Forensic Science International: Genetics 46 (2020) 102258

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Paternal and maternal mutations in X-STRs: A GHEP-ISFG collaborative study

Nádia Pinto[a,b,c,*], Vânia Pereira[d], Carmen Tomas[d], Silvia Loiola[e], Elizeu F. Carvalho[e], Nidia Modesti[f], Mariana Maxzud[f], Valeria Marcucci[g], Hortensia Cano[g], Regina Cicarelli[h], Bianca Januario[h], Ana Bento[i], Pedro Brito[i], Germán Burgos[j], Elius Paz-Cruz[k], Laura Díez-Juárez[l], Silvia Vannelli[m], Maria de Lurdes Pontes[n], Gabriela Berardi[o], Sandra Furfuro[p], Alberto Fernandez[q], Denilce Sumita[r], Cecilia Bobillo[s], Maria Gabriela García[t], Leonor Gusmão[e]

[a] Institute of Pathology and Molecular Immunology from University of Porto (IPATIMUP), Portugal
[b] Instituto de Investigação e Inovação em Saúde, I3S, Universidade do Porto, Portugal
[c] CMUP, Centro de Matemática da Universidade do Porto, Portugal
[d] Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark
[e] Laboratório de Diagnóstico por DNA (LDD), Universidade do Estado do Rio de Janeiro, Brazil
[f] Centro de Genética Forense, Poder Judicial de Córdoba, Argentina
[g] Laboratorio Regional de Investigación Forense, Tribunal Superior de Justicia de Santa Cruz, Argentina
[h] UNESP-Universidade Estadual Paulista, Faculdade de Ciências Farmacêuticas, Laboratório de Investigação de Paternidade-NAC, São Paulo, Brazil
[i] Instituto Nacional de Medicina Legal e Ciências Forenses, I.P., Serviço de Genética e Biologia Forenses, Delegação do Centro, Portugal
[j] Escuela de Medicina, Facultad de Ciencias de la Salud, Universidad de Las Américas (UDLA), Quito, Ecuador
[k] Laboratorio de ADN de la Fiscalía General del Estado, Quito, Ecuador
[l] Departamento de Biología, Servicio de Criminalística de la Guardia Civil, Spain
[m] Laboratorio Regional de Genética Forense, Poder Judicial de Río Negro, Argentina
[n] Instituto Nacional de Medicina Legal e Ciências Forenses, I.P., Serviço de Genética e Biologia Forenses, Delegação do Norte, Portugal
[o] PRICAI-Fundación Favaloro, Buenos Aires, Argentina
[p] Laboratorio de Análisis de ADN Facultad de Ciencias Médicas Universidad Nacional de Cuyo, Mendoza, Argentina
[q] LabGenetics: Laboratorio de Genética Clínica S.L., Madrid, Spain
[r] Genomic Engenharia Molecular Molecular, São Paulo, Brazil
[s] Laboratorio de Genética Forense, Poder Judicial de la Provincia de La Pampa, Argentina
[t] Laboratorio MANLAB, Área de Filiaciones, Buenos Aires, Argentina

ARTICLE INFO

Keywords:
Argus kit
Mutation rate
X chromosome
Population database

ABSTRACT

The GHEP-ISFG organized a collaborative study to estimate mutation rates for the markers included in the Investigator Argus X-12 QS kit Qiagen. A total of 16 laboratories gathered data from 1,612 father/mother/daughter trios, which were used to estimate both maternal and paternal mutation rates, when pooled together with other already published data. Data on fathers and mothers' age at the time of birth of the daughter were also available for ~93 % of the cases. Population analyses were computed considering the genetic information of a subset of 1,327 unrelated daughters, corresponding to 2,654 haplotypes from residents in several regions of five countries: Argentina, Brazil, Ecuador, Portugal and Spain. Genetic differentiation analyses between the population samples from the same country did not reveal signs of significant stratification, although results from Hardy-Weinberg and linkage disequilibrium tests indicated the need of larger studies for Ecuador and Brazilian populations. The high genetic diversity of the markers resulted in a large number of haplotype combinations, showing the need of huge databases for reliable estimates of their frequencies.

It should also be noted the high number of new alleles found, many of them not included in the allelic ladders provided with the kit, as very diverse populations were analyzed. The overall estimates for locus specific average mutation rates varied between 7.5E-04 (for DXS7423) and 1.1E-02 (for DXS10135), the latter being a troublesome figure for kinship analyses. Most of the found mutations (~92 %) are compatible with the gain or loss of a single repeat. Paternal mutation rates showed to be 5.2 times higher than maternal ones. We also found that older fathers were more prone to transmit mutated alleles, having this trend not been observed in the case of the mothers.

* Corresponding author at: IPATIMUP/i3S, R. Alfredo Allen, 4200-135 Porto, Portugal.
E-mail addresses: npinto@ipatimup.pt, nmgapinto@gmail.com (N. Pinto).

https://doi.org/10.1016/j.fsigen.2020.102258
Received 19 June 2019; Received in revised form 6 October 2019; Accepted 29 January 2020
Available online 05 February 2020
1872-4973/

1. Introduction

The analysis of markers located in the X chromosome can be useful to solve complex kinship cases in forensic genetics [1–5]. For this purpose, different genotyping methods have been described, including sets of X-chromosomal specific STRs, Indels and SNPs (e.g [6–14]).

In forensic genetics, the most widely used X-STR kit is the Investigator® Argus X-12 kit (Qiagen, Hilden, Germany), which comprises 12 markers organized in 4 linkage groups. Population data for this kit have been gathered from many populations around the world (e.g. [15–27]), although haplotype reference databases are still lacking for forensic use in many countries.

Because linked loci are more prone to show linkage disequilibrium (LD) in populations, a proper application of marker sets in the same chromosome usually requires much larger population databases than for unlinked markers. In the presence of LD, the need of large databases also increases with the inclusion of highly variable STRs, since it is necessary to estimate frequencies for many rare haplotypes that are absent or poorly represented in reference population samples.

Large databases are also important to detect LD between markers [5,28]. Since loci included in the Argus kit are closely located (three markers in each linkage group), and the X chromosome only recombines in female meiosis, it is expected to find higher LD levels between markers. Using haplotypic data from Swedish males, Kling et al. [28] showed a high probability of not detecting LD in samples of less than 400 male individuals, and LD could not be detected for linkage group 2 in 17 % of the cases where samples with 600 males were considered.

Moreover, for the application of X-STRs in kinship analyses, it is important to have reliable estimates of recombination and mutation rates [4,29]. In both cases, it is necessary to study allele/haplotype segregation between relatives. Mutation rates are usually estimated through the proportion of allele transfers resulting in mendelian incompatibilities in duos or trios from paternity cases or using information from large pedigrees. As in the current study, most approaches rely on the number of observed rather than occurring mutations. Indeed, these estimates are conservative as mutations can occur without leading to Mendelian incompatibilities – the so called “hidden” or “covered” mutations. For biases affecting these estimates, statistical adjustments were presented for autosomal markers [30–32]. The bias is greater when duos, instead of trios, are analysed and, in the case of X-chromosomal transmission, bias is greater when estimating mutation in mother-daughter than in father-daughter or mother-son duos [54].

Germline mutations were investigated for different X-STRs, and population and marker-specific rates were reported, although, in most cases, a limited amount of data prevents statistically accessing significant differences among markers or populations (e.g. [25,33–36]). To date, few data are available for the full set of 12 X-STRs from the Argus kit. Two studies reported average mutation rates of 3.3×10^{-3} and 5.1×10^{-3} per allele transmission in 513 (mother/son, mother/daughter) and 345 (father-daughter) duos, respectively [25,37]. The different observed values can be explained by a significantly higher male than female germline mutation rate, which was found by Tomas et al. [25].

For the reasons mentioned above, available data on X-STRs of the Argus kit are still insufficient to obtain good estimates of forensically relevant parameters.

Therefore, the Spanish and Portuguese speaking working group of the International Society for Forensic Genetics (GHEP-ISFG) organized a collaborative study to collect data for the 12 X-STR markers included in the Investigator Argus X-12 QS kit. This study aimed at improving the quality of the estimates that are relevant in forensic genetics, namely haplotype frequencies in different populations, LD between markers, and locus-specific mutation rates.

2. Material and methods

After approval at the general assembly, a working group was created with the purpose of carrying out the present study, which was open to all members of the GHEP-ISFG. The participating laboratories had to: (i) present a certificate of the proficiency test of the GHEP-ISFG, showing correct results for the Argus kit markers; and (ii) use the latest version of the Argus kit - Investigator® Argus X-12 QS (Qiagen) - to genotype at least 70 father/mother/daughter trios.

Table 1
Population of origin and distribution of the 1,612 analyzed trios.

	Population	Nº of trios [a)]	Nº of unrelated haplotypes
Africa	Somalia - General population	73	b)
America	Greenland - General population	104	b)
	Argentina - Santa Cruz	140	250
	Argentina – Mendoza	70	140
	Argentina - Rio Negro	74	142
	Argentina - La Pampa	69	136
	Argentina - Buenos Aires	70	140
	Argentina – Cordoba	143	286
	Brazil - Rio de Janeiro	149	266
	Brazil - General population	70	140
	Brazil - São Paulo	141	282
	Ecuador - General population	180	300
Europe	Portugal - General population	140	278
	Portugal - North region	72	144
	Spain - General population	106	150
	Other populations	11	c)
	Total	1,612	2,654

a) Includes 97 family clusters; b) Haplotype data from Somalia and Greenland were not considered for population analyses since they are partially published [25]; c) Haplotype data were not considered for population analyses since concerns to few information scattered by several populations.

2.1. Samples collection and genotyping

A total of 16 laboratories participated in this study, providing data from 1,612 father/mother/daughter trios. The trios were collected from 15 populations from Africa, America and Europe (Table 1). Most samples belonged to paternity cases, although some healthy volunteers were also included in the samples from Santa Cruz (Argentina), Rio Negro (Argentina), and Spain. In most cases, the samples encompassed residents in each of the 15 analyzed populations. In samples from paternity investigations, the biological relationship was previously confirmed using autosomal STRs ($LR > 10^5$). Each laboratory ensured the anonymization of the samples and the accomplishment of the legal and ethical requirements for their use in this research project.

All samples were genotyped using the Investigator Argus X-12 QS (Qiagen), following manufacturer’s instructions (Investigator® Argus X-12 QS Handbook), except for 299 samples where laboratories already had profiles obtained with the older version of the kit. In these cases, only profiles that were heterozygote for the markers DXS10101, DXS10146 and DXS10148 were accepted. The remaining samples had to be retyped with the new version of the kit to avoid null alleles. Indeed, according to the manufacturer, additional primers were included in the new version to prevent the high frequency of null alleles observed for these three markers in African populations [38].

2.2. Data analysis

A total of 1,612 father/mother/daughter trios, genotyped for 12 X-STRs, were collected and used to estimate locus specific mutation rates. Confidence intervals for mutation rates were estimated from the binomial standard deviation. Information on the parents’ age at the time of birth of the daughter was available for ~93 % of the trios. A total of

1,327 unrelated females were selected among the daughters, and gametic phase was determined using the haplotype information from the father (Table 1). These unrelated females were used for population analyses and to calculate forensic statistics. In each family cluster, only one (randomly selected) daughter was considered for population analyses.

Estimation of allele and haplotype frequencies and pairwise F_{ST} genetic distances were calculated using the software Arlequin ver 3.5.2.2 [39]. The same software was used to test population differentiation, Hardy-Weinberg equilibrium and pairwise linkage disequilibrium between loci. The significance level of 0.05 was adjusted by applying the Bonferroni's correction for multiple tests, namely by considering the number of markers used to test Hardy-Weinberg equilibrium (0.05/12), and the total number of pairwise comparisons in population differentiation (0.05/78) and LD (0.05/66) tests.

Statistics for forensic efficiency evaluation, namely mean exclusion chance in trios involving daughters (MECT) as well as in father/daughter duos (MECD), power of discrimination in females (PDF) and in males (PDM) were calculated using the formulae from Desmarais et al. [40].

3. Results and discussion

3.1. Population genetics analysis

A sample of 1,327 unrelated females was selected for population analyses. By selecting the daughters from the studied trios, it was possible to determine the genotype gametic phase, and to create a database comprising 2,654 unrelated haplotypes.

Pairwise comparisons for the 12 X-STRs between all samples showed low, non-statistically significant differences among populations from the same country, namely Argentina, Brazil and Portugal ($F_{ST} \leq 0.0023$; $p \geq 0.0239$; Supplementary Table S1A). Non-significant differences were also observed between the samples from Portugal and Spain ($F_{ST} \leq -0.0012$; $p \geq 0.5388$). Similar results were obtained when performing the test for markers in each linkage group (LG1 to LG4), separately (Supplementary Tables S1B-E). When comparing samples from different countries, the larger distances were observed for markers at LG2 and LG3, and Ecuador was the population that most departure from the remaining.

Therefore, Hardy-Weinberg equilibrium and LD tests were performed after pooling the samples from the same country, as well as by joining Portuguese and Spanish samples in a single population from Iberia.

3.1.1. Hardy-Weinberg equilibrium analysis

No statistically significant deviations from the Hardy-Weinberg equilibrium (HWE) expectations were detected in the Argentinian sample (Supplementary Table S2), indicating no signs of population stratification for the significance level considered, which supports the use of a single database for the analysed markers.

Concerning the samples from Brazil, a significant departure from the HWE was observed at DXS10134 locus (Supplementary Table S2). However, the observed deviation was not significant when analysing separately the three samples from Brazil (general population), Rio de Janeiro and São Paulo. Therefore, these results indicate that larger studies on Brazilian populations are required to investigate if a common database can be used for these markers.

Ecuador did not reveal statistically significant deviations from HWE, for the 12 X-STRs.

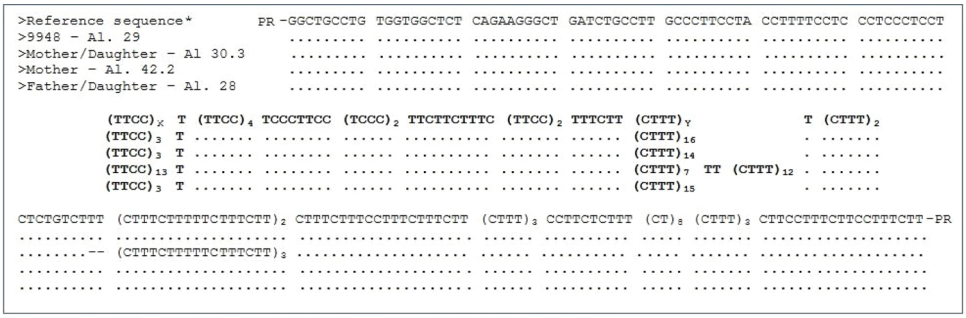
The HWE test showed a statistically significant deviation for DXS10079 in the Iberian population sample (Supplementary Table S2), due to a lower frequency of heterozygotes than the expected (observed 0.787 and expected 0.828). When the test was performed for the Spanish and Portuguese populations separately, the observed and expected values of heterozygosity were almost the same (0.787 and 0.838 for Spain, 0.787 and 0.825 for Portugal), although the deviation was not statistically significant due to a lower sample size. Therefore, the results indicate that the observed excess of homozygotes is not due to differences between Spain and Portugal and can most likely be explained by the presence of undetected null alleles in Iberian populations.

3.1.2. Linkage disequilibrium analysis

Pairwise linkage disequilibrium analysis was performed in all populations: Argentina with a total of 1,094 haplotypes, Brazil 688 haplotypes, Ecuador 300 haplotypes, and Iberia 572 haplotypes. The results are shown in Supplementary Table S3.

Table 2
Forensic efficiency statistics for each linkage group and overall values for the Argus system: Mean exclusion chance in trios involving daughters (MECT) and in father/daughter duos (MECD); power of discrimination in males (PDM) and in females (PDF).

	MECT	MECD	PDF	PDM
Argentina				
LG1	99.615 %	99.236 %	99.997 %	99.617 %
LG2	99.337 %	98.691 %	99.991 %	99.341 %
LG3	98.590 %	97.254 %	99.963 %	98.608 %
LG4	99.427 %	98.866 %	99.994 %	99.430 %
Total	**99**.9999998 %	**99.999**997 %	**99.999999999999**994 %	**99**.9999998 %
Brazil				
LG1	99.708 %	99.418 %	99.998 %	99.709 %
LG2	99.380 %	98.774 %	99.992 %	99.384 %
LG3	99.237 %	98.494 %	99.989 %	99.243 %
LG4	99.570 %	99.147 %	99.996 %	99.572 %
Total	**99.99999**994 %	**99.9999**991 %	**99.9999999999999**994 %	**99.99999**994 %
Ecuador				
LG1	99.165 %	98.353 %	99.986 %	99.171 %
LG2	98.917 %	97.874 %	99.977 %	98.928 %
LG3	96.666 %	93.710 %	99.805 %	96.756 %
LG4	99.069 %	98.168 %	99.983 %	99.078 %
Total	**99.999**997 %	**99.9996** %	**99**.9999999999990 %	**99.999**997 %
Iberia				
LG1	99.631 %	99.266 %	99.997 %	99.632 %
LG2	99.080 %	98.190 %	99.984 %	99.088 %
LG3	98.924 %	97.887 %	99.978 %	98.935 %
LG4	99.396 %	98.805 %	99.993 %	99.400 %
Total	**99**.9999998 %	**99.999**997 %	**99.999999999999**93 %	**99**.9999998 %



*Genome build Dec.2013, GRCh38/hg38 - chrX:150403779-150404155;
Note: The repeat sequence (in bold) is described in [42].

Fig. 1. Sequencing results obtained for DXS10146 alleles in a trio showing a Mendelian incompatibility due to a null allele shared by the mother and the daughter.

Associations between markers from the same linkage group were detected in all populations, although in Ecuador it was only detected inside LG1. This can be explained by its lower sample size, which reduces the chance of detecting the presence of LD [28]. Moreover, Ecuador revealed significant associations in 3 pairwise comparisons involving markers from different linkage groups. LD between non-linked markers can occur due to population stratification, which is also associated with an excess of homozygotes. The average value of observed heterozygotes in the Ecuadorian sample was indeed lower than the expected for a population in Hardy-Weinberg equilibrium (0.771 and 0.796, respectively), although deviations were not statistically significant, which can, once more, be due to the small sample size (150 genotypes). In conclusion, the overall results do not allow excluding the presence of a genetic stratification inside the sample from Ecuador, that needs to be investigated in more detail in larger samples with well-defined origin.

3.1.3. Haplotype frequencies and forensic parameters

In Supplementary Table S4, we present the haplotype frequencies per linkage group for each population group: Argentina, Brazil, Ecuador and Iberia. The parameters of forensic interest, calculated for each linkage group and the overall values for the 12 X-STRs included in the Investigator Argus X-12 QS kit, are presented in Table 2. The population from Brazil showed the highest values of mean exclusion chance and discrimination capacity, followed by Argentina and Iberia, with similar results. Ecuador was the population with the lowest values of forensic efficiency for the investigated set of X-STR markers.

3.2. Analysis of mutations

In a total of 38,688 allele transfers, 157 Mendelian incompatibilities were found, one of which explained either by mutation or by the presence of a silent allele (Supplementary Table S5). In this case, at DXS10146 locus, the mother was 42.2 and both the father and the daughter were 28. Therefore, a mutation from an allele 42.2 to 28 would be necessary to explain the genotypic configurations, being more plausible to assume the presence of a null allele. To test this hypothesis, the DXS10146 locus was amplified in singleplex, using the PCR amplification protocol described in [41]. The use of the new primers revealed no Mendelian incompatibilities among the trio, with both the mother and the daughter being heterozygotes for the DXS10146 locus. Sequencing results are presented in Fig. 1, showing a TT deletion and a 17 bp insertion, at the repeat flaking region, 8 bases downstream of the repeat.

Our results were analysed together with others already published for 9 out of the 12 X-STRs from this study, for father/mother/daughter trios [33,42–53] (Table 3). Although there are other reports on X-STR mutations, they consider only duos or a mixture of different pedigrees (not susceptible of being discriminated), preventing a joint analysis with our data.

Table 3
Number of mutations considering the parental origin and the minimal number of gains (+) or losses (−) of repeats or bases that can explain the inconsistencies between parental and filial genotypes. Data from our study and from others previously published [33,42–53] were analysed together.

Most Parsimonious Parental origin	Type of mutation						Total
	The same (repeat variation)					Different (bp variation)	
	{−2}	{−1}	{+1}	{+2}	{−1} V {+1}	{+0.1}	
Paternal	2	75	54	2	0	1	134
Maternal	0	8	15	1	1	1	26
Undetermined	0	4	11	0	6	0	21
Total	2	87	80	3	7	2	181

Out of the 181 mutations, 174 are compatible with the gain or loss of a single repeat, 5 with the gain or loss of two repeats, and two by the gain of one base pair. In 7 cases, inconsistencies are explainable by either the gain or loss of a single repeat. It is worth nothing that, in accordance with the other mutation rate studies, we assumed single-step mutations whenever they explain the genotypic configurations, which means that multistep mutations may be underestimated [43]. In a recent study, simulations of two-step mutations in 8 X-STRs showed that 14 % of these would be attributed to a single-step mutation [54].

Considering the parental origin, we found 5.2 more paternal than maternal mutations. A total of 134 out of the 181 genotypic configurations were compatible with a paternal mutation, 26 with a maternal one, and for 21 cases both origins are possible (Table 3). In humans, it is well known that mutations are more frequent in paternal than in maternal germline [55], and X-STR studies agree with this [25,43,46]. However, the increase of mutation rate with age is higher in males than in females, and thus, the ratio between paternal and maternal mutations is age-dependent [56]. This prevents a straightforward comparison of the estimate of ratio of male to female mutations we observed with those from studies with no age information available.

In our dataset, we found one case of simultaneous paternal and maternal mutations at the same locus and, in five trios, mutations were observed in two different loci.

3.2.1. Mutation rates and parental age

From the 1,612 analyzed trios, age information at the year of the daughter's birth was obtained for 1,492 mothers and 1,488 fathers. The age distribution, as well as the proportion of cases resulting in at least one mutation, are presented in Fig. 2. Globally, fathers tend to be older than mothers (average age at time of the birth of the daughter ~30.1

and ~25.8 years, respectively), and show higher mutation rates, the likelihood of which seems to increase also with the age. Particularly, note the case of the fathers between 51 and 55 years-old, for which at least one undoubted paternal mutation was observed in nearly 14 % of the meiosis. In mothers, no correlation between age and mutation rate was detected, maybe because it is too subtle to be perceived from the currently available data, as suggested by Ségurel et al. [56].

The increase with age of male germline mutations is well documented (e.g. [55–57]) and it should be accounted for likelihood calculations in kinship cases. Nevertheless, as for other parameters that also influence mutation rates in STRs (namely repeat structure and allele length, and the sex), a large amount of information is required for age specific estimates, by joining data from different studies. For the X-STRs included in this study, this is still not possible due to low amount and large heterogeneity of the information published on mutation rates.

3.2.2. Locus-specific mutations

After pooling our data with other available in the literature for father/mother/daughter trios [33,42–53], mutation rates and their 95 % confidence intervals were estimated for each marker (Table 4). Table 4 also shows the expected paternal and maternal mutation rates estimated resorting to trios and after the proportional distribution of the mutations with undetermined origin.

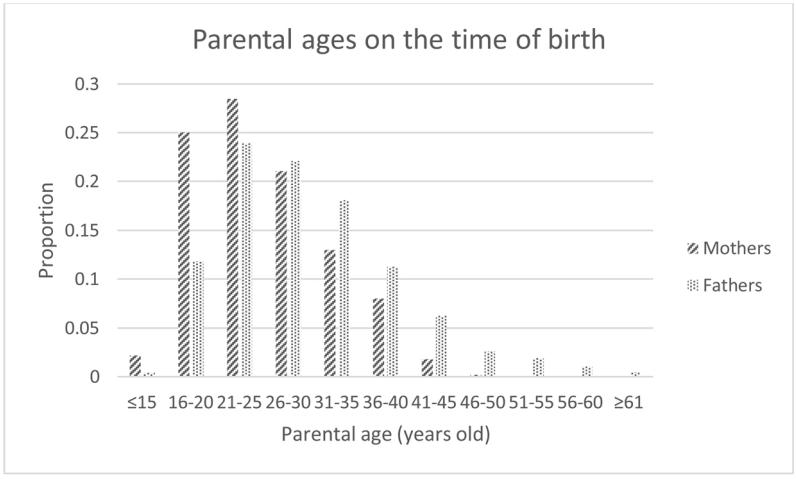
Detailed information on paternal and maternal allele transmissions in our samples is provided in supplementary material (Tables S5–S7). Globally, observed mutation rates varied from 7.5E-04 (for DXS7423) to 1.1E-02 (for DXS10135).

Note that the 21 observed mutations with undetermined origin (more than 10 % of the total) are not detected when father/daughter or mother/daughter duos are analyzed. Indeed, it is known that the use of duos instead of trios increases the possibility of hidden mutations and, consequently, resulting in a greater bias in estimates of actual mutation rates. As expected, in our data, the number of observed mutations in trios were higher or equal to those found in duos (Table S6), since the consideration of the other parent may uncover mutations. The greatest difference between the mutation rates estimated through duos or trios was achieved for the HPRTB marker, for both males and females. For this marker the estimated mutation rate almost doubled when trios instead of duos were considered.

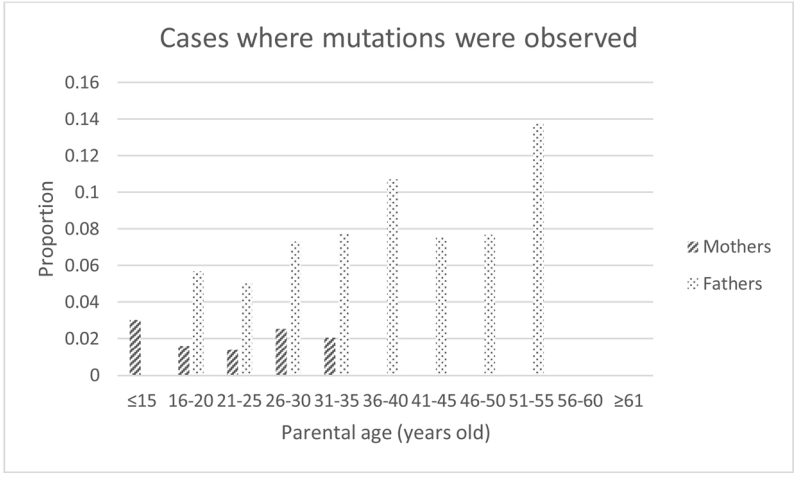
4. Conclusions

The present collaborative study of the GHEP-ISFG working group allowed the compilation of a large amount of population and segregation data for the 12 X-STRs most widely used in forensics, contributing with valuable information for their application in kinship investigations.

In this study, haplotype data were obtained for different populations from Argentina, Brazil, Ecuador, Portugal and Spain. Differentiation analyses, Hardy-Weinberg equilibrium and linkage disequilibrium tests did not reveal signs of significant stratification in Argentina, supporting the use of a national database. The same was found between Portuguese



a.



b.

Fig. 2. a. Distribution of the parental ages at the time of the daughter's birth for the cases where such information was available. **b.** Proportion of cases where the meiosis resulted in at least one mutation for both males and females. Note that only mutations where the parental origin was possible to determine were considered. This implies that proportions presented are necessarily conservative.

Table 4
Estimates for locus specific mutation rates for male and female meiosis, considering the analysis of trios. The total number of 2*1612 meiosis considered in this study were gathered with other published data [33,42–53]. A level of confidence equal to 95 % were considered in the calculations.

# Markers	#Meiosis	Mutation Rate				Paternal Mutation Rate			Maternal Mutation Rate		
		# Mutations	MR	Lower	Upper	PMR	Lower	Upper	MMR	Lower	Upper
DXS10103	3224	6	1.9E-03	6.8E-04	4.0E-03	3.7E-03	1.4E-03	8.1E-03	0.0E+00	0.0E+00	2.3E-03
DXS8378	4082	5	1.2E-03	4.0E-04	2.9E-03	7.3E-04	5.3E-05	3.1E-03	7.3E-04	5.3E-05	3.1E-03
DXS10101	3452	11	3.2E-03	1.6E-03	5.7E-03	5.7E-03	2.7E-03	1.1E-02	7.1E-04	1.8E-05	3.6E-03
DXS10134	3524	17	4.8E-03	2.8E-03	7.7E-03	6.6E-03	3.4E-03	1.2E-02	3.0E-03	9.8E-04	7.0E-03
DXS10074	3776	19	5.0E-03	3.0E-03	7.8E-03	1.0E-02	6.1E-03	1.6E-02	0.0E+00	0.0E+00	2.0E-03
DXS7132	4726	31	6.6E-03	4.5E-03	9.3E-03	1.2E-02	7.7E-03	1.7E-02	1.5E-03	3.0E-04	4.1E-03
DXS10135	3452	37	1.1E-02	7.6E-03	1.5E-02	1.7E-02	1.2E-02	2.5E-02	4.2E-03	1.7E-03	8.5E-03
DXS7423	4024	3	7.5E-04	1.5E-04	2.2E-03	9.9E-04	1.2E-04	3.6E-03	5.0E-04	1.3E-05	2.8E-03
DXS10146	3224	9	2.8E-03	1.3E-03	5.3E-03	4.2E-03	1.6E-03	8.8E-03	1.4E-03	1.7E-04	4.8E-03
DXS10079	3534	22	6.2E-03	3.9E-03	9.4E-03	1.1E-02	6.7E-03	1.7E-02	1.4E-03	1.7E-04	4.7E-03
HPRTB	4132	7	1.7E-03	6.8E-04	3.5E-03	2.0E-03	5.5E-04	5.1E-03	1.4E-03	2.3E-04	4.2E-03
DXS10148	3224	14	4.3E-03	2.4E-03	7.3E-03	6.8E-03	3.4E-03	1.2E-02	1.9E-03	3.8E-04	5.4E-03

and Spanish samples, which supports the use of a single database for Iberia. For Brazil and Ecuador, the results evidenced the need of larger databases to investigate the genetic diversity and a possible stratification of the populations.

In all the studied populations, the statistical parameters of forensic relevance showed a high discrimination capacity of the full set of 12 X-STRs, both in males and females, as well as high values of a priori exclusion chance in paternity father/daughter duos and father/mother/daughter trios.

The inclusion in this study of very diverse population samples allowed the detection of a high number of new alleles, many of them not yet included in the allelic ladders provided with the kit. The high diversity of the markers led to a large number of haplotype combinations, highlighting the need of huge databases to obtain reliable estimates of haplotype frequencies. Indeed, the maximum number of observed haplotypes in the 4 population samples varied from 248 to 478, for LG3 and LG1, respectively. When considering the proportion of different haplotypes in the total sample (number of different haplotypes/total number of samples), these varied between 44%–67% for LG1, 30%–53%, for LG2, 22%–36%, for LG3, and 37%–58% for LG4, showing that other studies are still necessary to achieve more accurate estimates of haplotype frequencies.

In this work, we estimated paternal and maternal locus specific mutation rates considering mendelian incompatibilities both in duos and trios. Except for haploid genetic transmission, it is known that estimates of mutation rates through the proportion of observed incompatibilities are conservative, due to the possibility of hidden mutations, and the likelihood of this event may decrease when considering trios instead of duos. In this study, 18 mutations (more than 10 % of the total) would not be detected if only duos were analysed. When simplified algebraic formulas are used, the possibility of mutation is only considered when an incompatibility is observed. In this case, depending if duos or trios are under analysis, the rate of incompatibilities estimated through duos or trios should be used, respectively. Notwithstanding, specific software is nowadays available to compute likelihood ratios in kinship analysis using X-chromosomal markers, including mutation, among others [28,58]. In this tool, the likelihood of a specific parental allele mutating to a specific filial one is considered for all the possible allelic transitions (even in the absence of mendelian incompatibilities). Therefore, in this case, the most accurate estimate of the average mutation rate – the one estimated through trios, along with the most biologically reliable mutation model (in the light of the state of the art), should be used in any case.

The results obtained revealed higher estimates for mutation rates in male than in female meiosis, with no intersection of the confidence intervals in three cases for both duos and trios ($p = 0.05$; DXS10074, DXS10135 and DXS10079). Also, the proportion of mendelian incompatibilities explained by a paternal mutation tend to increase with the age of the father. The DXS10135 marker showed the highest mutation rate, with order of magnitude equal to -2, which is above the desirable for kinship investigations. Most of the found mutations (96 %) can be explained by the gain or loss of a single repeat, and only one mutation between alleles belonging to different microvariant classes were observed.

Declaration of Competing Interest

None.

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.

We deeply thank the fruitful comments of an anonymous reviewer.

NP was supported by a Short-Term Fellowship grant from the International Society for Forensic Genetics (ISFG).

This work was partially financed by FEDER—Fundo Europeu de Desenvolvimento Regional funds through the COMPETE 2020—Operacional Programme for Competitiveness and Internationalisation (POCI), Portugal 2020; Portuguese funds through FCT—Fundação para a Ciência e a Tecnologia/Ministério da Ciência, Tecnologia e Inovação in the framework of the Decreto-Lei nº 57/2016 de 29 de Agosto (NP work contract) and the projects "Institute for Research and Innovation in Health Sciences" (POCI-01-0145-FEDER-007274) and "Center of Mathematics of the University of Porto" (UID/MAT/00144/2019), and by *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* – CAPES (Finance Code 001). LG was supported by *Conselho Nacional de Desenvolvimento Científico e Tecnológico* – CNPq (ref. 305330/2016-0) and Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro – FAPERJ (CNE-2018). RC was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (ref. 302198/2016-4), Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP (2016/10138-8) and PADC-FCFAr (2017/04-I). GB was supported by Dirección General de Investigación through project MED.GBF.18.06. EP participation was supported by Dirección Nacional de Investigaciones de la Fiscalía General del Estado, Ecuador.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.fsigen.2020.102258.

References

[1] V. Pereira, L. Gusmão, Types of genomes, sequences and genetic markers, in: A. Amorim, B. Budowle (Eds.), Handbook of Forensic Genetics. Biodiversity and Heredity in Civil and Criminal Investigation, World Scientific, New Jersey, 2016.

[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.

[3] N. Pinto, L. Gusmão, A. Amorim, X-chromosome markers in kinship testing: a generalisation of the IBD approach identifying situations where their contribution is crucial, Forensic Sci. Int. Genet. 5 (2011) 27–32.

[4] D. Kling, Curiosities of X chromosomal markers and haplotypes, Int. J. Legal Med. 132 (2018) 361–371.

[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.

[6] C. Tomas, J.J. Sanchez, A. Barbaro, C. Brandt-Casadevall, A. Hernandez, M.B. Dhiab, M. Ramon, N. Morling, X-chromosome SNP analyses in 11 human Mediterranean populations show a high overall genetic homogeneity except in North-west Africans (Moroccans), BMC Evol. Biol. 8 (2008) 75.

[7] L. Gusmão, P. Sánchez-Diz, C. Alves, I. Gomes, M.T. Zarrabeitia, M. Abovich, et al., A GEP-ISFG collaborative study on the optimization of an X-STR decaplex: data on 15 Iberian and Latin American populations, Int. J. Legal Med. 123 (2009) 227–234.

[8] L. Li, C. Li, S. Zhang, S. Zhao, Y. Liu, Y. Lin, Analysis of 14 highly informative SNP markers on X chromosome by TaqMan SNP genotyping assay, Forensic Sci. Int. Genet. 4 (2010) e145–e148.

[9] N.S. Freitas, R.L. Resque, E.M. Ribeiro-Rodrigues, J.F. Guerreiro, et al., X-linked insertion/deletion polymorphisms: forensic applications of a 33-markers panel, Int. J. Legal Med. 124 (2010) 589–593.

[10] T.M. Diegoli, M.D. Coble, Development and characterization of two mini-X chromosomal short tandem repeat multiplexes, Forensic Sci. Int. Genet. 5 (2011) 415–421.

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

[12] R. Pereira, V. Pereira, I. Gomes, C. Tomas, N. Morling, A. Amorim, M.J. Prata, A. Carracedo, L. Gusmão, A method for the analysis of 32 X chromosome insertion deletion polymorphisms in a single PCR, Int. J. Legal Med. 126 (2012) 97–105.

[13] T. Oki, T. Hayashi, M. Ota, H. Asamura, Development of multiplex assay with 16 SNPs on X chromosome for degraded samples, Leg. Med. (Tokyo) 14 (2012) 11–16.

[14] E. Prieto-Fernández, M. Baeta, C. Núñez, M.T. Zarrabeitia, R.J. Herrera, J.J. Builes, et al., Development of a new highly efficient 17 X-STR multiplex for forensic purposes, Electrophoresis 37 (2016) 1651–1658.

[15] C. Bini, L.N. Riccardi, S. Ceccardi, F. Carano, S. Sarno, D. Luiselli, et al., Expanding X-chromosomal forensic haplotype frequencies database: Italian population data of four linkage groups, Forensic Sci. Int. Genet. 15 (2015) 127–130.

[16] J. Xing, A. Adnan, A. Rakha, K. Kasim, et al., Genetic analysis of 12 X-STRs for forensic purposes in Liaoning Manchu population from China, Gene 683 (2019) 153–158.

[17] R. Tao, J. Zhang, Y. Bian, R. Dong, et al., Investigation of 12 X-STR loci in Mongolian and Eastern Han populations of China with comparison to other populations, Sci. Rep. 8 (2018) 4287.

[18] F. Guo, Genetic polymorphism of 17 autosomal STR loci in the Lahu ethnic minority from Yunnan Province, Southwest China, Forensic Sci. Int. Genet. 31 (2017) e52–e53.

[19] I. Cortés-Trujillo, F. Zuñiga-Chiquette, B. Ramos-González, M.L. Chávez-Briones, et al., Allele and haplotype frequencies of 12 X-STRs in Mexican population, Forensic Sci. Int. Genet. 38 (2019) e11–e13.

[20] J.M. Salvador, D.L.T. Apaga, F.C. Delfin, G.C. Calacal, et al., Filipino DNA variation at 12 X-chromosome short tandem repeat markers, Forensic Sci. Int. Genet. 36 (2018) e8–e12.

[21] G. Mršić, P. Ozretić, J. Crnjac, S. Merkaša, et al., Expanded Croatian 12 X-STR loci database with an overview of anomalous profiles, Forensic Sci. Int. Genet. 34 (2018) 249–256.

[22] 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.

[23] I. Veselinović, D. Vapa, M. Djan, N. Veličković, et al., Genetic analysis of 12 X-STR loci in the Serbian population from Vojvodina Province, Int. J. Legal Med. 132 (2018) 405.

[24] L. Cainé, S. Costa, M.F. Pinheiro, Population data of 12 X-STR loci in a North of Portugal sample, Int. J. Legal Med. 127 (2013) 63–64.

[25] C. Tomas, V. Pereira, N. Morling, Analysis of 12 X-STRs in greenlanders, Danes and somalis using argus X-12, Int. J. Legal Med. 126 (2012) 121–128.

[26] 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.

[27] 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.

[28] 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.

[29] T.M. Diegoli, H. Rohde, S. Borowski, M. Krawczak, M.D. Coble, M. Nothnagel, Genetic mapping of 15 human X chromosomal forensic short tandem repeat (STR) loci by means of multi-core parallelization, Forensic Sci. Int. Genet. 25 (2016) 39–44.

[30] P. Vicard, A.P. Dawid, A statistical treatment of biases affecting the estimation of mutation rates, Mutat. Res. 547 (2004) 19–33.

[31] P. Vicard, A.P. Dawid, J. Mortera, S.L. Lauritzen, Estimating mutation rates from paternity casework, Forensic Sci. Int. Genet. 2 (2008) 9–18.

[32] K. Slooten, F. Ricciardi, Estimation of mutation probabilities for autosomal STR markers, Forensic Sci. Int. Genet. 7 (3) (2013) 337–344.

[33] 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. Legal Med. 122 (2008) 347–351.

[34] K.J. Shin, B.K. Kwon, S.S. Lee, J.E. Yoo, M.J. Park, U. Chung, et al., Five highly informative X-chromosomal STRs in Koreans, Int. J. Legal Med. 118 (2004) 37–40.

[35] J. Edelmann, R. Szibor, DXS101: a highly polymorphic X-linked STR, Int. J. Legal Med. 114 (2001) 301–304.

[36] M. Poetsch, H. Petersmann, A. Repenning, E. Lignitz, Development of two pentaplex systems with X-chromosomal STR loci and their allele frequencies in a northeast German population, Forensic Sci. Int. 155 (2005) 71–76.

[37] M. García, L. Gusmão, C. Catanesi, G. Penacino, N. Pinto, Mutation rate of 12 X-STRs from Investigator Argus X-12 kit in argentine population, Forensic Sci. Int. Genet. Suppl. Ser. 6 (2017) e562–e564.

[38] M. Scherer, M. König, M. Bussmann, A. Prochnow, R. Peist, Development and validation of the new investigator® argus X-12 QS kit, Forensic Sci. Int. Genet. Suppl. Ser. 5 (2015) e256–e257.

[39] 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.

[40] D. Desmarais, Y. Zhong, R. Chakraborty, C. Perreault, L. Busque, Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA), J. Forensic Sci. 43 (1998) 1046–1104.

[41] 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.

[42] M. Chen, H. Ren, Z. Liu, J. Zhao, C. Chen, Y. Shi, et al., 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. Legal Med. (2019), https://doi.org/10.1007/s00414-019-02047-8 [Epub ahead of print].

[43] T.M. Diegoli, A. Linacre, M.S. Schanfield, M.D. Coble, Mutation rates of 15 X chromosomal short tandem repeat markers, Int. J. Legal Med. 128 (4) (2014) 579–587.

[44] T. Nishi, A. Kurosu, Y. Sugano, J. Kaminiwa, Y. Sekine, S. Yano, et al., Application of a novel multiplex polymerase chain reaction system for 12 X-chromosomal short tandem repeats to a Japanese population study, Leg. Med. (Tokyo) 15 (1) (2013) 43–46.

[45] Q.L. Liu, D.J. Lu, L. Quan, Y.F. Chen, M. Shen, H. Zhao, et al., Development of multiplex PCR system with 15 X-STR loci and genetic analysis in three nationality populations from China, Electrophoresis 33 (8) (2012) 1299–1305.

[46] 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 (6) (2012) e155–e156.

[47] D. Becker, H. Rodig, C. Augustin, J. Edelmann, F. Götz, S. Hering, et al., Population genetic evaluation of eight X-chromosomal short tandem repeat loci using Mentype Argus X-8 PCR amplification kit, Forensic Sci. Int. Genet. 2 (1) (2008) 69–74.

[48] Q.L. Liu, D.J. Lv, X.L. Wu, H.Y. Sun, X.Y. Wu, H.L. Lu, Development of a five ChX STRs loci typing system, Int. J. Legal Med. 122 (3) (2008) 261–265.

[49] M.A. Tariq, O. Ullah, S.A. Riazuddin, S. Riazuddin, Allele frequency distribution of 13 X-chromosomal STR loci in Pakistani population, Int. J. Legal Med. 122 (6) (2008) 525–528.

[50] 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 (2) (2007) 201–204.

[51] K. 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 (2–3) (2005) 173–180.

[52] 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.

[53] 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.

[54] 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 (1) (2019) 718–720, https://doi.org/10.1016/j.fsigss.2019.10.150.

[55] M. Ohno, Spontaneous de novo germline mutations in humans and mice: rates, spectra, causes and consequences, Genes Genet. Syst. 94 (1) (2019) 13–22.

[56] L. Ségurel, M.J. Wyman, M. Przeworski, Determinants of mutation rate variation in the human germline, Annu. Rev. Genomics Hum. Genet. 15 (2014) 47–70.

[57] A. Kong, M.L. Frigge, G. Masson, S. Besenbacher, P. Sulem, G. Magnusson, et al., Rate of de novo mutations and the importance of father's age to disease risk, Nature 488 (7412) (2012) 471–475.

[58] D. Kling, A. Tillmar, T. Egeland, P. Mostad, A general model for likelihood computations of genetic marker data accounting for linkage, linkage disequilibrium, and mutations, Int. J. Legal Med. 129 (5) (2015) 943–954.