Forensic Science International: Genetics 76 (2025) 103232

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

X-chromosomal STRs: Metapopulations and mutation rates

L. Gusmão [a,1], S. Antão-Sousa [b,c,d,1], M. Faustino [c,d], M.A. Abovich [e], D. Aguirre [f], R. Alghafri [g], C. Alves [b], A. Amorim [b,c,d], C. Arévalo [h], L. Baldassarri [i], C. Barletta-Carrillo [j], G. Berardi [k], C. Bobillo [l], L. Borjas [m], D.F. Braganholi [n], A. Brehm [o], J.J. Builes [f], L. Cainé [p,q], E.F. Carvalho [a], M. Carvalho [r], L. Catelli [s], R.M.B. Cicarelli [n], A. Contreras [t], D. Corach [l], F.G. Di Marco [u], M.V. Diederiche [v], P. Domingues [a], M. Espinoza [w], J.M. Fernandéz [x], M.G. García [u], O. García [y], A. Gaviria [z], I. Gomes [b,c], D. Grattapaglia [aa], J. Henao [ab], A. Hernandez [ac], A.A. Ibarra [ad], G. Lima [p], I.M. Manterola [ae], C. Marrero [af], J.A. Martins [ag], L. Mendoza [f], A. Mosquera [ah], E.C. Nascimento [ai], V. Onofri [aj], M.M. Pancorbo [ak], J.J. Pestano [al], G. Plaza [am], M.J. Porto [r], Y.C. Posada [ad], M.L. Rebelo [p], E. Riego [an], R. Rodenbusch [ao], A. Rodríguez [w], A. Rodríguez [ah], P. Sanchez-Diz [ap], S. Santos [aq], F. Simão [a], L.M. Siza Fuentes [ar], D. Sumita [as], C. Tomas [at], U. Toscanini [k], A. Trindade-Filho [au], C. Turchi [av], C. Vullo [s], I. Yurrebaso [y], V. Pereira [at,1], N. Pinto [b,aw,*,1]

[a] DNA Diagnostic Laboratory, State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil
[b] Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), Portugal
[c] Instituto de Investigação e Inovação em Saúde (i3S), Porto, Portugal
[d] Departamento de Biologia, Faculdade de Ciências da Universidade do Porto (FCUP), Porto, Portugal
[e] Banco Nacional de Datos Genéticos, Buenos Aires, Argentina and Sección Histocompatibilidad, Unidad Inmunología e Histocompatibilidad, Hospital General de Agudos Dr. Carlos G. Durand, CABA, Buenos Aires, Argentina
[f] Laboratorio Genes SAS, Medellín, Colombia
[g] General Department of Forensic Sciences and Criminology, Dubai Police General Head Quarters, Dubai, United Arab Emirates
[h] Laboratorio Biología-ADN, Comisaría General de Policía Científica, Madrid, Spain and Instituto Universitario de Investigación en Ciencias Policiales (IUICP), Universidad de Alcalá de Henares, Madrid, Spain
[i] Laboratorio di Genetica Forense de la Università Cattolica del Sacro Cuore di Roma, Rome, Italy
[j] Laboratorio de Genética Humana, Universidad Nacional Mayor de San Marcos, Lima, Peru
[k] PRICAI - Fundación Favaloro, Buenos Aires, Argentina
[l] Servicio de Huellas Digitales Genéticos (SHDG) and Cátedra de Genética y Bioquímica Molecular, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires, Argentina
[m] Laboratorio de Genética Molecular, Unidad de Genética Médica, Facultad de Medicina, Universidad del Zulia, Zulia, Venezuela
[n] Laboratório de Investigação de Paternidade-NAC, Faculdade de Ciências Farmacêuticas, Universidade Estadual Paulista (UNESP), Araraquara, São Paulo, Brazil
[o] Laboratório de Genética Humana, Universidade da Madeira, Campus da Penteada, Funchal, Portugal
[p] Serviço de Genética e Biologia Forenses, Instituto Nacional de Medicina Legal e Ciências Forenses, I.P. – Delegação do Norte, Porto, Portugal
[q] Faculdade de Medicina da Universidade do Porto, Portugal
[r] Serviço de Genética e Biologia Forenses, Instituto Nacional de Medicina Legal e Ciências Forenses, I.P. – Delegação do Centro, Coimbra, Portugal
[s] DNA Forensic Laboratory, Argentinean Forensic Anthropology Team (EAAF), Córdoba, Argentina
[t] Laboratorio Regional de Genética Forense - Poder Judicial de Rio Negro, Rio Negro, Argentina
[u] Laboratorio ManLab, Area de Filiaciones, Buenos Aires, Argentina
[v] Departamento de Ciências Biológicas, Universidade Estadual de Santa Cruz - UESC, Ilhéus, Bahia, Brazil
[w] Departamento de Ciencias Forenses, Sección de Bioquímica, Unidad de Genética Forense, Poder Judicial, San José, Costa Rica
[x] Departamento de Biología, Servicio de Criminalística, Dirección General de la Policía y la Guardia Civil, ámbito Guardia Civil, Spain
[y] Sección de Genética Forense, Area de Laboratorio Ertzaintza, Bizkaia, Spain
[z] Laboratorio de Genética Molecular and Hemocentro Nacional - Cruz Roja Ecuatoriana, Quito, Ecuador
[aa] Heréditas Tecnologia em Análise de DNA, Brasília, Brazil
[ab] Laboratorio de Genética Médica, Universidad Tecnológica de Pereira, Facultad de Ciencias de la Salud, Pereira, Colombia
[ac] Instituto Nacional de Toxicología y Ciencias Forenses, Delegación de Canarias, Santa Cruz de Tenerife, Spain
[ad] Laboratorio IdentiGEN - Universidad de Antioquia, Medellín, Colombia
[ae] Servicio Genómica - SGIker - Universidad del País Vasco (UPV-EHU), Bilbao, Spain

* Corresponding author at: Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), Portugal.
E-mail addresses: npinto@i3s.up.pt, nmgapinto@gmail.com (N. Pinto).

https://doi.org/10.1016/j.fsigen.2025.103232
Received 6 September 2024; Received in revised form 15 January 2025; Accepted 26 January 2025
Available online 27 January 2025

[af] Laboratorio Genomik C.A., Valencia, Venezuela
[ag] Research Centre for Biochemistry and Molecular Biology at the Medical School of São José do Rio Preto (FAMERP), São José do Rio Preto, São Paulo, Brazil
[ah] Forensic Genetics Unit, University of Santiago de Compostela, Santiago de Compostela, Spain
[ai] Coordenação de Genética Forense, Departamento de Polícia Técnica da Bahia, Salvador, Brazil
[aj] Legal Medicine Unit, Azienda Ospedaliero-Universitaria delle Marche, Ancona, Italy
[ak] Banco de ADN, Universidad del País Vasco (UPV/EHU), Vitoria, Gasteiz, Spain
[al] Laboratorio de Genética Forense, Facultad de Medicina, Universidad de Las Palmas de Gran Canaria, Las Palmas, Spain
[am] NEODIAGNOSTICA, SL, Lleida, Spain
[an] Unidad de Parentesco e Identificación Humana por ADN, Referencia Laboratorio Clínico, Dominican Republic
[ao] Laboratório PeritosLab Forense, Porto Alegre, Brazil
[ap] NASERTIC, Navarra, Spain
[aq] Human and Medical Genetics Laboratory, Federal University of Pará, Belém, Brazil
[ar] Genetica Molecular de Colombia Ltda, Bogota, Colombia
[as] Genomic Engenharia Molecular Ltda., São Paulo, Brasil
[at] Section of Forensic Genetics - Department of Forensic Medicine, Faculty of Health and Medical Sciences – University of Copenhagen, Copenhagen, Denmark
[au] Instituto de Pesquisa de DNA Forense – Polícia Civil do Distrito Federal, Brasília, Brazil
[av] Section of Legal Medicine, Department of Biomedical Sciences and Public Health, Polytechnic University of Marche, Ancona, Italy
[aw] Centro de Matemática da Universidade do Porto (CMUP), Porto, Portugal

ARTICLE INFO

Keywords:
Forensic genetics
Kinship analysis
Population stratification
Mutation rates
X-STRs

ABSTRACT

The analysis of STRs located on the X chromosome has been one of the strategies used to address complex kinship cases. Its usefulness is, however, limited by the low availability of population haplotype frequency data and lack of knowledge on the probability of mutations. Due to the large amount of data required to obtain reliable estimates, it is important to investigate the possibility of grouping data from populations with similar profiles when calculating these parameters. To better understand the partition of genetic diversity among human populations for the X-STRs most used in forensics, an analysis was carried out based on data available in the literature and new data (23,949 haplotypes in total; from these 10,445 new) obtained through collaborative exercises within the Spanish and Portuguese Working Group of the International Society for Forensic Genetics. Based on the available population data, a similarity in X-STR profiles was found in European populations, and in East Asian populations, except for some isolates. A greater complexity was found for African, South American, and South and Southeast Asian populations, preventing their grouping into large metapopulations. New segregation data on 2273 father/mother/daughter trios were also obtained, aiming for a more thorough analysis of X-STR mutation rates. After combining our data with published information on father/mother/daughter trios, no mutations were detected in 13 out of 37 loci analyzed. For the remaining loci, mutation rates varied between 2.68×10^{-4} (DXS7133) and $1.07x10^{-2}$ (DXS10135), being 5.2 times higher in the male (4.16×10^{-3}) than in the female (8.01×10^{-4}) germline.

1. Introduction

In the last two decades, studies have been performed to investigate the relevance of using X-chromosomal specific short tandem repeat markers (X-STRs) in forensic genetics [1–3]. The presence of a single X chromosome in males, which is only transmitted to the daughters, increases the informative power of X versus autosomal loci in some specific kinship investigations [1]. However, if the case requires the identification of a male individual the X-chromosomal markers have lower discrimination power than the autosomal ones. Furthermore, in male/female mixtures, the presence of up to two alleles in females and only one allele in males decreases the chance of retrieving a male profile, despite increasing the probability of recovering female alleles [2]. Thus, X-chromosomal markers are essentially useful in establishing the analysis of biological kinships, especially in complex situations of identification through relatives, where autosomal markers are not able to produce conclusive results. In the simplest case of paternity (father--daughter duos, with access to the putative father), when compared to equally diverse autosomal STRs, the X-STRs have a higher mean exclusion chance. Indeed, whenever the transmission is not interrupted by a father-son relationship, the mean exclusion chance is the same or higher for X-chromosomal markers than for autosomal ones [1]. Thus, X-STRs have been shown to be useful as a complementary tool in paternity cases where few genetic inconsistencies are observed for autosomal STRs. Similarly, these markers are very useful to solve paternity cases in which two questioned fathers are related as father and son, since the two men have different X chromosomes. Another advantage of the use of X-chromosomal markers in kinship analysis is in situations where the exclusion power is null for autosomal markers, e.g. in an indirect paternity investigation with only access to the putative paternal grandmother. In this case, there is an X-chromosomal allele that is necessarily transmitted to the granddaughter (through the father), and the absence of this sharing indicates a Mendelian incompatibility, only explained by mutation. The same applies to cases in which the paternal sisterhood or half-sisterhood is investigated, since the father necessarily transmits his single X chromosome to all the daughters. Other examples where X-STRs surpass the informative power of autosomal STRs to resolve kinship can be found in several publications [2,4–6].

The recognition of the usefulness of X-chromosomal markers applied to complex kinship analyses prompted the development of genotyping methodologies for X-STRs, and the screening of allele/haplotype frequency distributions in many populations worldwide. Alongside the development of genotyping multiplexes based on PCR-CE methodologies (e.g. [7–15]), X-STRs have also been incorporated into forensic marker panels designed for MPS (e.g. [16]). In parallel, interpretation software was developed to account for the specific characteristics associated with X-chromosomal loci, namely the specific mode of inheritance, mutation, linkage, and linkage disequilibrium between loci [17].

For any genetic marker to be used in forensic casework, it is crucial to know the probability of observing a given genetic profile in the relevant population. The evaluation of forensic genetic evidence is based on likelihood ratio (LR) principles, considering the probability of the observed profile(s) under two alternative and mutually exclusive hypotheses [18]. This probability is usually calculated based on population

[1] These authors have contributed equally to this work

allele frequencies, being, however, more complex when it comes to the combined analysis of markers in linkage disequilibrium (LD). Since marker independence can no longer be assumed, the final LR cannot be obtained by multiplying the individual values computed for each locus. The most direct implication for the use of markers in LD in kinship analyses is the need to estimate haplotype frequencies in the population, rather than single locus allele frequencies [5].

The underlying difficulties in estimating haplotype frequencies are notorious for non-recombining markers such as those from the Y chromosome and mitochondrial DNA (mtDNA), and different approaches have been proposed to calculate the probability of the evidence based on these markers [19–21]. Although to a lesser extent, the difficulty of estimating haplotype frequencies is also applied to X-STRs that are in LD, since the expected number of allelic combinations usually exceeds the hundreds [22]. Therefore, very large population samples are needed for a proper estimation of the haplotype frequencies in a population.

Another important parameter to consider in the application of genetic markers in kinship analyses is the mutation rate. Due to the low frequency and multiple factors associated with the probability of mutation of a specific allele, the amount of data needed to obtain reliable estimates is very high. The studies carried out so far on X-STRs show a mutational behavior similar to that of autosomal STRs. X-chromosomal marker mutation rates exhibit great variation between and within STRs that depends on the number and structure of the repetitive motif, varying also with the gender and age of the parent that underwent the mutation [23].

In summary, the usefulness of X-STRs is well recognized in kinship analyses and there are well-established genotyping and result interpretation methodologies. However, although there are many reports on X-STR allele/haplotype frequencies in several populations, the information available is still fragmentary both in terms of the populations analyzed and the markers included [3]. Moreover, the mutational data available is still scarce to encompass the variation that is known to exist between and within X-STR loci.

This study aimed to generate and compile data on X-STRs in different populations worldwide, as well as data on segregation analyses in father/mother/daughter trios. These data comprised 19 X-STRs widely used in forensic genetics, which have been described in two PCR multiplexes with three overlapping loci: one with 10 X-STRs (X-Decaplex) and the other with 12 X-STRs (Argus X-12).

A joint analysis of the new data obtained with those available in the literature allowed for the comparison of different populations, and the establishment of diversity and substructure levels. Additionally, more thorough mutation rates were estimated based on segregation analyses on father/mother/daughter trios.

2. Materials and methods

2.1. Sampling and genotyping methods

Data for this study was obtained through collaborative works of the Spanish and Portuguese Working Group of the International Society for Forensic Genetics (GHEP-ISFG). A total of 42 laboratories participated by producing data on 1395 father/mother/daughter trios genotyped for the X-Decaplex, and 878 trios genotyped for the Investigator Argus X-12 QS (Qiagen) loci. Only the daughters were used for population analyses, after inferring the gametic phase using the information from their fathers. Samples from 6066 unrelated males typed for the X-Decaplex were also included.

Samples were obtained from former paternity cases or healthy volunteers. In samples from paternity investigations, the biological relationship was previously confirmed using autosomal STRs (LR > 10,000). Each laboratory ensured the anonymization of the samples and the accomplishment of the legal and ethical requirements for their use in this research project.

The 10 loci included in the X-Decaplex were genotyped as described in Gusmão et al. [10], and those of the kit Investigator Argus X-12 QS (Qiagen) following the manufacturer's instructions (Investigator® Argus X-12 QS Handbook).

2.2. Population data and analyses

Population analyses were performed after joining the new data from the present study with the previously published ones, where information for X-Decaplex haplotypes or Argus Linkage Groups (LG1 to LG4) was available (Supplementary data). The final dataset included 187 population samples from 53 countries in Africa, America, Europe, and Asia (see Fig. 1).

For the X-Decaplex markers, a total of 12,440 haplotypes were compiled [14,24–39], including 8911 from this study (Supplementary Table S1). From the literature, we only included data from complete X-Decaplex male profiles. Data from Gusmão et al. [10] was not considered, since the male profiles overlap with the fathers of the daughters included in the present study, whose gametic phase was inferred using the paternal information.

For the Argus X-12, a total of 11,509 full haplotypes were collected [12,23,40–84], including 1534 new haplotypes. From the literature, we have only included data from male profiles with haplotypic information for the LGs. Some publications report the frequencies of LG haplotypes, but do not include the full 12 X-STR haplotypes. Therefore, sample size varied for the different LGs: 17,814 for LG1 and LG2; 17,349 for LG3; and 16,732 for LG4 (Supplementary Table S1).

Allele and haplotype frequencies were calculated using the software Arlequin ver 3.5.2.2 [85]. The same software was used for genetic structure analyses, based on population pairwise comparisons using F_{ST} values and corresponding non-differentiation p-values. Pairwise linkage disequilibrium was also tested using Arlequin ver 3.5.2.2 [85]. In population differentiation and LD tests, Bonferroni's correction was used to adjust the significance level, dividing 0.05 by the total number of pairwise comparisons.

2.3. Segregation data and mutation analyses

Data from 3748 father/mother/daughter trios was compiled from the literature [9,13,23,86–102] and joined to the 2273 trios from this study to estimate X-STR mutation rates (Supplementary Table S2). Segregation data from father/daughter or mother/son duos were not considered in this study, since the frequency of hidden mutations is higher in these cases, when compared to father/mother/daughter trios, and the different types of family data cannot be combined [103]. Confidence intervals for mutation rates were estimated from the binomial standard deviation and chi-square and Fisher's exact tests were used to assess the difference between the mean values of two subsets of data (the latter used in the cases involving categories with no observations). In all the cases a significance level $\alpha = 0.05$ was considered.

3. Results and discussion

3.1. Genetic structure analyses

With the aim of investigating the levels of genetic differentiation within and between countries as well as continents, pairwise comparisons were performed between available population samples for the studied X-STRs. The full list of haplotypes is provided as Supplementary Data. Since the set of samples studied for the X-Decaplex and the Argus X-12 loci do not overlap, analyses were performed separately for each one of these sets and the results interpreted based on the overall findings. Populations from the same country or continental region were grouped, unless they showed large (≥1 %) and statistically significant F_{ST}s for the studied markers.

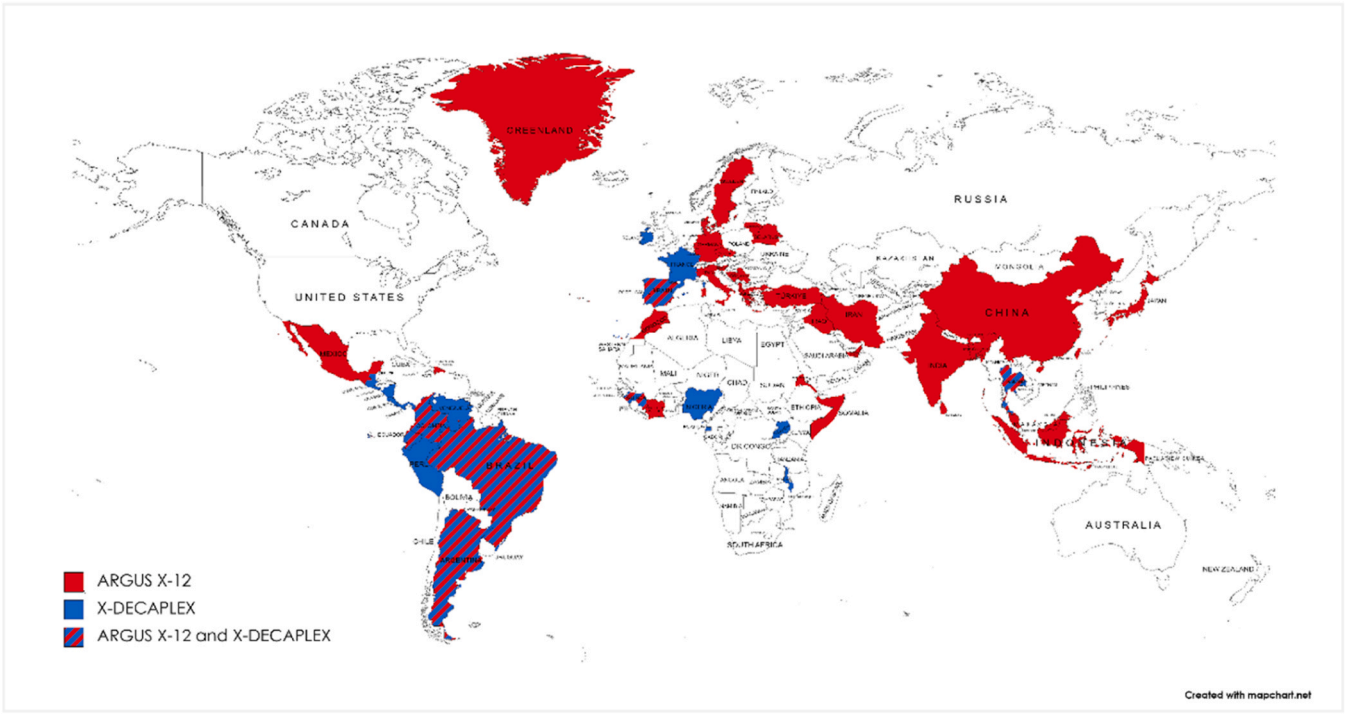

Fig. 1. Final dataset used in population analyses, corresponding to published and new data from this study, obtained using X-Decaplex and Argus X-12 PCR-CE multiplexes.

3.1.1. Pairwise population comparisons based on the X-Decaplex data

A hierarchical analysis was performed per country or continent, depending on the number of representative populations.

Pairwise comparisons, showed non-significant F_{ST} values between populations from the West-Central African region (Supplementary Table S3). Although not statistically significant, the F_{ST} between Guinea-Bissau and Equatorial Guinea was relatively high, which can be due to the small sample size of the latter. Malawi, the only population from Southeast Africa, presented low F_{ST}s with all populations from the West-Central African region, except Guinea-Bissau. As expected, both Morocco, in North Africa, and the Eastern population of a Nilotic-speaking group from Uganda (Karimojong) presented significant differences between them and with the remaining populations, except with Equatorial Guinea, most likely for the reasons mentioned above.

Concerning the two Asian populations from Macau and Thailand, a low, non-significant, F_{ST} was observed (F_{ST}=0.0006; $p = 0.40253$).

When comparing populations from Central America, the Native American groups from Guatemala and El Salvador showed non-significant differences (Supplementary Table S4; Supplementary Figure S1). In pairwise F_{ST} analyses among the admixed populations, all values were below 1 %, except between Guatemala and Panama. Between these two countries, the non-differentiation probability was not significant, although the F_{ST} was relatively high (0.01416). In contrast, a significant differentiation was observed between Costa Rica and Nicaragua, but for a low F_{ST} value, due to the large sample sizes.

Among the 10 population samples from Argentina, the three Native groups showed significant differences among them and with the remaining samples from non-Native (admixed) populations (Supplementary Table S5; Supplementary Figure S2). Except in comparisons involving the Northeastern populations of Corrientes and Misiones, all pairwise F_{ST}s were low ($F_{ST}\leq 0.0053$), with no statistically significant differentiation detected among the populations from the Central and South regions, for the studied markers.

The 14 population samples from Brazil were analyzed considering the main geographic regions of the country: Southeast (including Araraquara, Belo Horizonte, São Paulo, Rio de Janeiro and Vitoria), Northeast (Bahia), North (Belem), Central-West (Brasilia, Mato Grosso, Mato Grosso do Sul), and South (Parana, Santa Catarina and Rio Grande do Sul). Although statistically significant differences were observed in some pairwise comparisons of populations from the same region, the F_{ST}s were all below 1 % (Supplementary Table S6; Supplementary Figure S3). The same was observed between samples from Southeast, Central-West, and South regions. However, statistically significant F_{ST}s, above 1 %, were observed in pairwise comparisons involving populations from the North and Northeast regions.

In the comparison of the Colombian samples, high F_{ST} values were observed between Chocó and the remaining populations (Supplementary Table S7; Supplementary Figure S4), which is somehow expected due to the high African ancestry reported for this region in the Pacific Coast [104]. The same trend was observed for Bolivar, although with smaller distances. Lower F_{ST}s were observed among the remaining populations, all located in the Andean region. However, all pairwise comparisons involving Bogotá resulted in F_{ST}s slightly above 1 % ($0.0102 \leq F_{ST} \leq 0.0109$).

The three population samples from Ecuador presented large, statistically significant F_{ST}s, as expected considering that they correspond to different ethnic groups, namely the Kichwa, the Waorani and an Admixed population sample (known as Mestizos). The F_{ST} values were 0.04407 between Kichwa and Mestizos ($p < 0.000005$), 0.06778 between Waorani and Mestizos ($p < 0.000005$), and 0.05898 between Kichwa and Waorani ($p = 0.00099$).

Between the two population samples from Venezuela (Central region and Maracaibo), a low, non-significant, F_{ST} was observed (F_{ST}=0.0021; $p = 0.0584$).

When comparing European populations within and between countries, only the Portuguese Gypsies and the Canary Island populations showed significant differences in some pairwise comparisons, with F_{ST} values above 1 % (Supplementary Table S8). F_{ST} values among the remaining populations from Portugal, Italy, Ireland, Spain, and France were below 1 % (Supplementary Table S8).

Based on the above-described results, a new pairwise F_{ST} analysis was performed after grouping country/continent neighboring

populations with no evidence of presenting large distances (F_{ST}<0.01). The results obtained are displayed in Supplementary Table S9 and represented in the MDS plot in Fig. 2. In Fig. 2, the smaller and more isolated populations, including the Native American groups from Wichi, Colla, Toba, Kichwa and Waorani, as well as the Portuguese Gypsies, are spread in the MDS plot, not grouping with each other nor with other populations. Some groups (based on small F_{ST}s) can be discerned. Concerning the populations from Central and South America, it can be seen that Argentina, Venezuela, Colombia (Andes), Brazil (North region) and Central American Admixed populations are located between Europe and a cluster of populations with high Native American ancestry (from Peru and Ecuador), which are close to the Central American Natives. The remaining samples from South America are located between Europe and Africa: (i) Brazil Center and South regions clustering with Bogotá, close to Europe; (ii) Brazil Northeast region is slightly away towards Africa; and (iii) Chocó (Colombia) is closer to Africa than to Europe. The dispersion on the plot of the admixed populations from Central and South America is in accordance with the different levels of African, European, and Native American ancestries previously reported for these populations [104–110]. As for the population of the Canary Islands, it is far from the Western European group, not showing a significant differentiation with Morocco. This result is supported by previous studies showing a North African genetic inheritance in this archipelago [111], especially evident when considering mtDNA [112]. Since the X-chromosomal population background has a greater maternal than paternal contribution, it is therefore expected that these markers show greater affinity to North African populations when compared to the autosomal ones.

3.1.2. *Pairwise population comparisons based on the Argus X-12 data*

As for the X-Decaplex, pairwise F_{ST} analyses at different populational levels were performed using Argus X-12 data, to understand which samples could be grouped in a metapopulation, assuming a low F_{ST} and/or non-significant differentiation probability.

Different populations/ethnic groups reported for the same country were pooled whenever they were shown not to be significantly different in the original publication. For instance, the samples from the 3 ethnic groups from Morocco were grouped in a single population (called Morocco), since no significant differentiation was reported among them in the original study [83]. The same was done for the Sri-Lankan ethnicities studied by Perera et al. [80], since low genetic distances were observed among them (F_{ST}s $\leq$ 0.00375). The Zhuang and Mulao from the Chinese province of Guangxi were also joined, based on the results from Xiao et al. [75]. In accordance with the findings from Cortés-Trujillo et al. [69], a single population sample from Mexico was considered, including data from 7 different country regions. Also, following the original publication, a single population sample from Iran was considered, including the Persians, Lurs, Kurds and Azeris ethnic groups [51]. The samples from Sardinian open populations were grouped since no significant differentiation was detected among them, in contrast with the isolated populations that were more heterogeneous and were considered separately [64]. The nine subpopulations from Greece were pooled according to the results from the original study [50].

Concerning samples from the same population/country/region reported in different studies, a first analysis was performed to see if they were not significantly different.

When comparing haplotype frequencies between the different Jewish groups, no statistically significant differences were observed for pairwise F_{ST} values for the 4 LGs, after applying Bonferroni correction for ten comparisons inside each LG (LG1, $p \geq 0.0894$; LG2, $p \geq 0.0120$; LG3, $p \geq 0.1674$; LG4, $p \geq 0.0080$). For this study, these samples were pooled in a single group. However, high F_{ST} values were found in some comparisons, showing that studies with larger samples are needed to determine the true degree of differentiation between them. In fact, these population samples were compared in a previous study for a higher number of X-chromosomal markers, and significant differences were reported between some of them [59].

The 18 populations from Asia were compared inside three geographic regions: East, Southeast and South Asia, and grouped whenever genetic distances were below the defined threshold of F_{ST}

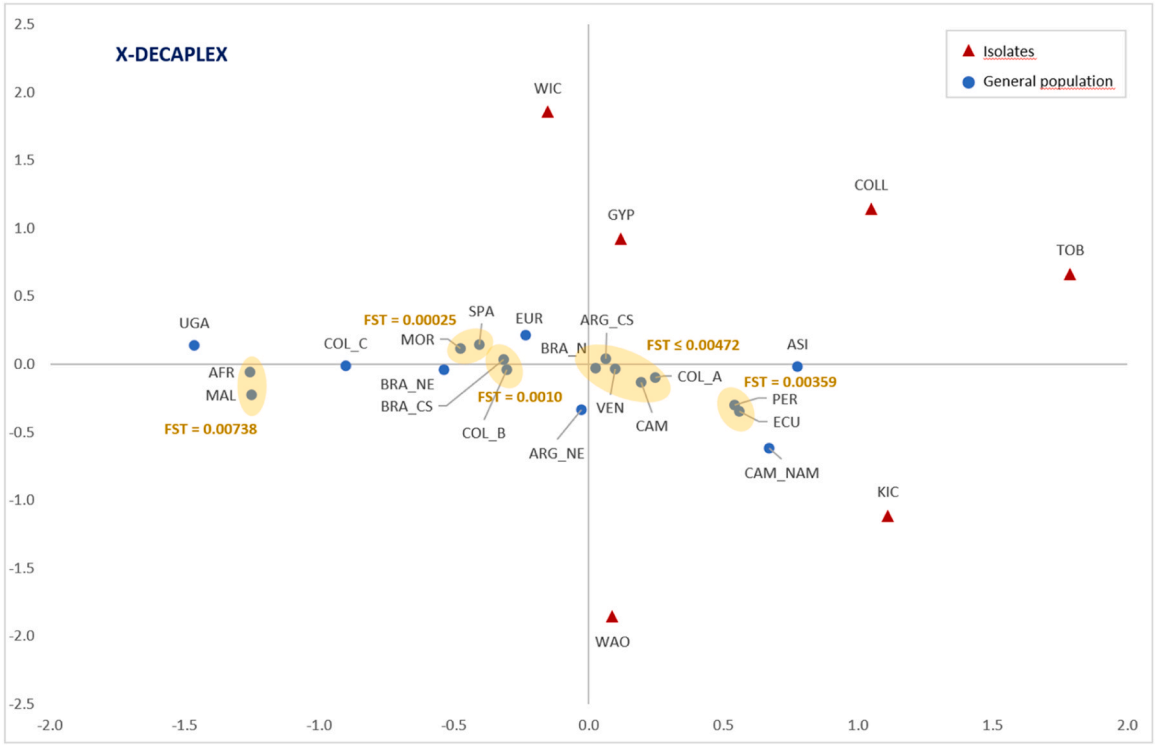

Fig. 2. MDS plot of pairwise genetic distances (F_{ST}) after population grouping, based on the X-Decaplex. Inside the yellow circles are populations with low (non-significant) distances. Population codes are described in Table S9.

(lower than 1 %) or non-statistically significant (Supplementary Figure S5). Although belonging to different ethnic groups (Manchu, Shenyang and Han, from Liaoning, and Korean from Jilin), the 4 population samples from the Northeast region of China were grouped since low, non-significant, pairwise F_{ST}s were detected for the 4 LGs ($F_{ST} \leq 0.0023$). This sample from Northeast China was further compared with samples from North (Inner Mongolia), Northwest (Xinjiang), Eastern (Shanghai) and South (Guangxi) regions. Small values of genetic distance were observed among them, with a single exception at LG2 between Xinjiang (Uyghur) and Shenyang populations (F_{ST}=0.0120). When comparing China, Japan and Taiwan, although F_{ST}s above 1 % were detected for LG4 in comparisons involving Taiwan, no significant non-differentiation *p*-values were obtained among them ($p \geq 0.1305$). Therefore, we have considered all these populations together as representing East Asia. Concerning populations from Southeast Asia (Malaysia, Philippines, Thailand and Borneo), samples from Thailand were joined ($-0.0021 \leq F_{ST} \leq 0.0122$; $p \geq 0.0057$), and significant F_{ST}s were obtained in the remaining pairwise comparisons, for at least one LG. South Asia was represented by population samples from India, Sri-Lanka and Bangladesh. The two population samples from Bangladesh were pooled [48,63], since no significant F_{ST}s were found between them ($F_{ST} \leq 0.0036$; $p \geq 0.1828$). In this group, F_{ST}s above 1 % and statistically significant *p*-values were only observed for LG3, between the Jat Sikh population from Punjab (India) and the Bangladesh and India Bhil.

Concerning South America, the results of differentiation analyses between samples from the 23 Argentinean provinces allowed grouping them in four regions, namely: Northwest (Catamarca, Jujuy, La Rioja, Salta, Santiago del Estero and Tucumán), Northeast (Chaco, Corrientes, Formosa and Misiones), Central (Buenos Aires, Cordoba, Entre Rios, La Pampa, Mendoza, San Juan, San Luis, and Santa Fe), and South (Chubut, Neuquén, Rio Negro, Santa Cruz, and Tierra del Fuego) (Supplementary Table S10). The three Brazilian population samples from the Southeast region (Supplementary Table S1) showed non-significant differences in all LGs ($F_{ST} \leq 0.0049$; $p \geq 0.0276$). When comparing the four Ecuadorian population samples (Supplementary Table S1), high F_{ST} values (from 0.0194 to 0.0259) were observed in some comparisons between the sample from general population described in Pinto et al. [23] and the three population groups included in Flores-Espinoza et al. [78], although not statistically significant (above the significance level of 0.0083), most probably due to the small sample size (n = 22).

For the European samples, a first pooling between populations from the same country was performed whenever non-significant F_{ST}s were observed. Samples from the 4 regions of Croatia (North, South, East and West) were pooled ($F_{ST} \leq 0.0097$; $p \geq 0.0126$); all Portuguese samples were treated as a single population ($F_{ST} \leq 0.0059$; $p \geq 0.0541$); Italy isolated populations were maintained separated from the open populations; and Spain was divided in mainland and Balearic Islands. A second comparison was carried out among all European populations (see Supplementary Table S11). Based on the F_{ST} results and corresponding non-differentiation probabilities, population samples were further grouped as follows: East Europe (Belarus, Czech Republic, Lithuania, and Slovenia), Southeast Europe (Albania, Croatia, Greece, and Serbia), and West Europe (Denmark, Germany, Italy, Portugal, Spain, and Sweden).

For the African dataset available, most pairwise comparisons involving Morocco, Somalia and Eritrea showed significant differences

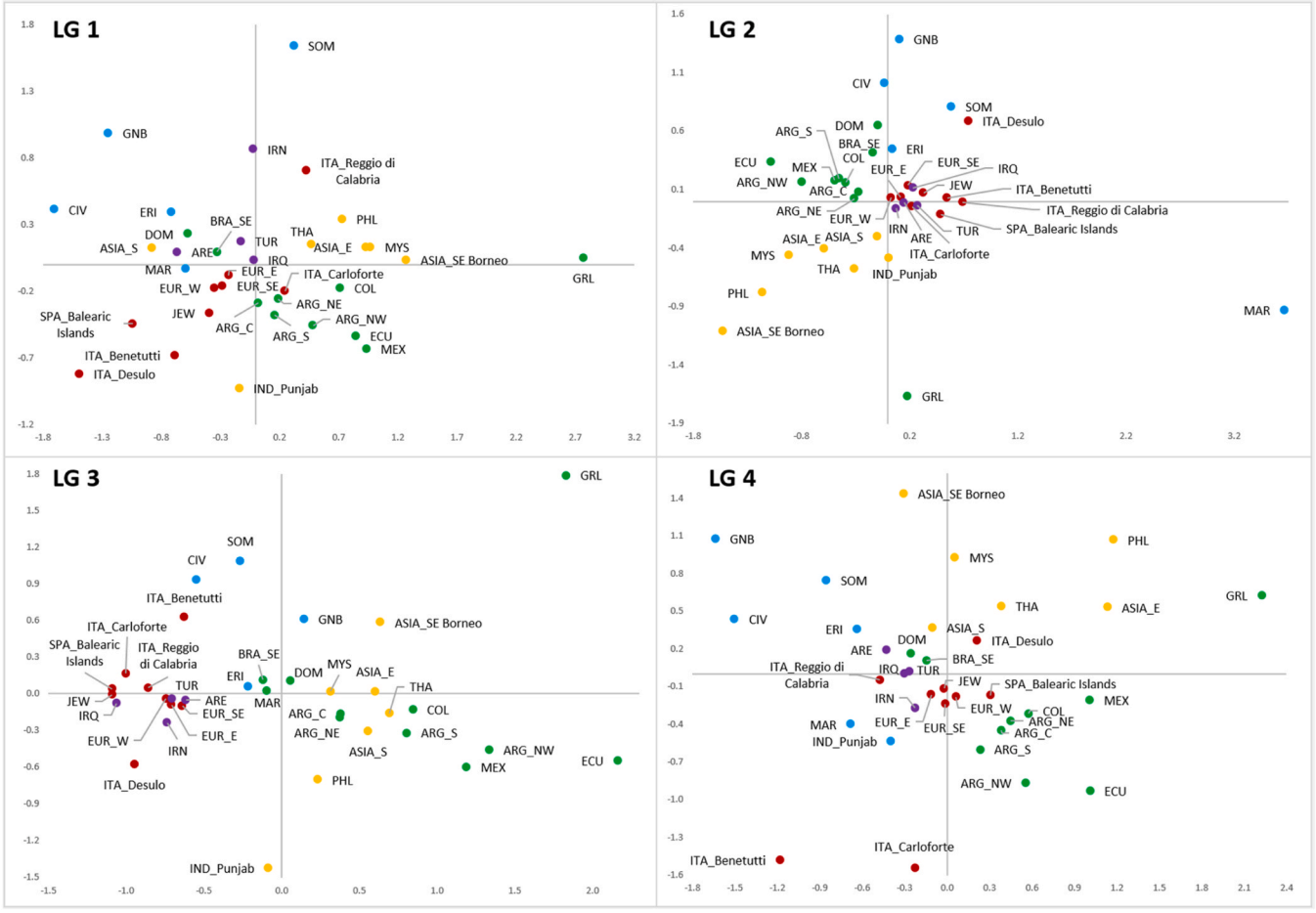

Fig. 3. MDS plot of pairwise genetic distances (F_{ST}) after population grouping, for the four LGs of the ARGUS X-12, including populations from Asia (yellow), Europe (red), Africa (blue), Middle East (purple) and South America (green). Population codes are described in Table S13.

(Supplementary Table S12). Between the 2 Western African populations from Guinea-Bissau and the Ivory Coast, differentiation was detected at LG 4, but not for the other linkage groups.

Based on the above-described results, a new pairwise F_{ST} analysis was performed after grouping country/continent neighboring populations with no evidence of presenting statistically significant differentiation or large distances (F_{ST}>0.01). The results obtained are displayed in Supplementary Table S13 and represented in the MDS plots in Fig. 3. Samples from West, Southeast and East Europe are grouped in all LG MDS plots, showing non-significant differentiation among them ($F_{ST}\leq$0.00049). These European samples showed non-statistically significant differences with Balearic Islands (Spain), although for larger distances ($F_{ST}\leq$0.0070). Population isolates from Sardinia (Italy) showed significant differences with the European group populations for at least one LG. The four populations from the Middle East present low (non-statistically significant) F_{ST}s in all pairwise comparisons, except for LG1 between Turkey and Iran, and between Dubai and Iran. Concerning the Asian continent, large significant differences were found among East, Southeast and South regions, as well as within Southeast and South regions. The samples from different Argentinean regions showed non-significant differences with all F_{ST} values below 0.0074. Nonetheless, as previously observed for the X-Decaplex, South America shows a high variation on the genetic background of populations from different countries, preventing their grouping into a single metapopulation.

3.2. Segregation and mutation analyses

Among the 2273 father/mother/daughter trios analyzed in this study (1395 analyzed for the X-Decaplex and 878 for the Argus X-12), 145 Mendelian incompatibilities were detected. A total of 19 X-STRs were analyzed, and marker specific allele transfers varied between 873 and 2273. Incompatibilities were not detected at DXS9898 (N = 1395). For the remaining X-STRs, the number of incompatibilities varied from one, at DXS10103 (N = 878) and DXS7133 (N = 1395), to 28 at DXS7132 (N = 2266) (Supplementary Table S14). Two of the 145 incompatibilities can be explained by null alleles at DXS6789 and GATA172D05, respectively, since both mother and daughter show a single allele at these loci. In these cases, when calculating the LR in kinship analyses it is important to consider both the mutation rate and the frequency of null alleles to reconcile the apparent inconsistency under the assumption of relatedness. In the population analyses, 10 null alleles were detected, seven of which at GATA172D05 locus, two at DXS7133 and one at DXS10101 (see Supplementary Data). In most cases, the population frequency was low (namely, for DXS7133 the frequency was 1/1623 in Portugal and 1/3038 in Brazil; for GATA172D05 it was 1/3038 in Brazil, 1/1952 in Spain; and for DXS10101, 1/871 in Portugal), except for GATA172D05 in Italy (5/406). Regarding the frequency of X-STR null alleles, it should be noted that these are always detected in male profiles. However, when father, mother, daughter trios are analyzed to establish the gametic phase in the daughter, and all of them are apparent homozygous for the same allele, null alleles may go unnoticed, as the mother could be in fact heterozygous and have transmitted the null allele. In this study, since the haplotypes of the daughters were inferred based on those of the parents, it is possible that the frequency of null alleles is slightly underestimated.

The remaining incompatibilities can be explained by mutation, with one trio showing a genotypic configuration compatible with both a maternal and a paternal mutation at DXS10101 (Supplementary Table S14).

It is worth noting that total or partial (maternal) uniparental disomies (UPD) were ruled out by analyzing the genotypic and haplotypic configuration of the trios for which Mendelian incompatibilities were found. Indeed, compensatory mechanisms occurring at early stages of development may lead to euploid individuals with a UPD – see, e.g. [113]. This is relevant for the topic under study because it is expected that individuals with a UPD exhibit mendelian incompatibilities with parental genotypic configurations, especially for highly polymorphic markers as those analyzed. This may be confused with the occurrence of a mutation. Assuming that two identical by descent (IBD) alleles are identical by state, a total uniparental disomy implies the same IBD state of the alleles for all markers, which can be either IBD, and thus autozygous (isodisomy) or non-IBD (heterodisomy, only possible in the case of a maternal disomy). On the other hand, partial uniparental disomy (possible only for maternal inheritance in the case of the X chromosome) exhibit both IBD states across the chromosomes. None of the haplotypic configurations of the individuals of the trios where mendelian incompatibilities were observed supported the presence of total or partial (maternal) uniparental disomy.

3.2.1. Locus specific mutation rates

Locus specific mutation rates were calculated after combining our data with published information on father/mother/daughter trios [9,13,23,72,86–102] (Table 1). Data was obtained for 37 X-STRs, including the 19 from this study. The average mutation rate for the 37 X-STRs was 2.827×10^{-3}, being higher when just considering the Argus-X12 (4.021×10^{-3}) than the X-Decaplex markers (1.635×10^{-3}). No mutations were detected in 13 loci, for 10 of which less than 1000 allelic transmissions were investigated ($100 \leq N \leq 950$). In these 10 cases, the upper limit of the 95 % confidence interval (CI) is above the average mutation rate for all loci, supporting that the null estimate may be a consequence of the low sample size. Contrarily, the results obtained for

Table 1
Locus specific mutation rates calculated after combining new and published data on father/mother/daughter trios for 37 X-STRs, including the 19 from this study.

Marker	Total				
	No. of Meiosis	No. of Mutations	Mutation Rate	Lower CI (95 %)	Upper CI (95 %)
DXS10074	6096	28	0.0046	0.0031	0.0066
DXS10075	620	3	0.0048	0.0010	0.0141
DXS10079	5594	39	0.0070	0.0050	0.0095
DXS101	2104	1	0.0005	0.0000	0.0026
DXS10101	5434	14	0.0026	0.0014	0.0043
DXS10103	4980	7	0.0014	0.0006	0.0029
DXS10134	5582	23	0.0041	0.0026	0.0062
DXS10135	5426	58	0.0107	0.0081	0.0138
DXS10146	4978	26	0.0052	0.0034	0.0076
DXS10147	874	0	0.0000	0.0000	0.0042
DXS10148	4978	22	0.0044	0.0028	0.0067
DXS10159	116	0	0.0000	0.0000	0.0313
DXS6789	5332	11	0.0021	0.0010	0.0037
DXS6793	100	0	0.0000	0.0000	0.0362
DXS6795	950	0	0.0000	0.0000	0.0039
DXS6797	168	0	0.0000	0.0000	0.0217
DXS6800	904	0	0.0000	0.0000	0.0041
DXS6801	720	0	0.0000	0.0000	0.0051
DXS6803	1694	3	0.0018	0.0004	0.0052
DXS6807	196	0	0.0000	0.0000	0.0186
DXS6809	4220	11	0.0026	0.0013	0.0047
DXS6810	216	1	0.0046	0.0001	0.0255
DXS7130	834	0	0.0000	0.0000	0.0044
DXS7132	10,936	59	0.0054	0.0041	0.0070
DXS7133	3738	1	0.0003	0.0000	0.0015
DXS7423	9544	6	0.0006	0.0002	0.0014
DXS7424	1950	2	0.0010	0.0001	0.0037
DXS8377	330	2	0.0061	0.0007	0.0217
DXS8378	9662	9	0.0009	0.0004	0.0018
DXS981 (STRX1)	1604	0	0.0000	0.0000	0.0023
DXS9895	156	0	0.0000	0.0000	0.0234
DXS9898	3552	0	0.0000	0.0000	0.0010
DXS9902	3880	11	0.0028	0.0014	0.0051
GATA165B12	1738	0	0.0000	0.0000	0.0021
GATA172D05	4142	4	0.0010	0.0003	0.0025
GATA31E08	4580	3	0.0007	0.0001	0.0019
HPRTB	6958	9	0.0013	0.0006	0.0025

the other three X-STRs (DXS9898, DXS981, and GATA165B12), for which no mutations were detected, support low mutation rates, being included within the 7th first ranked markers with lower 95 % upper limit interval.

For the remaining loci for which mutations were observed, rate estimates showed to be variable, with the lowest ($2.675x10^{-4}$ for DXS7133) being 40 times lower than the largest ($1.069x10^{-2}$ for DXS10135), and the 95 % lower limit interval of the latter being 5.5 times greater than the upper limit of the first.

3.2.2. Paternal and maternal mutation rates

The expected male mutation bias in humans was estimated to be around 6 [114], although varying with age. It is, however, noteworthy that male/female bias can be accentuated in the case of haplodiploid genomes since maternal transmissions are more prone to hidden mutations than paternal ones, as previously shown using simulated data [115]. Among the 143 mutations observed in this study, and assuming the most parsimonious event, 110 were compatible with paternal origin, 15 with maternal one, and 18 were indetermined (Supplementary Table S14). When combining new and data collected from the literature, 260 out of the 353 mutations observed were shown to be compatible with paternal origin, and 50 with maternal one (Supplementary Table S15). The estimated mutation rate was thus 5.2 times higher in males than in females, a gender bias that has been frequently described in humans, as well as in other mammals [114,116], due to the higher number of cell divisions in the male than in the female germline. The differential between the estimated mutation rates depending on the most parsimonious parental origin is represented in Fig. 4 for the markers for which more than 1000 meiosis were observed.

The parental origin of the remaining mutations (43 out of the 353) could not be ascertained, as the genotypic configurations could be attributed either to a paternal or a maternal event. This means that these mutations would not have been detected if father-daughter or mother-daughter duos had been analyzed. Therefore, if hidden mutations are ignored, the average mutation rate in father-daughter duos (4.164×10^{-3}) is 5.2 times greater than the one in mother-daughter duos (8.007×10^{-4}). As before, caution should be taken in the weighing of this gender imbalance, since hidden mutations are more likely in mother-daughter than in father-daughter duos [115].

3.2.3. Parental age and mutation rates

The parental age of the individuals was gathered from 3391 trios from this study and from Pinto et al. [23] (Supplementary Table S16 and Fig. 5). For both fathers and mothers, the age class with more subjects is the one considering individuals older than 20, and younger than 26. Statistically significant results were obtained when grouping fathers in two classes according to their age: younger or older than 20, 25, 30 or 35 years old ($p \leq 0.0370$). Nonetheless, the differences observed were low, with age increasing rates varying between 1.4 and 1.9 times (Fig. 5 and Supplementary Table S17). It is generally accepted that the likelihood of a paternal mutation occurring is positively correlated with the age of the father [117]. Although our data supports this trend for age classes between 20 and 35 years old, no support was obtained for other classes, possibly due to small sample sizes associated with large confidence intervals. No statistically significant differences were found when considering maternal mutation rates according to age classes (Supplementary Table S17). In any case, the impact of the paternal age on the mutation rate is expected to be low, in agreement with previous results obtained for Y-STRs [118].

3.2.4. Single and multistep mutation rates

For 340 of the 353 mutations analyzed it was possible to ascertain the number of step-changes involved (Supplementary Table S18). Mutations compatible with changes regarding a non-integer number of repeats were also observed (N = 3). The number of mutations compatible with single-step repeat changes (N = 331) was 55.2 times greater (chi-square statistic = 204.192; $p < 0.00001$) than those involving multiple-steps (N = 6) (Table 2). This unbalance tends to be overestimated since multi-step mutations may be interpreted as single step ones (assuming this to be the most parsimonious reasoning), as previously shown using simulated data [119]. However, it is worth highlighting that this overestimation is smaller when analyzing haplodiploid

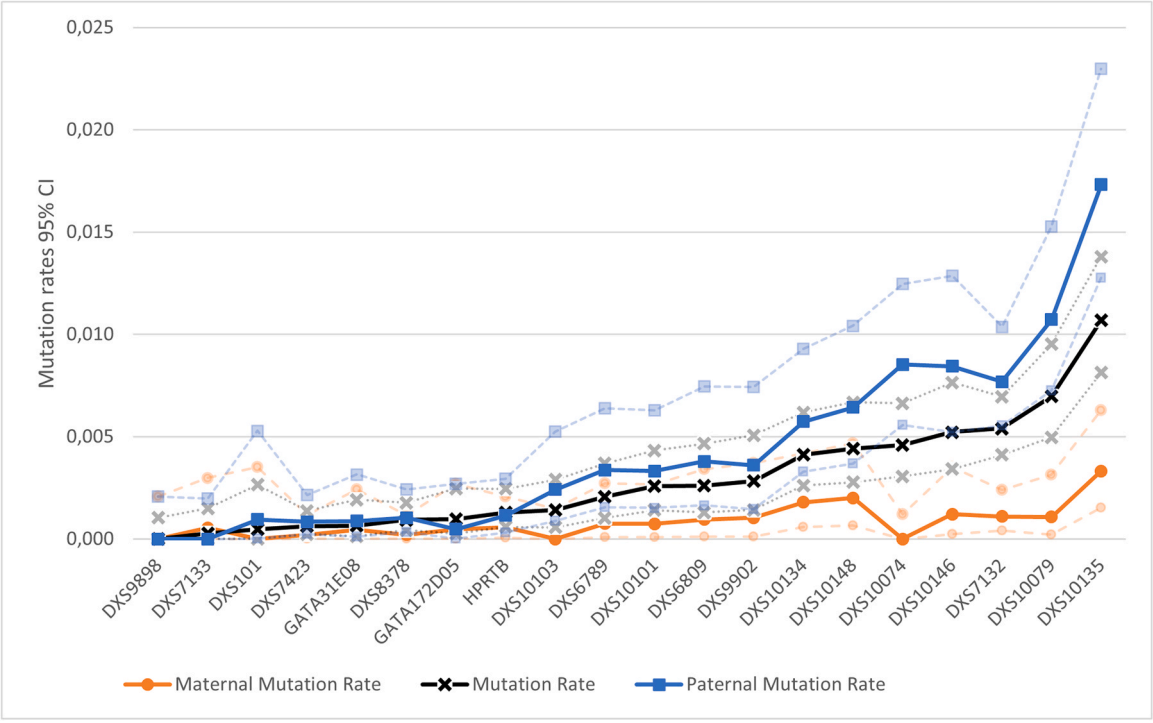

Fig. 4. Overall, maternal and paternal mutation rates for the set of markers for which more than 1000 allelic transmissions were observed, and the corresponding 95 % CIs (faded lines, detailed information is provided in Supplementary Table S15).

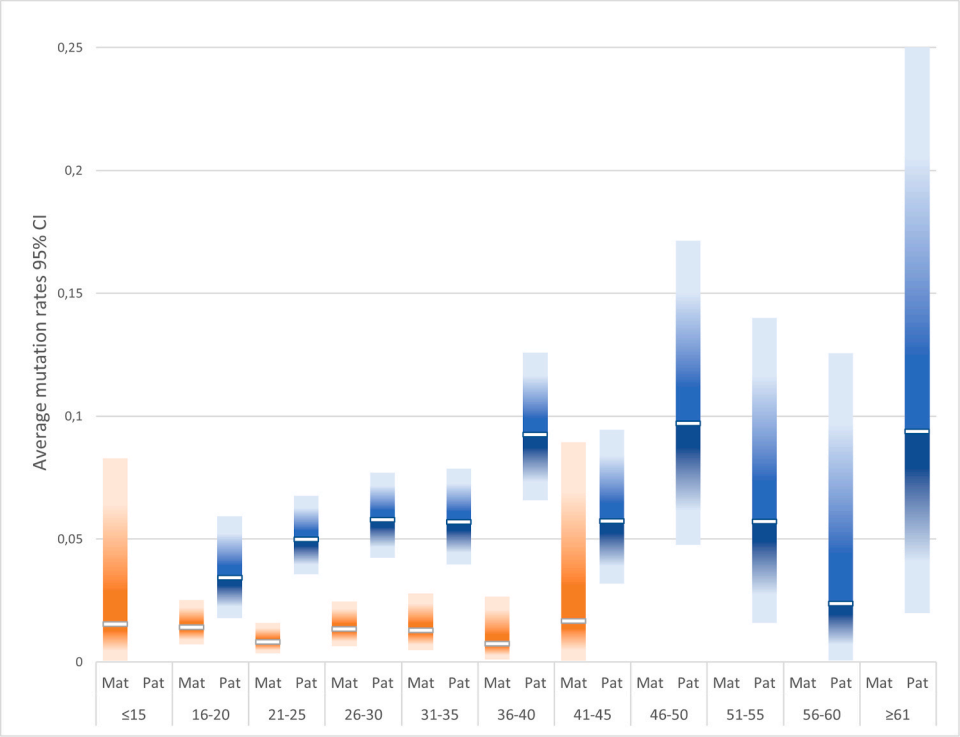

Fig. 5. Average values of paternal (Pat) and maternal (Mat) mutation rates in different age groups, and the corresponding 95 % CIs (detailed information is provided in Supplementary Table S17).

Table 2
Number of mutations for which information on the number of apparent mutational steps involved was available. NI represents mutations non compatible with mutations involving an integer number of repeats.

Marker	Total						
	No. of mutations	Mutational steps					NI
		{+ 1}	{-1}	{-1}V{+ 1}	> {+ 1}	< {-1}	
HPRTB	9	6	3	0	0	0	0
DXS7132	59	26	28	4	1	0	0
DXS7133	1	1	0	0	0	0	0
GATA172D05	4	3	1	0	0	0	0
DXS8378	9	2	4	3	0	0	0
DXS7423	6	2	4	0	0	0	0
DXS6809	11	5	5	1	0	0	0
DXS6789	11	6	5	0	0	0	0
GATA31E08	3	0	3	0	0	0	0
DXS9902	10	4	4	2	0	0	0
DXS10135	58	25	31	0	0	1	1
DXS10074	28	11	16	0	1	0	0
DXS10101	14	8	5	0	0	1	0
DXS10134	23	10	11	0	1	0	1
DXS10148	22	9	12	0	0	1	0
DXS10079	39	17	22	0	0	0	0
DXS10103	7	4	3	0	0	0	0
DXS10146	26	11	14	0	0	0	1
Total	340	150	171	10	3	3	3

instead of diploid markers [119]. No gender association was found between the apparent parental origin of the mutation, and the type of occurring changes involving either single or multiple repeats ($p = 0.9058$). No significant differences were found when comparing the number of mutations involving either the gain or loss of repeats ($p = 0.4117$), even when considering either paternal ($p = 0.0808$) or maternal ($p = 0.2204$) mutations, separately. As previously shown for Y-STRs [118], the trend to gain or lose repeats is correlated with the number of repeats of the parental allele. A similar analysis is not straightforward in the case of haploidploid transmission, as the transmitted maternal allele cannot be ascertained.

3.2.5. *Metapopulation specific mutation rates*

After pooling datasets based on the results of population differentiation analyses, metapopulation specific mutation rates were estimated for the set of markers for which more than 600 meiosis were observed, for at least one pair of metapopulations (Fig. 6). The following metapopulation groups were defined: Europe, East Asia, Brazil, and Ecuador and Peru (see Supplementary Table S19).

Although mutation rate estimates varied among metapopulations,

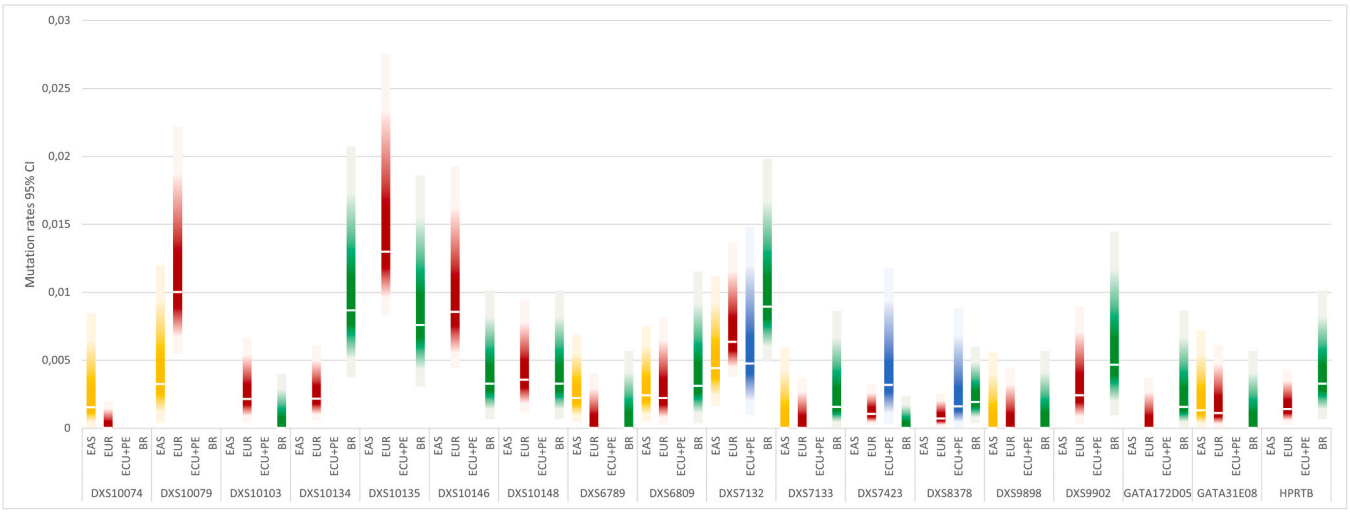

Fig. 6. Mutation rates when considering the data grouped by metapopulation, for the set of markers for which at least 600 meiosis were analyzed for at least one metapopulation pair. EUR: Europe, EAS: East Asia, ECU+PE: Ecuador and Peru, and BR: Brazil.

the large 95 % CI intervals did not allow excluding sample size effects. In fact, statistically significant differences were only observed for the marker DXS10134, considering Europe and Brazil ($p = 0.013838$, Supplementary Table S19). For all the other markers no statistically significant differences were observed, and no metapopulation showed consistently the greater or the lower mutation rates, which will depend on the allelic frequency distributions of the populations for the different markers.

4. Conclusions

The analysis of X-STRs is one of the strategies used to tackle complex kinship cases. Its usefulness is, however, limited by the low availability of population allele/haplotype frequency data and lack of knowledge about the probability of mutation occurrence. Within the scope of this study, the population data published for the two multiplex systems most used in forensics were compiled, as well as data were generated for a significant number of haplotypes: nearly 25,000 haplotypes of either X-Decaplex or Argus X-12 kits (available as supplementary data), for various populations. The comparative population analysis performed demonstrated significant heterogeneity of populations in Africa, America, Middle East, and Southeast and South Asia. There are few data for African and South and Southeast Asian populations, and the differentiation found within these regions can be explained by the representation of distant populations and/or different ethnicities. Much more data is available for South America, and the heterogeneity observed is explained by different levels of intercontinental admixture. It is known that in South America there was a sex-biased process of admixture that led to a high European paternal contribution in most countries. Native and African maternal inheritance is, however, more heterogeneous, which explains greater differentiation regarding X-chromosomal than autosomal or Y-chromosomal markers. In contrast, populations from Europe and East Asia showed shorter genetic distances, except for some isolates, anticipating the possibility of generating metapopulations within these continents, allowing better estimates of LD and haplotype frequencies inside LGs to be obtained.

Compiled and new father-mother-daughter data allowed us to obtain more precise locus-specific estimates, also taking into consideration the parental origin and age, the type of mutation, and the metapopulation under analysis. Our data supports both the greater prevalence of paternal mutations, over the maternal ones, and the occurrence of single step mutations, over those involving a greater number of repeats. Nevertheless, caution should be taken in both cases when quantifying these biases, since in haplodiploid transmission it is not possible to ascertain which maternal allele was transmitted. In any case, this bias is less accentuated for X-chromosomal than for autosomal markers. Concerning paternal age effect on mutation rate, a significant correlation was detected between 20 and 35 years old, but data proved to be insufficient for inferences for older ages, due to large 95 % CIs.

Regarding kinship analyses with X chromosomal markers, theoretical frameworks were already established for LR computations for both euploid [4] and aneuploid individuals [120], and devoted software for the first was already developed [17]. In most investigations, the population profile is considered when calculating allele/haplotype frequencies, but the mutation rates used are usually independent of the population. It is known, however, that the different factors that affect the probability of mutation are not independent of the population and, even, that the allelic distribution profile affects the frequency of hidden mutations. Limitations in the estimates per metapopulation were also encountered, on one hand due to the lack of sufficient data and, on the other, due to the complexity of some regions with a high population substructure.

Due to the scarcity of the available data, there is currently no model that takes into account all the variables expected to affect X-STR mutation rates (e.g., repeat composition, number of repeats, age of the father, etc.). Among the models implemented in FamLinkX [17], the one that best fits the biological phenomenon, including the variables so far demonstrated to be correlated with the frequency of mutations is the "Extended model". This model, unlike the others implemented in this software, allows accounting for the probability of both different step mutations (defined by a “range”) and of mutations involving a non-integer number of repeats (denoted as “rate 2”). Nonetheless, and unfortunately, the lack of currently available data does not allow accurate estimates of these parameters.

In summary, although providing a large amount of new data, this study highlights the limitations of a still fragmented representation of worldwide populations for the X-STRs used in forensics. The available data also proved to be insufficient to evaluate the effect of all parameters that are expected to affect mutation rates. Therefore, further population and segregation studies are still needed to fill this gap, particularly for populations on understudied continents.

CRediT authorship contribution statement

Oscar Garcia: Writing – review & editing, Validation, Methodology, Data curation; **Adriana Ibarra:** Writing – review & editing, Validation,

Methodology, Data curation; **Alexis Hernandez:** Writing – review & editing, Validation, Methodology, Data curation; **Irati Manterola:** Writing – review & editing, Validation, Methodology, Data curation; **Gabriela Lima:** Writing – review & editing, Validation, Methodology, Data curation, Conceptualization; **Iva Gomes:** Writing – review & editing, Validation, Methodology, Data curation; **Anibal Gaviria:** Writing – review & editing, Validation, Methodology, Data curation; **Julieta Henao:** Writing – review & editing, Validation, Data curation; **Dario Grattapaglia:** Writing – review & editing, Validation, Methodology, Data curation; **Cristina Marrero:** Writing – review & editing, Methodology, Data curation; **Chiara Turchi:** Writing – review & editing, Validation, Methodology, Data curation; **Aluisio Trindade-Filho:** Writing – review & editing, Validation, Methodology, Data curation; **Inaki Yurrebaso:** Writing – review & editing, Validation, Methodology, Data curation; **Carlos Vullo:** Writing – review & editing, Validation, Methodology, Data curation; **Marcia Diederiche:** Writing – review & editing, Validation, Methodology, Data curation; **Florencia di Marco:** Writing – review & editing, Validation, Methodology, Data curation; **Marta Espinoza:** Writing – review & editing, Validation, Methodology, Data curation; **Patricia Domingues:** Writing – review & editing, Validation, Methodology, Data curation; **Nadia Pinto:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization; **Regina Cicarelli:** Writing – review & editing, Validation, Methodology, Data curation; **Vania Pereira:** Writing – review & editing, Writing – original draft, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation; **Laura Catelli:** Writing – review & editing, Validation, Methodology, Data curation; **Daniel Corach:** Writing – review & editing, Validation, Methodology, Data curation; **Anahí Contreras:** Writing – review & editing, Validation, Methodology, Data curation; **Maria G Garcia:** Writing – review & editing, Validation, Methodology, Data curation; **Juan M. Fernandéz:** Writing – review & editing, Validation, Methodology, Data curation; **Eileen Riego:** Writing – review & editing, Validation, Methodology, Data curation; **Anayanci Rodríguez:** Writing – review & editing, Validation, Methodology, Data curation; **Rodrigo Rodenbusch:** Writing – review & editing, Validation, Methodology, Data curation; **Carmen Tomas:** Writing – review & editing, Validation, Methodology, Data curation; **Danilo Braganholi:** Writing – review & editing, Validation, Methodology, Data curation; **Denilce Sumita:** Writing – review & editing, Validation, Methodology, Data curation; **Lisbeth Borjas:** Writing – review & editing, Validation, Methodology, Data curation; **Juan J Builes:** Writing – review & editing, Validation, Methodology, Data curation; **Ulises Toscanini:** Writing – review & editing, Validation, Methodology, Data curation; **Antonio Brehm:** Writing – review & editing, Validation, Methodology, Data curation; **Sidney Santos:** Writing – review & editing, Validation, Methodology, Data curation; **Claudia Barletta-Carrillo:** Writing – review & editing, Validation, Methodology, Data curation; **Amelia Rodríguez:** Writing – review & editing, Validation, Methodology, Data curation; **Luz Siza-Fuentes:** Writing – review & editing, Validation, Methodology, Data curation; **Cecilia Bobillo:** Writing – review & editing, Supervision, Methodology, Data curation; **Filipa Simão:** Writing – review & editing, Validation, Methodology, Data curation; **Gabriela Berardi:** Writing – review & editing, Validation, Methodology, Data curation; **Elizeu F Carvalho:** Writing – review & editing, Validation, Methodology, Data curation; **Laura Cainé:** Writing – review & editing, Validation, Methodology, Data curation; **Monica Carvalho:** Writing – review & editing, Validation, Methodology, Data curation; **Libardo Mendoza:** Writing – review & editing, Validation, Methodology, Data curation; **Joyce Martins:** Writing – review & editing, Validation, Methodology, Data curation; **Maria Porto:** Writing – review & editing, Validation, Methodology, Data curation; **Mariel Abovich:** Writing – review & editing, Validation, Methodology, Data curation; **Gerard Plaza:** Writing – review & editing, Validation, Methodology, Data curation; **Marisa Faustino:** Writing – review & editing, Validation, Methodology, Data curation; **Maria Rebelo:** Writing – review & editing, Validation, Methodology, Data curation; **Rashed Alghafri:** Writing – review & editing, Validation, Methodology, Data curation; **Yeny Posada:** Writing – review & editing, Validation, Methodology, Data curation; **Diana Aguirre:** Writing – review & editing, Validation, Methodology, Data curation; **Valerio Onofri:** Writing – review & editing, Validation, Methodology, Data curation; **Ana Mosquera:** Writing – review & editing, Validation, Methodology, Data curation; **José Pestano:** Writing – review & editing, Validation, Methodology, Data curation; **Sofia Antão-Sousa:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Formal analysis, Data curation; **Marian Pancorbo:** Writing – review & editing, Validation, Methodology, Data curation; **Leonor Gusmão:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization; **Antonio Amorim:** Writing – review & editing, Validation, Methodology, Data curation; **Cintia Alves:** Writing – review & editing, Validation, Methodology, Data curation; **Laura Baldassarri:** Writing – review & editing, Validation, Methodology, Data curation; **Cristina Arévolo:** Writing – review & editing, Validation, Methodology, Data curation. **E.C. Nascimento:** Validation, Methodology, Data curation. **P. Sanchez-Diz:** Writing – review & editing, Validation, Methodology, Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank the support of Qiagen, as reduced prices in the purchase of the Investigator Argus X-12 QS were provided for the participants in this study.

L.G. is supported by Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), Brazil (CNE-2022) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil (ref. 315007/2023-0). FEDER- Fundo Europeu de Desenvolvimento Regional through the COMPETE - Operational Program for Competitiveness and Internationalization POCI, Portugal, and FCT- Fundação para a Ciência e a Tecnologia, Ministério da Ciência, Tecnologia e Inovação supported this work through "Institute for Research and Innovation in Health Sciences" POCI-01–0145- FEDER-007274). NP (2022.04997.CEECIND) and MF (2021.08783. BD) are funded by FCT.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.fsigen.2025.103232.

References

[1] N. Pinto, L. Gusmão, T. Egeland, A. Amorim, Paternity exclusion power: comparative behaviour of autosomal and X-chromosomal markers in standard and deficient cases with inbreeding, Forensic Sci. Int. Genet. 7 (2013) 290–295, https://doi.org/10.1016/j.fsigen.2012.12.002.

[2] T.M. Diegoli, Forensic typing of short tandem repeat markers on the X and Y chromosomes, Forensic Sci. Int. Genet. 18 (2015) 140–151, https://doi.org/10.1016/j.fsigen.2015.03.013.

[3] I. Gomes, N. Pinto, S. Antão-Sousa, V. Gomes, L. Gusmão, 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.

[4] N. Pinto, L. Gusmão, A. Amorim, X-chromosome markers in kinship testing: a generalisation of the IBD approach identifying situa-tions where their contribution is crucial, Forensic Sci. Int. Genet. 5 (2011) 27–32, https://doi.org/10.1016/j.fsigen.2010.01.011.

[5] A.O. Tillmar, D. Kling, J.M. Butler, W. Parson, M. Prinz, P.M. Schneider, T. Egeland, L. Gusmão, 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.
[6] V. Pereira, L. Gusmão, The X-chromosomal STRs in forensic genetics: X chromosome STRs, in: E. Pilli, A. Berti (Eds.), Forensic DNA Analysis: Technological Development and Innovative Applications, Apple Academic Press Inc, 2021, pp. 91–112.
[7] C. Bini, S. Ceccardi, G. Ferri, S. Pelotti, M. Alù, E. Roncaglia, G. Beduschi, L. Caenazzo, E. Ponzano, P. Tasinato, C. Turchi, L. Buscemi, M. Mazzanti, A. Tagliabracci, C. Toni, I. Spinetti, R. Domenici, S. Presciuttini, Development of a heptaplex PCR system to analyse X-chromosome STR loci from five Italian population samples, A Collab. Study, Forensic Sci. Int 153 (2005) 231–236, https://doi.org/10.1016/j.forsciint.2005.05.013.
[8] H. Asamura, H. Sakai, K. Kobayashi, M. Ota, H. Fukushima, MiniX-STR multiplex system population study in Japan and application to degraded DNA analysis, Int. J. Leg. Med. 120 (2006) 174–181, https://doi.org/10.1007/s00414-005-0074-6.
[9] S. Turrina, R. Atzei, G. Filippini, D. De Leo, Development and forensic validation of a new multiplex PCR assay with 12 X-chromosomal short tandem repeats, Forensic Sci. Int. Genet. 1 (2007) 201–204, https://doi.org/10.1016/j.fsigen.2007.01.011.
[10] L. Gusmão, P. Sánchez-Diz, C. Alves, I. Gomes, M.T. Zarrabeitia, M. Abovich, I. Atmetlla, C. Bobillo, L. Bravo, J. Builes, L. Cainé, R. Calvo, E. Carvalho, M. Carvalho, R. Cicarelli, L. Catelli, D. Corach, M. Espinoza, O. García, M. Malaghini, J. Martins, F. Pinheiro, M.J. Porto, E. Raimondi, J.A. Riancho, A. Rodríguez, A. Rodríguez, B. Rodríguez Cardozo, V. Schneider, S. Silva, C. Tavares, U. Toscanini, C. Vullo, M. Whittle, I. Yurrebaso, A. Carracedo, A. Amorim, A GEP-ISFG collaborative study on the optimization of an X-STR decaplex: data on 15 Iberian and Latin American populations, Int. J. Leg. Med. 123 (2009) 227–234, https://doi.org/10.1007/s00414-008-0309-4.
[11] E.M. Ribeiro Rodrigues, F.P. Leite, M.H. Hutz, T. Palha, A.K. Ribeiro dos Santos, S.E. Santos, A multiplex PCR for 11 X chromosome STR markers and population data from a Brazilian Amazon Region, Forensic Sci. Int Genet 2 (2008) 154–158, https://doi.org/10.1016/j.fsigen.2007.10.179.
[12] J. Edelmann, S. Lutz-Bonengel, J. Naue, S. Hering, X-chromosomal haplotype frequencies of four linkage groups using the Investigator Argus X-12 Kit, Forensic Sci. Int Genet 6 (2012) e24–34, https://doi.org/10.1016/j.fsigen.2011.01.001.
[13] Q.L. Liu, D.J. Lu, L. Quan, Y.F. Chen, M. Shen, H. Zhao, Development of multiplex PCR system with 15 X-STR loci and genetic analysis in three nationality populations from China, Electrophoresis 33 (2012) 1299–1305, https://doi.org/10.1002/elps.201100558.
[14] E. Prieto-Fernández, M. Baeta, C. Núñez, M.T. Zarrabeitia, R.J. Herrera, J. J. Builes, M.M. de Pancorbo, Development of a new highly efficient 17 X-STR multiplex for forensic purposes, Electrophoresis 37 (2016) 1651–1658, https://doi.org/10.1002/elps.201500546.
[15] Y. Zhang, Z. Yu, X. Mo, X. Zhao, W. Li, H. Liu, C. Liu, R. Wu, H. Sun, Development and validation of a new 18 X-STR typing assay for forensic applications, Electrophoresis 42 (2021) 766–773, https://doi.org/10.1002/elps.202000168.
[16] A.C. Jäger, M.L. Alvarez, C.P. Davis, E. Guzmán, Y. Han, L. Way, P. Walichiewicz, D. Silva, N. Pham, G. Caves, J. Bruand, F. Schlesinger, S.J.K. Pond, J. Varlaro, K. M. Stephens, C.L. Holt, Developmental validation of the MiSeq FGx forensic genomics system for targeted next generation sequencing in forensic DNA casework and database laboratories, Forensic Sci. Int. Genet. 28 (2017) 52–70, https://doi.org/10.1016/j.fsigen.2017.01.011.
[17] 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, https://doi.org/10.1016/j.fsigen.2015.02.007.
[18] D.W. Gjertson, C.H. Brenner, M.P. Baur, A. Carracedo, F. Guidet, J.A. Luque, R. Lessig, W.R. Mayr, V.L. Pascali, M. Prinz, P.M. Schneider, N. Morling, ISFG: Recommendations on biostatistics in paternity testing, Forensic Sci. Int. Genet. 1 (2007) 223–231, https://doi.org/10.1016/j.fsigen.2007.06.006.
[19] T. Egeland, A. Salas, Estimating haplotype frequency and coverage of databases, PLoS One 3 (2008) e3988, https://doi.org/10.1371/journal.pone.0003988.
[20] M.M. Andersen, A. Caliebe, K. Kirkeby, M. Knudsen, N. Vihrs, J.M. Curran, Estimation of Y haplotype frequencies with lower order dependencies, Forensic Sci. Int. Genet. 46 (2020) 102214, https://doi.org/10.1016/j.fsigen.2019.102214.
[21] M.M. Andersen, D.J. Balding, Assessing the forensic value of DNA evidence from Y chromosomes and mitogenomes, Genes (Basel) 12 (2021) 1209, https://doi.org/10.3390/genes12081209.
[22] D. Kling, Curiosities of X chromosomal markers and haplotypes, Int. J. Leg. Med. 132 (2018) 361–371, https://doi.org/10.1007/s00414-017-1612-8.
[23] N. Pinto, V. Pereira, C. Tomas, S. Loiola, E.F. Carvalho, N. Modesti, M. Maxzud, V. Marcucci, H. Cano, R. Cicarelli, B. Januario, A. Bento, P. Brito, G. Burgos, E. Paz-Cruz, L. Díez-Juárez, S. Vannelli, M.L. Pontes, G. Berardi, S. Furfuro, A. Fernandez, D. Sumita, C. Bobillo, M.G. García, L. Gusmão, Paternal and maternal mutations in X-STRs: a GHEP-ISFG collaborative study, Forensic Sci. Int. Genet. 46 (2020) 102258, https://doi.org/10.1016/j.fsigen.2020.102258.
[24] U. Toscanini, L. Gusmão, G. Berardi, E. Raimondi, Genetic data of 10 X-STR in two Native American populations of Argentina, Forensic Sci. Int. Genet. Suppl. Ser. 2 (2009) 405–406, https://doi.org/10.1016/j.fsigss.2009.08.170.
[25] C. Valente, I. Gomes, V. Pereira, A. Amorim, L. Gusmão, M.J. Prata, Association between STRs from the X chromosome in a sample of Portuguese Gypsies, Forensic Sci. Int. Genet. Suppl. Ser. 2 (2009) 391–393, https://doi.org/10.1016/j.fsigss.2009.08.113.
[26] 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. 3 (2009) e127–8, https://doi.org/10.1016/j.fsigen.2008.10.009.
[27] J.A. Martins, J.C. Costa, G.G. Paneto, R.F. Figueiredo, L. Gusmão, P. Sánchez-Diz, A. Carracedo, R.M.B. Cicarelli, Genetic profile characterization of 10 X-STRs in four populations of the southeastern region of Brazil, Int. J. Leg. Med. 124 (2010) 427–432, https://doi.org/10.1007/s00414-010-0478-9.
[28] V. Pereira, L. Gusmão, C. Valente, R. Pereira, J. Carneiro, I. Gomes, N. Morling, A. Amorim, M.J. Prata, Refining the genetic portrait of Portuguese Roma through X-chromosomal markers, Am. J. Phys. Anthropol. 148 (2012) 389–394, https://doi.org/10.1002/ajpa.22061.
[29] L.A. Glesmann, P.F. Martina, C.I. Catanesi, Genetic Variation of X-STRs in the Wichi Population from Chaco Province, Argentina, Hum. Biol. 85 (2013) 687–698, https://doi.org/10.3378/027.085.0503.
[30] M. Lopez, A. Manrique, D.P. Aguirre, Y. Puerto, L. Gusmão, J.J. Builes, Genetic data of 10 X-STR in an Afro-descendant population sample of the Department of Chocó—Colombia, Forensic Sci. Int. Genet. Suppl. Ser. 5 (2015) e506–e507, https://doi.org/10.1016/j.fsigss.2015.09.200.
[31] E. Prieto-Fernández, A. Díaz-de Usera, M. Baeta, C. Núñez, F. Chbel, S. Nadifi, K. Rouault, C. Férec, O. Hardiman, F. Pinheiro, M.M. de Pancorbo, A genetic overview of Atlantic coastal populations from Europe and North-West Africa based on a 17 X-STR panel, Forensic Sci. Int. Genet. 27 (2017) 167–171, https://doi.org/10.1016/j.fsigen.2016.11.011.
[32] J.A. Martins, D.P. Martins, C.I.F. Oliveira-Brancati, J. Martinez, R.M.B. Cicarelli, D.R.S. Souza, Genetic characterization of an X-STR decaplex system in the State of Mato Grosso, Brazil: distribution, forensic efficiency and population structure, Int. J. Leg. Med. 131 (2017) 1523–1530, https://doi.org/10.1007/s00414-017-1578-6.
[33] A.L. Rincón, D.P. Aguirre, L. Mendoza, J.A. Camacho, J.J. Builes, Genetic data of 10 X-STRs in a population sample of the Tunja city, Department of Boyacá-Colombia, Forensic Sci. Int. Genet. Suppl. Ser. 6 (2017) e458–e459, https://doi.org/10.1016/j.fsigss.2017.09.179.
[34] M. Baeta, E. Prieto-Fernández, C. Núñez, T. Kleinbielen, P. Villaescusa, L. Palencia-Madrid, O. Alvarez-Gila, B. Martínez-Jarreta, M.M. de Pancorbo, Study of 17 X-STRs in Native American and Mestizo populations of Central America for forensic and population purposes, Int. J. Leg. Med. 135 (2021) 1773–1776, https://doi.org/10.1007/s00414-021-02536-9.
[35] C. Gomes, A. Amorim, V.O. Okolie, C. Vullo, L. Gusmão, I. Gomes, Genetic insight into Nigerian population groups using an X-chromosome decaplex system, Forensic Sci. Int. Genet. Suppl. Ser. 1 (2019) 501–503, https://doi.org/10.1016/j.fsigss.2019.10.067.
[36] B. Martínez, J.J. Builes, L. Gusmão, Y. Puerto, L. Caraballo, M.L. Bravo, Genetic data of 10 X-STR in a Colombian population of Bolivar Department, Forensic Sci. Int. Genet. Suppl. Ser. 3 (2011) e59–e60, https://doi.org/10.1016/j.fsigss.2011.08.029.
[37] M. Baeta, C. Núñez, J.M. Aznar, C. Sosa, Y. Casalod, M. Bolea, F. González-Andrade, M.M. de Pancorbo, B. Martínez-Jarreta, Analysis of 10 X-STRs in three population groups from Ecuador, Forensic Sci. Int. Genet. 7 (2013) e19–20, https://doi.org/10.1016/j.fsigen.2012.08.004.
[38] M. Baeta, M.J. Illescas, L. García, C. Núñez, E. Prieto-Fernández, S. Jiménez-Moreno, M.M. de Pancorbo, Iberian allele frequency database for 10 X-STRs, Forensic Sci. Int. Genet. 19 (2015) 76–78, https://doi.org/10.1016/j.fsigen.2015.06.009.
[39] L.M. Cainé, M.T. Zarrabeitia, J.A. Riancho, M.F. Pinheiro, Genetic data of a Brazilian population sample (Santa Catarina) using an X-STR decaplex, J. Forensic Leg. Med. 17 (2010) 272–274, https://doi.org/10.1016/j.jflm.2010.04.002.
[40] 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. Leg. Med. 125 (2011) 749–752, https://doi.org/10.1007/s00414-011-0591-4.
[41] A.O. Tillmar, Population genetic analysis of 12 X-STRs in Swedish population, Forensic Sci. Int. Genet. 6 (2012) e80–e81, https://doi.org/10.1016/j.fsigen.2011.07.008.
[42] 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.
[43] S. Uchigasaki, J. Tie, D. Takahashi, Genetic analysis of twelve X-chromosomal STRs in Japanese and Chinese populations, Mol. Biol. Rep. 40 (2013) 3193–3196, https://doi.org/10.1007/s11033-012-2394-1.
[44] M. Samejima, Y. Nakamura, P. Nambiar, K. Minaguchi, Genetic study of 12 X-STRs in Malay population living in and around Kuala Lumpur using Investigator Argus X-12 kit, Int. J. Leg. Med. 126 (2012) 677–683, https://doi.org/10.1007/s00414-012-0705-7.
[45] A. Zidkova, P. Capek, A. Horinek, P. Coufalova, Investigator® Argus X-12 study on the population of Czech Republic: comparison of linked and unlinked X-STRs for kinship analysis, Electrophoresis 35 (2014) 1989–1992, https://doi.org/10.1002/elps.201400046.
[46] X.N. Li, J.L. Zheng, J. Yao, M. Ding, H. Pang, B.J. Wang, Population data of 12 X-chromosome STR loci in Yanbian Korean samples from China, Forensic Sci. Int. Genet. 13 (2014) 143–144, https://doi.org/10.1016/j.fsigen.2014.07.012.
[47] J.C. Pinto, V. Pereira, S.L. Marques, A. Amorim, L. Alvarez, M.J. Prata, Mirandese language and genetic differentiation in Iberia: a study using X chromosome markers, Ann. Hum. Biol. 42 (2015) 20–25, https://doi.org/10.3109/03014460.2014.944215.
[48] V. Pereira, C. Tomas, J.J. Sanchez, D. Syndercombe-Court, A. Amorim, L. Gusmão, M.J. Prata, N. Morling, The peopling of Greenland: further insights from the analysis of genetic diversity using autosomal and X-chromosomal

markers, Eur. J. Hum. Genet. 23 (2015) 245–251, https://doi.org/10.1038/ejhg.2014.90.
[49] K. Rębała, S.A. Kotova, V.I. Rybakova, T.V. Zabauskaya, A.A. Shyla, A.A. Spivak, I.S. Tsybovsky, Z. Szczerkowska, Variation of X-chromosomal microsatellites in Belarus within the context of their genetic diversity in Europe, Forensic Sci. Int. Genet. 16 (2015) 105–111, https://doi.org/10.1016/j.fsigen.2014.12.011.
[50] C. Tomas, I. Skitsa, E. Steinmeier, L. Poulsen, A. Ampati, C. Børsting, N. Morling, Results for five sets of forensic genetic markers studied in a Greek population sample, Forensic Sci. Int. Genet. 16 (2015) 132–137, https://doi.org/10.1016/j.fsigen.2015.01.001.
[51] L. Poulsen, M.S. Farzad, C. Børsting, C. Tomas, V. Pereira, N. Morling, Population and forensic data for three sets of forensic genetic markers in four ethnic groups from Iran: Persians, Lurs, Kurds and Azeris, Forensic Sci. Int. Genet. 17 (2015) 43–46, https://doi.org/10.1016/j.fsigen.2015.03.010.
[52] P. Shrivastava, T. Jain, U. Gupta, V.B. Trivedi, Genetic polymorphism study on 12 X STR loci of investigator Argus X STR kit in Bhil tribal population of Madhya Pradesh, India, Leg. Med. (Tokyo). 17 (2015) 214–217, https://doi.org/10.1016/j.legalmed.2014.11.004.
[53] L. Poulsen, C. Tomas, K. Drobnič, V. Ivanova, H.S. Mogensen, A. Kondili, P. Miniati, D. Bunokiene, J. Jankauskiene, V. Pereira, N. Morling, NGMSElect™ and Investigator® Argus X-12 analysis in population samples from Albania, Iraq, Lithuania, Slovenia, and Turkey, Forensic Sci. Int. Genet. 22 (2016) 110–112, https://doi.org/10.1016/j.fsigen.2016.02.004.
[54] K. Vongpaisarnsin, A. Boonlert, K. Rasmeepaisarn, P. Dangkao, Genetic variation study of 12 X chromosomal STR in central Thailand population, Int. J. Leg. Med. 130 (2016) 1497–1499, https://doi.org/10.1007/s00414-016-1363-y.
[55] 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.
[56] J. Crnjac, P. Ozretić, S. Merkaš, M. Ratko, M. Lozančić, M. Korolija, M. Popović, G. Mršić, Investigator Argus X-12 study on the population of northern Croatia, Genet. Mol. Biol. 40 (2017) 80–83, https://doi.org/10.1590/1678-4685-GMB-2015-0261.
[57] G. Mršić, P. Ozretić, J. Crnjac, S. Merkaš, I. Račić, S. Rožić, V. Sukser, M. Popović, M. Korolija, Analysis of 12 X-STR loci in the population of south Croatia, Mol. Biol. Rep. 44 (2017) 183–189, https://doi.org/10.1007/s11033-017-4096-1.
[58] G. Mršić, P. Ozretić, J. Crnjac, S. Merkaš, V. Sukser, I. Račić, S. Rožić, L. Barbarić, M. Popović, M. Korolija, Expanded Croatian 12 X-STR loci database with an overview of aberrant profiles, Forensic Sci. Int. Genet. 34 (2018) 249–256, https://doi.org/10.1016/j.fsigen.2018.03.004.
[59] J.F. Ferragut, K. Bentayebi, R. Pereira, J.A. Castro, A. Amorim, C. Ramon, A. Picornell, Genetic portrait of Jewish populations based on three sets of X-chromosome markers: Indels, Alu insertions and STRs, Forensic Sci. Int. Genet. 31 (2017) e5–e11, https://doi.org/10.1016/j.fsigen.2017.09.008.
[60] J.F. Ferragut, M. Bassitta, V. Torrens, V. Albeza, N. Acreche, J.A. Castro, C. Ramon, A. Picornell, Analysis of 21 X-chromosome polymorphisms in urban and rural populations in Salta province (north-western Argentina), Int. J. Leg. Med. 133 (2019) 1043–1047, https://doi.org/10.1007/s00414-018-1970-x.
[61] J.F. Ferragut, K. Bentayebi, A. Barbaro, M. Ramírez, A.Y. Saguillo, C. Ramon, A. Picornell, Exploring the Western Mediterranean through X-chromosome, Int. J. Leg. Med. 135 (2021) 787–790, https://doi.org/10.1007/s00414-020-02498-4.
[62] F. Guo, Population genetic data for 12 X-STR loci in the Northern Han Chinese and StatsX package as tools for population statistics on X-STR, Forensic Sci. Int. Genet. 26 (2017) e1–e8, https://doi.org/10.1016/j.fsigen.2016.10.012.
[63] A. Sufian, M.I. Hosen, K. Fatema, T. Hossain, M.M. Hasan, A.K. Mazumder, S. Akhteruzzaman, Genetic diversity study on 12 X-STR loci of investigator® Argus X STR kit in Bangladeshi population, Int. J. Leg. Med. 131 (2017) 963–965, https://doi.org/10.1007/s00414-016-1513-2.
[64] C. Robino, D. Lacerenza, S. Aneli, C. Di Gaetano, G. Matullo, R. Robledo, C. Calò, Allele and haplotype diversity of 12 X-STRs in Sardinia, Forensic Sci. Int. Genet. 33 (2018) e1–e3, https://doi.org/10.1016/j.fsigen.2017.12.002.
[65] I. Veselinović, D. Vapa, M. Djan, N. Veličković, T. Veljović, G. Petrić, Genetic analysis of 12 X-STR loci in the Serbian population from Vojvodina Province, Int. J. Leg. Med. 132 (2018) 405–408, https://doi.org/10.1007/s00414-017-1677-4.
[66] 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.
[67] R. Tao, J. Zhang, Y. Bian, R. Dong, X. Liu, C. Jin, R. Zhu, S. Zhang, C. Li, Investigation of 12 X-STR loci in Mongolian and Eastern Han populations of China with comparison to other populations, Sci. Rep. 8 (2018) 4287, https://doi.org/10.1038/s41598-018-22665-43.
[68] J.M. Salvador, D.L.T. Apaga, F.C. Delfin, G.C. Calacal, S.E. Dennis, M.C.A. De Ungria, Filipino DNA variation at 12 X-chromosome short tandem repeat markers, Forensic Sci. Int. Genet. 36 (2018) e8–e12, https://doi.org/10.1016/j.fsigen.2018.06.008.
[69] I. Cortés-Trujillo, F. Zuñiga-Chiquette, B. Ramos-González, M.L. Chávez-Briones, K.L. Islas-González, D.A. Betancourt-Guerra, R. Peralta-Coria, G. Martínez-Cortés, H. Rangel-Villalobos, Allele and haplotype frequencies of 12 X-STRs in Mexican population, Forensic Sci. Int. Genet. 38 (2019) e11–e13, https://doi.org/10.1016/j.fsigen.2018.10.012.
[70] J. Xing, A. Adnan, A. Rakha, K. Kasim, A. Noor, J. Xuan, X. Zhang, J. Yao, D. McNevin, B. Wang, Genetic analysis of 12 X-STRs for forensic purposes in Liaoning Manchu population from China, Gene 683 (2019) 153–158, https://doi.org/10.1016/j.gene.2018.10.020.
[71] Y. Li, J. Zeng, Y. Fan, G. He, M. Xie, T. Gao, S. Wang, J. Wu, Genetic portrait and phylogenetic analysis of an Aksu Uyghur population based on the 19 X-STR system, Int. J. Leg. Med. 133 (2019) 91–93, https://doi.org/10.1007/s00414-018-1861-1.
[72] M.G. García, C.I. Catanesi, G.A. Penacino, L. Gusmão, N. Pinto, X-chromosome data for 12 STRs: Towards an Argentinian database of forensic haplotype frequencies, Forensic Sci. Int. Genet. 41 (2019) e8–e13, https://doi.org/10.1016/j.fsigen.2019.04.005.
[73] E. Kalaoglu, M. Ay, A. Ulubay, H. Canan, A. Serin, B. Alper, Population data of 12 X-STRs in Turkey, Forensic Sci. Int. Genet. Suppl. Ser. 7 (2019) 632–634, https://doi.org/10.1016/j.fsigss.2019.10.118.
[74] S. Hering, A. Klimova, J. Edelmann, German population data for 18 X-STRs: a hexaplex PCR adding two clusters of X-STRs to the Argus X-12 set and expanding the German haplotype databases, Int. J. Leg. Med. 134 (2020) 2061–2062, https://doi.org/10.1007/s00414-020-02306-z.
[75] C. Xiao, S. Li, X. Zhang, X. Yang, C. Liu, L. Chen, Population genetic analysis of Chinese Zhuang and Mulao minorities using AGCU-X19 STR kit, Int. J. Leg. Med. 134 (2020) 501–503, https://doi.org/10.1007/s00414-019-02004-5.
[76] S. Kakkar, P. Shrivastava, V. Sahajpal, A. Sharma, N. Devi, S.P. Mandal, R. K. Kumawat, Genetic diversity of X-STR markers in Jat Sikh population of Punjab, India and its comparison with other 39 global populations, Ann. Hum. Biol. 47 (2020) 490–497, https://doi.org/10.1080/03014460.2020.1772876.
[77] 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.
[78] R. Flores-Espinoza, E. Paz-Cruz, V.A. Ruiz-Pozo, M. Lopez-Carrera, A. Cabrera-Andrade, L. Gusmão, G. Burgos, Investigating genetic diversity in admixed populations from Ecuador, Am. J. Phys. Anthropol. 176 (2021) 109–119, https://doi.org/10.1002/ajpa.24341.
[79] N. Perera, R. Wijithalal, G. Galhena, G. Ranawaka, Linkage, recombination and mutation rate analyses of 16 X-chromosomal STR loci in Sri Lankan Sinhalese pedigrees, Int. J. Leg. Med. 136 (2022) 415–422, https://doi.org/10.1007/s00414-021-02762-1.
[80] N. Perera, G. Galhena, G. Ranawaka, X-chromosomal STR based genetic polymorphisms and demographic history of Sri Lankan ethnicities and their relationship with global populations, Sci. Rep. 11 (2021) 12748, https://doi.org/10.1038/s41598-021-92314-9.
[81] H.M. Hakim, H.O. Khan, S.A. Ismail, J. Lalung, A.E. Kofi, M.Y. Aziz, S. Pati8, B. R. Nelson, G.K. Chambers, H.A. Edinur, Population data and genetic characteristics of 12 X-STR loci using the Investigator® Argus X-12 Quality Sensor kit for the Kedayan population of Borneo in Malaysia, Int. J. Leg. Med. 135 (2021) 1433–1435, https://doi.org/10.1007/s00414-021-02577-0.
[82] S. Khacha-ananda, P. Mahawong, Genetic analysis of 12 X-short tandem repeats loci in a northern Thai population, Med. Sci. Law. 61 (2021) 34–43, https://doi.org/10.1177/0025802420965000.
[83] K. Bentayebi, A. Picornell, M. Bouabdeallah, D. Squalli, M. Misericòrdia, S. Amzazi, Genetic diversity of 12 X-chromosomal short tandem repeats in the Moroccan population, Forensic Sci. Int. Genet. 6 (2011) e48–e49, https://doi.org/10.1016/j.fsigen.2011.03.008.
[84] C. Bini, L.N. Riccardi, S. Ceccardi, F. Carano, S. Sarno, D. Luiselli, S. Pelotti, Expanding X-chromosomal forensic haplotype frequencies database: Italian population data of four linkage groups, Forensic Sci. Int. Genet. 15 (2015) 127–130, https://doi.org/10.1016/j.fsigen.2014.11.008.
[85] 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.
[86] J. Edelmann, R. Szibor, DXS101: a highly polymorphic X-linked STR, Int. J. Leg. Med. 114 (2001) 301–304, https://doi.org/10.1007/s004140000188.
[87] J. Edelmann, S. Hering, E. Kuhlisch, R. Szibor, Validation of the STR DXS7424 and the linkage situation on the X-chromosome, Forensic Sci. Int. 125 (2002) 217–222, https://doi.org/10.1016/s0379-0738(02)00005-1.
[88] D. Huang, Q. Yang, C. Yu, R. Yang, Development of the X-linked tetrameric microsatellite markers HumDXS6803 and HumDXS9895 for forensic purpose, Forensic Sci. Int. 133 (2003) 246–249, https://doi.org/10.1016/s0379-0738(03)00077-x.
[89] D. Athanasiadou, B. Stradmann-bellinghausen, C. Rittner, K.W. Alt, P. M. Schneider, Development of a quadruplex PCR system for the genetic analysis of X-chromosomal STR loci, Int. Congr. Ser. 1239 (2003) 311–314, https://doi.org/10.1016/S0531-5131(02)00479-X.
[90] S. Turrina, D. De Leo, Population genetic comparisons of three X-chromosomal STRs (DXS7132, DXS7133 and GATA172D05) in North and South Italy, Int. Congr. Ser. 1261 (2004) 302–304, https://doi.org/10.1016/S0531-5131(03)01626-1.
[91] K.A. Tabbada, M.C. De Ungria, L.P. Faustino, D. Athanasiadou, B. Stradmann-Bellinghausen, P.M. Schneider, Development of a pentaplex X-chromosomal short tandem repeat typing system and population genetic studies, Forensic Sci. Int. 154 (2005) 173–180, https://doi.org/10.1016/j.forsciint.2004.10.005.
[92] W. Pepinski 1, M. Skawronska, A. Niemcunowicz-Janica, E. Koc-Zorawska, J. Janica, I. Soltyszewski, Polymorphism of four X-chromosomal STRs in a Polish population sample, Forensic Sci. Int. 151 (2005) 93–95, https://doi.org/10.1016/j.forsciint.2004.09.079.
[93] W. Pepinski, A. Niemcunowicz-Janica, M. Skawronska, E. Koc-Zorawska, J. Janica, J. Berent, I. Soltyszewski, X-chromosomal polymorphism data for the ethnic minority of Polish Tatars and the religious minority of Old Believers

residing in northeastern Poland, Forensic Sci. Int. Genet. 1 (2007) 212–214, https://doi.org/10.1016/j.fsigen.2007.01.007.
[94] M.A. Tariq, O. Ullah, S.A. Riazuddin, S. Riazuddin, Allele frequency distribution of 13 X-chromosomal STR loci in Pakistani population, Int. J. Leg. Med. 122 (2008) 525–528, https://doi.org/10.1007/s00414-008-0263-1.
[95] A. Pico, A. Castillo, C. Vargas, A. Amorim, L. Gusmão, Genetic profile characterization and segregation analysis of 10 X-STRs in a sample from Santander, Colombia, Int. J. Leg. Med. 122 (2008) 347–351, https://doi.org/10.1007/s00414-007-0215-1.
[96] Q. Liu, D. Lv, L. Wu, H. Sun, X. Wu, H. Lu, Development of a five ChX STRs loci typing system, Int. J. Leg. Med. 122 (2008) 261–265, https://doi.org/10.1007/s00414-008-0223-9.
[97] 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, https://doi.org/10.1016/j.fsigen.2007.08.013.
[98] A. Nadeem, M.E. Babar, M. Hussain, M.A. Tahir, Development of pentaplex PCR and genetic analysis of X chromosomal STRs in Punjabi population of Pakistan, Mol. Biol. Rep. 36 (2009) 1671–1675, https://doi.org/10.1007/s11033-008-9367-4.
[99] S. Tetzlaff, R. Wegener, I. Lindner, Population genetic investigation of eight X-chromosomal short tandem repeat loci from a northeast German sample, Forensic Sci. Int. Genet. 6 (2012) e155–e156, https://doi.org/10.1016/j.fsigen.2012.03.007.
[100] T. Nishi, A. Kurosu, Y. Sugano, J. Kaminiwa, Y. Sekine, S. Yano, K. Honda, Application of a novel multiplex polymerase chain reaction system for 12 X-chromosomal short tandem repeats to a Japanese population study, Leg. Med. 15 (2013) 43–46, https://doi.org/10.1016/j.legalmed.2012.07.001.
[101] T.M. Diegoli, A. Linacre, M.S. Schanfield, M.D. Coble, Mutation rates of 15 X chromosomal short tandem repeat markers, Int. J. Leg. Med. 128 (2014) 579–587, https://doi.org/10.1007/s00414-014-1016-y.
[102] M. Chen, H. Ren, Z. Liu, J. Zhao, C. Chen, Y. Shi, L. Jia, F. Cheng, T. Chen, Q. Fan, Y. Yang, Y. Liu, G. Zhang, J. Yan, Genetic polymorphisms and mutation rates of 16 X-STRs in a Han Chinese population of Beijing and application examples in second-degree kinship cases, Int. J. Leg. Med. 134 (2020) 163–168, https://doi.org/10.1007/s00414-019-02047-8.
[103] S. Antão-Sousa, E. Conde-Sousa, L. Gusmão, A. Amorim, N. Pinto, Underestimation and misclassification of mutations at X chromosome STRs depend on population's allelic profile, Forensic Sci. Int. Genet. Suppl. Ser. 7 (2019) 718–720, https://doi.org/10.1016/j.fsigss.2019.10.150.
[104] H. Ossa, J. Aquino, R. Pereira, A. Ibarra, R.H. Ossa, L.A. Pérez, J.D. Granda, M. C. Lattig, H. Groot, E. Fagundes de Carvalho, L. Gusmão, Outlining the ancestry landscape of Colombian admixed and Native populations, PLoS One 11 (2016) e0164414, https://doi.org/10.1371/journal.pone.0164414.
[105] S. Avena, M. Via, E. Ziv, E.J. Pérez-Stable, C.R. Gignoux, C. Dejean, S. Huntsman, G. Torres-Mejía, J. Dutil, J.L. Matta, K. Beckman, E.G. Burchard, M.L. Parolin, A. Goicoechea, N. Acreche, M. Boquet, C. Ríos Part Mdel, V. Fernández, J. Rey, M. C. Stern, R.F. Carnese, L. Fejerman, Heterogeneity in genetic admixture across different regions of Argentina, PLoS One 7 (2012) e34695, https://doi.org/10.1371/journal.pone.0034695.
[106] F.S.N. Manta, R. Pereira, R. Vianna, A.R.B. Araújo, D.L.G. Gitaí, D.A. Silva, E. V. Wolfgramm, I.M. Pontes, J.I. Aguiar, M.O. Moraes, E.F. Carvalho, L. Gusmão, Revisiting the genetic ancestry of Brazilians using autosomal AIM-Indels, PLoS One 8 (2013) e75145, https://doi.org/10.1371/journal.pone.0075145.
[107] R. Santangelo, F. González-Andrade, C. Børsting, A. Torroni, V. Pereira, N. Morling, Analysis of ancestry informative markers in three main ethnic groups from Ecuador supports a trihybrid origin of Ecuadorians, Forensic Sci. Int. Genet. 31 (2017) 29–33, https://doi.org/10.1016/j.fsigen.2017.08.012.
[108] M. Muzzio, J.M.B. Motti, P.B.P. Sepulveda, M.C. Yee, T. Cooke, M.R. Santos, V. Ramallo, E.L. Alfaro, J.E. Dipierri, G. Bailliet, C.M. Bravi, C.D. Bustamante, E. E. Kenny, Population structure in Argentina, PLoS One 13 (2018) e0196325, https://doi.org/10.1371/journal.pone.0196325.
[109] F. Messina, T. Di Corcia, M. Ragazzo, C.S. Mellado, I. Contini, P. Malaspina, B. M. Ciminelli, O. Rickards, C. Jodice, Signs of continental ancestry in urban populations of Peru through autosomal STR loci and mitochondrial DNA typing, PLoS One 13 (2018) e0200796, https://doi.org/10.1371/journal.pone.0200796.
[110] A.K. Zambrano, A. Gaviria, S. Cobos-Navarrete, C. Gruezo, C. Rodríguez-Pollit, I. Armendáriz-Castillo, J.M. García-Cárdenas, S. Guerrero, A. López-Cortés, P. E. Leone, A. Pérez-Villa, P. Guevara-Ramírez, V. Yumiceba, G. Fiallos, M. Vela, C. Paz-Y-Miño, The three-hybrid genetic composition of an Ecuadorian population using AIMs-InDels compared with autosomes, mitochondrial DNA and Y chromosome data, Sci. Rep. 9 (2019) 9247, https://doi.org/10.1038/s41598-019-45723-w.
[111] B. Guillen-Guio, J.M. Lorenzo-Salazar, R. González-Montelongo, A. Díaz-de Usera, I. Marcelino-Rodríguez, A. Corrales, A. Cabrera de León, S. Alonso, C. Flores, Genomic Analyses of Human European Diversity at the Southwestern Edge: Isolation, African Influence and Disease Associations in the Canary Islands, Mol. Biol. Evol. 35 (2018) 3010–3026, https://doi.org/10.1093/molbev/msy190.
[112] R. Fregel, A.C. Ordóñez, J. Santana-Cabrera, V.M. Cabrera, J. Velasco-Vázquez, V. Alberto, M.A. Moreno-Benítez, T. Delgado-Darias, A. Rodríguez-Rodríguez, J. C. Hernández, J. Pais, R. González-Montelongo, J.M. Lorenzo-Salazar, C. Flores, M.C. Cruz-de-Mercadal, N. Álvarez-Rodríguez, B. Shapiro, M. Arnay, C. D. Bustamante, Mitogenomes illuminate the origin and migration patterns of the indigenous people of the Canary Islands, PLoS One 14 (2019) e0209125, https://doi.org/10.1371/journal.pone.0209125.
[113] P. Nakka, S. Pattillo Smith, A.:H. O'Donnell-Luria, K.F. McManus, 23andMe Research Team, J.L. Mountain, S. Ramachandran, J.F. Sathirapongsasuti, Characterization of Prevalence and Health Consequences of Uniparental Disomy in Four Million Individuals from the General Population, Am. J. Hum. Genet. 7 (2019) 921–932, https://doi.org/10.1016/j.ajhg.2019.09.016.
[114] L.D. Hurst, H. Ellegren, Sex biases in the mutation rate, Trends Genet. 14 (1998) 446–452, https://doi.org/10.1016/s0168-9525(98)01577-7.
[115] S. Antão-Sousa, E. Conde-Sousa, L. Gusmão, A. Amorim, N. Pinto, How often have X- and autosomal-STRs mutations equivocal parental origin been assigned? Forensic Sci. Int. Genet. Suppl. Ser. 8 (2022) 99–101, https://doi.org/10.1016/j.fsigss.2022.09.035.
[116] M.A.W. Sayres, K.D. Makova, Genome analyses substantiate male mutation bias in many species, BioEssays 33 (2011) 938–945, https://doi.org/10.1002/bies.201100091.
[117] P. Forster, C. Hohoff, B. Dunkelmann, M. Schürenkamp, H. Pfeiffer, F. Neuhuber, B. Brinkmann, Elevated germline mutation rate in teenage fathers, Proc. R. Soc. B 282 (2015) 20142898, https://doi.org/10.1098/rspb.2014.2898.
[118] S. Antão-Sousa, L. Gusmão, N.M. Modesti, S. Feliziani, M. Faustino, V. Marcucci, C. Sarapura, J. Ribeiro, E. Carvalho, V. Pereira, C. Tomas, M.M. de Pancorbo, M. Baeta, R. Alghafri, R. Almheiri, J.J. Builes, N. Gouveia, G. Burgos, M.L. Pontes, A. Ibarra, C.V. da Silva, R. Parveen, M. Benitez, A. Amorim, N. Pinto, Microsatellites' mutation modeling through the analysis of the Y-chromosomal transmission: results of a GHEP-ISFG collaborative study, Forensic Sci. Int. Genet. 69 (2023) 102999, https://doi.org/10.1016/j.fsigen.2023.102999.
[119] S. Antão-Sousa, E. Conde-Sousa, L. Gusmão, A. Amorim, N. Pinto, How frequently are Autosomal and X-STRs multistep mutations perceived as single-step? Forensic Sci. Int. Genet. Suppl. Ser. 8 (2022) 165–166, https://doi.org/10.1016/j.fsigss.2022.10.022.
[120] M. Faustino, L. Gusmão, A. Amorim, D. Kling, N. Pinto, A mathematical framework for genetic relatedness analysis involving X chromosome aneuploidies, Forensic Sci. Int. Genet. 74 (2025) 103128, https://doi.org/10.1016/j.fsigen.2024.103128.