



Haplotype data and forensic evaluation of 23 Y-STR and 12 X-STR loci in eight ethnic groups from Eritrea

Carla Bini¹ · Stefania Sarno² · Elisabetta Tangorra¹ · Alessandra Iuvaro¹ · Sara De Fanti^{2,3} ·
Yohannes Ghebremedhin Tseghereda⁴ · Susi Pelotti¹ · Donata Luiselli^{2,5}

Received: 3 August 2020 / Accepted: 14 October 2020 / Published online: 22 October 2020
© The Author(s) 2020

Abstract

Eritrea is a multi-ethnic country of over 3 million of people consisting of different ethnic groups, having each its own language and cultural tradition. Due to the lack of population genetic data for markers of forensic interest, in this study, we analyzed the genetic polymorphisms of 23 Y-chromosome STR loci and of 12 X-chromosome STR loci in a sample of 255 unrelated individuals from 8 Eritrean ethnic groups, with the aim to generate a reference haplotype database for anthropological and forensic applications. X- and Y-chromosomes markers may indeed offer information especially in personal identification and kinship testing, when relying on the availability of large local population data to derive sufficiently accurate frequency estimates. The population genetic analyses in the Eritrean sample for both the two set of Y- and X-STR markers showed high power of discrimination both at country-based and population levels. Comparison population results highlight the importance of considering the ethnic composition within the analyzed country and the necessity of increasing available data especially when referring to heterogeneous populations such as the African ones.

Keywords Y-STRs · X-STRs · Eritrean ethnic groups · Forensic parameters · Population genetics

Eritrea is a multi-ethnic country located in the Horn of Africa with nine officially recognized ethnic groups in its population

of over 3 million people [1]. Between Afro-Asiatic communities, Tigrinya make up about 55% of the population, and the

Carla Bini and Stefania Sarno contributed equally to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00414-020-02446-2>) contains supplementary material, which is available to authorized users.

✉ Carla Bini
carla.bini@unibo.it

✉ Susi Pelotti
susi.pelotti@unibo.it

Stefania Sarno
stefania.samo2@unibo.it

Elisabetta Tangorra
elisabetta.tangorra@unibo.it

Alessandra Iuvaro
ales.iuv@gmail.com

Sara De Fanti
sara.defanti2@unibo.it

Yohannes Ghebremedhin Tseghereda
seghereda@gmail.com

Donata Luiselli
donata.luiselli@unibo.it

- ¹ Department of Medical and Surgical Sciences, Unit of Legal Medicine, University of Bologna, via Imerio, 49, 40126 Bologna, Italy
- ² Department of Biological, Geological and Environmental Sciences-Lab. of Molecular Anthropology & Centre for Genome Biology, University of Bologna, Via Selmi 3, 40126 Bologna, Italy
- ³ Interdepartmental Centre “Alma Mater Research Institute on Global Challenges and Climate Change (Alma Climate)”, University of Bologna, Bologna, Italy
- ⁴ Jomo Kenyatta University and Agriculture and Technology (JKUAT), P.O. Box 62 000, Nairobi 00200, Kenya
- ⁵ Department of Cultural Heritage, Ravenna Campus, University of Bologna, Via degli Ariani 1, 48121 Ravenna, Italy

Tigre constitute around 30% of the residents, whereas Rashaida are one of the smallest ethnic groups representing ~2% of the population. Most of the rest of the population belong to the other Afro-Asiatic-speaking communities of the Cushitic branch, such as the Afar, Bilen, Hedareb, and Saho. The Kunama and Nara populations represent small Nilo-Saharan ethnic groups in the country [2].

Population genetic data for markers of forensic interest are still limited for some populations, in particular of African origin. As concern the Eritrean population, a previous study on the Y-chromosome variability in East-Africa countries analyzed a total of 161 individuals belonging to 5 different ethnic groups of Eritrea [3], while, to the best of our knowledge, no data are available for X-chromosome markers. X- and Y-chromosomes markers may provide useful information especially in personal identification and kinship testing, when relying on the availability of large local population data to derive sufficiently accurate frequency estimates.

With the aim to generate a relevant reference database for anthropological and forensic statistical evaluations, we analyzed the genetic polymorphisms of 23 Y-chromosome STR loci and of 12 X-chromosome STR loci in a sample of 255 unrelated individuals from 8 Eritrean ethnic groups. In particular, the Eritrean dataset was sampled based on ethnolinguistic information and consists of Afar-Cushitic ($n = 21$), Bilen-Cushitic ($n = 15$), Hedareb-Cushitic ($n = 15$), Kunama Nilo-Saharan ($n = 12$), Nara Nilo-Saharan ($n = 33$), Saho-Cushitic ($n = 21$), Tigre-Semitic ($n = 62$), and Tigrinya-Semitic ($n = 76$) groups.

DNA anonymous samples had been stored in the laboratory without any individual associated information, and the markers analyzed in this study identify only the geographical distribution of genetic lineages. The present research was approved by the Bioethical Committee of the University of Bologna.

All samples were amplified for the 23 Y-STRs included in the PowerPlex® Y23 System (Promega) as well as for the 12 X-STRs loci included in the Investigator Argus X-12 kit (Qiagen), following the manufacturer recommended protocols. PCR products were separated and detected on an ABI PRISM 310 Genetic Analyzer using POP-4 polymer; alleles were called and binned by GeneMapper ID v3.2 software (Thermo Fisher Scientific) according to the manufacturer's instructions.

Haplotype data were submitted to the Y-chromosomal Haplotype Reference Database (YHRD, www.yhrd.org) [4], and the following accession numbers were assigned: YA004649 for Afar, YA004650 for Bilen, YA004651 for Hedareb, YA004652 for Saho, YA004653 for Kunama, YA004654 for Nara, YA004655 for Tigre, and YA004656 for Tigrinyas.

Arlequin v3.5.1.3 software [5] was used to calculate Y-chromosome standard diversity index, haplotype diversity

(HD), average gene diversity over loci, and mean number of pairwise differences, and the forensic parameters, match probability (MP) and discrimination capacity (DC). Pairwise genetic distances (R_{st} values) based on Y-chromosome STR data were calculated between the Eritrean population and a wide set of reference comparison groups [6] and used to generate a multidimensional scaling (MDS) plot with the software R (library MASS) [7].

For X-chromosome loci, allele frequencies, haplotype frequencies for each linkage group (LG), and forensic efficiency parameters (gene diversity, GD; polymorphism information content, PIC; power of discrimination, PD; mean exclusion chance, MEC) were generated using the StatsX v2.0 software as described in [8]. Linkage disequilibrium (LD) in the male population sample was estimated by using the Arlequin software [5]. Inter-population F_{st} genetic distances based on haplotype frequencies between the Eritrean dataset and 13 comparison populations from Europe, Asia, and Africa [9–18] were integrated for each of the four X-STRs linkage groups and graphically represented by using the R software package DISTATIS [19].

In our population sample, we found 238 different Y-haplotypes, of which 221 were unique, with the other 17 instead shared by pairs of two subjects (Table S1). Excepting a single case of sharing between Hedareb and Kunama populations, all the subjects that shared the same haplotype also belonged to the same ethnic group. Allele frequencies for the 255 Eritrean samples are reported in Table S2. One null allele at the DYS19 locus and one duplication at the DYS439 locus were observed in two different individuals belonging respectively to the Nara and Tigre ethnic groups. The microvariant alleles 15.2, 17.2, 18.2, 19.2, 20.2, and 21.2, not included in the bin set of the allelic ladder, were observed at locus DYS458 in twenty-nine samples. The microvariant allele 17.1 at locus DYS385a/b and the rare variant allele 4 at locus DYS643 were respectively found in two different samples from Tigre. All these variants are already reported in locus information on YHRD [4].

Diversity indexes and forensic parameters for the PYPY23 markers set are shown in Table S3. Overall, the typing of Y-chromosome loci revealed low cases of haplotype sharing, resulting in relative high values of haplotype diversity and good discrimination capacity. At a population level, the Eritrean ethnic group that showed the lower values of both HD and DC was the Cushitic-speaking population of Saho, as already reported in a previous study [3].

The data for the Eritrea population newly reported in the present study were compared with those of other 20 Eastern African population ethnic groups from Eritrea, Ethiopia, Djibouti, and Kenya already available from the literature. When compared by using all the 21 Y-STRs loci common to both studies, the considered countries showed significant R_{st} genetic distances among each other, thus confirming the presence of a notable genetic sub-structure within East Africa

as described previously [3]. Importantly, significant *Rst* values were observed not only at a country-based level but also at a population-level when considering pairwise comparisons among single ethnic groups (Table S4). While not excluding different sample sizes and sampling strategies, part of the differences observed between the Eritrean samples presented by this and the previous study could however be driven by the outlying genetic composition of the Saho ethnic group. Indeed, the genetic differences at a country-based level between the two Eritrean population samples become close to zero and not significant once excluding this ethnic group from the comparison. Accordingly, when we directly compared the haplotype composition of single ethnic groups which were sampled in both of the studies, significant genetic differences were found among the two Saho samples, but not in the comparisons between the two Kunama, the two Nara or the two Tigre collected populations. This may be related to the peculiar genetic composition already described for the Saho sample [3], that was indeed shown to be characterized by the lowest values of intra-population diversity parameters and by high frequencies (~88%) of a peculiar Y-chromosome haplogroup (E-V22) which is relatively uncommon in the other African populations.

A MDS analysis was further performed on the linearized *Rst* genetic distances calculated by considering the Eritrean population and 77 comparison groups from Africa, Europe, and the Middle East for which a comparable level of analysis for all the 23 Y-chromosome STR loci was available [6]. The first MDS dimension clearly separates the African groups from the non-African ones (Fig. S1). Along this axis of variation, the Eritrean population particularly clusters with the East African Kenyan Maasai, being located in an intermediate position between populations from Central and South Africa (Yoruba, Zimbabwe, and Xhosa) on one hand and those from the Middle East and South-Eastern Europe on the other hand. A North-to-South gradient of genetic variation finally characterizes the pattern observed within Europe along the second MDS dimension.

As concern X-STR results, twenty-two out of the 255 analyzed males showing a locus dropout—of which 16 samples for the DXS10148 and 6 for the DXS10146 locus, respectively—were excluded from statistical analyses. Presumably, these silent alleles are due to one or more mutations in the primer binding sites, which reduce the efficiency of the PCR reaction, as previously observed for the African populations [11, 14]. Indeed, amplification performed with higher DNA concentration showed small off-ladder peaks in DXS10148 system that need to be further investigated, by also considering the new sequence variant identified for this locus [20]. Furthermore, a total of 32 off-ladder alleles in the linkage groups I, II, and IV were detected and designated according to their base pair sizes (Table S5). Overall, the allele and haplotype frequencies for the X-STRs loci and corresponding

linkage groups (LGs) are detailed in Table S6 and Table S7, respectively.

The number of observed haplotypes and the forensic parameters calculated for each of the four LGs are reported in Table S8. The most informative linkage group is LG1 (PIC = 0.9936) including the DXS10135 marker that showed the highest forensic efficiency (PIC = 0.94) with 30 different alleles typed. The least polymorphic locus was instead the DXS7423 counting for only 5 alleles. As expected, evidence of significant linkage disequilibrium (LD) was found between markers within the same LG, but significant associations were also observed between loci of LG2/LG4 and LG3/LG4 groups and more precisely for DXS10135-DXS7423 and for DXS10079-DXS10146 pairs (Table S9) probably related to the population history, such as the presence of population substructure, non-random mating, or local genetic drift. Haplotype frequencies for the four X-chromosome linkage groups were finally used to summarize the relationships between Eritreans and other populations from Europe, Asia, and Africa. The DISTATIS plot in Fig. S2 generally shows a pattern of genetic structuring that overall resembles the geographic distribution of the considered populations according to their continent of origin, placing the Eritreans between European and North-African populations on one hand and the other West-African groups on the other and thus reflecting the pattern also shown by PPY23 system analysis.

In conclusion, the present study provides a contribution for population databasing, adding new allele and haplotype frequency estimates at both Y-STR and X-STR loci from a population not easily accessible for sampling. The typing of 255 male individuals largely extends previous Y-chromosome STRs data available for the Eritrean population and constitutes a first reference dataset for X-chromosome STRs useful for statistical evaluation in forensic casework. Overall, our results remark the importance of implementing population databases by including representative data from local populations, especially when referring to heterogeneous groups such as the African ones.

This manuscript follows the recommendations of the ISFG on the use of YSTRs and X-STRs in forensic analysis [21, 22] and the guidelines for publication of population data requested by the journal [23].

Acknowledgements Open access funding provided by Alma Mater Studiorum - Università di Bologna within the CRUI-CARE Agreement.

Authors' contributions Carla Bini: Conceptualization, formal analysis, and writing—original draft preparation.

Stefania Sarno: Formal analysis and writing—reviewing and editing.

Elisabetta Tangorra: Validation.

Alessandra Iuvaro: Investigation.

Sara De Fanti: Resources.

Tseghereda Yohannes: Resources.

Susi Pelotti: Supervision, project administration, and writing—reviewing and editing.

Donata Luiselli: Supervision and writing—reviewing and editing.

Compliance with ethical standards

The study was conducted in compliance with ethical standards and was approved by the Bioethical Committee of the University of Bologna.

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- United Nations, Department of Economic and Social Affairs, Population Division (2019) World Population Prospects Volume II: Demographic Profiles (ST/ESA/SER.A/427). United Nations, New York
- Eritrea Population (2020) <https://worldpopulationreview.com/countries/eritrea-population>. Accessed on 28 Jul 2020
- Iacovacci G, D'Atanasio E, Marini O et al (2017) Forensic data and microvariant sequence characterization of 27 Y-STR loci analyzed in four Eastern African countries. *Forensic Sci Int Genet* 27:123–131. <https://doi.org/10.1016/j.fsigen.2016.12.015>
- Willuweit S, Roewer L (2015) The new Y chromosome haplotype reference database. *Forensic Sci Int Genet* 15:43–48. <https://doi.org/10.1016/j.fsigen.2014.11.024>
- Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10:564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>
- Purps J, Siegert S, Willuweit S, Nagy M, Alves C, Salazar R, Angustia SMT, Santos LH, Anslinger K, Bayer B, Ayub Q, Wei W, Xue Y, Tyler-Smith C, Bafalluy MB, Martínez-Jarreta B, Egyed B, Balitzki B, Tschumi S, Ballard D, Court DS, Barrantes X, Bäßler G, Wiest T, Berger B, Niederstätter H, Parson W, Davis C, Budowle B, Burri H, Borer U, Koller C, Carvalho EF, Domingues PM, Chamoun WT, Coble MD, Hill CR, Corach D, Caputo M, D'Amato ME, Davison S, Decorte R, Larmuseau MHD, Ottoni C, Rickards O, Lu D, Jiang C, Dobosz T, Jonkisz A, Frank WE, Furac I, Gehrig C, Castella V, Grskovic B, Haas C, Wobst J, Hadzic G, Drobnic K, Honda K, Hou Y, Zhou D, Li Y, Hu S, Chen S, Immel UD, Lessig R, Jakovski Z, Ilievska T, Klann AE, García CC, de Knijff P, Kraaijenbrink T, Kondili A, Miniati P, Vouropoulou M, Kovacevic L, Marjanovic D, Lindner I, Mansour I, al-Azem M, Andari AE, Marino M, Furfuro S, Locarno L, Martín P, Luque GM, Alonso A, Miranda LS, Moreira H, Mizuno N, Iwashima Y, Neto RSM, Nogueira TLS, Silva R, Nastainczyk-Wulf M, Edelmann J, Kohl M, Nie S, Wang X, Cheng B, Núñez C, Pancorbo MM, Olofsson JK, Morling N, Onofri V, Tagliabracci A, Pamjav H, Volgyi A, Barany G, Pawlowski R, Maciejewska A, Pelotti S, Pepinski W, Abreu-Glowacka M, Phillips C, Cárdenas J, Rey-Gonzalez D, Salas A, Brisighelli F, Capelli C, Toscanini U, Piccinini A, Piglionica M, Baldassarra SL, Ploski R, Konarzewska M, Jastrzebska E, Robino C, Sajantila A, Palo JU, Guevara E, Salvador J, Ungria MCD, Rodriguez JJR, Schmidt U, Schlauderer N, Saukko P, Schneider PM, Sirker M, Shin KJ, Oh YN, Skitsa I, Ampati A, Smith TG, Calvit LS, Stenzl V, Capal T, Tillmar A, Nilsson H, Turrina S, de Leo D, Verzeletti A, Cortellini V, Wetton JH, Gwynne GM, Jobling MA, Whittle MR, Sumita DR, Wolańska-Nowak P, Yong RYY, Krawczak M, Nothnagel M, Roewer L (2014) A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. *Forensic Sci Int Genet* 12:12–23. <https://doi.org/10.1016/j.fsigen.2014.04.008>
- Venables WN, Ripley BD (2002) Modern applied statistics with S, 4th edn. Springer, New York
- Lang Y, Guo F, Niu Q (2019) StatsX v2.0: the interactive graphical software for population statistics on X-STR. *Int J Legal Med* 133:39–44. <https://doi.org/10.1007/s00414-018-1824-6>
- Bini C, Riccardi LN, Ceccardi S, Carano F, Sarno S, Luiselli D, Pelotti S (2015) Expanding X-chromosomal forensic haplotype frequencies database: Italian population data of four linkage groups. *Forensic Sci Int Genet* 15:127–130. <https://doi.org/10.1016/j.fsigen.2014.11.008>
- Edelmann J, Lutz-Bonengel S, Naue J, Hering S (2012) X-chromosomal haplotype frequencies of four linkage groups using the Investigator Argus X-12 Kit. *Forensic Sci Int Genet* 6:e24–e34. <https://doi.org/10.1016/j.fsigen.2011.01.001>
- Tomas C, Pereira V, Morling N (2012) Analysis of 12 X-STRs in Greenlanders, Danes and Somalis using Argus X-12. *Int J Legal Med* 126:121–128. <https://doi.org/10.1007/s00414-011-0609-y>
- Zidkova A, Capek P, Horinek A, Coufalova P (2014) Investigator1 Argus X-12 study on the population of Czech Republic: comparison of linked and unlinked X-STRs for kinship analysis. *Electrophoresis* 35:1989–1992. <https://doi.org/10.1002/elps.201400046>
- Tillmar AO (2012) Population genetic analysis of 12 X-STRs in Swedish population. *Forensic Sci Int Genet* 6:e80–e81. <https://doi.org/10.1016/j.fsigen.2011.07.008>
- Elakkary S, Hoffmeister-Ullrich S, Schulze C, Seif E, Sheta A, Hering S, Edelmann J, Augustin C (2014) 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* 11:26–30. <https://doi.org/10.1016/j.fsigen.2014.02.007>
- Pasino S, Caratti S, Del Pero M, Santovito A, Torre C, Robino C (2011) Allele and haplotype diversity of X-chromosomal STRs in Ivory Coast. *Int J Legal Med* 125:749–752. <https://doi.org/10.1007/s00414-011-0591-4>
- Afonso Costa H, Morais P, Vieira da Silva C, Matos S, Marques Santos R, Espinheira R, Costa Santos J, Amorim A (2014) X-chromosome STR markers data in a Cabo Verde immigrant population of Lisboa. *Mol Biol Rep* 41:2559–2569. <https://doi.org/10.1007/s11033-014-3114-9>
- Uchigasaki S, Tie J, Takahashi D (2013) Genetic analysis of twelve X-chromosomal STRs in Japanese and Chinese populations. *Mol Biol Rep* 40:3193–3196. <https://doi.org/10.1007/s11033-012-2394-1>
- Samejima M, Nakamura Y, Nambiar P, Minaguchi K (2012) Genetic study of 12 X-STRs in Malay population living in and around Kuala Lumpur using Investigator Argus X-12 kit. *Int J Legal Med* 126:677–683. <https://doi.org/10.1007/s00414-012-0705-7>

19. Abdi H, Williams LJ, Valentin D, Bennani-Dosse M (2012) STATIS and DISTATIS: optimum multi-table principal component analysis and three-way metric multidimensional scaling. *Wiley Interdiscip Rev Comput Stat* 4:124–167. <https://doi.org/10.1002/wics.198>
20. Gomes I, Brehm A, Gusmão L, Schneider PM (2016) New sequence variants detected at DXS10148, DXS10074 and DXS10134 loci. *Forensic Sci Int Genet* 20:112–116. <https://doi.org/10.1016/j.fsigen.2015.10.005>
21. Gusmão L, Butler JM, Carracedo A, Gill P, Kayser M, Mayr WR, Morling N, Prinz M, Roewer L, Tyler-Smith C, Schneider PM (2006) DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. *Int J Legal Med* 120:191–200. <https://doi.org/10.1007/s00414-005-0026-1>
22. Tillmar AO, Kling D, Butler JM, Parson W, Prinz M, Schneider PM, Egeland T, Gusmão L (2017) DNA Commission of the International Society for Forensic Genetics (ISFG): guidelines on the use of X-STRs in kinship analysis. *Forensic Sci Int Genet* 29:269–275. <https://doi.org/10.1016/j.fsigen.2017.05.005>
23. Parson W, Roewer L (2010) Publication of population data of linearly inherited DNA markers in the *International Journal of Legal Medicine*. *Int J Legal Med* 124:505–509. <https://doi.org/10.1007/s00414-010-0492-y>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.