Forensic Science International: Genetics 31 (2017) 89–94

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Short communication

# Genetic characterization of Guinea-Bissau using a 12 X-chromosomal STR system: Inferences from a multiethnic population

Iva Gomes[a], Plácido J.P. Pereira[b], Sonja Harms[a], Andréa M. Oliveira[c], Peter M. Schneider[a], António Brehm[b,*]

[a] Institute of Legal Medicine, Faculty of Medicine, University of Cologne, Cologne, Germany
[b] Human Genetics Laboratory, University of Madeira, Campus of Penteada, 9020-105 Funchal, Portugal
[c] Genetic Department, Institute of Biology, State University of Rio de Janeiro (UERJ), Brazil

ARTICLE INFO

*Keywords:*
Guinea-Bissau
West Africa
X- STRs
Investigator Argus X-12
Forensic genetics

ABSTRACT

A male West African sample from Guinea-Bissau (West-African coast) was genetically analyzed using 12 X chromosomal short tandem repeats that are grouped into four haplotype groups. Linkage disequilibrium was tested ($p \leq 0.0008$) and association was detected for the majority of markers in three out of the four studied haplotype clusters. The sample of 332 unrelated individuals analyzed in this study belonged to several recognized ethnic groups ($n = 18$) which were used to evaluate the genetic variation of Guinea-Bissau's population. Pairwise genetic distances ($F_{ST}$) did not reveal significant differences among the majority of groups. An additional 110 samples from other countries also belonging to West Africa were as well compared with the sample of Guinea-Bissau. No significant differences were found between these two groups of West African individuals, supporting the genetic homogeneity of this region on the X chromosome level. The generation of over 100 DNA West African sequences provided new insights into the repeat sequence structure of some of the present X-STRs. Parameters for forensic evaluation were also calculated for each X-STR, supporting the potential application of these markers in typical kinship scenarios. Also, the high power of discrimination values for samples of female and male origin observed in this study, confirms the usefulness of the present X-STRs in identification analysis.

## 1. Introduction

The X chromosome has distinct genetic features when compared to the autosomes and to the Y-chromosome. X chromosome STRs are hemizygous in males, thus behaving as a haplotype such as the Y chromosome, but in females they resemble autosomes due to recombination. Owing to the way in which it is inherited, the X chromosome is a rich resource of easily accessible genetic data, and therefore provides a unique tool for population-genetic studies [1]. X-STRs seem to be more effective than autosomal markers in the identification of female traces in male contamination [2]. In several other occasions, like case-studies involving identification of skeletons or victims of mass disasters where only remote relatives are available and females play a key role in identification, X-STRs can also be more efficient than autosomal ones [3]. The same is true in deficiency paternity cases, for example, relationship can be excluded when comparing potential half-sisters, even without the presence of both parents.

Four linkage groups of X chromosome markers are recognized in the commonly used commercial X-STR kit, the Investigator Argus X-12 kit [4]: linkage group 1: DXS10148, DXS10135 and DXS8378; linkage group 2: DXS7132, DXS10079 and DXS10074; linkage group 3: DXS10103, HPRTB and DXS10101; linkage group 4: DXS10146, DXS10134 and DXS7423. Each group of markers can be treated as a haplotype which may provide useful information for human identification, especially for paternity or other kinship testing [3]. On the other hand, it has been shown that STRs are also useful markers to infer ethnicity although the degree of correct assignment varies considerably depending on the degree of historical and ethnical relationships [5].

Guinea-Bissau is a very small western-African country, about 37000 $km^2$, with 1.6 million inhabitants, speaking about 25 different languages, and an equal number of recognized ethnic groups [6–8]. Animists constitute the vast majority of the population (55%) followed by Islamites (40%). Settling started around 40000 years before present [9]. Around 9000 YBP several Neolithic cultures flourished in the region bringing in peoples from both Sub-Sahara and North Africa origin. Around 4000 YBP the region received the income of Bantu peoples,

* Corresponding author.
*E-mail address:* brehm@uma.pt (A. Brehm).

http://dx.doi.org/10.1016/j.fsigen.2017.08.016
Received 4 November 2016; Received in revised form 14 August 2017; Accepted 17 August 2017
Available online 24 August 2017
1872-4973/ 

mostly coming from the Gulf of Guinea area. In the 9th century there is a documented admixture with Berbers following the arrival of the first pastoral Peuls or Fulbe (here designated by Fula). In the 15th century most of the different groups were already in place [8,10–12]. The Balanta constitute the major ethnic group from Guinea-Bissau (30%) and have an oral tradition of being originally from East Africa (Sudan and Ethiopia) migrating to West Africa between the 10–14th centuries. During the 16th century they settled in the territory which is today Guinea-Bissau. Mandinga (Mandenka) is one if not the major ethnic group in West Africa with an estimated population of 11 million. They descend from the Mali Empire (12–13th centuries) and were originally animists later Islamized by the Fula [8,10–12]. They belong to the biggest ethnolinguistic group of West Africa – the Mandé (https://www.ethnologue.com/). This is the main ethnic group from which most of the slaves to the Americas came from. The Fula (Fulbe, Peuls) constitute today the second ethnic group in number (25%). Totally Islamized, they are a heterogeneous group, originally coming from the Sahel region. The Fula Islamized the Mandinga and both pushed all other groups towards the coastal and southern regions of today's Guinea-Bissau. Around the 17th century they already dominated the whole area [11–13]. Manjacos, Mancanhas, Papel and Brame are closely related groups (some even claim that they are just different clans from the same ethnic group) [14]. This animist group accounts for 24% of the population inhabiting coastal areas but is considered autochthonous at least since the 15th century [11]. In spite of the relatively small population, Guinea-Bissau has today a huge collection of ethnic groups, speaking different dialects and languages and most importantly with different religious beliefs and traditions (see Supplementary Figs. S1 and S2 and Supplementary table S1 for a comprehensive review).

X-STR population genetic data is still limited for some populations, in particular of African origin. We therefore aimed to present the results obtained for 12 X-STRs for a male population sample from Guinea-Bissau. The genetic variation of X-STRs complements previous studies done on the same population group using other markers namely HLAs, mtDNA, Y-Chromosome SNPs and STRs, and autosomal STRs [15–19]. Moreover the typing of specific sub-Saharan populations such as the Guinean-Bissau presented here has an obvious forensic interest in other Western countries, such as Portugal, where a significant immigration population exists representing 4% of the total population [20,21] and for which ethnic specific databases can be extremely valuable.

## 2. Materials and methods

### 2.1. Samples, DNA extraction and quantification

Blood samples were obtained from 332 unrelated and healthy male individuals from Guinea-Bissau (West African coast, Senegambia region). All individuals gave informed consent and were submitted to an interview in order to be linked to a given ethnic group (Supplementary table S1). A complete description of the Guinean ethnic groups, languages and linguistic affiliation according to major language groups, as well as religious beliefs and geographical distribution, has been published in detail elsewhere [18] but updated here [see Supplementary Figs. S1, S2]. Only individuals that could unambiguously trace the origin of their parents and ascendants back to three generations as belonging to the same ethnic group were used in the present study. Finally, 18 groups were used for the data analyses (Supplementary table S1). An additional 110 individuals (31 females and 79 males) belonging to other countries from West Africa (namely from Senegal, Gambia, Guinea, Serra Leo, Liberia, Ghana, Togo, Benin, Nigeria and Niger) were also included in this study for population comparisons. Only samples obtained with informed consent for research purposes were selected.

DNA from the Guinea-Bissau samples was extracted from whole blood samples using a traditional salting-out methodology. The other West African samples were extracted using the EZ1 Investigator kit and the EZ1 Advanced automated nucleic acid purification system according to the manufacturer's protocol (Qiagen GmbH, Hilden, Germany). DNA concentrations were estimated using the NanoDrop® 2000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, USA) following the manufacturer's protocol.

### 2.2. DNA amplification and detection of alleles

PCR amplification of X-STR markers was performed using the Investigator® Argus X-12 kit according to the manufacturer's instructions (Qiagen GmbH, Hilden, Germany) [4]. The following twelve X-STRs were amplified: DXS10103, DXS8378, DXS7132, DXS10134, DXS10074, DXS10101, DXS10135, DXS7423, DXS10146, DXS10079, HPRTB and DXS10148. These 12 loci are distributed along the X chromosome and grouped under four linkage groups: linkage group 1 comprises DXS10148, DXS10135 and DXS8378; linkage group 2 comprises DXS7132, DXS10079 and DX10074; linkage group 3 is formed by DXS10103, HPRTB and DXS10101 and finally linkage group 4 by DXS10146, DXS10134 and DXS7423. Fragment length determination and allele designation were performed using the allelic ladder, internal size standard and positive controls included in the Argus X-12 kit. Detection of amplified fragments was performed by standard capillary electrophoresis using an ABI PRISM® 3130 Genetic Analyzer (Life Technologies, Darmstadt, Germany) and analyzed with GeneMapper® ID Software v3.2.

### 2.3. Sequencing of X-STR alleles

A total of 105 samples were directly sequenced by traditional Sanger sequencing due to the identification of silent (null), rare/new or uncertain genotypes of some alleles such as intermediate alleles (Supplementary table S2). Sequences were obtained for the following 6 X-STRs: DXS10101, DXS10148, HPRTB, DXS10079, DXS10146 and DXS10135 (Supplementary table S2). All details regarding the sequencing method and conditions used were the same as the ones described in the sequencing study of Gomes et al. [22]. In this previous study [22], sequencing primers used for DXS10148 were the same used in the present work. Sequencing primers for the other 5 loci (DXS10101, DXS10135, DXS10146, DXS10079 and HPRTB) were either redesigned or chosen form the literature if suitable. Descriptions of all primers used for the sequencing reactions as well as references of nomenclatures used for allele assignment are listed in Supplementary table S2.

### 2.4. Statistical analysis

Basic genetic parameters such as allele and haplotype frequencies, gene diversities (same as the power of discrimination in males; $PD_M$) and linkage disequilibrium (LD) testing between loci were determined using the Arlequin 3.5.2.2 software [23]. In addition, pairwise genetic distances ($F_{ST}$) were also calculated among the Guinean-Bissau ethnic groups as well as with an additional population sample from West Africa using the same software tool. $F_{ST}$ is based on the number of different alleles and was chosen as an alternative to $R_{ST}$ which takes into account the "stepwise mutation model" behind the formation of STRs. The main reason is that the majority of X-STRs here studied have many intermediate alleles and at high frequencies. Non-consensus alleles likely originate from an insertion/deletion rather than a single step mutation and therefore $R_{ST}$ is not appropriate. Forensic efficiency parameters such as expected probability of exclusion in trios with female offspring ($PE_T$); expected probability of exclusion in father/daughter duos ($PE_D$); power of discrimination in females ($PD_F$) and power of discrimination in males ($PD_M$) were calculated using the formulae according to Desmarais et al. [24]. In addition, the expected probability of exclusion in two half-sisters ($PE_{HS}$) has also been calculated according to the formula in Gomes et al. [25].

## 3. Results and discussion

### 3.1. Genetic variation

In the present work, a population of 332 male individuals from the Western African country of Guinea-Bissau was studied that can be subdivided into 18 ethnic groups. From a forensic genetic point of view it is important to study the genetic affinity among these groups that compose Guinea-Bissau in order to infer about the use of local or global databases for human identification cases. On the other hand, from a population genetic perspective, it is interesting to investigate the genetic relationships of these groups and if the language affinities that subdivide the people of this country are similar to their genetic relationships. Therefore, all data analyses were performed for the global sample set of Guinea-Bissau as well as per ethnic group due to this potential interest in local databases that can allow future additional population comparisons and sample size enlargements of the main population and respective subgroups. Thus, allele frequencies for all of the 12 loci studied and haplotype frequencies for the linkage groups 1–4 were estimated for the global sample of 332 males as well as for the different ethnic groups (Supplementary tables S3 and S3a). A wide allelic distribution for the majority of loci was observed, yielding 217 different alleles for al 12 loci. Of the total number of alleles found 40 (18.4%) were private alleles appearing in only one ethnic group. DXS10101 contributed with the highest percentage of private alleles (28%). Four loci showed an unusually high proportion of alleles with long repeat motifs (> 35): DXS10135 (4.5%), DXS10148 (18%), DXS10146 (21%) and DXS10134 (43%). Pasino et al. [26] also found at DXS10148 11.5% of such long repeat alleles in a male sample from the Ivory Coast, which they explained as a potential marker of sub-Saharan origin. High gene diversities (Table 1: same as the power of discrimination in males; $PD_M$) were observed for the majority of loci which support the high degree of polymorphism among the multiplex system. DXS10135 and DXS10146 were the most polymorphic markers with diversities of 95% and 94%, respectively. Except for DXS8378 (68%) and DXS7423 (54%) displaying the lowest gene diversities, all other X-STRs had diversities above 73%.

As the Investigator Argus X-12 system is composed of four linkage groups that can be useful for haplotype analysis such as in specific X-STR based kinship testing, haplotype frequencies were also estimated for these clusters of potentially linked markers (Supplementary table S3a). Eight hundred thirty haplotypes were recovered in the four linkage groups, 255 of them common to more than one ethnic group. This means that 69.3% of the haplotypes were private and observed only in one ethnic group, although this may of course in part be due to the limited number of individuals tested from each of these groups. Linkage group 1 yielded 16.8% (44 in 262) of haplotypes common to more than one ethnic group, linkage group 2, 45.5% (81 in 178), linkage group 3, 39.4% (71 in 180) and linkage group 4 28.1% (59 in 210). For the global sample of Guinea-Bissau population, the highest average haplotype gene diversity (0.854) was observed for linkage group 1 (DXS8378-DXS10135-DXS10148) and the lowest average gene diversity (0.794) was obtained for linkage group 4 (DXS10134-DXS7423-DXS10146) (Supplementary table S3a).

### 3.2. Detection of new sequence variations

Previously undetected sequence variations were identified for DXS10148, DXS10074 and DXS10134 in a recent study, conducted on some of the West African individuals as well as on samples from the Middle East (Iraq) [22]. A major contribution of the present study is the description of new sequencing data of samples from 105 males from West Africa offering insights into the sequence variation and structure of the X-STRs sequenced here (Supplementary table S2). No sequence data obtained in this study (Supplementary table S2) overlaps with data from the previous study mentioned above [22].

### 3.3. Silent and new (rare) alleles

The straightforward identification of silent alleles in males is due to the presence of only one single X chromosome. This allowed the detection of several alleles with primer site mutations at DXS10148 and DXS10146 loci, with (high) frequencies of 19% and 4%, respectively, in the Guinean-Bissau individuals. All of these samples were directly sequenced to identify the primer-binding mismatches responsible for the allele dropouts and to allow genotype assignment at these loci (Supplementary table S2). This was particularly important because of the high frequency of silent alleles detected in the present sample set and therefore enabling the use of the full genotype data of the 332 individuals for the analyses.

Silent alleles at DXS10148 were detected in 63 male individuals (Supplementary table S2). Sequencing results revealed the same mutations and structure changes reported in Gomes et al. [22]: a C → T transition in the 9th nucleotide counting from the end of the repeat (reverse strand) (Supplementary table S2). This position corresponds to the 2nd base of the 3′ end of the forward primer sequence (forward strand) published in Hundertmark et al. [27]. Additionally, four samples displayed other point mutations at DXS10148: two observed at allele 40.1 and two at allele 41.1 (Supplementary table S2).

Silent alleles were also observed in 26 individuals at DXS10146 (Supplementary table S2). Analyses revealed, in all samples sequenced, a CTTT deletion between 15 and 18 nucleotides from the 3′ end of the reverse primer as described by Edelmann et al. [28], which is most likely responsible for a primer binding mismatch and consequently the silent alleles. The primer sequences used for amplification of the Investigator Argus X-12 markers are not publicly available, nevertheless,

**Table 1**
Forensic evaluation parameters for the 12 X-STRs in the Guinea-Bissau population.

| Locus | $PE_T$ | $PE_D$ | $PE_{HS}$ | $PD_F$ | $PD_M$ | | $PE_T$ | $PE_D$ | $PE_{HS}$ | $PD_F$ | $PD_M$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| DXS8378 | 0.613 | 0.466 | 0.242 | 0.830 | 0.680 | | | | | | |
| DXS10135 | 0.952 | 0.910 | 0.823 | 0.996 | 0.954 | LG1 | 0.995 | 0.990 | 0.979 | 0.999 | 0.995 |
| DXS10148 | 0.915 | 0.851 | 0.729 | 0.989 | 0.920 | | | | | | |
| DXS7132 | 0.685 | 0.545 | 0.320 | 0.883 | 0.729 | | | | | | |
| DXS10074 | 0.879 | 0.793 | 0.633 | 0.978 | 0.889 | LG2 | 0.992 | 0.984 | 0.969 | 0.999 | 0.992 |
| DXS10079 | 0.803 | 0.688 | 0.486 | 0.948 | 0.824 | | | | | | |
| DXS10103 | 0.697 | 0.558 | 0.334 | 0.890 | 0.738 | | | | | | |
| DXS10101 | 0.906 | 0.836 | 0.699 | 0.986 | 0.913 | LG3 | 0.992 | 0.984 | 0.968 | 0.999 | 0.992 |
| HPRTB | 0.769 | 0.644 | 0.429 | 0.931 | 0.797 | | | | | | |
| DXS10134 | 0.870 | 0.780 | 0.613 | 0.975 | 0.881 | | | | | | |
| DXS7423 | 0.495 | 0.351 | 0.164 | 0.742 | 0.558 | LG4 | 0.993 | 0.987 | 0.973 | 0.999 | 0.993 |
| DXS10146 | 0.933 | 0.878 | 0.768 | 0.992 | 0.936 | | | | | | |

**$PE_T$**: expected probability of exclusion in trios with female offspring; **$PE_D$**: expected probability of exclusion in father/daughter duos; **$PE_{HS}$:** expected probability of exclusion in half-sisters; **$PD_F$**: power of discrimination in females; **$PD_M$**: power of discrimination in males. **LG1-4**: linkage groups 1 to 4, respectively.

most likely the mutations detected at DXS10148 and DXS10146 are responsible for the frequently observed allelic dropouts. These mutations are not novel as these have been already observed in other studies, e.g. [29,30]: in a study on an Egyptian population also using the Argus X-12 kit mutations were reported for both DXS10148 and DXS10146 [29]. In the study of Tomas et al. [30] all of the silent alleles described, with the exception of one Danish individual, were also from African origin (Ghana and Somalia). Therefore, it can be assumed that the single base mutation at DXS10148 and the CTTT deletion at DXS10146 are informative markers of African ancestry due to their high prevalence in this group. However, further studies would be necessary to confirm this hypothesis since not much frequency data has been collected on samples of African origin for X-STRs, particularly for DXS10148 and DXS10146 loci.

The X-STR population study of Guinea-Bissau also revealed several new rare alleles at DXS10101, HPRTB, DXS10079, DXS10146 and DXS10135 loci that, to our knowledge, have not been previously described. All alleles were sequenced and the respective DNA repeat structures presented in Supplementary table S2. The DXS10101 (33.1), HPRTB (12.1) and DXS10079 (15.1) loci each presented a new, rare intermediate allele in the population sample set. Sequencing results revealed the respective mutations responsible for these variants (Supplementary table S2). Eleven males also presented rare, non-consensus, alleles at the DXS10135 locus. Sequencing data showed that five of these individuals presented the same mutation: a TCT deletion 12 bp upstream from the repeat region. Despite the fact that some of these rare alleles (12.1, 15.1, 18.1 and 24.1) have also been found in other populations of African origin [e.g.,26,31] no report on the TCT deletion, to our awareness, has been given.

The Investigator Argus X-12 kit used for the population genetic study of Guinea-Bissau was the first version of this X-STR kit [4]. A second version of the multiplex kit has been developed recently and launched by Qiagen, the Investigator Argus X-12 QS [32]. This kit offers new primer designs for some markers that will amplify some of the silent alleles mainly detected in samples of African ancestry.

### 3.4. Linkage disequilibrium

Linkage disequilibrium (LD) is the allelic dependency between alleles at different loci [33] that can be the result of physical proximity between loci or can be due to population substructure. As such LD is an important aspect that should always be taken into account in both population and forensic genetic studies because if markers are linked these cannot be treated independently but should be analyzed together. This is even more important for the X chromosome as all markers are located on the same chromosome and in particular for the Argus X-12 due to the four linkage group situation. Therefore, an exact test for linkage disequilibrium (LD) was performed for all pairs of markers to access the level of association between the X-STR markers in the present sample set of Guinea-Bissau. LD results for the 66 pairwise comparisons (Supplementary table S4) showed after Bonferroni correction ($p \leq 0.0008$) highly significant $p$ values ($p \leq 10^{-5}$) for 8 pairs of loci that are included in three of the four linkage groups: DXS10135-DXS8378 and DXS10148-DXS8378 from linkage group 1; DXS10103-HPRTB; DXS10103-DXS10101 and HPRTB-DXS10101 from linkage group 3 and DXS10146-DXS10134, DXS10146-DXS7423 and DXS10134-DXS7423 from linkage group 4. LD among markers of linkage groups 3 and 4 was still observed between all three X-STRs that compose the clusters. Linkage group 2 was the only group (DXS7132-DXS10079-DXS10074) where no association between markers could be detected. Therefore, in practice, it would be only necessary to use haplotype frequencies for the markers that revealed positive association in cases of identification or kinship testing of individuals from Guinea-Bissau.

It could be expected that LD would be present among all loci within each linkage group but this is not always the case. Similar results were obtained by Pasino and colleagues in an Ivory Coast population [26], showing a limited extent of LD. This lack of LD may be due to a longer evolutionary history of Sub-Saharan populations that allowed divergence among these loci due to recombination [34]. Another possible explanation for the non-detection of LD in linkage group 2 could be due to the low power of estimation of an exact test of LD demonstrated by Kling et al. [35]. In this study authors showed that linkage group 2 had the lowest power of estimation in several size samples by this type of test when compared to the other linkage groups [35]. On the other hand, as known, the lack of association between some of the markers inside linkage groups has advantages as it contributes to the increased power of discrimination of this multiplex in the present population.

### 3.5. Forensic efficiency

Forensic statistical evaluation parameters were calculated for the global sample and are shown in Table 1. The forensic application of the present X-STRs was assessed by the calculation of the expected probabilities of exclusion in trios with female offspring ($PE_T$) and in father/daughter duos ($PE_D$) as well at the individual level by calculating the power of discrimination in females ($PD_F$) and in males ($PD_M$). All loci used in this work revealed to be highly polymorphic and useful for identification purposes. In fact, in kinship analysis cases the several linkage groups all showed powers of discrimination of > 99.1% for testing samples of female and male origin. In the Guinean-Bissau population, the average probability of excluding a random man as the father in paternity testing (when both mother and daughter are analyzed) was > 99.1%. As one of the significant features of X-STRs and which autosomal STRs alone do not have is the potential to exclude the relationship between two putative half-sisters. Therefore, we have also estimated the expected probability of exclusion in half-sisters ($PD_{HS}$) [25] as being one example of an important scenario of application of X-STRs. It is worth noting that several relationships that make up the so-called “deficient paternity testing” are also possible where the use of X chromosome polymorphisms may be of use. For example, the software, FamLinkX [35], provides functions in likelihood calculation for family relationships/pedigrees using X-chromosomal genetic marker data that serves as a useful free resource for these other type of analyses.

Results indicate that all loci either as individual markers or as linkage groups (depending on the needed application of the present X-STRs) provide highly discriminatory genetic information that is appropriate for the purposes of forensic identification and paternity testing involving a female child.

### 3.6. Population variability among Guinea-Bissau ethnic groups

Comparisons between the 18 ethnic groups were performed for the full set of markers (AMOVA and genetic distance analysis, $F_{ST}$s) and for haplotype groups (genetic distance analysis, $F_{ST}$s) (Supplementary table S5). After applying Bonferroni’s correction for multiple testing ($p \leq 0.003$) results revealed significant genetic distances between Mancanha and Bijago ($F_{ST} = 0.028$; $p \leq 10^{-5}$); Mancanha and Balanta-Mané ($F_{ST} = 0.064$; $p \leq 10^{-5}$) and Mancanha and Manjaco ($F_{ST} = 0.019$; $p \leq 10^{-5}$). Balanta-Mané also showed differentiation, with an equal genetic distance, when compared to Fula and Fula-Fula ($F_{ST} = 0.047$; $p \leq 10^{-5}$). Bijaco and Fula-Forro were also significantly different ($p \leq 10^{-5}$) with an $F_{ST} = 0.024$. AMOVA was performed by grouping ethnic groups according to their linguistic affiliations. Clustering groups whose language is genealogically more closely related (depicted in Fig. S2) showed that genetic substructuring among populations is non-existent. In fact, only 0.15% of the total variation can be attributed to “among group” differences with 99.85% of the total variation being attributed to “within group” differences (total $F_{ST} = 0.002$; $p$ non-significant). This result was not different from the total $F_{ST}$ obtained ($F_{ST} = 0.002$; $p$ non-significant) when no clustering of the populations is performed based on linguistics and the 18 original

ethnic groups are considered. AMOVA results in this case also revealed that 99.74% of variation is best attributed to within population variation and only 0.26% to among populations. At the haplotype level, for linkage groups 1 and 2, no significant differentiation among any of the groups was observed (Supplementary table S5). For linkage group 3 significant differences were observed only between Bijago and Mancanha ($F_{ST} = 0.061$; $p \leq 10^{-5}$) (which was also detected at the $F_{ST}$ global analysis); and between Bijago and Mandinga ($F_{ST} = 0.068$; $p \leq 10^{-5}$). For linkage group 4, significant population differentiation was observed between Mandinga and Mancanha ($F_{ST} = 0.073$; $p \leq 10^{-5}$) and between Fula-Fula and Fula ($F_{ST} = 0.043$; $p \leq 10^{-5}$).

Only two populations revealed differences both at the global $F_{ST}$ analysis and haplotype level (Bijago and Mancanha) which could be a potential indication of the need of use of different local databases for these two groups. Nevertheless, from a global perspective, no major differentiation among groups is observed as grouping appears to be erratic for the several population differentiation methodologies. It should be noted that the clustering of the 18 groups has its limitations as the number of individuals in each group (n ≤ 48) might not be sufficiently large enough to detect real differences among them therefore caution is needed in the interpretation of these results. Obtaining more samples from African populations is very difficult and unfortunately the increase of the groups that compose the present dataset is not currently possible. Therefore, these observations on population variability among the present Guinean-Bissau ethnic groups stand as preliminary data rather than definite conclusions.

### *3.7. Population variability among Guinea-Bissau and other West African countries*

In addition to the population sample studied of 332 individuals from Guinea-Bissau, an additional 110 samples available from other countries dispersed over West Africa, were genotyped for the Investigator Argus X-12 markers. This set of samples contained 31 females and 79 individuals from Senegal, Gambia, Guinea, Serra Leo, Liberia, Ghana, Togo, Benin, Nigeria and Niger. Since each of these West African countries do not have a number of samples representative enough at a local level but on the other hand do not have individuals from Guinea-Bissau, these samples were grouped and used for population comparisons. Female and male data were combined after an exact test for population differentiation (p ≤ 0.05) did not reveal any differences among gender groups (data not shown).

Genetic distances were obtained between both populations ("Guinea-Bissau" vs. "West African-other"). No significant differences were found between these two groups of West African individuals ($F_{ST} = 0.002$; $p = 0.018$) [when applying Bonferroni correction for multiple testing ($p \leq 0.016$)] supporting the genetic homogeneity of this region at the X chromosome level. This result further supports the lack of significant differentiation among the Guinea-Bissau ethnic groups also observed in this study.

## 4. Concluding remarks

In conclusion, this study is the first major genetic variation study on X-STRs in samples from Guinea-Bissau West African country contributing with allelic and haplotype frequency estimations as well as with significant X-STR sequencing data. The present West African samples provide interesting findings particularly at the DNA sequence level and also regarding the linkage disequilibrium status among the established linkage groups. From a population genetic point of view, 18 ethnic groups were considered in the present sample allowing the assessment of population variation in Guinea-Bissau. However, no significant differentiation among the majority of Guinean ethnic groups was detected (except between Bijago and Mancanha) and therefore a common frequency database for identification purposes could be considered for the global population of Guinea-Bissau. Nevertheless, the low number of individuals in each group may have not allowed for the detection of genetic differences and further studies at this level are needed if more samples become available. Other West African samples included in this work did not show significant genetic differentiation with the sample from Guinea-Bissau, revealing a genetic homogeneity of West African countries at the X chromosome level.

From a forensic perspective, the Argus X-12 kit proved to be extremely robust and informative in the present population. It is therefore a useful tool in forensic practice, particularly in "deficient paternity" and other kinship cases.

## Conflict of interest statement

The authors declare no conflict of interests.

## 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.2017.08.016.

## References

[1] S.F. Schaffner, The X chromosome in population genetics, Nat. Rev. Genet. 5 (2004) 43–51.
[2] R. Szibor, M. Krawczak, S. Hering, J. Edelmann, E. Kuhlisch, D. Krause, Use of X-linked markers for forensic purposes, Int. J. Legal Med. 117 (2003) 67–74.
[3] R. Szibor, X-chromosomal markers: past, present and future, Forensic Sci. Int. Genet. 1 (2007) 93–99.
[4] Investigator® Argus X-12 Handbook, Qiagen, Hilden, Germany, 2013 (June).
[5] M. Graydon, F. Cholette, L.K. Ng, Inferring ethnicity using 15 autosomal STR loci-comparisons among populations of similar and distinctly different physical traits, Forensic Sci. Int. Genet. 3 (2009) 251–254.
[6] A. Carreira, Aspectos históricos da evolução do Islamismo na Guiné Portuguesa (achegas para o seu estudo) Boletim Cultural da Guiné Portuguesa, Bissau Porto: Impr. Portuguesa 84 (1966) 405–455.
[7] F.P. Chagas, Povos da Guiné Portuguesa, Annaes do Clube Militar Naval 41 (1910) 86–96.
[8] R. Pelissier, História da Guiné. Portugueses e Africanos na Senegâmbia Editorial Estampa, Lisboa 2 (1997) 1841–1936.
[9] H. Alimen, Evolution du climat et des civilisations depuis 40000 ans du nord au sud du Sahara occidental, (Premières conceptions confrontées aux données récentes) Bull, L'Assoc. Franc. L'Étude Quaternaire 4 (1987) 215–227.
[10] M. Jao, Relações entre fulas e mandingas nos espaços Gabú e Forreá, Soronda: revista de estudos guineenses INEP, Bissau 3 (2002) 5–27.
[11] J.M. Moreira, Os Fulas da Guiné Portuguesa na panoramica geral do mundo Fula, Boletim Cultural da Guiné Portuguesa XIX (1964) 417–432.
[12] A. Teixeira da Mota, Guiné Portuguesa, Agencia Geraldo Ultramar, Lisboa, 1954.
[13] A. Carreira, F.R. Quintino, Antroponímia da Guiné Portuguesa, Memórias da Junta de Investigação do Ultramar Lisboa vol. 49 and 52, JIU, Lisboa, 1964.
[14] A. Marques Basso, Aspectos do problema da semelhança da língua dos papéis, manjacos e bramos, Boletim cultural da Guiné portuguesa II (5) (1947) 77–109.
[15] H. Spínola, J. Bruges-Armas, D. Middleton, A. Brehm, HLA polymorphisms in Cabo Verde and Guiné-Bissau inferred from sequence-based typing, Hum. Immunol. 66 (2005) 1082–1092.
[16] A. Rosa, A. Brehm, T. Kivisild, E. Metspalu, R. Villems, MtDNA profile of West Africa Guineans: towards a better understanding of the Senegambia region, Ann. Hum. Genet. 68 (2004) 340–352.
[17] A. Rosa, C. Ornelas, A. Brehm, R. Villems, Population data on 11 Y-chromosome STRs from Guiné-Bissau, Forensic Sci. Int. 157 (2006) 210–217.
[18] A. Rosa, C. Ornelas, M.A. Jobling, A. Brehm, R. Villems, Y-chromosomal diversity in the population of Guinea-Bissau: a multiethnic perspective, BMC Evol. Biol. 7 (2007) 124.
[19] R. Gonçalves, J. Jesus, A.T. Fernandes, A. Brehm, Genetic profile of a multi-ethnic population from Guiné-Bissau (west African coast) using the new PowerPlex 16 System kit, Forensic Sci. Int. 129 (2002) 78–80.
[20] A População Estrangeira em Portugal-2011, Instituto Nacional de Estatística (Statistics Portugal), 2012, www.ine.pt.
[21] Relatório de Imigração, Fronteiras e Asilo, Serviço de Estrangeiros e Fronteiras (Foreigners and Borders Service), (2015) (2015), www.sef.pt.
[22] I. Gomes, A. Brehm, L. Gusmão, P.M. Schneider, New sequence variants detected at DXS10148, DXS10074 and DXS10134 loci, Forensic Sci. Int. Genet. 20 (2016) 112–116.
[23] L. Excoffier, H.E. Lischer, 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.
[24] D. Desmarais, Y. Zhong, R. Chakraborty, C. Perreault, L. Busque, Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA), J. Forensic Sci. 43 (1998) 1046–1049.
[25] I. Gomes, V. Pereira, V. Gomes, M.J. Prata, N. Pinto, A. Carracedo, A. Amorim,

L. Gusmão, The Karimojong from Uganda: genetic characterization using an X-STR decaplex system, Forensic Sci. Int. Genet. 4 (2009) e127–128.
[26] S. Pasino, S. Caratti, M. Del Pero, A. Santovito, C. Torre, C. Robino, Allele and haplotype diversity of X-chromosomal STRs in ivory coast, Int. J. Legal Med. 125 (2011) 749–752.
[27] 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 (2008) 489–492.
[28] 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 (2008) 41–46.
[29] S. Elakkary, S. Hoffmeister-Ullerich, 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. 11 (2014) 26–30.
[30] 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.
[31] D. Becker, H. Rodig, C. Augustin, J. Edelmann, F. Götz, S. Hering, R. Szibor, W. Brabetz, Population genetic evaluation of eight X-chromosomal short tandem repeat loci using Mentype Argus X-8 PCR amplification kit, Forensic Sci. Int. Genet. 2 (2008) 69–74.
[32] Investigator® Argus X-12 QS Kit Handbook, Qiagen, Hilden, Germany, 2015 (June).
[33] A.O. Tillmar, P. Mostad, T. Egeland, B. Lindblom, G. Holmlund, K. Montelius, Analysis of linkage and linkage disequilibrium for eight X-STR markers, Forensic Sci. Int. Genet. 3 (2008) 37–41.
[34] S. Tishkoff, S. Williams, Genetic analysis of African populations: human evolution and complex disease, Nat. Rev. Genet. 3 (2002) 611–621.
[35] D. Kling, B. Dell'Amico, A.O. Tillmar, FamLinkX − implementation of a general model for likelihood computations for X-chromosomal marker data, Forensic Sci. Int. Genet. 17 (2015) 1–7.