A reference database of forensic autosomal and gonosomal STR markers in the Tigray population of Ethiopia

A reference database of forensic autosomal and gonosomal STR markers in the Tigray population of Ethiopia

K. Haddish [a 1 2], E. Chierto [b 1], G. Di Vella [b], D. Lacerenza [b], S. Raddi [b], S. Aneli [c d], A.L. Bogale [e], E. Kidane [e], A. Yizengaw [e], Y. Getaneh [e], G. Tewelemedhin [a], H.R.S. Kumar [a 3], C. Robino [b*].

[a] Department of Forensic Medicine, University of Mekelle, Mekelle, Ethiopia

[b] Department of Public Health Sciences and Pediatrics, University of Turin, Turin, Italy

[c] Department of Medical Sciences, University of Turin, Turin, Italy

[d] Department of Biology, University of Padua, Padua, Italy

[e] Ethiopian Public Health Institute, Addis Ababa, Ethiopia

[*] Correspondence to: Department of Public Health Sciences and Pediatrics, University of Turin, Corso Galileo Galilei 22, 10126 Turin, Italy. E-mail address: carlo.robino@unito.it (C. Robino).

[1]These authors contributed equally to the study.

[2] Present Address: Department of Biotechnology, Daegu University, Gyeongsan, South Korea.

[3] Present Address: Government Medical College Barmer, Rajasthan University of Health Sciences, Rajasthan, India.

Highlights

- Forensic STR variation was investigated in the Tigray population of Ethiopia.
- Mean F_{ST} at aSTR loci between Tigray and other Horn of Africa populations was 0.003.
- High internal Y-STR haplotype diversity was found in the Tigray population.
- A novel null allele at locus DXS10146 with population frequency > 1% was identified.
- Data can aid genetic investigations of human trafficking in the Mediterranean Sea.

Abstract

Allele frequencies of 21 autosomal STR markers (AmpF/STR GlobalFiler) and haplotype frequencies of 27 Y- and 12 X-STR markers (AmpF/STR YFiler Plus and Investigator Argus X-12, respectively) were investigated in the Tigray population of Ethiopia, representing the main population group in the Tigray regional state of Ethiopia and neighboring Eritrea. For autosomal STR allele frequencies, the average random match probability in the Tigray sample was 2.1×10^{-27}. The average locus by locus F_{ST} distance calculated comparing autosomal STR allele frequencies from Tigray and from a broad regional reference dataset currently available for the Horn of Africa was 0.003. The Tigray male sample displayed high Y-STR diversity, with complete individualization of haplotypes using the AmpF/STR YFiler Plus panel. Analysis of molecular variance did not detect significant heterogeneity between Y-STR haplotypes observed in the present study and those previously reported in the literature for other Tigray population samples from Ethiopia and Eritrea. Study of the X-STR landscape in Tigray evidenced several distinctive features including: the molecular characterization of a novel null allele at locus DXS10146 with frequency > 1%; allele dependency between loci within linkage groups I and III; significant differences in haplotype distribution compared to other Horn of Africa populations, that should be taken into account in kinship analysis. The collected data can be used as a reference STR database by local forensic genetics services and in genetic identification procedures of victims of human trafficking in the Mediterranean Sea, which frequently involve individuals originating from the Horn of Africa.

Keywords: Horn of Africa Short tandem repeat X chromosome Y chromosome Population genetics.

1. Introduction

Evaluation of the weight of evidence in forensic identification and kinship testing by means of short tandem repeat (STR) DNA polymorphisms requires allele/haplotype frequency information in the relevant population. Autosomal STRs (aSTRs) included in standard forensic panels, like the Combined DNA Index System (CODIS) and the European Standard Set (ESS), can be considered genetically independent of one another [1]. Consequently, random match probability (RMP) and likelihood ratio (LR) calculations are done according to the "product rule", by multiplying the individual allele frequencies in the profile in question. Whenever allele frequencies are drawn from broad population groups that might be unrepresentative of the subpopulation from which the person of interest originates, RMP and LR values can be corrected by an adjustment factor (θ, F_{ST}) expressing the correlation/co-ancestry of randomly chosen alleles within subpopulation relative to the entire population [2]. Due to completely linked inheritance on the non-recombining part of the Y chromosome, the product rule cannot be applied to Y-STRs. For the interpretation of Y-STR matches, therefore, the availability of detailed reference databases describing haplotype variation within local populations is of vital importance [3]. Population-specific haplotype data obtained from hemizygous males are also crucial for X-STR analysis, since population history and sub-structure can lead to allelic dependency between alleles at different X-STR loci (linkage disequilibrium, LD) [4]. In general, a higher degree of LD is expected for X-STRs compared to autosomal markers, since X chromosome recombination only occurs in women and because of the lower mutation rates observed in the female germline [5]. Moreover, since only three X chromosome copies are present in the population for every four autosomes, stronger effects of genetic drift are also expected [6].

Africa is the most genetically diverse region in the world and harbors extensive population structure, due to the continuous presence of modern humans in the past 300 thousand years in association with complex demographic events including migration, admixture and possible archaic introgression [7], [8].

As in other fields of genetics and genomics research [9], despite the constantly increasing number and range of available forensic STR allele/haplotype data from global populations, studies addressing the African continent remain limited in terms of density and sample size. As a contribution to the knowledge of forensic STR diversity in Africa, we focused on Ethiopia, a nation characterized by the highest genetic heterogeneity within the continent. Previous high-density SNP-array studies have shown that Ethiopian populations, taken together, span half of the genetic diversity observed in all African populations [10] and display the highest proportions of continental unshared and novel genetic variants [11].

In particular, the Tigray population investigated in the present study is, according to the 2007 Population and Housing Census [12], the fourth largest ethnic group in the Federal Democratic Republic of Ethiopia after Amhara, Oromo and Somali. It counts about 4.5 million individuals and represents the prevailing population (> 95%) in the regional state of Tigray, located in Northern Ethiopia. Tigray is also the major ethnic group of the neighboring State of Eritrea [13].

Availability of geographically detailed allele and haplotype diversity data for autosomal and gonosomal STR markers is a stepping stone to the development of forensic services willing to provide DNA-based identification and kinship testing in African countries [14]. Collection of forensic STR variation data in the Horn of Africa, however, can also have a larger resonance by contributing to the genetic identification of victims and the reunification of separated family members in maritime accidents involving migrants trying to reach Europe from Northern Africa [15]. As a matter of fact, the Horn of Africa, and specifically Eritrea, is presently one of the main regions of origin of migrant victims of human trafficking across the Mediterranean Sea [16], [17].

The aim of the present study was to compile allele and haplotype frequency data of STRs commonly used for identification purposes in a large Tigray population sample. Data were made available to the forensic community through public, quality controlled repositories such as STRidER and the Y-STR Haplotype Reference Database (YHRD). STR diversity in Tigray was investigated in the context of human genetic variation within the Horn of Africa. Extent of population structure as detected with the different categories of STRs markers, depending on their specific pattern of inheritance, was assessed. Possible adjustment strategies when calculating match probabilities (aSTRs) and in kinship analysis (X-STRs) in cases involving individuals from the Horn of Africa were discussed.

2. Material and methods

2.1. Ethical requirements and sample collection

In compliance with ethical guidelines for publication of research on genetics of biological material [18], authorization of research methods and development of the study was obtained from Mekelle University Research Ethics Review Committee (ERC 0841/2016). Approved procedures and protocols were followed in the study.

All samples were collected by means of sterile buccal swabs (LP Italiana Spa, Milan, Italy) from adult volunteer donors (students and staff of Mekelle University, Ethiopia). All participants signed the approved consent form indicating fully informed consent. For aSTR analysis, samples were obtained from 236 donors of both sexes (65 women and 171 men) who self-reported as unrelated and having all the four grandparents of Tigray origin. Male donors of the aSTR study were complemented with: 75 additional male donors who self-reported as unrelated and having the paternal grandfather of Tigray origin (Y-STR study, n = 246); 77 additional male donors who self-reported as unrelated and having the two maternal grandparents of Tigray origin (X-STR study, n = 248). Among additional male donors for gonosomal STR analysis, 51 were eligible for both the Y- and X-STR studies. Geographic location of Tigray within the Horn of Africa and composition of sample cohorts for aSTR, Y-STR and X-STR studies are depicted in Supplementary material, Fig. S1.

2.2. DNA extraction, PCR amplification and genotyping

Buccal swabs were processed using the ChargeSwitch gDNA Normalized Buccal Cell kit (Invitrogen), that provides purified DNA at a normalized concentration of 1–3 ng/µl. Multiplex amplification of aSTRs (AmpF*l*STR GlobalFiler kit, ThermoFisher Scientific), Y-STRs (AmpFlSTR YFiler Plus kit, ThermoFisher Scientific) and X-STRs (Investigator Argus X-12 kit, Qiagen) was performed on the GeneAmp PCR System 9700 (ThermoFisher Scientific) using 1 µl of each DNA extract, according to the manufacture's instructions. Amplified DNA was detected and separated using the ABI Prism 3500 XL Genetic Analyzer (Thermo Fisher Scientific). STR genotyping was performed with GeneMapper ID-X v1.4 software (Thermo Fisher Scientific).

Samples with null alleles at loci DXS10146 and DXS10148 were reamplified with the Investigator Argus X-12 QS kit (Qiagen). Those still displaying DXS10146 amplification failure afterwards were subjected to single locus PCR amplification and sequencing with primers described by Tomas et al. [19] and the BigDye Terminator v 3.1 kit (Thermo Fisher Scientific).

2.3. Statistical analysis

RMPs of aSTR profiles, calculated with allele frequencies from the present study (cognate database) and from geographically close populations or broad group of individuals assumed to be generically representative of sub-Saharan Africa (non-cognate databases), were obtained using

EuroForMix software version 2.1 [20]. In RMP calculations with EuroForMix the value of F_{ST} was taken into account according to the Balding-Nichols formula [21].

Arlequin software version 3.5 [22] was used to perform: Test of Hardy–Weinberg equilibrium (HWE) in aSTR; calculation of pairwise F_{ST} genetic distances (aSTRs, X-STRs, Y-STRs); calculation of haplotype diversity (X-STRs and Y-STRs); test of LD (X-STRs); mean number of allelic pairwise differences and analysis of molecular variance (AMOVA) in Y-STRs. In population comparisons by means on F_{ST} and AMOVA, multi-copy Y-STR loci (DYS385 and DYF387S1) were excluded from calculations and the number of repeat units in DYS389I was subtracted from DYS389II. Standard forensic genetics parameters calculated in Y-STRs also included discrimination capacity (DC), estimated as number of distinct haplotypes divided by total number of haplotypes, and haplotype match probability (HMP), corresponding to the sum of squares of the haplotype frequencies. Matrixes of Slatkin's linearized pairwise genetic distances were calculated from haplotype frequencies with Arlequin software and represented by multidimensional scaling (MDS) analysis using the *cmdscale* function as implemented in stats package of R v.4.0.0 [23].

LR calculations for X-STR markers in kinship cases were performed with software FamlinkX [24], applying recombination fractions and mutation rates described by Nothnagel et al. [25].

2.4. Quality control (QC)

According to guidelines for the publication of population genetic data [26], aSTR and Y-STR data were submitted to STRidER [27] and YHRD [28]. After completion of QC process, they were assigned accession numbers STR000344 and YA004670, respectively.

3. Results and discussion

3.1. aSTRs

aSTR allele frequencies in the Tigray population sample calculated after STRidER QC are displayed in Supplementary material, Table S1. Several alleles not included in the AmpF*l*STR GlobalFiler allelic ladder (OL) were observed and are highlighted in Table S1. Notably, OL alleles made up > 5% of total observed alleles at loci D18S51 and SE33. All aSTRs were in HWE, after Bonferroni correction for multiple testing ($\alpha = 0.002$). The average RMP in the Tigray population sample, assuming $F_{ST} = 0$, was 2.1×10^{-27}. We tested the impact on RMP calculations when using the Tigray (cognate) allele frequency database rather than non-cognate databases. The chosen datasets were: a study of 799 Eastern Africans from Somalia, Ethiopia and Eritrea [29] that represents, to our knowledge, the largest reference dataset from the Horn of Africa currently available in the literature [30]; sub-Saharan Africans (n = 103) from the human genome diversity panel (HGDP) [31]; U.S. African Americans (n = 342) [32]. A parameter *d* was then calculated, corresponding to log10 of the ratio between RMP estimates obtained from each aSTR profile in the Tigray population sample when using cognate (Tigray) and non-cognate allele frequencies [33]. Values of $d > 0$ indicate an overestimation of the weight of the evidence when reference non-cognate databases were used for calculations. In order to compensate for bias due to rare alleles, the frequency of alleles only detected in the Tigray population was set in non-cognate databases to $1/2n_{min}$, where n_{min} is the number of individuals in the smallest database used in the calculation [34]. Distribution of *d* values applying different F_{ST} adjustments is reported in Table 1.

F_{ST} adjustments required to make the RMP calculations unbiased were in agreement with mean locus by locus F_{ST} distances calculated between Tigray and other reference populations, which were lower for the Horn of Africa (0.003), intermediate for African Americans (0.009), and higher for HGDP sub-Saharan Africans (0.013). It could be seen that the application of a 0.01 F_{ST} correction to RMP calculations using the Horn of Africa database was necessary in order for the risk of non-conservative estimations to become almost negligible (13.1% of samples with $1 > d > 0$, corresponding to a less than 10-fold overestimation of the weight of evidence). A larger

F_{ST} adjustment (0.03) was required when broad African or African American databases were applied, with 2.97% and 0.42% of samples, respectively, displaying a 1000-fold overestimation (d > 3) of the rarity of the observed STR profile with F_{ST} = 0.01. The larger proportion of non-conservative results in HGDP sub-Saharan Africans, compared to African Americans, could be explained by the combined effect of genetic heterogeneity in continental African populations [7] and the admixed nature of African Americans and Tigray, both displaying a strong Eurasian genetic component as a consequence of recent demographic events connected to the history of the slave trade, in the case of African Americans [35], and to ancient backflow into Eastern Africa for Tigray [36]. The obtained results are in line with previous studies indicating that a 0.03 F_{ST} correction is sufficiently large to be almost always conservative, when forensic RMP calculations make use of broad collections of individuals sharing African ancestry as reference population [37], [38]. As previously stated, one of the goal of the present study was to provide genetic data helping the identification of victims of human trafficking across the Mediterranean Sea. In humanitarian procedures, setting the threshold for the posterior probability of identity is a matter of weighing the costs of false decisions [39], since deviation of only a few orders of magnitude in the calculated LRs can have a significant impact on the possibility to achieve identification, especially in the case of low template DNA samples and when only indirect comparison (kinship testing) is possible [30]. Geographical origin within Africa of migrants involved in maritime incidents is often reasonably known a priori [17]. In such cases, application of a 0.01 rather than 0.03 F_{ST} adjustment in combination with a regional population dataset seems adequate, at least for Horn of Africa populations. It could be seen that, when using a regional Horn of Africa reference dataset for Tigray samples, a 0.03 F_{ST} correction completely eliminated the occurrence of d > 0 (i.e. overestimation of evidence), but also generated a mean d value of −3.4, corresponding to an over 1000-fold underestimation of evidence on average. This could lead to a potentially large number of false-negative comparisons, which is hardly acceptable in a disaster victim identification context.

3.2. Y-STRs

Y-STR haplotypes observed in 246 Tigray male individuals are listed in Supplementary material, Table S2. Di-and tri-allelic patterns, OL and null alleles are highlighted in Table S2 and were all sporadic except for intermediate alleles at DYS458 (~ 14%) known to be signature variants of Y-haplogroup J1-M267, which is widespread in the Middle East, Northern and Eastern Africa [40]. The combination of loci included in the AmpF*l*STR YFiler Plus panel provided complete individualization of haplotypes in the Tigray population (haplotype diversity and DC equal to 1.000). HMP was 0.004, with a mean number of allelic pairwise differences between samples of 18.3 (± 8.1 SD). Previously observed high intra-population Y-STR haplotype diversity in the Tigray ethnic group [41], [42] was therefore confirmed. Given the admixed nature of the Tigray population and the geographic position of Ethiopia, bridging Africa and the Arabic peninsula [10], [11], [35], [36], Y-STR population comparisons were carried out with available data from the Horn of Africa [41], [42], Northern Africa [43] and the Middle Eastern (Saudi Arabia) [44] (see Supplementary material, Table S3A for population details). Haplotypes observed in Tigray were not shared with any other population including the Eritrean sample described in [42], which had been analyzed with the PowerPlex Y23 panel, thus limiting comparisons to a subset of twenty Y-STR loci overlapping between genotyping kits. The matrix of pairwise F_{ST} distances was calculated and is provided in Supplementary material, Table S3B. It could be seen that, on average, lower pairwise F_{ST} values were observed between Tigray and other Horn of Africa populations (0.011), compared to Northern Africa (0.095) and Saudi Arabia (0.101). Notable exceptions within the Horn of Africa were Saho (0.260), Ethiopian Somali (0.092), Afar (0.129) and Djibuti Somali (0.241) from [41]. These populations are known to be organized in patrilinear and patrilocal clans [45], displaying low internal haplotype diversity making them the most divergent populations within the Horn of Africa [41]. In contrast, not significant (negative) F_{ST} values between the Tigray sample from the present study and Saho and Afar from [42] were

observed. This could be partly due to the reduced set of Y-STR markers available for comparison with [42]. However, dissimilarities in patrilinear descent structure of local clans in the different Saho sample sets analyzed in [41] and [42], which displayed a significant pairwise genetic distance between each other, could not be excluded.

Interestingly, contrasting pairwise F_{ST} estimates between Tigray samples from the present study and those included in [41] and [42] were obtained: 0.060 and statistically significant in the first case, negative and not statistically significant in the second case. A likely explanation of this observation could be sampling bias. In this study a rather large (n = 248) and diversified population of students and staff members being drawn to Mekelle University from across the regional state of Tigray was analyzed, whereas the sample in [41] was limited in size (n = 28) and mostly came from small rural areas, with reduced haplotype diversity and possible recent paternal co-ancestry.

Not statistically significant F_{ST} distances between the Tigray sample and Nara [41], [42] and Kunama [42] were also observed. These two populations belong to the Nilo-Saharian linguistic family distinct from the Afro-Asiatic linguistic family of Semitic-speaking Tigray, but they are the closest to Tigray in geographical terms. Lack of differentiation in Y-chromosome haplogroup distribution between Nara, Kunama and neighboring Afro-Asiatic populations had been previously reported and attributed to an underlying common origin, predating the demographic expansion and linguistic divergence of early Horn of Africa populations, in combination with more recent admixture [46], possibly favored by matrilineality peculiar to these two ethnic groups [41]. MDS plots obtained from Slatkin's linearized F_{ST} distances are shown in Fig. 1b.

It could be seen that the Tigray sample was part of a cluster that included all other populations from the Horn of Africa except for Saho (EA-SA1), Djibuti Somali (EA_SD), Ethiopian Somali (EA_SE), and Afar (EA-AF1) samples from [41]. Saho, Afar and Somali populations, as previously mentioned, are known to be the most divergent within the Horn of Africa. Moreover, in these populations, significant genetic distances between distinct sample sets collected from different geographical locations were reported [41], [42]. A similar pattern was observed here with Saho (EA_SA2) and Afar (EA_AF2) samples from [42] that, in contrast to EA-SA1 and EA-AF1, were located within the main cluster of Horn of Africa populations. This cluster occupied an intermediate position between the Saudi Arabians and Northern Africans samples, who were distributed according to geography (longitude), with Algerian/Moroccan Berbers and Saudi Arabians at opposite sides, reflecting a documented east-to-west increase of autochthonous genetic component, with corresponding decrease of Near Eastern Arabic genetic component, in Northern Africa [47] (Fig. 1a). The intermediate position of Horn of Africa samples was consistent with previously described Y-chromosomal signatures of bidirectional gene flow between the Horn of Africa and nearby regions through the Nile corridor to the North [48] and the Bab-el-Mandeb strait to the East [49]. MDS plots redrawn by considering only populations from the Horn of Africa (Fig. 1b) further enhanced the clustering between most of the Eritrean and Ethiopian populations, including different Tigray sample sets, and the outlier position of EA_SA1, EA_SD, EA_SE and EA_AF1. No clear pattern of differentiation was observed in the MDS plot according to linguistic affiliation or nationality of the tested populations. This was confirmed by hierarchical AMOVA analysis (Supplementary material, Table S4) that evidenced significant Y-STR haplotype heterogeneity among populations within groups (8.58%, $p < 0.001$) but not between groups (1.07%) when populations were grouped by language family/subfamily. Clear limitations in the power to detect signs of linguistic stratification, as revealed by whole-genome analysis of the Ethiopian gene pool [10], could be expected for the small set of fast-evolving Y-STR markers investigated in the present study. However results were in line with Y-chromosome haplogroup data [46] and with a more recent whole genome study of Ethiopians that reported some heterogeneity within linguistic groups and the occurrence of genetic similarities between populations belonging to different language classifications, explained by recent intermixing [50]. A

pattern similar to that found for linguistic affiliation could be seen when populations were grouped by nationality (Djibuti, Eritrea, Ethiopia), with 10.35% of variance among populations within groups ($p < 0.001$) and −1.78% between groups. No evidence of geographical population substructure by Y-chromosome analysis following recent political events [51], such as the independence of Eritrea from Ethiopia in 1993, was therefore detected. On the contrary, no significant differences between populations within groups (−1.15%) but significant heterogeneity between groups (14.42%, $p = 0.001$) was found when samples from different nations were grouped according to ethnicity, with the Tigray group including samples from Ethiopia analyzed in the present study and in [41] and from Eritrea described in [41] and [42]. However, when AMOVA analysis according to ethnicity was performed after removing the highly divergent Afar, Saho and Somali populations, the partition of variance between remaining groups (Tigray, Kunama, Nara and Tigre) became not significant (1.99%). Overall, the obtained results suggested that Y-STR diversity in the Horn of Africa at inter- and intrapopulation level is shaped by a combination of factors including geographic distance and culture. Limited variation was observed between neighboring populations despite different linguistic affiliation (i.e. Tigray, Nara and Kunama). Y-haplotype structuring was evident in strictly patrilinear groups, such as the Afar, Saho and Somali, but weak in populations, like Tigray, who are not organized in patrilineages and clans [45].

3.3. X-STRs

Haplotype and allele frequencies in the Tigray sample are displayed in Supplementary material, Table S5A and B, respectively. Common OL alleles were observed at several loci, as previously reported for Eritrean populations [42], with an overall frequency of 20.0% at DXS10135, 18.3% at DXS10148, 5.8% at DXS1046, and 5.2% at DXS10134. Null alleles were sporadic with the notable exception of loci DXS10146 (3.6%) and DXS10148 (2.4%). Frequent amplification failure at both loci had been previously described in Eastern [19], Northern [52], and Western [53] African populations analyzed with the Investigator Argus X-12 kit. Sequencing analysis carried out in these studies indicated that, in all populations, the same primer binding site mutations were the likely cause of the observed silent alleles [19], [52], [53]. They were, namely, a [G → A] transition at position 9270946 of Genome Reference Consortium Human Build 38 (GRCh38) in DXS10148 and a [CTTT] deletion from position 150404032 to 150404035 (GRCh38) in DXS10146. Tigray samples displaying null alleles at DXS10146 and DXS10148 loci were reamplified with the recently introduced Investigator Argus X-12 QS kit, which includes new primer pairs designed to overcome amplification failure at DXS1046 and DXS10148 loci [54]. DXS10146 and DXS10148 genotypes obtained with Investigator Argus X-12 and Investigators Argus X-12 QS kits are reported in Supplementary material, Table S6A. It could be seen that, while DXS10148 genotyping with the Investigator Argus X-12 QS kit was successful in all tested samples, three Tigray samples still displayed null alleles at the DXS10146 locus. Single locus PCR amplification and sequencing with primers external to those supposedly included in the Investigator Argus X-12 QS kit indicated that a novel point mutation, a [C → T] transition at position 150404028 of GRCh38, could explain the persistent amplification failure observed in these samples (Fig. 2). Updated haplotype and allele frequencies in the Tigray sample after re-amplification with the Investigator Argus X-12 QS kit are displayed in Supplementary material, Table S6B and C, respectively.

As shown in Supplementary material, Table S7, pairwise test of LD between the twelve X-STR markers after Bonferroni correction for multiple testing ($\alpha = 0.0008$) showed significant LD between DXS10135 and DXS10148 markers located within linkage group (LG) I and the three markers of LG III (DXS10101-HPRTB; DXS10101-DX10103, HPRTB-DX10103).

X-STR haplotype diversity in the Tigray population was compared with available data from Northern (Egypt) [52], Western (Guinea Bissau) [53] and Eastern (Eritrea, Somalia) [42], [55], Africa, and the Middle East (United Arab Emirates, UAE) [56]. It could be seen that haplotype diversity values at the four LGs observed in Tigray (LG I: 0.998; LG II: 0.995; LG III: 0.992; LG IV: 0.997) were in line with those previously observed in neighboring Eritrean samples (LG I: 0.998;

LG II: 0.992; LG III: 0.994; LG IV: 0.997) [42] and constantly higher than those found in the Somali population (LG I: 0.996; LG II: 0.991; LG III: 0.990; LG IV: 0.995) [55]. Haplotype sharing, calculated as the number of Tigray haplotypes shared with reference population divided by number of different haplotypes in the reference population, is shown in Supplementary material, Table S8. Eritreans were the population that shared the highest percentage of haplotypes with Tigray at each LG. Somali and Western Africans displayed the highest, and constantly statistically significant, pairwise F_{ST} values with Tigray, whereas the shortest and always not significant distances were observed with Eritrea and Egypt (Supplementary Table S8). This was also reflected in the corresponding MDS plot obtained from mean Slatkin's linearized distances calculated across the four LGs, shown in Supplementary material, Fig. S2. It could be seen that the Tigray sample was almost superimposable with the neighboring Eritrea sample, suggesting an overall genetic homogeneity, and that they both formed a cluster with samples from Northern Africa and the Middle East, clearly separated from Somali and Western African populations. The outlier position of the Somali in intracontinental X-STR comparisons with other Horn of Africa [42], Northern African [42] and Western African [19] populations had been previously reported. As for Y-STRs, this is the likely effect of deviation from panmixia (consanguinity), which was shown to be maximum in Somali among African populations [57] and that could also explain the higher level of LD observed in Somali X-STR haplotypes [55], [57] compared to Tigray (present study) and Eritrea [42].

X-STR markers are principally employed as integrative markers in complex kinship analysis [4], [54]. To evaluate the effect on kinship calculations of LD structure and different Horn of Africa haplotype databases, LR values were redetermined in a small set (n = 12) of archival kinship cases involving Ethiopian, Eritrean and Somali individuals that had all been originally assessed using Somali [55] haplotype frequencies, which were the only available at the time of analysis. Characteristics of the tested kinship cases are outlined in Supplementary material, Table S9. FamlinkX software, which was used for LR calculations, does not offer the possibility to adjust for population substructure applying a F_{ST} correction. On the other hand, it allows to directly calculate from X-STR haplotype data a parameter "lambda", used to give prior weight to unobserved haplotypes and to adjust the frequency estimation of observed haplotypes, thus taking LD structure within LGs into account [24]. Small values of lambda result in a high weight of the observed data (the population specific haplotype list), while high values of lambda correspond to greater weight given to the expected haplotype frequencies derived from allele frequencies. LG-specific lambda values calculated in the Tigray, Eritrean [42] and Somali [55] populations are given in Table S9. FamlinkX includes two main LR calculations models, one that considers LD ("LR exact"), and one, "LR(LE)", that disregards LD and haplotype observations. Results of LR calculations are listed in Table S9. It could be seen that, on average, a slight increase in LR values (~ 20-fold) was observed when applying to kinship cases the cognate (nation-specific) rather than non-cognate haplotype database and the "LR exact" model, but several cases in which the use of different haplotype databases lead to a > 100-fold divergence in LR estimates were noted (Fig. S3). When applying the "LR(LE)" model, on the contrary, differences between cognate and non-cognate LR values were less pronounced and constantly within two orders of magnitude (Fig. S3). While more conservative, the use of the "LR(LE)" model appears questionable in presence of allelic dependencies in the tested populations, a phenomenon observed in the present study and in other Horn of Africa populations [42], [55]. Further collection of haplotype data in this area, covering previously untested populations, is therefore advisable for more reliable LR estimations in kinship cases requiring X-STR analysis.

4. Concluding remarks

The reference database of aSTR allele frequencies, Y-STR and X-STR haplotype frequencies established in the Tigray population of Ethiopia will hopefully contribute to the local development of forensic genetics services, assisting weight of evidence calculations in identification and kinship cases. Identification procedures of victims of human trafficking in the Mediterranean Sea will also benefit from the availability of novel data on forensic STR variation in the Horn of Africa, a geographic area that has been little studied in the past and is currently a major point of origin of individuals involved in migration accidents.

Acknowledgments

This work was supported by: Fondazione CRT, Italy, WWS2 Project mobility grant to C.R.; Fondazione CRT, Italy, Ordinary Grant Program 2017.0980 to G.D.V.

References

[1] K.L. O'Connor, C.R. Hill, P.M. Vallone, J.M. Butler, Linkage disequilibrium analysis of D12S391 and vWA in U.S. population and paternity samples, Forensic Sci. Int. Genet. 5 (2011) 538–540, https://doi.org/10.1016/j.fsigen.2010.09.003.

[2] K.E. Holsinger, B.S. Weir, Genetics in geographically structured populations: defining, estimating and interpreting F(ST), Nat. Rev. Genet. 10 (2009) 639–650, https://doi.org/10.1038/nrg2611.

[3] L. Roewer, Y chromosome STR typing in crime casework, Forensic Sci. Med. Pathol. 5 (2009) 77–84, https://doi.org/10.1007/s12024-009-9089-5.

[4] A.O. Tillmar, D. Kling, J.M. Butler, W. Parson, M. Prinz, P.M. Schneider, T. Egeland, L. Gusm˜ ao, DNA Commission of the International Society for Forensic Genetics (ISFG): guidelines on the use of X-STRs in kinship analysis, Forensic Sci. Int. Genet. 29 (2017) 269–275, https://doi.org/10.1016/j.fsigen.2017.05.005.

[5] W.H. Li, S. Yi, K. Makova, Male-driven evolution, Curr. Opin. Genet. Dev. 12 (2002) 650–656, https://doi.org/10.1016/s0959-437x(02)00354-4.

[6] S.F. Schaffner, The X chromosome in population genetics, Nat. Rev. Genet. 5 (2004) 43–51, https://doi.org/10.1038/nrg1247.

[7] S. Fan, D.E. Kelly, M.H. Beltrame, M.E.B. Hansen, S. Mallick, A. Ranciaro, J. Hirbo, S. Thompson, W. Beggs, T. Nyambo, S.A. Omar, D.W. Meskel, G. Belay, A. Froment, N. Patterson, D. Reich, S.A. Tishkoff, African evolutionary history inferred from whole genome sequence data of 44 indigenous African populations, Genome Biol. 20 (2019) 82, https://doi.org/10.1186/s13059-019-1679-2.

[8] B. Lorente-Galdos, O. Lao, G. Serra-Vidal, G. Santpere, L.F.K. Kuderna, L.R. Arauna, K. Fadhlaoui-Zid, V.N. Pimenoff, H. Soodyall, P. Zalloua, T. Marques-Bonet, D. Comas, Whole-genome sequence analysis of a Pan African set of samples reveals archaic gene flow from an extinct basal population of modern humans into subSaharan populations, Genome Biol. 20 (2019) 77, https://doi.org/10.1186/ s13059-019-1684-5.

[9] G. Sirugo, S.M. Williams, S.A. Tishkoff, The missing diversity in human genetic studies, Cell 177 (2019) 26–31, https://doi.org/10.1016/j.cell.2019.02.048.

[10] L. Pagani, T. Kivisild, A. Tarekegn, R. Ekong, C. Plaster, I. Gallego Romero, Q. Ayub, S.Q. Mehdi, M.G. Thomas, D. Luiselli, E. Bekele, N. Bradman, D. J. Balding, C. Tyler-Smith, Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool, Am. J. Hum. Genet. 91 (2012) 83–96, https://doi.org/10.1016/j.ajhg.2012.05.015.

[11] D. Gurdasani, T. Carstensen, F. Tekola-Ayele, L. Pagani, I. Tachmazidou, K. Hatzikotoulas, S. Karthikeyan, L. Iles, M.O. Pollard, A. Choudhury, G.R. S. Ritchie, Y. Xue, J. Asimit, R.N. Nsubuga, E.H. Young, C. Pomilla, K. Kivinen, K. Rockett, A. Kamali, A.P. Doumatey, G. Asiki, J. Seeley, F. Sisay-Joof, M. Jallow, S. Tollman, E. Mekonnen, R. Ekong, T. Oljira, N. Bradman, K. Bojang, M. Ramsay, A. Adeyemo, E. Bekele, A. Motala, S.A. Norris, F. Pirie, P. Kaleebu, D. Kwiatkowski, C. Tyler-Smith, C. Rotimi, E. Zeggini, M.S. Sandhu, The African genome variation project shapes medical genetics in Africa, Nature 517 (2015) 327–332, https://doi. org/10.1038/nature13997.

[12] Census 2007. 〈https://www.statsethiopia.gov.et/census-2007-2/〉. (Accessed 16 March 2021).

[13] EASO Issues Country of Origin Information report on Eritrea: Country Focus, European Asylum Support Office, 2016. 〈https://www.easo.europa.eu/news -events/easo-issues-country-origin-information-report-eritrea-country-focus〉. (Accessed 16 March 2021).

[14] T. Tau, A. Wally, T.P. Fanie, G.L. Ngono, S.W. Mpoloka, S. Davison, M.E. D'Amato, Genetic variation and population structure of Botswana populations as identified with AmpFLSTR Identifiler short tandem repeat (STR) loci, Sci. Rep. 7 (2017) 6768, https://doi.org/10.1038/s41598-017-06365-y.

[15] S.T.D. Ellingham, P. Perich, M. Tidball-Binz, The fate of human remains in a maritime context and feasibility for forensic humanitarian action to assist in their recovery and identification, Forensic Sci. Int. 279 (2017) 229–234, https://doi. org/10.1016/j.forsciint.2017.07.039.

[16] International Organization for Migration. 〈https://www.iom.int/newsdesk/ 20170721〉. (Accessed 16 March 2021).

[17] L. Olivieri, D. Mazzarelli, B. Bertoglio, D. De Angelis, C. Previder`e, P. Grignani, A. Cappella, S. Presciuttini, C. Bertuglia, P. Di Simone, N. Polizzi, A. Iadicicco, V. Piscitelli, C. Cattaneo, Challenges in the identification of dead migrants in the Mediterranean: the case study of the Lampedusa shipwreck of October 3rd 2013, Forensic Sci. Int. 285 (2018) 121–128, https://doi.org/10.1016/j. forsciint.2018.01.029.

[18] M.E. D'Amato, M. Bodner, J.M. Butler, L. Gusm˜ ao, A. Linacre, W. Parson, P. M. Schneider, P. Vallone, A. Carracedo, Ethical publication of research on genetics and genomics of biological material: guidelines and recommendations, Forensic Sci. Int. Genet. 48 (2020), 102299, https://doi.org/10.1016/j.fsigen.2020.102299.

[19] C. Tomas, V. Pereira, N. Morling, Analysis of 12 X-STRs in Greenlanders, Danes and Somalis using Argus X-12, Int. J. Leg. Med. 126 (2012) 121–128, https://doi.org/ 10.1007/s00414-011-0609-y.

[20] Ø. Bleka, G. Storvik, P. Gill, EuroForMix: an open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts, Forensic Sci. Int. Genet. 21 (2016) 35–44, https://doi.org/10.1016/ j.fsigen.2015.11.008.

[21] D.J. Balding, R.A. Nichols, DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands, Forensic Sci. Int. 64 (1994) 125–140, https://doi.org/10.1016/0379-0738(94) 90222-4.

[22] L. Excoffier, H.E.L. 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, https://doi.org/10.1111/j.1755-0998.2010.02847.x.

[23] R. Core Team, R: The R Project for Statistical Computing, 2017. 〈https://www. r-project.org/〉. (Accessed 16 March 2021).

[24] D. Kling, A. Tillmar, T. Egeland, P. Mostad, A general model for likelihood computations of genetic marker data accounting for linkage, linkage disequilibrium, and mutations, Int. J. Leg. Med. 129 (2015) 943–954, https://doi. org/10.1007/s00414-014-1117-7.

[25] M. Nothnagel, R. Szibor, O. Vollrath, C. Augustin, J. Edelmann, M. Geppert, C. Alves, L. Gusm˜ ao, M. Vennemann, Y. Hou, U.-D. Immel, S. Inturri, H. Luo, S. Lutz-Bonengel, C. Robino, L. Roewer, B. Rolf, J. Sanft, K.-J. Shin, J.E. Sim, P. Wiegand, C. Winkler, M. Krawczak, S. Hering, Collaborative genetic mapping of 12 forensic short tandem repeat (STR) loci on the human X chromosome, Forensic Sci. Int. Genet. 6 (2012) 778–784, https://doi.org/10.1016/j.fsigen.2012.02.015.

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

[27] M. Bodner, I. Bastisch, J.M. Butler, R. Fimmers, P. Gill, L. Gusmao, ˜ N. Morling, C. Phillips, M. Prinz, P.M. Schneider, W. Parson, Recommendations of the DNA commission of the international society for forensic genetics (ISFG) on quality control of autosomal short tandem repeat allele frequency databasing (STRidER), Forensic Sci. Int. Genet. 24 (2016) 97–102, https://doi.org/10.1016/j. fsigen.2016.06.008.

[28] L. Roewer, The Y-chromosome haplotype reference database (YHRD) – publicly available reference and research datasets for the forensicinterpretation of Ychromosome STR profiles. Handbook of Forensic Genetics, World Scientific, Europe, 2015, pp. 231–248, https://doi.org/10.1142/9781786340788_0012.

[29] B. Dupuy, M. Stenersen, D. Kling, Frequency data for 35 autosomal STR markers in a Norwegian, an East African, an East Asian and Middle Asian population and simulation of adequate database size, Forensic Sci. Int. Genet. Suppl. Ser. 4 (2013) e378–379.

[30] B. Bertoglio, P. Grignani, P. Di Simone, N. Polizzi, D. De Angelis, C. Cattaneo, A. Iadicicco, P. Fattorini, S. Presciuttini, C. Previder`e, Disaster victim identification by kinship analysis: the Lampedusa October 3rd, 2013 shipwreck, Forensic Sci. Int. Genet. 44 (2020), 102156, https://doi.org/10.1016/j.fsigen.2019.102156.

[31] C. Phillips, L. Fernandez-Formoso, M. Garcia-Magarinos, ˜ L. Porras, T. Tvedebrink, J. Amigo, M. Fondevila, A. Gomez-Tato, J. Alvarez-Dios, A. Freire-Aradas, A. Gomez-Carballa, A. Mosquera-Miguel, A. Carracedo, M.V. Lareu, Analysis of global variability in 15 established and 5 new European standard set (ESS) STRs using the CEPH human genome diversity panel, Forensic Sci. Int. Genet. 5 (2011) 155–169, https://doi.org/10.1016/j.fsigen.2010.02.003.

[32] C.R. Steffen, M.D. Coble, K.B. Gettings, P.M. Vallone, Corrigendum to “U.S. population data for 29 autosomal STR loci”, Forensic Sci. Int. Genet. 31 (2017) e36–e40, https://doi.org/10.1016/j.fsigen.2017.08.011.

[33] D.E. Krane, R.W. Allen, S.A. Sawyer, D.A. Petrov, D.L. Hartl, Genetic differences at four DNA typing loci in Finnish, Italian, and mixed Caucasian populations, Proc. Natl. Acad. Sci. USA 89 (1992) 10583–10587, https://doi.org/10.1073/ pnas.89.22.10583.

[34] A.L. Lowe, A. Urquhart, L.A. Foreman, I.W. Evett, Inferring ethnic origin by means of an STR profile, Forensic Sci. Int. 119 (2001) 17–22, https://doi.org/10.1016/ s0379-0738(00)00387-x.

[35] S.A. Tishkoff, F.A. Reed, F.R. Friedlaender, C. Ehret, A. Ranciaro, A. Froment, J. B. Hirbo, A.A. Awomoyi, J.-M. Bodo, O. Doumbo, M. Ibrahim, A.T. Juma, M. J. Kotze, G. Lema, J.H. Moore, H. Mortensen, T.B. Nyambo, S.A. Omar, K. Powell, G.S. Pretorius, M.W. Smith, M.A. Thera, C. Wambebe, J.L. Weber, S.M. Williams, The genetic structure and history of Africans and African Americans, Science 324 (2009) 1035–1044, https://doi.org/10.1126/science.1172257.

[36] H.R.S. Kumar, K. Haddish, D. Lacerenza, S. Aneli, C. Di Gaetano, G. Tewelemedhin, R.V. Manukonda, N. Futwi, V. Alvarez-Iglesias, M. de la Puente, M. Fondevila, M. V. Lareu, C. Phillips, C. Robino, Characterization of ancestry informative markers in the Tigray population of Ethiopia: a contribution to the identification process of dead migrants in the Mediterranean Sea, Forensic Sci. Int. Genet. 45 (2020), 102207, https://doi.org/10.1016/j.fsigen.2019.102207.

[37] C.D. Steele, D.S. Court, D.J. Balding, Worldwide F(ST) estimates relative to five continental-scale populations, Ann. Hum. Genet. 78 (2014) 468–477, https://doi. org/10.1111/ahg.12081.

[38] J. Buckleton, J. Curran, J. Goudet, D. Taylor, A. Thiery, B.S. Weir, Populationspecific FST values for forensic STR markers: a worldwide survey, Forensic Sci. Int. Genet. 23 (2016) 91–100, https://doi.org/10.1016/j.fsigen.2016.03.004.

[39] B. Budowle, J. Ge, R. Chakraborty, H. Gill-King, Use of prior odds for missing persons identifications, Investig. Genet. 2 (2011) 15, https://doi.org/10.1186/ 2041-2223-2-15.

[40] G. Ferri, C. Robino, M. Alù, D. Luiselli, S. Tofanelli, L. Caciagli, V. Onofri, S. Pelotti, C. Di Gaetano, F. Crobu, G. Beduschi, C. Capelli, Molecular characterisation and population genetics of the DYS458.2 allelic variant, Forensic Sci. Int. Genet. Suppl. Ser. 1 (2008) 203–205, https://doi.org/10.1016/j.fsigss.2007.10.217.

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

[42] C. Bini, S. Sarno, E. Tangorra, A. Iuvaro, S. De Fanti, Y.G. Tseghereda, S. Pelotti, D. Luiselli, Haplotype data and forensic evaluation of 23 Y-STR and 12 X-STR loci in eight ethnic groups from Eritrea, Int. J. Leg. Med. 135 (2021) 449–453, https:// doi.org/10.1007/s00414-020-02446-2.

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

[44] Y.M. Khubrani, J.H. Wetton, M.A. Jobling, Extensive geographical and social structure in the paternal lineages of Saudi Arabia revealed by analysis of 27 YSTRs, Forensic Sci. Int. Genet. 33 (2018) 98–105, https://doi.org/10.1016/j. fsigen.2017.11.015.

[45] I.M. Lewis, P.A. Jewell, The peoples and cultures of Ethiopia [and discussion], Proc. R. Soc. Lond. B Biol. Sci. 194 (1976) 7–16.

[46] E.I. Gebremeskel, M.E. Ibrahim, Y-chromosome E haplogroups: their distribution and implication to the origin of Afro-Asiatic languages and pastoralism, Eur. J. Hum. Genet. EJHG 22 (2014) 1387–1392, https://doi.org/10.1038/ejhg.2014.41.

[47] L.R. Arauna, J. Mendoza-Revilla, A. Mas-Sandoval, H. Izaabel, A. Bekada, S. Benhamamouch, K. Fadhlaoui-Zid, P. Zalloua, G. Hellenthal, D. Comas, Recent historical migrations have shaped the gene pool of Arabs and Berbers in North Africa, Mol. Biol. Evol. 34 (2017) 318–329, https://doi.org/10.1093/molbev/ msw218.

[48] F. Cruciani, R. La Fratta, B. Trombetta, P. Santolamazza, D. Sellitto, E.B. Colomb, J.-M. Dugoujon, F. Crivellaro, T. Benincasa, R. Pascone, P. Moral, E. Watson, B. Melegh, G. Barbujani, S. Fuselli, G. Vona, B. Zagradisnik, G. Assum, R. Brdicka, A.I. Kozlov, G.D. Efremov, A. Coppa, A. Novelletto, R. Scozzari, Tracing past human male movements in northern/eastern Africa and western Eurasia: new clues from Y-chromosomal haplogroups E-M78 and J-M12, Mol. Biol. Evol. 24 (2007) 1300–1311, https://doi.org/10.1093/molbev/msm049.

[49] K.K. Abu-Amero, A. Hellani, A.M. Gonzalez, ´ J.M. Larruga, V.M. Cabrera, P. A. Underhill, Saudi Arabian Y-chromosome diversity and its relationship with nearby regions, BMC Genet. 10 (2009) 59, https://doi.org/10.1186/1471-2156- 10-59.

[50] S. Lopez, ´ A. Tarekegn, G. Band, L. van Dorp, N. Bird, S. Morris, T. Oljira, E. Mekonnen, E. Bekele, R. Blench, M.G. Thomas, N. Bradman, G. Hellenthal, Evidence of the interplay of genetics

and culture in Ethiopia, Nat. Commun. 12 (2021) 3581, https://doi.org/10.1038/s41467-021-23712-w.

[51] M. Kayser, O. Lao, K. Anslinger, C. Augustin, G. Bargel, J. Edelmann, S. Elias, M. Heinrich, J. Henke, L. Henke, C. Hohoff, A. Illing, A. Jonkisz, P. Kuzniar, A. Lebioda, R. Lessig, S. Lewicki, A. Maciejewska, D.M. Monies, R. Pawłowski, M. Poetsch, D. Schmid, U. Schmidt, P.M. Schneider, B. Stradmann-Bellinghausen, R. Szibor, R. Wegener, M. Wozniak, M. Zoledziewska, L. Roewer, T. Dobosz, R. Ploski, Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y-chromosome analysis, Hum. Genet. 117 (2005) 428–443, https://doi.org/10.1007/s00439-005-1333-9.

[52] 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, https://doi. org/10.1016/j.fsigen.2014.02.007.

[53] I. Gomes, P.J.P. Pereira, S. Harms, A.M. Oliveira, P.M. Schneider, A. Brehm, Genetic characterization of Guinea-Bissau using a 12 X-chromosomal STR system: inferences from a multiethnic population, Forensic Sci. Int. Genet. 31 (2017) 89–94, https://doi.org/10.1016/j.fsigen.2017.08.016.

[54] I. Gomes, N. Pinto, S. Ant˜ ao-Sousa, V. Gomes, L. Gusmao, ˜ A. Amorim, Twenty years later: a comprehensive review of the X chromosome use in forensic genetics, Front. Genet. 11 (2020) 926, https://doi.org/10.3389/fgene.2020.00926.

[55] D. Kling, B. Dell'Amico, P.J.T. Haddeland, A.O. Tillmar, Population genetic analysis of 12 X-STRs in a Somali population sample, Forensic Sci. Int. Genet. 11 (2014) e7–8, https://doi.org/10.1016/j.fsigen.2014.02.013.

[56] M.A. Almarri, R.A. Lootah, Allelic and haplotype diversity of 12 X-STRs in the United Arab Emirates, Forensic Sci. Int. Genet. 33 (2018) e4–e6, https://doi.org/ 10.1016/j.fsigen.2017.12.013.

[57] F.C. Ceballos, S. Hazelhurst, M. Ramsay, Runs of homozygosity in Sub-Saharan African populations provide insights into complex demographic histories, Hum. Genet. 138 (2019) 1123–1142, https://doi.org/10.1007/s00439-019-02045-1.

Table 1. Distribution of *d* values. No F_{ST} correction was applied to RMPs obtained with cognate Tigray allele frequencies.

Non-cognate database	F_{ST}	*d* > 3	*d* > 2	*d* > 1	*d* > 0	Mean *d*	± SD
Horn of Africa	0.00	0.85%	14.4%	41.4%	86.9%	0.935	0.956
Horn of Africa	0.01	0.0%	0.00%	0.00%	13.1%	-0.978	0.881
Horn of Africa	0.03	0.00%	0.00%	0.00%	0.00%	-3.386	1.094
HGDP sub-Saharan Africans	0.00	43.2%	68.6%	86.0%	96.2%	2.761	1.513
HGDP sub-Saharan Africans	0.01	2.97%	17.8%	42.8%	71.2%	0.652	1.381
HGDP sub-Saharan Africans	0.03	0.00%	0.00%	0.42%	5.08%	-2.082	1.429
African Americans	0.00	22.9%	51.3%	78.8%	93.2%	2.041	1.362
African Americans	0.01	0.42%	1.69%	14.8%	48.7%	-0.115	1.162
African Americans	0.03	0.00%	0.00%	0.00%	0.85%	-2.719	1.270

Fig. 1. MDS plots obtained from Slatkin's linearized F_{ST} distances calculated from Y-STR haplotypes of all tested populations (a), and within Horn of Africa populations subdivided by language group and nationality (b). Abbreviations for populations which were compared to TIGRAY (this study) were as follows: Central Saudi Arabian (SA-C), Eastern Saudi Arabian (SA_E), Northern Saudi Arabian (SA_N), Southern Saudi Arabian (SA_S), Western Saudi Arabian (SA_W) from [44]; Asni Berbers (Morocco) (NA_AB), Berbers from Bouhria (Morocco) (NA_BB), Egyptians from Bahariya (NA_EB), Egyptian Berbers from Siwa (NA_ES), Libyan Arabs (NA_LA), Libyan Jews (NA_LJ), Mozabite Berbers (Algeria) (NA_MB), Moroccan Jews (NA_MJ), Northern Egyptians (NA_NE), Ouarzazate Berbers (Morocco) (NA_OB), Souss Berbers (Morocco) (NA_SB) from [43]; Afar (EA_AF2), Bilen (EA_BI), Hidareb (EA_HI), Kunama (EA_KU2), Nara (EA_NA2), Saho (EA_SA2), Tigre (EA_TG), Tigray (EA_TY2) from [42]; Afar (EA_AF1), Amhara (EA_AM), Ethiopians Jews (EA_EJ), Cunama (EA_KU1), Nara (EA_NA1), Oromo (EA_OR), Saho (EA_SA1), Djibuti Somali (EA_SD), Ethiopian Somali (EA_SE), Tigray (EA_TY1), Wolayta (EA_WO) from [41]. Details of tested populations are given in Table S3A.

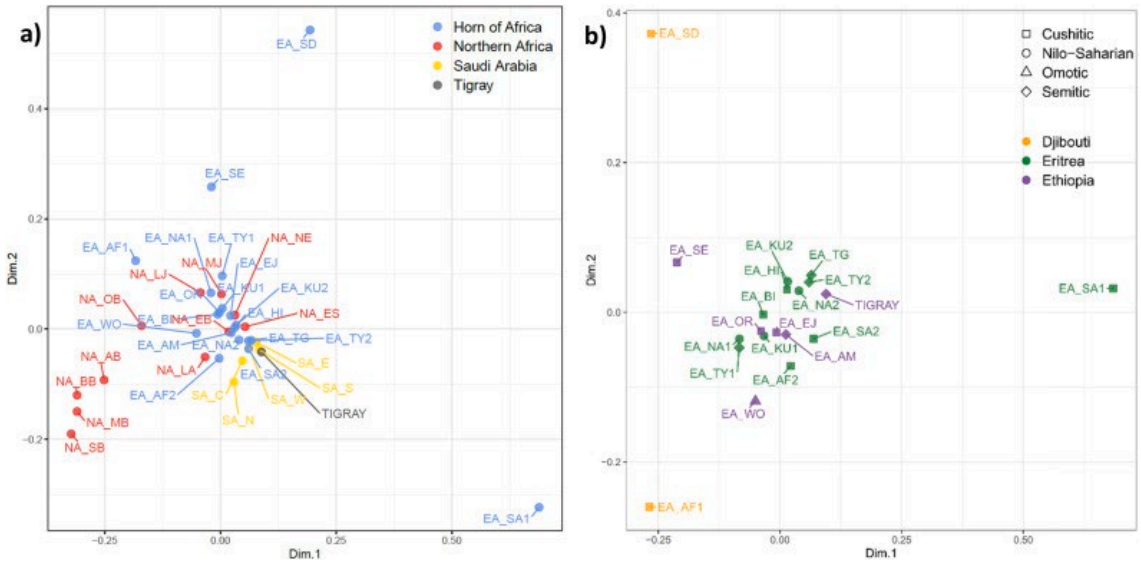

Fig. 2. Sequencing analysis of region 150404023–150404045 of GRCh38 in Tigray samples indicates that different primer binding site mutations can explain amplification failure observed at locus DXS10146. The sequence of a sample normally amplified with the Investigator Argus X-12 kit is shown in (a). A sample with a CTTT deletion (150404032–150404035), displaying a null allele after amplification with Investigator Argus X-12, but successfully amplified with Investigator Argus X-12 QS, is shown in (b). Finally, underlined in (c) is a base substitution (C→T) at position 150404028, that was observed in Tigray samples affected by amplification failure with both Investigator Argus X-12 and Investigator Argus X-12 QS kits.