Forensic Science International: Genetics 20 (2016) 112–116

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Short communication

New sequence variants detected at DXS10148, DXS10074 and DXS10134 loci

Iva Gomes[a,*], António Brehm[b], Leonor Gusmão[c,d,e], Peter M. Schneider[a]

[a] Institute of Legal Medicine, Faculty of Medicine, University of Cologne, Cologne, Germany
[b] Human Genetics Laboratory, University of Madeira, Funchal, Portugal
[c] DNA Diagnostic Laboratory (LDD), Institute of Biology, State University of Rio de Janeiro (UERJ), Brazil
[d] IPATIMUP, Institute of Pathology and Molecular Immunology of the University of Porto, Porto, Portugal
[e] Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal

ARTICLE INFO

Article history:
Received 9 June 2015
Received in revised form 29 September 2015
Accepted 15 October 2015
Available online 28 October 2015

Keywords:
Investigator Argus X-12
X-STRs
X chromosome
West Africa
Iraq
Guinea-Bissau

ABSTRACT

A great amount of population and forensic genetic data are available for X-STRs supporting the need for having a common and accurate nomenclature among laboratories allowing for better communication, data exchange, and data comparison. DXS10148, DXS10074 and DXS10134 are commonly used X-STRs particularly due to their inclusion in the commercial kit Investigator Argus X-12 (Qiagen). Samples from West Africa and Iraq were sequenced for all three X-STRs allowing the detection of new DNA sequence variants. At DXS10148, variation was detected at four bases downstream from the flanking region from the repeat motif. The sequence AAGG-AAAG has been detected for the first time as a varying $(AAGG\text{-}AAAG)_{1-3}$ motif, in the present work. One additional string when compared to the common one $(AAGG\text{-}AAAG)_2$ adds eight bases to the fragment size of the tetranucleotide STR. This means that 2 repeats are added in these cases to the fragment size of the allele, while the presence of only one copy will reduce the expected allele size by 2 repeats. At DXS10074 two varying stretches consisting of AC and AG dinucleotide repeats were observed in the upstream flanking region, six bases from the main repeat core that also influence the expected allele size. DXS10134 revealed a simpler nomenclature in the Guinea-Bissau sample set when compared to the previously described allele nomenclature. This detected new hidden variation also has impact on the actual allele nomenclature at this locus as it contributes to a new class of short alleles so far undetected in other studies.

1. Introduction

X-chromosomal STRs (X-STRs) have been broadly used over the last decade in forensic and population genetics and many worldwide population data have been produced. From *in-house* to commercially developed kits (e.g., [1–3]) a wide range of highly polymorphic X chromosome markers are now available for forensic identification and, in particular, for specific kinship analysis settings. The great amount of studies focusing on X-STRs supports the need for having a common and accurate nomenclature among laboratories allowing for better communication, data exchange, and data comparison. This has long been emphasized by several international DNA working groups (e.g., [4–7]), as well as by many other studies focusing on STRs (e.g., [8–11]); two of them specifically concentrated on nomenclature issues regarding X-STRs [8,9].

In this study, the comparison of the sequence data from different populations allowed for the observations of new sequence variations in the flanking regions of the allele repeat structures at DXS10148 and DXS10074 loci, and in the main repeat motif of DXS10134, which are presented here. We also discuss the implications of these findings for the originally proposed allele nomenclatures.

2. Material and methods

2.1. X-STR profiling and sample selection for sequencing

Samples were obtained under informed consent from West African countries (with the bulk of data originating from Guinea-Bissau) and from Iraq. Twelve X-STRs were genotyped using the Investigator Argus X-12 kit (Qiagen, Hilden, Germany), following the recommended protocol in the kit's manual [3].

* Corresponding author. Fax: +49 221 47888370.
E-mail address: iva.gomes@uk-koeln.de (I. Gomes).

http://dx.doi.org/10.1016/j.fsigen.2015.10.005
1872-4973/

During X-STR profiling a number of silent, rare or new alleles were detected. Consequently, traditional Sanger sequencing was performed to either identify or confirm the genotypes. In addition to the population samples, reference cell line DNA 9948 [12] was also sequenced for all three loci.

2.2. Sequencing of DXS10148, DXS10074 and DXS10134 alleles

Primers for singleplex amplification and sequencing reactions were either selected from the literature or newly designed using the online software Primer3 [13] (Table S1). A touchdown PCR (TD-PCR) protocol was applied [14] to avoid unspecific amplification products, observed initially by electrophoresis on a 1.5% agarose gel and ethidium bromide staining. All amplified products were purified with the PCR product cleanup ExoSAP-IT (USB Corporation) following the manufacturer's conditions. Forward and reverse sequencing reactions were performed using the BigDye Terminator v1.1 (Life Technologies) following the recommended protocol. Final sequenced products were purified with a Sephadex *in-house* filtration column protocol and detected in an ABI 3130 Genetic Analyser capillary electrophoresis system. Results were analysed with the Sequencing analysis software v.5.2 (Life Technologies).

3. Results and discussion

3.1. New sequence variant at DXS10148

The allele nomenclature used for allele assignment of DXS10148 was according to the study by Hundertmark et al. [15], which is the one used in the Argus X-12 manual [3]. The repeat structure is defined as $(GGAA)_4$-$(AAGA)_X$-$(AAAG)_Y$-N_8-$(AAGG)_2$. All tetrameric repeat blocks are considered for allele assignment, for example, an allele 22 has the following repeat motif: $(GGAA)_4$-$(AAGA)_{12}$-$(AAAG)_4$-N_8-$(AAGG)_2$.

During sequencing of samples with known genotypes, such as control DNA 9948 (allele 23) and two rare alleles from West Africa (alleles 29 and 32), a minus-two-repeat difference was detected when comparing genotypes to sequenced data (Table 1). For reference DNA 9948, sequencing results revealed the following allele structure $(GGAA)_4$-$(AAGA)_{11}$-$(AAAG)_4$-N_8-$(AAGG)_2$ which, according to the proposed nomenclature, corresponds to an allele 21 and not an allele 23 (Table 1). The same was detected in the sequences of the two West African samples: the alleles genotyped as 29 and 32 revealed DNA repeat sequences for alleles 27 and 30, respectively (Table 1). When analysing the flanking regions of the main repeat STR core of these samples an additional eight bases (AAGG-AAAG) were identified in all three samples (at the fourth base downstream from the repeat motif (Table 1 and Fig. S1)) when compared to the sequence in Hundertmark et al. [15]. Since DXS10148 displays a tetranucleotide repeat motif that adds eight bases to the fragment size of the STR, this means that two repeats are added to these alleles [samples with $(AAGG\text{-}AAAG)_3$ instead of $(AAGG\text{-}AAAG)_2$]. In an Algerian population study [16] it seems that in some alleles (e.g., 14, 23, 28 and 29) the sequenced fragment also does not correspond to the genotyped allele. However, without direct sequence comparisons it is not clear if the same variation is responsible for this difference.

Additional sequencing data generated and analysed for DXS10148 from two samples from Iraq (allele 25.2) revealed the allele repeat motifs that were expected based on the genotypes obtained by comparison with the Argus X-12 allelic ladder (Table 1). In the Iraqi samples the motif (AAGG-AAAG) was found in duplicate (Table 1, Fig. S2) and not in triplicate as in the West African samples and therefore no differences were detected between fragment sizes and sequenced alleles.

A subset of the West African samples from Guinea-Bissau was sequenced to establish the genetic basis for detecting null (or silent) alleles, hence without previously known genotypes. 62 out of the 63 samples sequenced from Guinea-Bissau displayed two copies of the sequence motif described above (as an example, see alleles 39.1 and 41.1 in Table 1). In one sample the presence of a single copy of the AAGG-AAAG motif was detected adding variation to this site (allele 41.1 in Table 1; Fig. S3).

The null alleles sequenced at DXS10148 displayed the same mutation in all cases: a G → A transversion in the 9th nucleotide counting from the beginning of the repeat. This position corresponds to the 2nd base of the 3′ end of the forward primer sequence published in Hundertmark et al. [15]. This mutation is most likely responsible for the silent alleles found at DXS10148 which also changes the structure of the repeat (see alleles 37.1, 38.1, 39.1 and 41.1 as examples, Table 1). The detection of null alleles at this locus as well as the single base mutation observed here has also been described previously in other studies (e.g., 17).

The results obtained in this study allowed the identification of the variable motif AAGG-AAAG adjacent to the core repeat region. This finding adds further variation to the relevant repeat region, with an important impact on the allele nomenclature as the allele designation based on fragment sizes does not always match the sequence-based allele structure. This is important in studies where a high frequency of silent alleles is present at this locus mostly seen

Table 1
Sequence structure variation observed for **DXS10148**.

Reference repeat structure described by Hundertmark et al. [15]:
$(GGAA)_4$-$(AAGA)_X$-$(AAAG)_Y$-N_8-$(AAGG)$ 2 *AAAG*(**AAGG-AAAG)$_2$** *GGAAA*

Allele structure variation observed in this study:
$(GGAA)_{2,4}$-$(AGAA\text{-}GGAA)_{0-1}$-$(AAGA)_X(AAAG)_{0,Y}(AAAG\text{-}A)_{0-2}(AAAG)_4$-$N_8$-$(AAGG)$ 2 AAAG **$(AAGG\text{-}AAAG)_{1-3}$** *GGAAA*

Al. gen.	Al. seq.	*n*	Al. type (pop)	Individual allele structures detected:
23	21	1	Ref. DNA 9948	$(GGAA)_4$-$(AAGA)_{11}$-$(AAAG)_4$-N_8-$(AAGG)_2$-*AAAG*-**$(AAGG\text{-}AAAG)_3$**
29	27	1	Rare (WAF)	$(GGAA)_4$-$(AAGA)_{17}$-$(AAAG)_4$-N_8-$(AAGG)_2$-*AAAG*-**$(AAGG\text{-}AAAG)_3$**
32	30	1	Rare (WAF)	$(GGAA)_4$-$(AAGA)_{20}$-$(AAAG)_4$-N_8-$(AAGG)_2$-*AAAG*-**$(AAGG\text{-}AAAG)_3$**
–	37.1	1	Null (WAF)	$(GGAA)_2$-**A**GAA-GGAA-$(AAGA)_{15}$-$(AAAG)_{12}$-A-$(AAAG)_4$-N_8-$(AAGG)_2$-*AAAG*-**$(AAGG\text{-}AAAG)_3$**
–	38.1	1	Null (WAF)	$(GGAA)_2$-**A**GAA-GGAA-$(AAGA)_{15}$-$(AAAG)_{13}$-A-$(AAAG)_4$-N_8-$(AAGG)_2$-*AAAG*-**$(AAGG\text{-}AAAG)_3$**
25.2	25.2	2	Rare (IRQ)	$(GGAA)_4$-$(AAGA)_{13}$-$(AAAG)$-A-$(AAAG)$-A-$(AAAG)_4$-N_8-$(AAGG)_2$-*AAAG*-**$(AAGG\text{-}AAAG)_2$**
–	39.1	2	Null (WAF)	$(GGAA)_2$-**A**GAA-GGAA-$(AAGA)_{15}$-$(AAAG)_{14}$-A-$(AAAG)_4$-N_8-$(AAGG)_2$-*AAAG*-**$(AAGG\text{-}AAAG)_2$**
–	41.1	1	Null (WAF)	$(GGAA)_2$-**A**GAA-GGAA-$(AAGA)_{16}$-$(AAAG)_{15}$-A-$(AAAG)_4$-N_8-$(AAGG)_2$-*AAAG*-**$(AAGG\text{-}AAAG)_1$**

Italic sequence (*AAAG*) in structures represent the immediate downstream flanking region of the STR. Bold sequences in allele structures represent the new variant detected. N_8 = AAGGAAAG. The bold and underlined **A** in alleles 37.1, 38.1, 39.1 and 41.1 represents the mutation (G → A transversion) most likely responsible for the allele drop-out at DXS10148. The non-bold, underlined A represents the mutation responsible for the intermediate allele types. Al. gen. = allele genotyped by capillary electrophoresis fragment size analysis; Al. seq. = allele sequenced; *n* = number observed; Al. type (pop) = type of allele and population sequenced in; WAF = West Africa; IRQ = Iraq.