

Reference database of hypervariable STR Loci in the Entre Ríos Province of Argentina

G.G. Martínez^{a,b,*}, L.C. Schaller^a, L.E. Vázquez^a, M. Bolea^b,
B. Martínez-Jarreta^b

^a *Servicio de Genética Forense, Superior Tribunal de Justicia, Entre Ríos, Argentina*

^b *Laboratorio de Identificación Humana y Genética Forense, Facultad de Medicina, Universidad de Zaragoza, España*

Abstract. Allele frequencies of twelve short tandem repeat (STR) loci, CSF1PO, TPOX, TH01, F13A01, FESFPS, vWA, D16S539, D7S820, D13S317, D5S818, F13B and LPL, were determined over six major regional groups in the Argentinean province of Entre Ríos. No deviation was observed in the total population analysed or in the subpopulation for all loci. Neither was there any evidence of allele correlation between loci. The combined matching probability and the combined mean exclusion chance in the Entre Ríos population were 2.44×10^{-13} and 0.99993, respectively. Frequencies, statistical parameters and phylogenetic inference based on distance matrix for all the population groups are provided. We analysed allele frequency distribution by Pairwise F_{ST} Genetic Distance to construct a tree based on the Neighbour-Joining method, and obtained one that coincides well with their geographical distribution. This study demonstrates that these loci are a useful and handy tool for forensic identification and parentage testing in this Argentinean province. © 2005 Elsevier B.V. All rights reserved.

Keywords: STR database; Entre Ríos; Argentine

1. Introduction

At present, STR genetic markers are mainly used in the forensic field for personal identification and paternity analysis. As recommended in the literature [1,2], a forensic lab must create an adequate database to calculate probabilities in these cases. The aim of this study is to analyse allelic distributions of 12 STR loci (CSF1PO, TPOX, TH01, F13A01,

* Corresponding author. Servicio de Genética Forense, Superior Tribunal de Justicia, Entre Ríos, Argentina. Tel.: +54 3434209x402; fax: +54 3434209x402.

E-mail address: martinezgg@arnet.com.ar (G.G. Martínez).

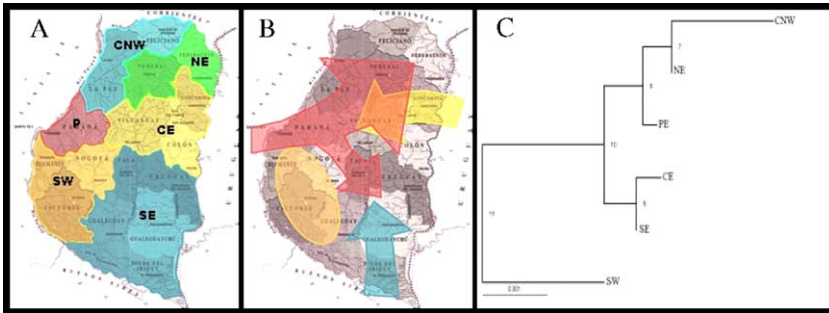


Fig. 1. (A) Map of Entre Ríos with regional groups marked by CNW (Centre North West), NE (North East), P (Paraná, department where the capital is located), CE (Centre East), SE (South East), and SW (South West). (B) Regional immigration influences based on physical connections (bridges over Paraná and Uruguay rivers) to the adjacent territory. (C) Neighbour-joining tree based on Coancestry Identity (pairwise F_{ST} distances) between six subpopulations over Entre Ríos province of Argentina.

FESFPS, vWA , D16S539, D7S820, D13S317, D5S818, F13B and LPL) over six regional population groups in the Argentine province of Entre Ríos, to establish a more representative STR database for these populations, which will allow us to analyse the cases requested by the court more accurately. For the present study we subdivided the Entre Ríos province into six groups, shown in Fig. 1A, and analysed these samples by Pairwise F_{ST} genetics distances [3] to obtain the Neighbour-Joining tree that best adjusts to the geographical distribution and the way in which the population has become established in the province over the last decades. In addition, we provide allele frequencies and statistical parameters for the Entre Ríos population and allele frequencies for the six regional groups analysed (upon request).

2. Materials and methods

DNA from blood samples was collected from 686 unrelated healthy donors from all over the province and grouped according to six previously established sub-regions referred to as CNW, NE, P, CE, SE, and SW, for Centre North West, North East, Paraná (department in which the capital is located), Centre East, South East, and South West respectively, in accordance with Fig. 1A. All twelve loci were genotyped according to manufacturer recommendations [4]. Statistical parameters were calculated with BDGen® 1.0 [5], GDA [6], TreeView 1.6.5 [7] and Power Marker 3.0 [8] software packages.

3. Results and discussion

We calculated the allele frequencies for these twelve STR loci in six regional groups (data not shown) and over the entire province (see statistical parameters in Table 1). Combined matching probability and combined mean exclusion chance in the Entre Ríos population were 2.44×10^{-13} and 0.99993, respectively. No significant deviation from (HWE) were observed among the Entre Ríos population, supporting the view that the use of the product rule would provide a good approximation of the rarity of a multiple locus profile.

No significant differences were found among the subpopulation studied. However, we evaluated their genetic relationship by constructing a Neighbour-Joining tree based on F_{ST} distances according

Table 1
Allele frequencies and statistical parameters for 12 STR loci in Entre Ríos population

Allele	CSF1PO	TPOX	TH01	F13A01	FESFPS	VWA	D16S539	D7S820	D13S317	D5S818	F13B	LPL
<i>n</i>	683	686	685	679	681	680	676	674	675	424	644	657
3.2				0.1966								
4				0.0663								
5		0.0007		0.1826								
6	0.0007	0.0036	0.2504	0.2202					0.0015		0.0629	
7	0.0022	0.0022	0.2445	0.2975				0.0126	0.0007	0.0519	0.0116	0.0030
8	0.0051	0.4665	0.1022	0.0052	0.0147		0.0214	0.1202	0.0793	0.0083	0.2120	0.0023
9	0.0146	0.0802	0.1569		0.0073		0.1561	0.1231	0.1481	0.0436	0.2966	0.0350
9.3			0.2409									
10	0.2489	0.0598	0.0051		0.2276		0.1006	0.2708	0.0600	0.0519	0.4146	0.4848
11	0.3038	0.3156		0.0007	0.4501		0.3025	0.2567	0.2385	0.3785	0.0023	0.2451
12	0.3653	0.0678		0.0022	0.2372	0.0007	0.2641	0.1736	0.3007	0.3314	0.0629	0.1956
13	0.0534	0.0036		0.0022	0.0565		0.1294	0.0356	0.1237	0.1226		0.0327
14	0.0044			0.0074	0.0059	0.0728	0.0251	0.0074	0.0452	0.0118		0.0015
15	0.0015			0.0081	0.0007	0.0882	0.0007		0.0022			
16				0.0096		0.2691						
17				0.0015		0.3132						
18						0.1728						
19						0.0691						
20						0.0132						
21						0.0007						
MF	0.0042	0.0042	0.0045	0.0045	0.0042	0.0044	0.0046	0.0047	0.0047	0.0069	0.0045	0.0043
PE	0.4523	0.4224	0.5717	0.5808	0.4349	0.5796	0.5852	0.6056	0.6183	0.4967	0.4314	0.4135
MP	0.1365	0.1678	0.0822	0.0801	0.1503	0.0765	0.0789	0.0728	0.0669	0.1254	0.1522	0.1663
Ho	0.6866	0.6867	0.7857	0.7850	0.6800	0.7473	0.8034	0.8164	0.8120	0.7242	0.6896	0.6605
He	0.7097	0.6686	0.7849	0.7869	0.6864	0.7822	0.7870	0.8001	0.8041	0.7252	0.6916	0.6648
PIC	0.6548	0.6153	0.7495	0.7535	0.6341	0.7498	0.7550	0.7705	0.7771	0.6812	0.6349	0.6119
DF	45	36	15	78	28	36	28	28	45	28	28	15
CST	52.3922	18.3139	18.7083	81.5283	13.1153	31.2023	23.3335	29.4127	43.6441	21.1711	9.6771	25.4381
CST <i>p</i>	0.2091	0.9937	0.2273	0.3701	0.9924	0.6961	0.7162	0.3918	0.5294	0.8181	0.8395	0.6039
LRT	26.1256	17.8901	15.9111	51.6296	13.1942	31.1130	19.2304	25.2544	39.0357	24.4199	11.0207	18.9059
LRT <i>p</i>	0.9890	0.9950	0.3880	0.9908	0.9920	0.7001	0.8909	0.6140	0.7216	0.6592	0.7511	0.9010
ET <i>p</i>	0.3984	0.8973	0.2632	0.4187	0.9113	0.2611	0.8494	0.3947	0.3759	0.3685	0.7621	0.4692

MF: minimum frequency ($\alpha=0.045$), PE: power of exclusion, MP: matching probability, Ho: observed heterozygosity, He: expected heterozygosity, PIC: polymorphism information content, DF: degree freedom for χ^2 and likelihood ratio test, CST: Chi Squared Test, CST *p*: *p* value for CST, LRT: Likelihood Ratio Test, LRT *p*: *p* value for LRT, ET *p*: Exact Test *p* value with 5000 iterations.

to Weir [3] (Fig. 1C). The genetic relationship in these six populations was consistent with their geographical distributions and historical migrations, represented by arrows in Fig. 1B.

References

- [1] H.A. Hammond, et al., Evaluation of 13 short tandem repeat loci for use in personal identification applications, *Am. J. Hum. Genet.* 55 (1994) 175–189.
- [2] National Research Council, *The Evaluation of Forensic DNA Evidence*, National Research Council, National Academy Press, Washington, DC, 1996.
- [3] B.S. Weir, *Genetic Data Analysis II*. Sinauer, Sunderland, Massachusetts, 1996, pp. 194–195.
- [4] Gene Print STR Systems Technical Manual (Silver Stain Detection) and Gene Print Fluorescent STR Systems Technical Manual, Promega.
- [5] B.S. Vázquez, et al., New database software with statistic analysis tool for human genotyping, VII Meeting of the Spanish–Portuguese Speaking Working Group for the International Society for Forensic Genetics, Barcelona, 2002, <http://www.simedico.com.ar>.
- [6] P.O., Lewis, D. Zaykin. Genetic data analysis: computer program for the analysis of allelic data. Version 1.0 (d16c). <http://lewis.eeb.uconn.edu/lewishome/software.html> (2001).
- [7] TreeView 1.6.5. Copyright Roderic D. M. Page. (2001), <http://taxonomy.zoology.gla.ac.uk/rod/rod.html>.
- [8] Power Marker 3.0 by Jack Liu, <http://www.powermarker.net>.