


ORIGINAL ARTICLE

Open Access



# Analysis of fortuitous matches in a STR genotype database from Mexico and its forensic efficiency parameters

Joaquín Alejandro Quiroz-Mercado<sup>1,2</sup>, Ricardo Jesus Ríos-Rivas<sup>1</sup>, Víctor Manuel Martínez-Sevilla<sup>3</sup>, Gabriela Chávez-Marín<sup>1</sup>, Hueman Jaimes-Díaz<sup>2</sup>, Juan Carlos Santiago-Hernández<sup>1</sup>, Rogelio Maldonado-Rodríguez<sup>2</sup> and Héctor Rangel-Villalobos<sup>3\*</sup> 

## Abstract

**Background:** Commercial kits for Human Identification (HID) purposes, based on 15 short tandem repeat (STR) regions, allow the resolution of most forensic and paternity cases. However, some pitfalls arise in situations such as identification of missing persons, disaster victims, and in motherless paternities. We describe nine fortuitous matches found during the inclusion of new DNA profiles and/or searching for missing persons within a Mexican STR database ( $N = 2000$ ). For these cases, we estimated both the likelihood ratio (LR) and the modified LR according to National Research Council (NRC) recommendations (i.e.,  $LR^{NRC}$ ).

**Results:** In most cases, low  $LR^{NRC}$  values ( $\sim 1$  or  $< 1$ ) and elevated presence of modal STR alleles ( $> 50\%$  loci) were indicative of fortuitous matches. We estimated a significant increase in the observed frequency of fortuitous matches in this database when using the PowerPlex 16 DNA-typing system. In addition, forensic efficiency parameters for this HID kit are reported for the State of Mexico based on a population sample selected from the STR database ( $n = 493$ ). We describe the population structure and relationships, which detail comparative homogeneity among Mexican-Mestizos from the Central region, excluding the Western region.

**Conclusions:** The results suggest that additional / more powerful HID systems should be used when searching for relatives within STR databases in order to enable definitive conclusions, which are also required in cases of motherless paternity and massive disasters. The genetic structure described herein allows forensic geneticists to use alternative STR population data during statistical interpretation, given that most Mexican populations do not have their own STR databases.

**Keywords:** STR, Database, Mexico, Fortuitous matching, False kinship

## Introduction

Short tandem repeat (STR) loci are the most informative genetic markers used worldwide for human identification (HID) purposes. Although commercial kits that analyze 15 STRs allow solving most forensic and paternity cases, some pitfalls arise in situations such as the identification of disaster victims, missing persons, and in motherless paternity cases (Wurmb-Schwark et al. 2006; Coletti et al. 2008; Borovko et al. 2009; Li et al. 2012;

Ziętkiewicz et al. 2012). In our experience, the lack of informative reference samples (first-degree relatives) is probably the most common problem when identifying unknown bodies. This is sometimes overcome by searching STR databases that include samples from missing persons and their relatives, criminals, victims, etc. (Álvarez-Cubero et al. 2012). A considerable number of countries have established local databases that include the 13 STR loci of the Combined DNA Index System (CODIS), the amelogenin locus for gender determination, and additional STRs depending on the HID kit employed (Collins et al. 2004). In 2011, the federal State of Mexico (located within the Central region, Mexico)

\* Correspondence: hrangel13@hotmail.com

<sup>3</sup>Instituto de Investigación en Genética Molecular, Universidad de Guadalajara (CUCiénega-UdeG), Av. Universidad #1115, Col. Paso Blanco, CP 47810 Ocotlán, Jalisco, Mexico

Full list of author information is available at the end of the article

established a civil and criminal STR database containing approximately 2000 DNA profiles analyzed with the commercial PowerPlex 16 System (Krenke et al. 2002). Although many Mexican populations have been analyzed with autosomal STRs to support forensic casework (Rubi-Castellanos et al. 2009; Salazar-Flores et al. 2015), only the following states/regions of the country have been studied with the PowerPlex 16 kit: Guanajuato, Veracruz, Nayarit, Yucatan, Mexico City, and the Western region (Rangel-Villalobos et al. 2010; González-Herrera et al. 2010; Ramírez-Flores et al. 2014; Martínez-Sevilla et al. 2016). Furthermore, the population structure and genetic relationships of the Mexican populations remain unexplored using PowerPlex 16 System.

Therefore, we analyzed this STR genotype database from the State of Mexico. The findings: *i*) report nine fortuitous matches detected during the addition of new DNA profiles to this genetic database, which suggest first-degree kinship; *ii*) evaluate the presence of modal alleles in these cases; *iii*) estimate the expected probability of finding these fortuitous matches; *iv*) estimate statistical parameters of forensic efficiency in this previously unstudied population (State of Mexico); and *v*) evaluate the structure and genetic relationships among Mexican populations previously studied with the PowerPlex 16 system.

## Material and methods

### DNA extraction

Bones analyzed in this work were previously decalcified with EDTA 0.5 M. Bone slices were treated with a Proteinase K Digestion Solution prepared with Bone Incubation Buffer and Proteinase K solution at 21 mg/mL (Promega Corp. Madison, WI). Similarly, tissues (mainly muscle) were digested with Proteinase K and digestion buffer, whereas saliva swabs were spun for DNA extraction from the resulting pellets. The final step of DNA extraction employed the DNA IQ System according to the manufacturer's instructions (Promega Corp., Madison, WI). Conversely, blood samples obtained from relatives, or reference samples included in the database, were placed on FTA cards and processed with FTA purification reagent (Whatman Inc., Clifton, NJ). One punch derived from FTA cards was used during PCR amplification as DNA sample. All individuals provided a signed, written informed consent in accordance with the ethical guidelines of the Helsinki Declaration. The anonymity of the recruited individuals was preserved. This project was authorized by the Ethical Committee Dirección de Servicios Periciales de la Fiscalía General de Justicia del Estado de México (PGJEM).

### STR genotyping

The PowerPlex 16 system was used according to the supplier's instructions (Promega Corp. Madison, WI).

Amplified products were run by capillary electrophoresis using the ABI Prism 310 Genetic Analyzer (Applied Biosystems, Foster City, CA). Allele calling was achieved by comparison with the allelic ladder provided in the kit, helped by GeneMapper software (version 3.2).

### STR database creation

DNA profiles based on the 15 STRs constituting the PowerPlex 16 system were individually uploaded into the Genetics Platform (Server Unix Solaris, Architecture Sparc64, Motor Rdbms Oracle, Enterprise Edition) developed by Grupo Empresarial Iberoamericano (GEI) (<http://geigen.mx>). In accordance with international recommendations (<http://enfsi.eu/documents/>), the inclusion criteria and upload process of DNA profiles to the STR database included quality control verification of the laboratory's procedures (ENFSI DNA Working Group 2014). The laboratory has participated in the quality control exercise organized by the Grupo Iberoamericano de Trabajo en Análisis de DNA (GITAD: <http://www.aicef.net/>). The STR genotype database primarily contains details on unknown bodies, people searching for missing relatives, and biological evidence from criminal cases. The sample's origin was recorded in the database for classification purposes. We carefully selected a subpopulation of unrelated individuals to estimate forensic parameters ( $n = 493$ ). For this purpose, we excluded genetic data from criminal samples, unknown bodies, and from individuals sharing surnames in order to avoid possible kinship in this subpopulation sample.

### Data analysis

Allele frequencies and the following statistical parameters of forensic importance were calculated using the Excel spreadsheet Powerstats (Tereba 2001): allele frequencies, minimum allele frequencies (MAF), probability of exclusion (PE), power of discrimination (PD), polymorphism information content (PIC), observed heterozygosity (Het), and typical paternity index (TPI). GDA (version 1.1) software was used to perform Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) tests (Lewis and Zaykin 2002). For the interpopulational analyses, we included STR datasets from four Mexican populations (Rangel-Villalobos et al. 2010; González-Herrera et al. 2010; Ramírez-Flores et al. 2014; Martínez-Sevilla et al. 2016) and from four main American ethnic populations (Hill et al. 2013). Geographic location of the Mexican populations is represented in Fig. 1. Analysis of molecular variance (AMOVA), Fst distances, and pairwise Fst *p*-values were estimated with Arlequin 3.1 software (Excoffier et al. 2005). Distances were represented



graphically in a multidimensional scaling (MDS) plot using the software SPSS (version 20.0 for Windows).

Likelihood ratios (LRs) representing paternity indexes with only one parent (motherless or fatherless) were computed either when the match was detected during the inclusion of new DNA profiles, or when searching missing persons within the STR database. Bayesian *posteriori* probabilities assuming equal *prior* probabilities (0.5) for the two alternative probabilities were considered for this purpose (Gjertson et al. 2007), using the allele frequencies reported here and the Familias 3 software (Kling et al. 2014). In addition, we estimated another LR, termed  $LR^{NRC}$ , by applying a correction factor according to the size of the database (i.e.,  $N = 2000$ ). This follows Recommendation 5.1 of the National Research Council (NRC II), which was endorsed by the FBI's DNA Advisory Board in the February 2000 recommendations on statistical approaches (National Research Council (NRC) Committee on DNA Forensic Science 1996). In order to check whether the observed cases showed agreement with expectations simply by chance, we estimated for each STR the matching probabilities between all possible genotypes sharing at least one allele.

For this purpose, HWE and equilibrium linkage assumptions were applied for estimating genotype and DNA profile frequency, respectively. The combined matching probability was obtained by applying the product rule and the Bonferroni correction according to the sample size ( $N = 2000$ ), using Microsoft Excel 2007.

## Results and discussion

### Observed cases of fortuitous matching

We report nine matches (presumed to be fortuitous) found during the search or inclusion of DNA profiles into the STR database of 2000 samples from the State of Mexico. According to our records, there were no biological relationships between these matching individuals.

### Case 1

When comparing the mother's DNA profile with the database, two positive matches suggesting first-degree kinship were found during the search for a missing daughter (Table 1). A singular difference between LR and  $LR^{NRC}$  values distinguishes the fortuitous match from the (presumed) real kinship

**Table 1** Cases of fortuitous matches (FM) with 15 STRs versus –presumably– real kinship between a mother (M) and/or father (F) searching a missing daughter (MD) and missing son (MS), respectively

STR	Case 1 <sup>a</sup>			Case 2 <sup>b</sup>			
	Fortuitous Match (FM1) <sup>e</sup>	Mother (M) <sup>c</sup>	Missing Daughter (MD)	Father (F1) <sup>c</sup>	Fortuitous Match (FM2) <sup>e</sup>	Father (F2) <sup>c</sup>	Missing Son (MS)
D3S1358	<u>15,17</u>	4,15	<b>15,17</b>	16,17	<u>14,15</u>	15,16	<b>15,16</b>
TH01	<u>7,7</u>	7,7	<b>7,7</b>	7,9.3	<u>6,6</u>	6,6	<b>6,6</b>
D21S11	<u>30,31.2</u>	30,30	<b>30,31.2</b>	31.2,32.2	<u>28,32.2</u>	30,32.2	<b>30,33.2</b>
D18S51	<u>14,19</u>	13,14	<b>14,14</b>	14 <sup>c</sup> ,14	<u>14,16</u>	14,15	<b>14,15</b>
Penta E	<u>15,19</u>	12,15	<b>12,12</b>	12 <sup>c</sup> ,16	<u>12,15</u>	15,16	<b>13,15</b>
D5S818	<u>11,11</u>	11,12	<b>11,11</b>	11,11	<u>11,12</u>	11,11	<b>11,12</b>
D13S317	<u>12,14</u>	13,14	<b>9,13</b>	9,12	<u>9,9</u>	9,9	<b>9,14</b>
D7S820	<u>11,11</u>	11,12	<b>11,11</b>	11,11	<u>8,11</u>	8,11	<b>8,8</b>
D16S539	<u>11,12</u>	10,12	<b>12,12</b>	11,12	<u>10,12</u>	12,12	<b>12,12</b>
CSF1PO	<u>11,12</u>	10,12	<b>10,10</b>	10,10 <sup>c</sup>	<u>11,12</u>	11,12	<b>11,12</b>
Penta D	<u>10,12</u>	9,12	<b>12,12</b>	10,12	<u>12,13</u>	10,13	<b>13,13</b>
AMEL	X,X	X,X	X,X	X,Y	X,Y	X,Y	X,Y
vWA	<u>16,17</u>	16,17	<b>16,16</b>	16,17	<u>17,18</u>	15,17	<b>15,17</b>
D8S1179	<u>10,13</u>	10,12	<b>12,13</b>	13,13	<u>13,13</u>	13,14	<b>14,14</b>
TPOX	<u>8,8</u>	8,11	<b>8,11</b>	8,8	<u>11,12</u>	11,12	<b>9,11</b>
FGA	<u>20,26</u>	22,26	<b>20,26</b>	20,23	<u>24,25</u>	24,25	<b>18,25</b>
LR, <i>LR<sup>NRC</sup></i>	<b>150.5, 0.08</b>				<b>2495.77, 1.25</b>		
LR, <i>LR<sup>NRC</sup></i>		<b>5248.7, 2.62</b>				<b>32961.9, 16.48</b>	
LR, <i>LR<sup>NRC</sup></i>		<b>57502.7, 28.75</b>					

<sup>a</sup>Case 1: Alleles matching between the mother (M) with the missing daughter (MD) and the fortuitous match (FM1) are indicated in bold and underlined, respectively

<sup>b</sup>Case 2: Alleles matching between the father (F) with the missing son (MS) and the fortuitous match (FM2) are indicated in bold and underlined, respectively

<sup>c</sup>Key alleles shared between the father (F) and MD to establish the real kinship in case 1 respect to FM1 sample

<sup>d</sup>Estimated PIs between the samples of the columns above described, before (in bold) and after applying NRC correction (italics) (Hill et al. 2013)

<sup>e</sup>Cells in gray indicate genotypes with modal alleles involved in the cases of fortuitous matches

regarding the missing daughter: the first does not support the maternity hypothesis ( $LR^{NRC} = 0.08$  versus  $LR = 2.62$ ). Eventually, the inclusion of the father confirmed which one was the fortuitous match by means of three inconsistencies (D18S51, Penta E, and CSF1PO), discarding undoubtedly the biological relationship of one sample (FM1, see Table 1). Simultaneously, the father’s inclusion increased the LR for the match with the other sample to get a more confident decision on returning the corpse to the (presumably) real relatives.

**Case 2**

Two positive matches were found during the search for a missing son when a father’s DNA profile was compared with the STR database (Table 2). Although both matches offered relatively low but positive  $LR^{NRC}$  values (1.25 versus 16.48), one was clearly larger when comparing uncorrected LR (2495.8 versus 32,961.9). Therefore, this difference seems helpful to make a decision and returning the corpse to the (presumed) biological father. In brief, low  $LR^{NRC}$  values (~1 or <1) appear useful for indicating a fortuitous

**Table 2** Fortuitous matching cases found during the inclusion of DNA profiles (PowerPlex® 16 System) to the STR database of the state of Mexico

STR	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9	Modal allele in matches (n; %) <sup>a</sup>	
D3S1358	16,16	15,16	15,18 15,15	15,16 15,15	15,19 15,15	16,17 16,16	15,17 16,17	15,15 15,16	<b>15</b> (6; 67%)
TH01	6,9,3	6,7	6,7 6,9,3	7,7 7,7	6,7 6,9	7,9,3 7,7	6,7 6,7	7,7 6,7	<b>7</b> (5; 56%)
D21S11	29,30	29,30	30,30 30,32.2	29,30 30,32.2	28,29 29,30	29,29 29,30	29,31.2 29,32.2	29,33.2 29,30	<b>30</b> (4; 44%)
D18S51	15,19	15,17	15,16 14,15	17,18 12,18	15,15 14,15	17,19 17,17	14,17 15,17	13,15 15,16	<b>14</b> (2; 22%)
Penta E	7,17	7,15	12,20 15,20	5,12 5,19	12,20 12,23	13,20 13,13	12,15 12,12	13,14 14,16	<b>12</b> (2; 22%)
D5S818	11,13	11,11	11,12 11,13	9,11 11,13	11,11 11,12	11,13 9,13	11,12 11,12	11,11 11,11	<b>11</b> (8; 89%)
D13S317	11,14	11,11	9,12 9,12	12,13 11,12	12,12 12,12	11,11 11,11	10,12 10,11	9,13 8,9	<b>9</b> (3; 33%)
D7S820	11,13	10,11	11,12 11,11	11,12 10,12	10,12 8,10	11,11 10,11	10,11 11,11	12,12 12,12	<b>11</b> (6; 67%)
D16S539	11,13	10,11	9,11 9,10	12,13 9,13	12,12 12,13	11,12 12,12	11,12 12,12	12,13 10,12	<b>12</b> (6; 67%)
CSF1PO	10,13	10,12	9,12 10,12	11,11 11,11	11,12 12,13	12,12 10,12	10,12 10,12	12,14 11,12	<b>12</b> (7; 78%)
Penta D	10,12	10,14	9,10 10,11	11,14 10,11	9,11 9,10	11,14 11,12	8,9 9,13	11,13 9,13	<b>10</b> (2; 22%)
AMEL	X,Y	X,Y	X,Y X,Y	X,Y X,Y	X,Y X,Y	X,X X,Y	X,Y X,Y	X,Y X,X	---
vWA	17,18	16,17	15,16 16,17	16,17 16,17	16,17 17,17	17,17 17,17	16,16 16,17	18,18 17,18	<b>16</b> (4; 44%)
D8S1179	10,14	10,13	12,13 12,12	10,13 13,15	13,15 13,15	10,16 15,16	13,14 13,13	10,13 10,13	<b>13</b> (5; 56%)
TPOX	8,9	8,8	8,8 8,11	8,10 8,8	8,11 11,12	10,11 11,11	8,10 8,12	8,8 8,10	<b>8</b> (6; 67%)
FGA	22,24	24,25	23,25 25,26	22,25 22,23	20,24 22,24	21,25 21,24	24,27 24,26	22,25 25,25	<b>25</b> (3; 33%)
Modal alleles <sup>b</sup>	<b>5</b>	<b>10</b>	<b>7</b>	<b>6</b>	<b>4</b>	<b>9</b>	<b>9</b>	<b>Mean = 7.7</b>	
LR	<b>113,5</b>	<b>110,3</b>	<b>4239,7</b>	<b>200,0</b>	<b>2,320,000</b>	<b>486,2</b>	<b>767,9</b>		
LR <sup>NRC</sup>	0.0567	0.055	2.1198	0.1	1160	0.2431	0.3839		

LR<sup>NRC</sup>: LR modified according with National Research Council recommendations (LR/database size)

<sup>a</sup>Number of times that modal alleles are involved in the nine cases of fortuitous matches (genotypes in gray cells)

<sup>b</sup>Number of modal alleles matching by case

match, whereas uncorrected LR comparison could be useful for making a final decision when two matching events are observed during the search for a missing person within a STR database. However, we must kept in mind that, whenever possible, additional relatives or markers should be included for further DNA analysis, besides to anchoring anthropological and/or circumstantial data to the DNA identification. Unfortunately, this was not possible when this case was solved some years ago.

**Cases 3–9**

Seven presumed fortuitous matches were detected between pairs of samples during the inclusion of DNA profiles into the STR database (Table 2). Comparison of surnames and/or available records indicated no biological relationship between matching individuals. Although in most cases (6/7; 85.7%) the low LR values allow deducing fortuitous kinship events (LR<sup>NRC</sup> < 2.2), one (case 7) displayed large LRs (LR = 2,320,000; LR<sup>NRC</sup> = 1160), suggesting a real biological relationship. Unfortunately, further biological and/or DNA samples were not available for additional genetic analyses (e.g., X-STRs, mtDNA, etc.) to confirm or discard this hypothesis. Again, low

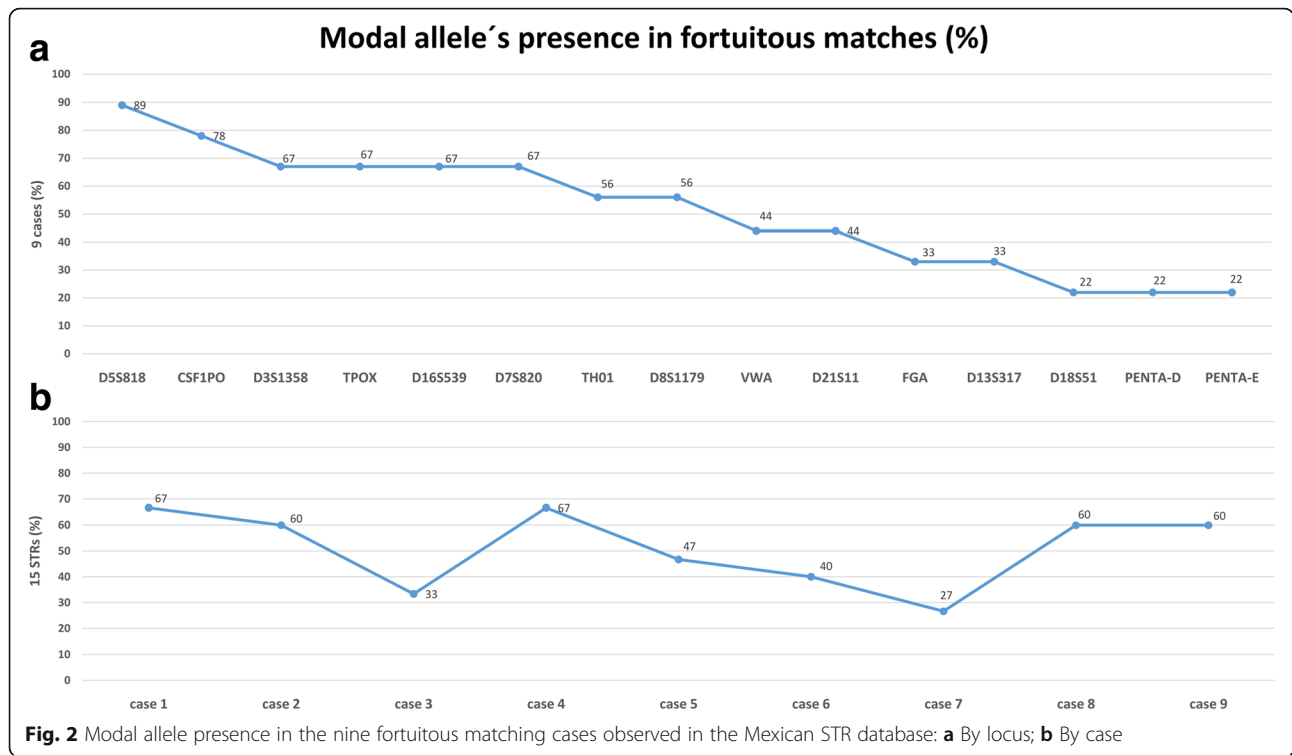
LR<sup>NRC</sup> values (~1 or <1) appear useful for establishing fortuitous matches.

**Presence of modal alleles in fortuitous matching**

We evaluated the presence of modal alleles in the nine fortuitous matches described here (Tables 1 and 2). By STR locus, the markers whose modal alleles were commonly found (>50%) were D5S818 (8/9), CSF1PO (7/9), D3S1358, D7S820, D16S539, and TPOX (6/9), as well as TH01 and D8S1179 (5/9) (Fig. 2a). To a lesser extent, the modal alleles of D21S11 and vWA were involved in four cases (4/9), followed by D13S317 and FGA (3/9). Conversely, the modal alleles of D18S51, Penta E, and Penta D were rarely observed in fortuitous matches (2/9).

Interestingly, in most cases (5/9; 55%) the majority of modal alleles for the 15 loci were observed in the fortuitous matches: 9 STRs in three cases, and 10 STRs in two cases. In the remaining four (Cases 3, 5, 6 and 7), seven or fewer STRs (<50% of STRs) were involved in fortuitous matching (Table 3; Fig. 2b). Although the small number of cases presented herein limits the application of statistical tests, the results suggest those STRs that could be involved in fortuitous matches by their allele distribution, confirming their elevated presence (>50% loci) in most of the reported fortuitous matches (>55% cases). In addition,





**Table 3** Expected probability for sharing one/two allele(s) between two unrelated individuals for the Powerplex 16 kit in the state of Mexico database (N = 2000)

STR loci	p (matching)
TH01	0.71892
PENTA E	0.32048
D21S11	0.53897
D18S51	0.41570
FGA	0.41145
PENTA D	0.53029
D8S1179	0.60275
CSF1PO	0.70407
D7S820	0.62973
D16S539	0.64500
D13S317	0.52484
D5S818	0.73017
VWA	0.68664
TPOX	0.77201
D3S1358	0.77629
Total probability (Expected)	$0.00030617 / 2000^b = 1.153081 \times 10^{-17}$ (1 in 6,532,422 random comparisons)
Probability (Observed)	0.000004977 (9 in 2001,000 <sup>a</sup> )

<sup>a</sup>All possible pairwise comparisons among DNA profiles in our database (N = 2000)

<sup>b</sup>Bonferroni correction according to the sample size

these results highlight that the inclusion of more powerful STR human identification systems (i.e., PowerPlex Fusion, GlobalFiler, Investigator 24plex QS kits) and lineage markers (e.g., Y-STRs, X-STRs, or mtDNA) can be critical to establishing confident biological relationships when searching STR databases for missing persons.

**Evaluating the probability of matching events in the database**

Fortuitous matches potentially have serious legal implications, mainly in motherless paternity tests (Poetsch et al. 2006), and during the search for missing persons in STR databases, given that, frequently, few first-degree relatives are available for the test (Ge et al. 2011). This problem could be accentuated in Mexico by the growing number of missing and killed persons during the last years, along with clandestine graves discovered in different parts of the country. Therefore, we estimated, for individual and combined STR loci, the matching probability between all possible genotypes sharing at least one allele ( $p = 0.00030617$ ). The corrected p-value according to the sample size ( $p = 0.00000015308$ ) indicated that one fortuitous match between two DNA profiles is expected in 6,532,422 comparisons for this database (Table 3). This frequency differs from that observed in our study ( $p = 0.000004$ ; Yate’s chi-square = 21.105). This finding is similar to motherless paternity cases described in Germany,

where 26 non-STR mismatches were observed between 336 children empirically compared with 348 men (Poetsch et al. 2006). Although we could not identify a direct explanation for this increment, it might involve the presence of neighboring indigenous individuals with larger inbreeding coefficient, who are constantly incorporated into the Mexican-Mestizo populations. This hypothesis would be in agreement with descriptions of Native American ancestry throughout the country (Moreno-Estrada et al. 2014).

#### Concluding remarks regarding fortuitous matches

As could be expected, many recommendations of the International Society of Forensic Genetics (ISFG) for disaster victim identification (DVI) are useful to promote the correct identification of missing persons (Prinz et al. 2007). However, some critical ISFG recommendations to avoid fortuitous matches are listed briefly herein: #4) Multiple direct references and samples from first-degree relatives should be collected for each missing person; #6) Use of additional typing systems, such as mtDNA, Y-chromosomal STRs or SNP markers; #10) DNA based identification should whenever possible be anchored by anthropological and/or circumstantial data, a second identification modality, or multiple DNA references; #11) Use of LR that permit DNA results to be combined among multiple genetic systems or with other non-DNA evidence. LR threshold should be determined for when DNA data alone can suffice for an identification, which will be based on the size and circumstances of the event (Prinz et al. 2007). It must be noticed that there are difficulties to implement some recommendations in Mexico and probably in another developing countries. For instance, whereas DNA testing nowadays constitutes the main tool for human identification, the use of anthropological and circumstantial data has become scarcely employed and underappreciated. However, as described in case 2 (Table 1), these data would have been helpful to get stronger conclusions; thus, they should be available whenever possible. In brief, as can be noted, strict application of these ISFG recommendations for DVI would have avoided the fortuitous matches described herein for missing persons identification.

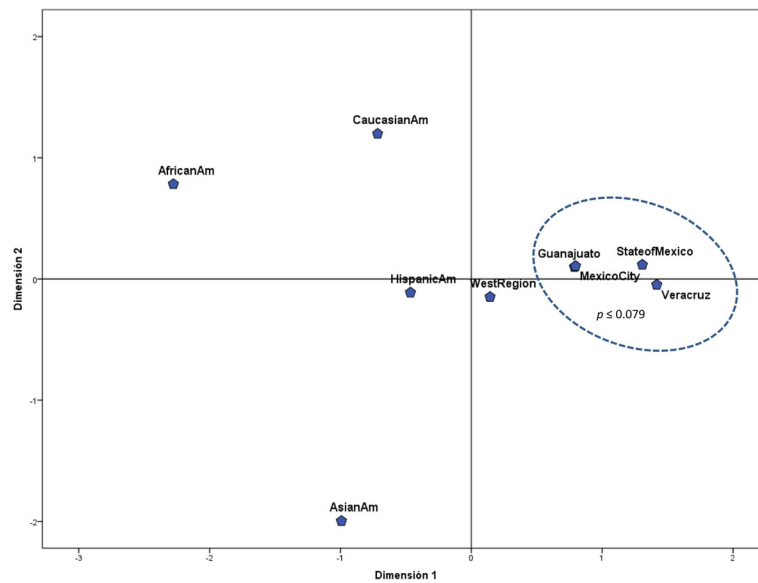
#### Forensic parameters of the STR database

In agreement with ISFG recommendations for DVI (Prinz et al. 2007), we report statistical parameters of forensic efficiency for the PowerPlex 16 kit when used in the State of Mexico population (Additional file 1: Table S1). The forensic parameters were estimated for each STR locus based on 493 unrelated DNA profiles carefully selected from the STR database according to the criteria above

described. HWE test showed that only D7S820 was in disequilibrium after applying the Bonferroni correction ( $p < 0.0033$ ). Similarly, only two LD cases were detected between D13S317/PENTA E and TH01/TPOX after exact tests (data not shown). Altogether, these isolated findings indicate that DNA profiles can be confidently estimated in the State of Mexico population. The combined power of discrimination (PD) and power of exclusion (PE) in this Mexican population were  $>99.9\%$ , which is sufficiently reliable to solve most forensic and paternity cases, respectively.

#### Interpopulation comparison

To our knowledge, this is the first time that all 15 STRs of the PowerPlex 16 System have been employed for this purpose, including five Mexican-Mestizo (Rangel-Villalobos et al. 2010; González-Herrera et al. 2010; Ramírez-Flores et al. 2014; Martínez-Sevilla et al. 2016) and four American populations (Hill et al. 2013) as reference. Interestingly, a central population cluster was formed by Mexico City, Veracruz, the State of Mexico (Fig. 3), and Guanajuato, the latter displaying no differentiation with Mexico City. Intergroup comparisons by  $F_{st}$  distances and  $F_{st}$   $p$ -values are indicated in Additional file 2: Table S2. Mexican-Mestizos from the Western region were closer to Hispanic Americans (probably due to their closer European ancestry) than the remaining Mexican populations, according to the genetic structure previously described in Mexican-Mestizos based on the 13 CODIS-STRs (Wurmb-Schwark et al. 2006; Rubi-Castellanos et al. 2009). AMOVA results suggest significant differentiation between Mexican and American populations ( $F_{st} = 1.45\%$ ;  $p = 0.0000$ ), which is approximately seven times larger than that observed among the five Mexican populations studied herein ( $F_{st} = 0.198\%$ ;  $p = 0.0000$ ), and 17 times larger than the interpopulation differentiation among the Central Mexican populations (excluding the Western region) ( $F_{st} = 0.198\%$ ;  $p = 0.0000$ ). In brief, based on the PowerPlex 16 system, our results support a relative homogeneity among Mexican populations from the Central region, excluding the Western region from this cluster. This conclusion applies to Mestizos (admixed), who constitute the largest proportion of the Mexican population ( $\sim 90\%$ ) (Rubi-Castellanos et al. 2009; Salazar-Flores et al. 2015), but is invalid for Native American groups, who display a particular genetic structure (Rangel-Villalobos et al. 2016). The genetic structure described herein is important in forensic casework because it allows geneticists using alternative STR population data, given that most of the Mexican populations do not have their own STR databases. This paper follows the guidelines for



**Fig. 3** MDS plot that shows genetic relationships between Mexican-Mestizo and four American populations based on the Powerplex 16 System

publication of population data requested by the journal (Carracedo et al. 2013).

## Conclusions

In brief, results emphasize the importance of analyzing a sufficient number of relatives and/or HID systems to reach reliable conclusions when searching for relatives in STR databases for DVI, missing persons identification, and motherless paternity cases. When this is not possible, the concomitant presence of low  $LR^{NRC}$  values ( $\sim 1$  or  $< 1$ ) and elevated presence of modal STR alleles should be analyzed to detect possible fortuitous matches. Interestingly, an increased frequency of fortuitous matches was observed in the studied Mexican STR database. Finally, we observed a relative homogeneity among Mexican-Mestizos of the Central region based on the PowerPlex 16 system.

## Additional files

**Additional file 1: Table S1.** Allele frequencies and statistical parameters of forensic efficiency of the Powerplex 16<sup>®</sup> System estimated in 493 individuals of a STR database from the State of Mexico. (DOCX 22 kb)

**Additional file 2: Table S2.** Fst distances (above diagonal) and Fst p-values (below diagonal) between Mexican and American populations based on Powerplex 16 genetic system. (DOCX 14 kb)

## Acknowledgements

We thank to COMECYT: 8BCD0013 for the Doctoral fellowship to J.A. Quiroz-Mercado.

## Funding

This work was supported by the Dirección de Servicios Periciales de la Procuraduría General de Justicia del Estado de México (PGJEM).

## Availability data and materials

The complete STR genotype database from the studied Mexican population is available upon request to the corresponding author.

## Authors' contributions

Conception of the project. QMJA, JDH, MRR, RVH; Population sampling and DNA Extraction. QMJA, RRR, CMG, SHJC; Genotyping process. QMJA, RRR, CMG, SHJC; Population data analysis. QMJA, MSVM, RVH; Discussion of results and writing of the manuscript. QMJA, MSVM, RVH. All authors read and approved the final manuscript.

## Ethics approval and consent to participate

All volunteers included in the population study signed a written informed consent according to the ethical guidelines of the Helsinki Declaration. The anonymity of the recruited individuals was preserved. This project was authorized by the Ethical Committee Dirección de Servicios Periciales de la Procuraduría General de Justicia del Estado de México (PGJEM).

## Consent for publication

Not applicable

## Competing interests

The authors declare that they have no competing interests.

## Publisher's Note

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

## Author details

<sup>1</sup>Forensic Genetics Laboratory, Instituto de Servicios Periciales, Fiscalía General de Justicia del Estado de México, 50090 Toluca, Mexico, Mexico. <sup>2</sup>Laboratory of Molecular Diagnostics, Department of Biochemistry, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional, 11340 Mexico City, Mexico. <sup>3</sup>Instituto de Investigación en Genética Molecular, Universidad de Guadalajara (CUCiénega-UdeG), Av. Universidad #1115, Col. Paso Blanco, CP 47810 Ocotlán, Jalisco, Mexico.

Received: 9 May 2017 Accepted: 12 November 2017

Published online: 06 December 2017

## References

Álvarez-Cubero MJ, Saiz M, Martínez-Gonzalez LJ, Alvarez JC, Eisenberg AJ, Budowle B, Lorente JA (2012) Genetic identification of missing persons: DNA



- analysis of human remains and compromised samples. *Pathobiology* 79(5): 228–238
- Borovko SR, Korban W, Kritskaya SV, Yeumenenka SA (2009) Missing people: problems of identification of unknown bodies using DNA database. *Forensic Sci Int Supp Ser* 2(1):260
- Carracedo A, Butler JM, Gusmão L, Linacre A, Parson W, Roewer L, Schneider PM (2013) New guidelines for the publication of genetic population data. *Forensic Sci Int Genet* 7:217–220
- Coletti A, Lancia M, Massetti S, Dobosz M, Carnevali E, Bacci M (2008) Considerations on a motherless paternity case with two related fathers: possible pitfalls. *Forensic Sci Int Supp Ser* 1(1):505–506
- Collins PJ, Hennessy LK, Leibelt CS, Roby RK, Reeder DJ, Foxall PA (2004) Developmental validation of a single-tube, amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: the AMPFLSTR® Identifier® PCR amplification kit. *J Forensic Sci* 49(6):1265–1277
- ENFSI DNA Working Group DNA-database management review and recommendations 2014. (<http://enfsi.eu/>)
- Excoffier L, Laval G, Scheneider S (2005) Arlequin ver. 3. 0: an integrated software package for population genetic analysis. *Evol Bioinformatics Online* 1:47–50
- Ge J, Budowle B, Chakraborty R (2011) Choosing relatives for DNA identification of missing persons. *J Forensic Sci* 56(Suppl 1):S23–S28
- Gjertson DW, Brenner C, Baur MP, Carracedo A, Guidet F, Luque JA et al (2007) ISFG: recommendations on biostatistics in paternity testing. *Forensic Sci Int Genet* 1(3–4):223–231
- González-Herrera L, Vega-Navarrete L, Roche-Canto C, Canto-Herrera J, Virgen-Ponce D, Moscoso-Caloca G et al (2010) Forensic parameters and genetic variation of 15 autosomal STR loci in Mexican-Mestizo populations from the states of Yucatan and Nayarit. *Open Forensic Sci J* 3:57–63
- Hill CR, Duewer DL, Kline MC, Coble MD, Butler JM (2013) US population data for 29 autosomal STR loci. *Forensic Sci Int Genet* 7(3):e82–e83
- Kling D, Tillmar AO, Egeland T (2014) Familias 3 extensions and new functionality. *Forensic Sci Int Genet* 13:121–127
- Krenke BE, Tereba A, Anderson SJ, Buel E, Culhane S, Finis CJ et al (2002) Validation of a 16-locus fluorescent multiplex system. *J Forensic Sci* 47(4): 773–785
- P.O. Lewis, D. Zaykin, Genetic data analysis, a computer program for the analysis of allelic data. Version 1.1, (2002). Free program distributed by the authors over the internet from the GDA home page. <https://phylogeny.uconn.edu/software/>
- Li L, Ge J, Zhang S, Guo J, Zhao S, Li C et al (2012) Maternity exclusion with a very high autosomal STRs kinship index. *Int J Legal Med* 126(4):645–648
- Martínez-Sevilla VM, Aguilar-Velázquez JA, Inclán-Sánchez A, Romero-Rentería O, Díaz-Navarro XX, Saiz M, Rangel-Villalobos H (2016) Parámetros forenses del sistema Powerplex® 21 Promega Corp en población mestiza del occidente de México. *Rev Esp Med Legal* 42(1):10–16
- Moreno-Estrada A, Gignoux CR, Fernández-López JC, Zakharia F, Sikora M, Contreras AV et al (2014) Human genetics. The genetics of Mexico recapitulates native American substructure and affects biomedical traits. *Science* 344:1280–1285
- National Research Council (NRC) Committee on DNA Forensic Science (1996) The evaluation of forensic DNA evidence. National Academic Press, Washington, DC
- Poetsch M, Lüdcke C, Repenning A, Fischer L, Mályusz V, Simeoni E et al (2006) The problem of single parent/child paternity analysis-practical results involving 336 children and 348 unrelated men. *Forensic Sci Int* 159(2–3):98–103
- Prinz M, Carracedo A, Mayr WR, Morling N, Parsons TJ, Sajantila A, Scheithauer R, Schmitter H, Schneider PM (2007) DNA Commission of the International Society for forensic genetics (ISFG): recommendations regarding the role of forensic genetics for disaster victim identification (DVI). *Forensic Sci Int Genet* 1:3–12
- Ramírez-Flores E, Saiz M, Villegas-Carmona D, Álvarez-Cubero MJ, Álvarez JC, Vega-Navarrete L (2014) Genetic variation of 24 STR loci in a Mexican-Mestizo population from Mexico DF. *Forensic Sci Int Genet* 10:e4–e6
- Rangel-Villalobos H, Martínez-Sevilla VM, Aguilar-Velázquez JA, Martínez-Cortes G, Sosa-Macias M, Gonzalez-Martin A (2016) Importance of the geographic barriers to promote gene drift and avoid (pre and post-Columbian) gene flow in Mexican native groups: evidence from forensic STR loci. *Am J Phys Anthropol* 160(2):298–316
- Rangel-Villalobos H, Rubi-Castellanos R, Morales-Vallejo ME, Molina-Araujo V, Licea-Cadena LA, Rizzo-Juárez S et al (2010) Admixture estimates and statistical parameters of forensic importance based on PowerPlex 16 system in Mexican-Mestizos from the states of Guanajuato (center) and Veracruz (east). *Forensic Sci Int Genet* 4(4):271–272
- Rubi-Castellanos R, Martínez-Cortés G, Muñoz-Valle JF, González-Martín A, Cerda-Flores R, Rangel-Villalobos H (2009) Pre-Hispanic Mesoamerican demography approximates the present-day ancestry of Mestizos throughout the territory of Mexico. *Am J Phys Anthropol* 139(3):284–294
- Salazar-Flores J, Zuñiga-Chiquette F, Rubi-Castellanos R, Alvarez-Miranda JL, Zetina-Hernández A et al (2015) Admixture and genetic relationships of Mexican-Mestizos regarding Latin American and Caribbean populations based on 13 CODIS-STRs. *Homo* 66(1):44–59
- A. Tereba, Tools for analysis of population statistics. Profiles in DNA: Promega Corp., (2001)
- Wurmb-Schwark VN, Mályusz V, Simeoni E, Lignitz E, Poetsch M (2006) Possible pitfalls in motherless paternity analysis with related putative fathers. *Forensic Sci Int* 159(2–3):92–97
- Ziętkiewicz E, Witt M, Daca P, Zebracka-Gala J, Goniewicz M, Jarzab B, Witt M (2012) Current genetic methodologies in the identification of disaster victims and in forensic analysis. *J Appl Genet* 53(1):41–60

**Submit your manuscript to a SpringerOpen® journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)