units (Table S1). However, when aligning human and chimpanzee sequences deposited at National Center for Biotechnology Information (www.ncbi.nlm.nih.gov), at bases 15 and 17 from the 5'end of the reverse primer hybridization site, two nucleotide differences were encountered (Table S1). This primer DNA mismatch did not inhibit primer annealing and subsequent amplification of the DXS8378 fragments when using the same human primers for amplification of the chimpanzee samples.

In summary, sequencing of DXS8378 alleles in all studied samples from the five different groups, as well as in the chimpanzees revealed a simple repetitive CTAT structure.

DXS9898

The tetranucleotide STR DXS9898 has been reported as presenting a simple structure, with non-consensus repeats and with an allele range of seven to 16 repeats [3, 7, 13, 17]. The non-consensus nature of this STR is in fact supported by an 8.3 allele that is present in high frequencies in most of the studied populations and represented by approximately 25% of the chromosomes in most Caucasian populations (e.g., [3, 12, 13, 18]).

Sequencing results reveal the presence of two initial TATC units that are interrupted by a trinucleotide invariable ATC motif (that is not being considered for allele nomenclature, as proposed by Hering et al. [13]). The repeat then continues with a variable number of TATCs. Regarding the non-consensus alleles sequenced in this work (alleles 8.3, 13.3, 14.3, and 15.3), the tandem TATC is again interrupted by another ATC trinucleotide motif (Table S2). Apart from this, no nucleotide sequence structure variations were observed inside the flanking or repeat regions, in or among any of the five different human groups.

DXS9898 was also found to be polymorphic among chimpanzees. Sequencing of the different fragment lengths

Table 1 Sequencing results for reference DNA samples 9948 (male) and 9947A (female)

Locus	Allele	Reference DNA 9948	Allele	Reference DNA 9947A[a]
DXS8378	11	$(CTAT)_{11}$	–	–
DXS9898	13	$(TATC)_2$-atc-$(TATC)_{11}$	–	–
GATA31E08	12	$(AGGG)_2$ $(AGAT)_{10}$	13	$(AGGG)_2$ $(AGAT)_{11}$
GATA172D05	6	tata $(TAGA)_6$	10	tata $(TAGA)_{10}$
DXS6789	20	$TATC(TATG)_9(TATC)_{10}$	–	–

[a] For reference sample 9947A, sequencing results were obtained only for the homozygous state

revealed the presence of seven different alleles, with four different structures, among the 12 primate chromosomes studied. Unlike the human population groups, data showed a more complex repeat composition (Table S2). The TATC tandem repeat is constantly interrupted by several point mutations which contributed to intra-allelic variation: fragments with equal base pair sizes revealed sequence structural differences.

Comparing *Homo sapiens* and *P. troglodytes* genomes, the upstream flanking sequence of the repeat regions differ by only one base mismatch at the first nucleotide of the 5'end forward primer, which did not interfere with fragment amplification in our study. Another difference was a four-base deletion –GTCT– in chimpanzees downstream of the repeat pattern (Table S2). The comparison of both genome sequences was also relevant to the understanding of the DXS9898 STR structure. The initial and constant $(TATC)_2$ ATC structure observed in humans was also invariant in chimpanzees. Findings showed that this motif remains conserved among the human groups, as well as between humans and chimpanzees, and among chimpanzees. Moreover, this unvarying structure seen throughout the 61 human samples (also observed in a German population sample by Hering and Szibor [13]) and the 12 chimpanzees chromosomes sequenced in this work predicts that no variation is expected to occur in humans. Furthermore, still in the human genome, the end of the TATC pattern is interrupted by nine nucleotides that have disrupted the repeat, most likely by an insertion of an A (TAATC) and an A→G transition (TGTC). This disruption is then followed by another ATC– $(TATC)_3$ motif, while in chimpanzees variation still occurs at this region (Table S2). Results from chimpanzee and human genome alignment emphasize this unstable region at DXS9898. Therefore, although no variation was found in humans at this region, the differences observed within chimpanzees and between humans and chimpanzees do not allow the probable occurrence of variation to be ruled out. The STR structure most probably evolved from a single array that included this region. Also, at this locus, observations point to the ATC sequence as the motif that constantly interrupts the repeat. Thus, the mutation process responsible for this structure type seems to be more complex than single base deletions that normally interrupt perfect repeats.

DXS6789

The microsatellite DXS6789 presents a compound sequence composition and with an allele range of 13 to 26 repeats (e.g., [14, 18, 20]). In the five human groups studied, results demonstrate two types of repetitive sequences (TATG and TATC; Table S3). For longer alleles presenting more than 17 repeats, an invariant TATC insertion is observed at the beginning of the variable sequence motif. This result was consistent in all populations, with no indication of a population-specific association. For the majority of the alleles sequenced and for the groups present in this work, other sequence polymorphisms were also observed. Alleles that are identical in size exhibited structural variations regarding the TATG/TATC proportion, revealing a hidden variation undetected by fragment size analysis (Table S3). The tandem TATG seems to have less variation than the TATC, showing an allele size range from six to 12 repeats. On the other hand, the higher apparent variation of the TATC sequence is supported by the 12 different repeats for this structure (four to 15 repetitions) that were encountered in the studied populations.

For analysis of the intra-allelic variation of the DXS6789 repeats, alleles were grouped into subtypes (s) according to the TATG pattern observed: s_0—$(TATG)_{12}$; s_1—$(TATG)_{11}$; s_2—$(TATG)_{10}$; s_3—$(TATG)_9$; s_4—$(TATG)_8$; s_5—$(TATG)_7$; s_6—$(TATG)_6$. The intra-allelic variation detected seems to be lower in the Caucasian group since nearly 80% of the samples have a constant ten TATG repeat motifs (s_2 allele subtype), while, in Africans, only nearly half of the samples showed this characteristic. Not surprisingly, Africans presented a higher intra-allelic variation at this locus. For both the Asian and Hispanic groups, about 62% of the samples show an s_2 allele subtype. An A→G transition was also detected in an African sample with 21 repeats that interrupts the TATC variable motif (Table S3). Despite this variation, no exclusive sequence structure was seen in any of the population groups.

When analyzing the chimpanzee sequence results at DXS6789, the TATG repetitive tract that is present in humans is absent in chimpanzees (Table S3). This fact supports a probable more recent emergence of the TATG tract than the TATC unit. Another difference is the interruption by eight nucleotides of the tandem (in most of the samples): $(TATC)_2$ TATT AATC $(TATC)_n$. Still in chimpanzees, further analysis revealed that an alternative STR structure could be considered at DXS6789 if a TCTA pattern is designated: $(TCTA)_2TTAA/TTTA(TCTA)_{9–10}$. Two initial TCTA copies are interrupted by a four-base TTAA (nine out of the 11 sequenced chromosomes); on the other hand, two samples showed a different structure with a TTTA disrupting the tandem. Although observations point out that both in humans and in chimpanzees the first tract that is repetitive is the TATC, the TCTA motif creates a simpler locus structure for chimpanzees. Two base mutations were also observed between primer annealing sequences of humans and chimpanzees (Table S4). However, these mismatches did not inhibit amplification in chimpanzees while using the same human amplification primer sequences.

GATA31E08

The X-linked STR, GATA31E08 has not been described in many population studies; Asian population data have initially been reported on Korean and Japanese groups [1, 20]. More recently, this locus was included in an X-STR decaplex optimized by the GEP-ISFG working group that was used for data collection and genetic analysis of several groups from Iberian and Latin American populations [12]. An allele range from six to 15 repeats has been observed for this polymorphic locus in the referred studies [1, 12, 20]. An allele 5, not previously described, was selected and also sequenced in African samples (L. Gusmão unpublished personal data). The available information in the literature regarding sequence nomenclature describes GATA31E08 as a simple repeat of AGAT variable units [20].

Sequencing data results of the different human groups revealed an additional variation at the immediate flanking region upstream to the repeat: a tetrameric AGGG is present with two or three copies (Table S4). Out of the 75 sequenced samples, seven exhibited three copies of AGGG; three were sequenced in African–Americans as well as in Caucasians and one in the Asian group (Macau). No samples containing three AGGG repeats were observed in the African group (Uganda; Table S4). The AGGG tract is then followed by a variable number of simple AGAT units. No other differences were observed in the flanking or repeat regions between or among the different population groups at GATA31E08.

Sequencing of nine chimpanzee alleles also revealed variation between *P. troglodytes* and *H. sapiens* genomes: only one copy of the AGGG tetranucleotide was found in chimpanzees at the beginning of the repeat region (Table S4). When compared with humans, chimpanzees' sequence structure also revealed an additional difference at the repeat: an extra sequence composition –AGAC– is present, interrupting the AGAT tract. The repeat then ends with one or two copies of this motif (AGAT; Table S4). At the downstream flanking region from the main stretch, one base nucleotide mutation difference was also observed between humans and chimpanzees. In addition, intra-allelic variation was also found among chimpanzees: alleles identical in fragment sizes showed different sequences compositions.

GATA172D05

GATA172D05 sequence structure was first described by Edelmann et al. [3] as a GATA simple repeat. The allele nomenclature was afterwards altered to a TAGA motif by Edelmann et al. [4] that added an extra repeat to the GATA172D05 alleles. Repeat variation was found between alleles. The immediate upstream TAGA tract flanking region has a TATA sequence that in three out of the 41 samples presented a T→G transversion (Table S5). This substitution converts the TA*T*A into a TA*G*A motif that, due to its adjacent location, becomes part of the repeat region. Considering the sequence of these alleles [...ta*G*a-$(TAGA)_n$...] as the ancestral state of a single repeat, an extra unit is gained. Furthermore, if the two GA bases before this structure [...*GA*taga-$(TAGA)_n$...] are also included, then a GATA repeat motif should be considered instead of TAGA as originally reported by Edelmann et al. [3].

When analyzing chimpanzee sequence, the simple uninterrupted TAGA repeat present in the human genome is absent in *P. troglodytes* (Table S5). A complex structure suggests a highly unstable region due to the numerous differences that distinguish the chimpanzee genome from humans at the sequence structure level of GATA172D05 (Table S5). In fact, although it was not possible to identify a repetitive type structure in chimpanzees, from the 14 alleles sequenced, three different structures were observed. The complex sequence structure composition at GATA172D05 in chimpanzees can be summarized as...TAGA $(TAAA)_{0–1}$ $(TAGA)_2$ TCTATA ACTATA $(TCTATA)_{2–3}$... This locus revealed many sequence structural differences between humans and chimpanzees: a simple repeat of TAGA units in humans showed a very complex and compound composition in chimpanzees.

Discussion

Implication in nomenclature

Although no specific ISFG guidelines on the use of X-chromosome markers in forensics analysis are available, the existing recommendations concerning autosomal STRs and more specifically on Y-STRs can be extended to X-STRs regarding locus structure and allele nomenclature designation [2, 6, 11]. In the context of this work, we underline some of those main suggestions important for nomenclature discussion: (1) the identification of regions that vary and that may likely vary should be supported by sequence analysis of individuals from different human populations, as well as by comparison with chimpanzee sequences; (2) the first repetitive motif in a simple or complex sequence structure that encloses the highest number of repeats should be defined; (3) alleles must be named in agreement with the total number of adjacent variant and non-variant repeats obtained from sequence data analysis; (4) non-variant repeats should be included in the allele nomenclature if the number of nucleotides that separate these structures from the main variable repeat is equal or less to the number of nucleotides of the repeat unit; (5) if a previously established nomenclature of an STR is not in accordance

with the ISFG guidelines but has been widely utilized, the nomenclature should not be altered to avoid unnecessary confusion. In light of these particular recommendations, we discuss the results obtained in this work and the implications for the nomenclature of the X-STRs DXS8378, DXS9898, GATA172D05, DXS6789, and GATA31E08. Final analyses on the STR structure and allele nomenclatures are summarized in Table 2.

DXS8378

The STR structure stability observed at DXS8378 suggests that no variation is expected to occur at this locus from the accumulation of new population data. This locus presents a simple structure composed by CTAT repeat units (Table 2). Therefore, allele nomenclature for this locus is straightforward, as described by Edelmann et al. [3].

DXS9898

Major findings at DXS9898 locus show that the $(TATC)_2$ ATC adjacent structure remains conserved in the humans and chimpanzees studied in this work. Although, based on these results, no variation is expected to occur, this non-variant repeat must remain included in the allele nomenclature since the entire structure could have evolved from a single array. On the other hand, the sequence variation found between humans and chimpanzees at the downstream region adjacent to the main stretch would suggest the inclusion of this unstable region for allele nomenclature. However, given that no variation was found among the human populations and since DXS9898 has been a broadly used STR, proposing a new nomenclature would produce unnecessary confusion (Table 2).

DXS6789

Genome comparisons between human populations and between both human and chimpanzee species did not reveal any sequence variations at this locus that would have an impact on the already described nomenclature (Table 2).

GATA31E08

The variation found between humans, as well as between humans and chimpanzees, strongly supports a nomenclature change at this locus. The variable AGGG adjacent region should be included in allele nomenclature adding two extra repeats to the previously described AGAT sequence structure [20]. Therefore, in this work, we propose a new nomenclature for the GATA31E08 locus as follows: $(AGGG)_n(AGAT)_m$ (Table 2). To our knowledge, this structure has not been previously described nor been included for allele nomenclature designation.

GATA172D05

The variation found at the upstream adjacent region of the TAGA repeat suggests an allele nomenclature change at GATA172D05 by adding an extra repeat. Nevertheless, according to the ISFG guidelines [2, 6, 11], the wide use of this locus in forensic and population genetics does not favor a new nomenclature proposal and therefore the established nomenclature should be the one considered [4] (Table 2).

Final remarks

Findings in this work resulted from extensive sequence data analysis that included different human populations, as well as chimpanzee samples for the DXS8378, DXS9898, GATA172D05, DXS6789, and GATA31E08 loci. This work demonstrates the importance of screening individuals from different populations and the use of the *P. troglodytes* genome sequence when establishing a new STR nomenclature. The chimpanzee genome has been completely sequenced and is accessible using sequence similarity search tools, e.g.,

Table 2 STR consensus structures taking into account the locus variation observed in this study between humans and chimpanzees and allele nomenclatures based on the ISFG guidelines

X-STR	STR structure	Allele nomenclature	References
DXS8378	$(CTAT)_n$	$(CTAT)_n$	[3]
DXS9898	$(TATC)_2$ ATC $(TATC)_m$ $(ATC)_{0–1}$ $(TATC)_n$ TAATC TGTC ATC $(TATC)_3$	$(TATC)_2$ atc $(TATC)_m$ $(ATC)_{0–1}$ $(TATC)_n$	[13]
DXS6789	$(TATC)_{0–1}$ $(TATG)_m$ $(TATC)_n$	$(TATC)_{0–1}$ $(TATG)_m$ $(TATC)_n$	[14]
GATA31E08	$(AGGG)_{2–3}$ $(AGAT)_n$	$(AGGG)_{2–3}$ $(AGAT)_n$	This study
GATA172D05	GATA T/G A $(TAGA)_n$ GCTATA TCAATA CCTATA TCTATA GATATA	$(TAGA)_n$	[4]

At DXS9898, lower case atc is not considered for allele nomenclature

the available online programs BLAST (www.ncbi.nlm.nih.gov/blast/Blast.cgi) and BLAT (www.genome.ucsc.edu), thus replacing the need for reference DNA. This approach has been demonstrated in several studies (e.g., [9, 10, 15, 16]) and is strongly suggested by the ISFG [11] since it is crucial to identifying regions that may vary. An example for this is the variable region that was detected in this study at GATA31E08 leading to a new nomenclature proposal. So far, the few sequence nomenclature studies available on X-STRs have only focused on the sequence analysis of a single population (e.g., [4, 13, 14, 20]) without the use of chimpanzee information. The present work emphasizes the importance of a more thorough study of the STR sequence variation at a specific locus before the establishment of a new nomenclature. Also, on a sequence structural level, our final results for the repeat regions do not support any ethnic differences between the studied populations that could justify the use of these markers for ethnic group differentiation in an anthropological context.

Acknowledgements We would like to thank Fr. Germano Serra and Magdalen Awor for the Uganda sampling. This work was partially supported by Fundação para a Ciência e a Tecnologia through grants SFRH/BD/21647/2005, SFRH/BD/30039/2006, and POCI/AFR/62242/2004 and by "Programa Operacional Ciência, e Inovação 2010," as well as by the Fundação Luso-Americana para o Desenvolvimento (FLAD).

References

1. Asamura H, Sakai H, Kobayashi K, Ota M, Fukushima H (2006) MiniX-STR multiplex system population study in Japan and application to degraded DNA analysis. Int J Legal Med 120:174–181
2. Bär W, Brinkmann B, Budowle B et al (1997) DNA recommendations: further report of the DNA commission of the ISFH regarding the use of short tandem repeat systems. Int J Legal Med 110:175–176
3. Edelman J, Hering S, Michael M et al (2001) 16 X-chromosome STR loci frequency data from a German population. Forensic Sci Int 124:215–218
4. Edelmann J, Deichsel D, Hering S, Plate I, Szibor R (2002) Sequence variation and allele nomenclature for the X-linked STRs DXS9895, DXS8378, DXS7132, DXS6800, DXS7133, GATA172D05, DXS7423 and DXS8377. Forensic Sci Int 129:99–103
5. Fracasso T, Schürenkamp M, Brinkmann B, Hohoff C (2008) An X-STR meiosis study in Kurds and Germans: allele frequencies and mutation rates. Int J Legal Med. doi:10.1007/s00414-008-0231-9
6. Gill P, Brinkmann B, d'Aloja E et al (1997) Considerations from the European DNA profiling group (EDNAP) concerning STR nomenclature. Forensic Sci Int 87:185–192
7. Gomes I, Alves C, Maxzud K et al (2007) Analysis of 10 X-STRs in three African populations. Forensic Sci Int Genet 1:208–211
8. Gomes I, Prinz M, Pereira P et al (2007) Genetic analysis of 3 US population groups using an X-chromosomal STR decaplex. Int J Legal Med 121:198–203
9. González-Neira A, Elmoznino M, Lareu MV, Sánchez-Diz P, Gusmão L, Prinz M, Carracedo A (2001) Sequence structure of 12 novel Y chromosome microsatellites and PCR amplification strategies. Forensic Sci Int 122:19–26
10. Gusmão L, González-Neira A, Alves C, Lareu M, Costa S, Amorim A, Carracedo A (2002) Chimpanzee homologous of human Y specific STRs. A comparative study and a proposal for nomenclature. Forensic Sci Int 126:129–136
11. Gusmão L, Butler JM, Carracedo A et al (2006) DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. Int J Legal Med 120:191–200
12. Gusmão L, Sánchez-Diz P, Alves C et al (2008) A GEP-ISFG collaborative study on the optimization of an X-STR Decaplex: Data on 15 Iberian and Latin American populations. Int J Legal Med (in press)
13. Hering S, Szibor R (2000) Development of the X-linked tetrameric microsatellite marker DXS9898 for forensic purposes. J Forensic Sci 45:929–931
14. Hering S, Kuhlisch E, Szibor R (2001) Development of the X-linked tetrameric microsatellite marker HumDXS6789 for forensic purposes. Forensic Sci Int 119:42–46
15. Kayser M, Caglià A, Corach D et al (1997) Evaluation of Y-chromosomal STRs: a multicenter study. Int J Legal Med 110:125–133 141–149
16. Lazaruk K, Wallin J, Holt C, Nguyen T, Walsh PS (2001) Sequence variation in humans and other primates at six short tandem repeat loci used in forensic identity testing. Forensic Sci Int 119:1–10
17. Lee HY, Park MJ, Jeong CK, Lee SY, Yoo JE, Chung U, Choi JH, Kim CY, Shin KJ (2004) Genetic characteristics and population study of 4 X-chromosomal STRs in Koreans: evidence for a null allele at DXS9898. Int J Legal Med 118:355–360
18. Pereira R, Gomes I, Amorim A, Gusmão L (2007) Genetic diversity of 10 X-chromosome STRs in northern Portugal. Int J Legal Med 121:192–197
19. Pico A, Castillo A, Vargas C, Amorim A, Gusmão L (2008) Genetic profile characterization and segregation analysis of 10 X-STRs in a sample from Santander, Colombia. Int J Legal Med 122:347–351
20. Shin SH, Yu JS, Park SW, Min GS, Chung KW (2005) Genetic analysis of 18 X-linked short tandem repeat markers in Korean population. Forensic Sci Int 147:35–41
21. Szibor R, Krawczak M, Hering S, Edelmann J, Kuhlisch E, Krause D (2003) Use of X-linked markers for forensic purposes. Int J Legal Med 117:67–74