



First-degree familial relationships coincidences in a population database of Jujuy (Argentina) compared with simulated populations

Juan José Martínez^a, Marcos Felipe Tula Molina^b, Luis Adrián Sánchez^b, María Inés Deus^b, María Isabel Ramella^b, Gladys Beatriz Maurín^b, Gustavo Martínez^c, Ariel Chernomoretz^{d,e,f}, María Soledad Escobar^g, Gustavo Sibilla^g, María Cecilia Miozzo^{b,*}

^a Instituto de Ecorregiones Andinas, CONICET and Universidad Nacional de Jujuy, San Salvador de Jujuy, Argentina

^b Laboratorio Regional de Genética Forense del NOA, Departamento Médico, Poder Judicial de Jujuy, San Salvador de Jujuy, Argentina

^c Servicio de Genética Forense y Registro Provincial de Datos Genéticos, Superior Tribunal de Justicia de la Provincia de Entre Ríos, Paraná, Argentina

^d Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina

^e Instituto de Física de Buenos Aires (IFIBA) UBA/CONICET, Argentina

^f Laboratorio de Biología de Sistemas Integrativa, Fundación Instituto Leloir, Argentina

^g Fundación Sadosky, Argentina

ARTICLE INFO

Keywords:

DNA databases
Simulated populations
Adventitious matches
Jujuy
Argentina
Low stringency search

ABSTRACT

Jujuy, a province from Northwestern Argentina, has a population with a great input of the Native American gene pool together with European and African contributions. For forensic purposes, a frequency population database was built by selecting 500 nonrelated individuals from Jujuy province and genotyping them with 21 autosomal STRs from the kit GlobalFiler™. The genetic profiles were incorporated in GENis (software developed in Argentina for storage and comparison of DNA profiles) to corroborate the lack of matches under high stringency. After this confirmation, we performed a searching strategy with low stringency, often used to find parent-child relationships. As a result, four (4) adventitious matches were found. One hundred (100) populations were simulated; with 500 random genetic profiles each one, using Jujuy allele frequencies. The coincidences for these populations were on average 2.07 (95% CI: 1.77–2.36) and there were no significant differences between Jujuy and the simulated populations. It is known that the number of adventitious matches increases when you go down with search stringency in DNA databases. Thus, care should be taken when searching for first-degree relationships, even with a high number of markers. LR calculations and complementary studies (such as extra autosomal markers, Y chromosome, and mitochondrial DNA) should be performed in this kind of search.

1. Introduction

Jujuy is an Andean province from Northwestern Argentina. It has a population with a great input of the native South American gene pool, together with European and African contributions [1,2]. For forensic purposes, a frequency population database was built by selecting 500 nonrelated individuals from Jujuy province and by genotyping them with 21 autosomal STRs. To corroborate the presence of repeated individuals, the genetic profiles were incorporated into GENis (<http://www.fundacionsadosky.org.ar/genis>), an open-source software developed in Argentina for storage and comparison of DNA profiles. Like CODIS, GENis follows the ENFSI (European Network of Forensic Science Institutes) recommendations for DNA database management [9] and provides optimized strategies for profile matching tasks. After the

confirmation that no matches were found in high stringency, we performed a searching strategy with low stringency, often used to find parent-child relationships. As opposed to what was expected, four (4) adventitious matches were found. To investigate if this result was caused by a specific genetic characteristic of the Jujuy population (e. g., endogamy), we performed a simulation-based approach to test the hypothesis that adventitious matches arose at random.

2. Material and methods

DNA from 500 buccal swab samples were extracted with DNA IQ™ Casework Pro Kit for Maxwell® 16 (Promega). Samples were genotyped for 21 autosomal STR with GlobalFiler™ PCR Amplification kit (Applied Biosystems). The bias from Hardy-Weinberg equilibrium (HWE) from

* Corresponding author.

E-mail address: cmiozzo@justiciajujuy.gov.ar (M.C. Miozzo).

<https://doi.org/10.1016/j.figss.2019.10.148>

Received 30 August 2019; Received in revised form 9 October 2019; Accepted 11 October 2019

Available online 19 October 2019

1875-1768/ © 2019 Published by Elsevier B.V.

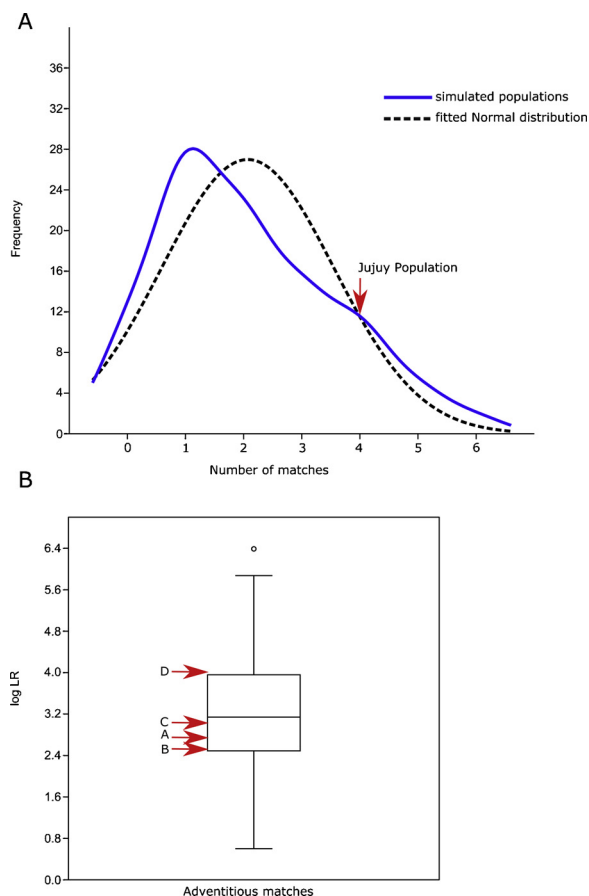


Fig. 1. (A) Histogram based on the number of matches found by GENis under low stringency search from simulated populations and the number of coincidences from the Jujuy database. (B) Box-and-whisker plot showing the range of variation of LR of adventitious matches from simulated populations. The box delimits the 25th and 75th quartiles, the horizontal bar indicates the median, and the whiskers extend to the most extreme data points. Red arrows indicate the values of log LR found in the Jujuy database (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

the Jujuy province database was estimated by means of the exact test in Arlequin 3.5 [3] and allele frequencies were estimated in GenAEx 6.502 [4,5]. Forensic population parameters were estimated with the software PowerStats v 1.2 [6]. One hundred (100) populations were simulated with the *forensim* package [7] for R [8]. For each one of them, the observed Jujuy province allele frequencies were used to simulate 500 random genetic profiles. Low stringency matching profiles were then sought using GENis. The results were recorded and its distribution was described. Finally, this distribution was compared with the observed value obtained in the Jujuy province database. The strength evaluation of a first-degree familial relationship between low-stringency matching profiles was assessed using LR statistics. The LRs were calculated with the software Familias version 3.2.6 [10].

3. Results and discussion

There were no deviations from HWE in the Jujuy population after Bonferroni correction (p-value 0.00238). The forensic parameters demonstrated that these 21 STR have a good performance for parentage testing and forensic casework.

Low values of likelihood ratios (LRs), for parent-child relationships, were obtained for the matches of the Jujuy population (A = 512, B = 302, C = 905 and D = 13,496). There were inconsistencies (exclusions) in all matches when more STRs were assayed, including Y

STRs in three of the four matches. This proved that the coincidences were adventitious in the Jujuy population.

The adventitious matches for the simulated populations were on average 2.07 (mean 95% CI: 1.77–2.36), even though 4 matches were found in 12 populations, 5 matches in 5 and there were 2 populations with 6 matches. As there were no significant differences between Jujuy and the simulated populations ($p > 0.05$) (Fig. 1A), the cause of adventitious matches due to a specific Jujuy population characteristic was disregarded. The LRs of 207 matches obtained from the 100 simulated populations, ranged mostly from 4 to 100,000, but there was one value above 2,000,000 (Fig. 1B). It is known that the number of adventitious matches increases when you go down with search stringency in DNA databases, but analyzing a high number of autosomal STRs (21), the 4 matches in the Jujuy population were unexpected. Thus, care should be taken with first-degree relationships searches, even with a high number of markers. LR calculations and complementary studies (like more autosomal markers, the Y chromosome, and mitochondrial DNA) should be performed in this kind of search.

4. Conclusions

DNA databases are one of the most powerful tools in the investigation of crimes, linking DNA profiles from crime-related biological trace material to each other and to the possible donors. They are vital for the resolution of serial crimes. Adventitious matches could arise when forensic genetic databases became larger, when the number of loci are insufficient, and also when you go down with search stringency. This may occur with the searching strategy called “familial searching”, allowed in some countries, where a search for matches with possible relatives of the donor of a crime scene associated DNA profile can be conducted [9]. Here we described a case study from a forensic database from Argentina. The next step is to shed light on the prediction of the number of adventitious coincidences in a low stringency search strategy.

Conflict of interest

None.

Acknowledgments

The authors would like thank to Cecilia Garcia for her assistance with the English; and Silvia Malnatti, Hugo Valdiviezo and Patricia Zurueta for administrative support.

References

- [1] S. Cardoso, L. Palencia-Madrid, L. Valverde, M.A. Alfonso-Sánchez, L. Gómez-Pérez, E. Alfaro, C.M. Bravi, J.E. Dipierri, J.A. Peña, M.M. Pancorbo, Mitochondrial DNA control region data reveal high prevalence of Native American lineages in Jujuy province, NW Argentina, *Forensic Sci. Int. Genet.* 7 (2013) e52–e55.
- [2] G. Martínez, A. Borosky, D. Corach, C. Llull, L. Locarno, M. Lojo, M. Marino, M.C. Miozzo, N. Modesti, C. Pacharoni, J.P. Pilili, M.I. Ramella, A. Sala, C. Schaller, C. Vullo, U. Toscanini, Investigator® HDplex (Qiagen) reference population database for forensic use in Argentina, *Forensic Sci. Int. Genet.* 26 (2017) 91–95.
- [3] L. Excoffier, G. Laval, S. Schneider, Arlequin ver. 3.0: an integrated software package for population genetics data analysis, *Evol. Bioinform.* 1 (2005) 47–50.
- [4] R. Peakall, P.E. Smouse, GENALEX 6: genetic analysis in excel. Population genetic software for teaching and research, *Mol. Ecol. Notes* 6 (2006) 288–295.
- [5] R. Peakall, P.E. Smouse, GenAEx 6.5: genetic analysis in excel. Population genetic software for teaching and research—an update, *Bioinformatics* 28 (2012) 2537–2539.
- [6] A. Tereba, Tools for analysis of population statistics, *Profiles DNA* 3 (1999) 14–16.
- [7] H. Hamed, *Forensim*: an open-source initiative for the evaluation of statistical methods in forensic genetics, *Forensic Sci. Int. Genet.* 5 (2011) 265–268.
- [8] R Development Core Team, R: A Language and Environment for Statistical Computing, (2014).
- [9] ENFSI DNA Working Group, DNA Database Management Review and Recommendations, (2017).
- [10] D. Kling, A.O. Tillmar, T. Egeland, Familias 3 — extensions and new functionality, *Forensic Sci. Int. Genet.* 13 (2014) 121–127.