



ELSEVIER

Contents lists available at [ScienceDirect](http://www.sciencedirect.com)

## Forensic Science International: Genetics

journal homepage: [www.elsevier.com/locate/fsig](http://www.elsevier.com/locate/fsig)

## Short communication

## New sequence variants detected at DXS10148, DXS10074 and DXS10134 loci

Iva Gomes<sup>a,\*</sup>, António Brehm<sup>b</sup>, Leonor Gusmão<sup>c,d,e</sup>, Peter M. Schneider<sup>a</sup><sup>a</sup> Institute of Legal Medicine, Faculty of Medicine, University of Cologne, Cologne, Germany<sup>b</sup> Human Genetics Laboratory, University of Madeira, Funchal, Portugal<sup>c</sup> DNA Diagnostic Laboratory (LDD), Institute of Biology, State University of Rio de Janeiro (UERJ), Brazil<sup>d</sup> IPATIMUP, Institute of Pathology and Molecular Immunology of the University of Porto, Porto, Portugal<sup>e</sup> 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-AAAG)<sub>1–3</sub> motif, in the present work. One additional string when compared to the common one (AAGG-AAAG)<sub>2</sub> 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.

© 2015 Elsevier Ireland Ltd. All rights reserved.

## 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](mailto:iva.gomes@uk-koeln.de) (I. Gomes).

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  $(AAGG-AAAG)_3$  instead of  $(AAGG-AAAG)_2$ ]. 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) <sub>2</sub> GGAAA				
Allele structure variation observed in this study: $(GGAA)_{2,4}-(AGAA-GGAA)_{0-1}-(AAGA)_x-(AAAG)_{0,y}-(AAAG-A)_{0-2}-(AAAG)_4-N_8-(AAGG)_2$ AAAG (AAGG-AAAG) <sub>1-3</sub> GGAAA				
Al. gen.	Al. seq.	n	Al. type (pop)	Individual allele structures detected:
23	21	1	Ref. DNA 9948	$(GGAA)_4-(AAGA)_{11}-(AAAG)_4-N_8-(AAGG)_2-AAAG-(AAGG-AAAG)_3$
29	27	1	Rare (WAF)	$(GGAA)_4-(AAGA)_{17}-(AAAG)_4-N_8-(AAGG)_2-AAAG-(AAGG-AAAG)_3$
32	30	1	Rare (WAF)	$(GGAA)_4-(AAGA)_{20}-(AAAG)_4-N_8-(AAGG)_2-AAAG-(AAGG-AAAG)_3$
–	37.1	1	Null (WAF)	$(GGAA)_2-AGAA-GGAA-(AAGA)_{15}-(AAAG)_{12}-A-(AAAG)_4-N_8-(AAGG)_2-AAAG-(AAGG-AAAG)_3$
–	38.1	1	Null (WAF)	$(GGAA)_2-AGAA-GGAA-(AAGA)_{15}-(AAAG)_{13}-A-(AAAG)_4-N_8-(AAGG)_2-AAAG-(AAGG-AAAG)_3$
25.2	25.2	2	Rare (IRQ)	$(GGAA)_4-(AAGA)_{13}-(AAAG)-A-(AAAG)-A-(AAAG)_4-N_8-(AAGG)_2-AAAG-(AAGG-AAAG)_2$
–	39.1	2	Null (WAF)	$(GGAA)_2-AGAA-GGAA-(AAGA)_{15}-(AAAG)_{14}-A-(AAAG)_4-N_8-(AAGG)_2-AAAG-(AAGG-AAAG)_2$
–	41.1	1	Null (WAF)	$(GGAA)_2-AGAA-GGAA-(AAGA)_{16}-(AAAG)_{15}-A-(AAAG)_4-N_8-(AAGG)_2-AAAG-(AAGG-AAAG)_1$

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.

in African populations [16–19], and requires sequencing to identify the genotype. For example, in the case of the West African samples, the two null alleles were sequenced as 37.1 and 38.1 and therefore the genotypes of these samples would be considered as such. However, if the genotypes were obtained by fragment length sizing (if no drop-out had occurred) the corresponding alleles would have been 39.1 and 40.1, respectively. This is because eight additional bases are present in these samples when compared to the reference structure [15] causing a shift of plus 2 repeats (Table 1). The major impact of this is the deviation in the allele frequency distribution that can be introduced in these populations, and consequently during the genetic distance comparisons with other population groups. Furthermore, when analysing the repeat motif sequence of DXS10148, the last AAGG of the repeat (considering the reference nomenclature in Hundertmark et al. [15]) followed by the first AAAG of the flanking sequence should also be included together with the detected variant as it is the same one and considered for the allele designation (Table 1). This means that an additional AAGG-AAAG should be added to the already detected one in this study and become (AAGG-AAAG)<sub>2-4</sub>.

### 3.2. New sequence variant at DXS10074

DXS10074 has been defined with the simple repeat structure (AAGA)<sub>n</sub> in the Argus X-12 manual. The given reference allele is an allele 14 displaying fourteen straight repeats as (AAGA)<sub>14</sub>. In the study by Hering et al. [20] two types of allele structures have been reported for DXS10074. In this latter work, short alleles are defined by containing the repeat block (AAGA)<sub>n</sub> in which *n* varies between 7 and 10 as (ACAC)<sub>2</sub>-(AGAG)<sub>4</sub>-AA-AAAG-(AAGA)<sub>7-10</sub>-AAGG-AAGA (sequences represented in italic are not included for allele designation). Longer alleles have been characterized by an additional 12 bases underlined in the following structure: (ACAC)<sub>2</sub>-(AGAG)<sub>4</sub>-AA-AAAG-(AAGA)<sub>10-18</sub>-AAGG-(AAGA)<sub>2</sub>-AAGG-AAGA. Due to this insertion the authors proposed that in these types the allele count should be *n*+3. According to this repeat motif definition [20], an allele 14 has the following structure: (AAGA)<sub>11</sub>-AAGG-(AAGA)<sub>2</sub> which is different than the simple (AAGA)<sub>14</sub> reported in the Argus X-12 kit [3]. Either a different motif is being considered or these are just two different structures for the same allele.

Six samples from West-Africa (Guinea-Bissau) were chosen for sequencing due to the detection of new (previously not described) alleles at this locus (Table 2). Interestingly, all of these alleles were non-consensus types ranging from 11.3 to 15.3 (except for 13.3 which was not observed) as well as a 13.2 allele. Results revealed that the sequence variations responsible for the

intermediate alleles (with exception of the 13.2 allele) were all detected in the main repeat motif AAGA (Table 2), but affecting different nucleotide positions, as either an A or a G has been lost, interrupting the core tandem repeat motif. On the other hand, the intermediate allele 13.2 revealed two additional nucleotides (AG) located outside of the core repeat in the upstream flanking region (Table 2). This region is composed of two stretches of dinucleotide repeats AC and AG motifs, respectively, reported as tetranucleotides [(ACAC)<sub>2</sub>-(AGAG)<sub>4</sub>] [20]. We prefer to consider these as dinucleotides repeats as variation at the dinucleotide level has been detected in this study in the following way: (AC)<sub>3-4</sub> and (AG)<sub>8-9</sub> (Table 2). Three of the samples plus control DNA 9948 have an (AC)<sub>4</sub>-(AG)<sub>8</sub> composition. This seems to also be the same for the 30 PCR amplicons sequenced by Hering et al. [20], as no variation was reported at this level. It appears that this is the most common composition of the dinucleotide stretches, at least in Europeans where more sequence data is available for comparison. For the additional samples, two samples had an (AC)<sub>3</sub>-(AG)<sub>9</sub> composition which means the loss of one AC dinucleotide is compensated by one additional AG, thus not changing the total amplicon size (Table 2). However, another allele displayed an (AC)<sub>4</sub>-(AG)<sub>9</sub> combination where the additional AG is responsible for the intermediate allele 13.2 as already described above.

The variation detected in the two dinucleotide AC and AG stretches at DXS10074 has not been considered for the allele designation in the originally recommended nomenclature [20]. This variable DNA region has also impact on the allele nomenclature as in some cases the intermediate x.2 alleles might be due to this variation such as the detected 13.2 allele in this work. Several intermediate alleles of this type have already been detected in other studies for DXS10074, e.g., in a Polish population study a 15.2 allele was described [21], a 16.2 allele in Algeria and in the Ivory Coast [16,18] or 14.2 and 19.2 in US populations [19]. Variation could possibly be occurring in the dinucleotide repeats; however, as no sequence data is available for these intermediate alleles in the mentioned studies this cannot be confirmed.

### 3.3. New sequence variant at DXS10134

The X-chromosome STR DXS10134 presents a highly complex allele sequence structure [3,22]: (GAAA)<sub>3</sub>-GAGA-(GAAA)<sub>4</sub>-AA-(GAAA)-GAGA-(GAAA)<sub>4</sub>-GAGA-(GACAGA)<sub>3</sub>-(GAAA)GTAA-(GAAA)<sub>3</sub>-AAA-[(GAAA)<sub>4</sub>-AAA]<sub>1-2</sub>-(GAAA)<sub>n</sub>. In this proposed sequence, sixteen GAAA repeats located adjacent to the main repetitive core (represented in italic) contribute to the allele designation. These 16 invariable repeats are added to the variable 3'-located GAAA repeats. To confirm doubts in the observed rare genotypes of two

**Table 2**  
Sequence structure variation observed at DXS10074 detected in the Guinea-Bissau population.

Reference repeat structures described by Hering et al. [20]:				
Short alleles: (ACAC) <sub>2</sub> -(AGAG) <sub>4</sub> -N <sub>6</sub> -(AAGA) <sub>7-10</sub> -AAGG-AAGA				
Long alleles: (ACAC) <sub>2</sub> -(AGAG) <sub>4</sub> -N <sub>6</sub> -(AAGA) <sub>10-18</sub> -AAGG-(AAGA) <sub>2</sub> -AAGG-AAGA				
Allele structure variation observed in this study:				
(AC) <sub>3-4</sub> -(AG) <sub>8-9</sub> -AAAAAG-(AAA) <sub>0-1</sub> -(AAGA) <sub>2</sub> -(AGA) <sub>0-1</sub> -(AAGA) <sub>n</sub> -(AAGG-(AAGA) <sub>2</sub> ) <sub>0-1</sub> -AAGG-AAGA				
Al. gen.	Al. seq.	<i>n</i>	Al. type	Individual allele structures detected:
18	18	1	Ref. DNA 9948	(AC) <sub>4</sub> -(AG) <sub>8</sub> -N <sub>6</sub> -(AAGA) <sub>15</sub> -(AAGG)-(AAGA) <sub>2</sub> -AAGG-AAGA
11.3	11.3	2	New	(AC) <sub>4</sub> -(AG) <sub>8</sub> -N <sub>6</sub> -(AAGA) <sub>2</sub> -AGA-(AAGA) <sub>9</sub> -AAGG-AAGA
12.3	12.3	1	New	(AC) <sub>4</sub> -(AG) <sub>8</sub> -N <sub>6</sub> -(AAGA) <sub>2</sub> -AGA-(AAGA) <sub>10</sub> -AAGG-AAGA
13.2	13.2	1	New	(AC) <sub>4</sub> -(AG) <sub>9</sub> -N <sub>6</sub> -(AAGA) <sub>13</sub> -AAGG-AAGA
14.3	14.3	1	New	(AC) <sub>3</sub> -(AG) <sub>9</sub> -N <sub>6</sub> -AAA-(AAGA) <sub>14</sub> -AAGG-AAGA
15.3	15.3	1	New	(AC) <sub>3</sub> -(AG) <sub>9</sub> -N <sub>6</sub> -AAA-(AAGA) <sub>15</sub> -AAGG-AAGA

Italic regions in reference repeat structures have not been considered for allele assignment. Bold sequences in individual allele structures represent the variable regions detected. Underlined sequences represent the sequence sites responsible for the intermediate allele types.

N<sub>6</sub>: AAAAAG; Al. gen. = allele genotyped by capillary electrophoresis fragment size analysis; Al. seq. = allele sequenced; Al. type = type of allele.

**Table 3**  
Sequence structure variation observed for **DXS10134**.

Reference repeat structures described by Edlmann et al. [22]: (GAAA) <sub>3</sub> -GAGA-(GAAA) <sub>4</sub> -AA-(GAAA)-GAGA-(GAAA) <sub>4</sub> -GAGA-(GACAGA) <sub>3</sub> -(GAAA)-GTAA-(GAAA) <sub>3</sub> -AAA-[(GAAA) <sub>4</sub> -AAA] <sub>1-2</sub> -(GAAA) <sub>n</sub>				
Allele structure variation observed in this study: (GAAA) <sub>3</sub> -GAGA-(GAAA) <sub>4</sub> -AA-(GAAA)-GAGA-(GAAA) <sub>3-4</sub> -[GAGA-(GACAGA) <sub>3</sub> -(GAAA)-GTAA-(GAAA) <sub>3</sub> -AAA-[(GAAA) <sub>4</sub> -AAA] <sub>1,3</sub> ] <sub>0-1</sub> -(GAAA) <sub>5</sub> (A) <sub>0-1</sub> (GAAA) <sub>n</sub>				
Al. gen.	Al. seq.	n	Al. type	Individual allele structures detected:
34	34	1	9948	(GAAA) <sub>3</sub> -GAGA-(GAAA) <sub>4</sub> -AA-(GAAA)-GAGA-(GAAA) <sub>4</sub> -GAGA-(GACAGA) <sub>3</sub> -(GAAA)-GTAA-(GAAA) <sub>3</sub> -AAA-[(GAAA) <sub>4</sub> -AAA] <sub>1</sub> -(GAAA) <sub>14</sub>
35	35	1	Rare (WAF)	(GAAA) <sub>3</sub> -GAGA-(GAAA) <sub>4</sub> -AA-(GAAA)-GAGA-(GAAA) <sub>4</sub> -GAGA-(GACAGA) <sub>3</sub> -(GAAA)-GTAA-(GAAA) <sub>3</sub> -AAA-[(GAAA) <sub>4</sub> -AAA] <sub>1</sub> -(GAAA) <sub>15</sub>
43.2	43.2	1	Rare (WAF)	(GAAA) <sub>3</sub> -GAGA-(GAAA) <sub>4</sub> -AA-(GAAA)-GAGA-(GAAA) <u>3</u> -GAGA-(GACAGA) <sub>3</sub> -(GAAA)-GTAA-(GAAA) <sub>3</sub> -AAA-[(GAAA) <u>4</u> -AAA] <u>3</u> -(GAAA) <sub>15</sub>
39.1	39.1	1	New (WAF)	(GAAA) <sub>3</sub> -GAGA-(GAAA) <sub>4</sub> -AA-(GAAA)-GAGA-(GAAA) <sub>4</sub> -GAGA-(GACAGA) <sub>3</sub> -(GAAA)-GTAA-(GAAA) <sub>3</sub> -AAA-[(GAAA) <sub>4</sub> -AAA] <sub>1</sub> -(GAAA) <u>5</u> -A-(GAAA) <sub>14</sub>
15	23	5	New (WAF)	(GAAA) <sub>3</sub> -GAGA-(GAAA) <sub>4</sub> -AA-(GAAA)-GAGA-(GAAA) <sub>15</sub>
16	24	1	New (WAF)	(GAAA) <sub>3</sub> -GAGA-(GAAA) <sub>4</sub> -AA-(GAAA)-GAGA-(GAAA) <sub>16</sub>

Bold, underlined sequences represent the variable regions observed in this study. Only the GAAA repeat motifs are considered for allele assignment [15].

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.

West African individuals (exhibiting alleles 35 and 43.2) sequencing was performed. Results for these samples showed the same type of allele structure compared to the above mentioned one and confirmed the respective genotypes (Table 3). Allele 43.2 showed two additional DNA sequence variations: one GAAA is missing in the adjacent sixteen GAAA repeat region and a total of three [(GAAA)<sub>4</sub>-AAA] structures are present instead of only one.

In six West African male individuals from Guinea-Bissau, off ladder alleles with very short fragment sizes were observed at DXS10134 that overlapped with the bin set of the adjacent smaller locus DXS7132 of the Argus X-12 kit (Fig. S4). The smallest allele in the Argus Investigator X-12 allelic ladder for DXS10134 has 28 repeats and the newly detected alleles based on the fragment sizes were estimated as alleles 15 and 16. To our knowledge these alleles have not been reported so far for this locus. To confirm these short fragment genotypes, all six samples were sequenced (Table 3). These alleles revealed a much simpler allele structure (GAAA)<sub>3</sub>-GAGA-(GAAA)<sub>4</sub>-AA-(GAAA)-GAGA-(GAAA)<sub>15-16</sub> compared to the highly complex one described above [3,22]. The allele sequence starts with the same structure but the subsequent tract [GAGA-(GACAGA)<sub>3</sub>-(GAAA)-GTAA-(GAAA)<sub>3</sub>-AAA-(GAAA)<sub>4</sub>-AAA] is missing in these samples (accounting for a loss of eight GAAA repeats when compared to the reference structure [22]). Thus according to the suggested allele nomenclature alleles 23 and 24 are assigned by sequencing data instead of alleles 15 and 16, respectively (Table 3). In addition, a new 39.1 allele, so far unreported, was found in the West African sample set from Guinea-Bissau (Table 3). This allele revealed the same complex sequence structure as the initially proposed one. Reference DNA 9948 was also sequenced for DXS10134 fitting as well into the reference allele nomenclature (Table 3).

In the population samples from Guinea-Bissau a considerably simpler structure was detected in for DXS10134 for much shorter amplicons than those observed in the allelic ladder of the Investigator Argus X-12 kit. Such short alleles have not been described elsewhere for DXS10134 not even in the few African groups that have been analyzed previously for this system [16–19,23–24]. Possibly this structure is population specific and confined to the region of Guinea-Bissau, since we did not find these short alleles in other West African populations. However, much more sequence data from other populations would be needed to support this hypothesis.

#### 4. Conclusions

Based on the data obtained in this work a more complex allele structure [(GGAA)<sub>4</sub>-(AAGA)<sub>x</sub>-(AAAG)<sub>y</sub>-N<sub>8</sub>-AAGG-(AAGG-AAAG)<sub>z</sub>] than previously described is present at DXS10148. The same was observed for DXS10074 that harbors two varying dinucleotide repeat stretches [(AC)<sub>x</sub>-(AG)<sub>y</sub>-N<sub>6</sub>-(AAGA)<sub>z</sub>] that were not included in the allele nomenclature.

Nevertheless, and despite the importance of these findings, according to the ISFG guidelines [4,6,7], if a previously established nomenclature of an STR is not in accordance with the ISFG recommendations but has been widely utilized, the nomenclature should not be altered to avoid unnecessary confusion. This is the case for DXS10148 and DXS10074 which, due to its inclusion in the only available X-STR commercial kit in the market, has been already widely used for population and forensic genetic applications. Recently, the manufacturer of this kit has critically revised the primers to improve the detection of previously observed null alleles which are common in the African population. Also, reference has been made in the manual to raise awareness regarding the unusual repeat structure of DXS10148.

The present findings demonstrate the importance of screening individuals from different world-wide populations before developing proposals for new STR nomenclatures. The ISFG DNA commission also strongly recommends [4,7] the screening of diverse population groups as well as the comparison with the *Pan troglodytes* genome sequence when establishing a new STR nomenclature since it is crucial to identifying regions that may vary as demonstrated by other studies [8–11]. The chimpanzee genome has been completely sequenced and is accessible using sequence similarity search tools, e.g., the online programs BLAST ([www.ncbi.nlm.nih.gov/blast/Blast.cgi](http://www.ncbi.nlm.nih.gov/blast/Blast.cgi)) and BLAT ([www.genome.ucsc.edu](http://www.genome.ucsc.edu)) thus replacing the need for reference DNA samples from this species.

#### Acknowledgements

The DNA sequencing work performed by Oscar Wienholt, Sonja Harms and Plácido Pereira is gratefully acknowledged. IG has received financial support from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 285487 (EUROFORGEN-NoE).

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.fsigen.2015.10.005>.

## References

- [1] L. Gusmão, P. Sánchez-Diz, C. Alves, I. Gomes, et al., 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.* 123 (3) (2009) 227–234.
- [2] I. Gomes, M. Prinz, R. Pereira, C. Meyers, R.S. Mikulasovich, A. Amorim, A. Carracedo, L. Gusmão, Genetic analysis of three US population groups using an X-chromosomal STR decaplex, *Int. J. Legal Med.* 121 (3) (2007) 198–203.
- [3] Investigator<sup>®</sup> Argus X-12 Handbook, June 2013, Qiagen, Hilden, Germany.
- [4] W. Bär, B. Brinkmann, B. Budowle, A. Carracedo, et al., DNA recommendations: further report of the DNA commission of the ISFH regarding the use of short tandem repeat systems, *Int. J. Legal Med.* 110 (1997) 175–176.
- [5] P. Gill, B. Brinkmann, E. d'Aloja, J. Andersen, et al., Considerations from the European DNA profiling group (EDNAP) concerning STR nomenclature, *Forensic Sci. Int.* 87 (1997) 185–192.
- [6] P.M. Schneider, Scientific standards for studies in forensic genetics, *Forensic Sci. Int.* 165 (2007) 238–243.
- [7] L. Gusmão, J.M. Butler, A. Carracedo, P. Gill, et al., DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis, *Forensic Sci. Int.* 157 (2–3) (2006) 187–197.
- [8] I. Gomes, M. Prinz, R. Pereira, E. Bieschke, W.R. Mayr, A. Amorim, A. Carracedo, L. Gusmão, X-chromosome STR sequence variation, repeat structure, and nomenclature in humans and chimpanzees, *Int. J. Legal Med.* 123 (2) (2009) 143–149.
- [9] I. Gomes, R. Pereira, W.R. Mayr, A. Amorim, A. Carracedo, L. Gusmão, Evaluation of DXS9902, DXS7132, DXS6809, DXS7133, and DXS7423 in humans and chimpanzees: sequence variation, repeat structure, and nomenclature, *Int. J. Legal Med.* 123 (5) (2009) 403–412.
- [10] L. Gusmão, A. González-Neira, C. Alves, M. Lareu, S. Costa, A. Amorim, A. Carracedo, Chimpanzee homologous of human Y specific STRs. A comparative study and a proposal for nomenclature, *Forensic Sci. Int.* 126 (2002) 129.
- [11] K. Lazaruk, J. Wallin, C. Holt, T. Nguyen, P.S. Walsh, Sequence variation in humans and other primates at six short tandem repeat loci used in forensic identity testing, *Forensic Sci. Int.* 119 (2001) 1–10.
- [12] R. Szibor, J. Edelmann, S. Hering, I. Plate, H. Wittig, L. Roewer, P. Wiegand, F. Cali, V. Romano, M. Michael, Cell line DNA typing in forensic genetics—the necessity of reliable standards, *Forensic Sci. Int.* 138 (2003) 37–43.
- [13] A. Untergrasser, I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm, S.G. Rozen, Primer3—new capabilities and interfaces, *Nucl. Acids Res.* 40 (15) (2012) e115.
- [14] D.J. Korbie, J.S. Mattick, Touchdown PCR for increased specificity and sensitivity in PCR amplification, *Nat. Protoc.* 3 (9) (2008) 1452–1456.
- [15] T. Hundertmark, S. Hering, J. Edelmann, C. Augustin, I. Plate, R. Szibor, The STR cluster DXS10148-DXS8378-DXS10135 provides a powerful tool for X-chromosomal haplotyping at Xp22, *Int. J. Legal Med.* 122 (6) (2008) 89–92.
- [16] A. Bekada, S. Benhamamouch, A. Boudjema, M. Fodil, S. Menegon, C. Torre, C. Robino, Analysis of 21 X-chromosomal STRs in an Algerian population sample, *Int. J. Legal Med.* 124 (4) (2010) 87–294.
- [17] C. Tomas, V. Pereira, N. Morling, Analysis of 12 X-STRs in Greenlanders, Danes and Somalis using Argus X-12, *Int. J. Legal Med.* 126 (2012) 121–128.
- [18] S. Pasino, S. Caratti, P. Del, M. ero, A. Santovito, C. Torre, C. Robino, Allele and haplotype diversity of X-chromosomal STRs in Ivory Coast, *Int. J. Legal Med.* 125 (5) (2011) 749–752.
- [19] T. Diegoli, A. Linacre, P.M. Vallonec, J.M. Butler, M.D. Coble, Allele frequency distribution of twelve X-chromosomal short tandem repeat markers in four U. S. population groups, *Forensic Sci. Int. Genet. Suppl. Ser.* (2011) e481–e483.
- [20] S. Hering, C. Augustin, J. Edelmann, M. Heidele, J. Dressler, H. Rodig, E. Kuhlisch, R. Szibor, DXS10079, DXS10074 and DXS10075 are STRs located within a 280-kb region of Xq12 and provide stable haplotypes useful for complex kinship cases, *Int. J. Legal Med.* 120 (6) (2006) 337–345.
- [21] M. Czarnogórska, M. Sanak, D. Piniewska, N. Polańska, A. Stawowiak, B. Opolska-Bogusz, Identification of rare genetic variants at DXS10074, DXS10079, DXS10146 and DXS10148 loci of investigator Argus X-12 multiplex in the south Polish population, *Arch. Med. Sadowej Kryminol.* 60 (4) (2010) 35–242.
- [22] J. Edelmann, S. Hering, C. Augustin, R. Szibor, Characterisation of the STR markers DXS10146, DXS10134 and DXS10147 located within a 79.1 kb region at Xq28, *Forensic Sci. Int. Genet.* 2 (1) (2008) 41–46.
- [23] S. Elakkary, S. Hoffmeister-Ullrich, C. Schulze, E. Seif, A. Sheta, S. Hering, J. Edelmann, C. Augustin, Genetic polymorphisms of twelve X-STRs of the investigator Argus X-12 kit and additional six X-STR centromere region loci in an Egyptian population sample, *Forensic Sci. Int. Genet.* (2014) 26–30.
- [24] K. Thiele, S. Löffler, J. Löffler, F. Günthner, K. Nitschke, J. Edelmann, R. Lessig, Population data of eight X-chromosomal STR markers in Ewe individuals from Ghana, *Forensic Sci. Int. Genet. Suppl. Ser.* (2008) 167–169.