POPULATION DATA

Allelic frequencies with 23 autosomic STRS in the Aymara population of Peru

Carlos David Neyra-Rivera [1] · Andres Ticona Arenas [2] · Edgardo Delgado Ramos [3] ·
Margarita Rosa Eugenia Velasquez Reinoso [3] · Bruce Budowle [4]

Abstract
Population data of the Aymara in the province of Puno were established for 23 autosomal STR markers. DNA was obtained from unrelated individuals ($n = 190$) who reside in three areas of the Floating Islands of Lake Titicaca, residents on the border with Bolivia and residents who are not from the border with Bolivia. The PENTA E marker presented the highest PD (0.9738), PIC (0.8793), and PM (0.7847) values. The combined PD was greater than 0.99999999 and the combined PE was 0.99999994. The largest distance, based on Fst values, was between the Aymara population and the Ashaninca population (0.04022), and the smallest distance was with the populations of Bolivia (0.00136) and Peru (0.00525).

Keywords Allele frequencies · Aymara population · Forensic genetics · STR markers

Forensic Genetics laboratories rely on allele frequencies from short tandem repeat loci (STRs), such as the 20 core loci recommended for CODIS (Combined DNA Index System) in the USA [1]. With the increase of core STR loci for CODIS, there has been development of expanded STR kits. Because of the additional loci, there is a need for supporting population data for casework analyses. Studies have been carried out on more than 20 STRs and with more than 100 individuals in mestizo populations in Peru [2–4], but not on any of the 76 native ethnic groups or populations united by existing

cultural and linguistic ties in Peru [5]. Among the 76 ethnic groups in Peru, the largest population is Quechua followed by Aymara [6]. The Aymara ethnic group is predominantly found in southeast Peru in the department of Puno and in the border areas with Bolivia [7] and has been studied with nuclear genes, mitochondrial DNA, and insertion-deletion markers [8–12] but not with autosomal STRs.

This study was approved by the ethics committee of the Institute of Tropical Medicine "Daniel Alcides Carrion" from the Universidad Nacional Mayor de San Marcos (Certificate of Approval CIEI-2018-015). Before participating in the study, all participants freely and voluntarily signed their informed consent. DNA was obtained from 190 unrelated adults from the Aymara population from the province of Puno. All participants were born in Puno as well as their parents and grandparents. The samples were taken from 37 individuals residing in the Puno Islands (Anapia, Suana, and Iscaya), 58 individuals residing in districts bordering Bolivia (Unicachi, Ollaraya, Tinicachi, and Yunguyo), and 95 individuals residing in districts that are not on the border with Bolivia (Zepita, Platería, Cuturapi, Juli, Pomata Huacullani, Copani, and Yunguyo). Samples were taken by finger puncture, and 4–6 drops of blood were placed on a Nucleid Card (Copan). These samples were stored in sterile, ambient conditions until use.

Prior to sample processing, the Nucleid Cards (Copan) were dried at 37 °C for 60 min. Then, a 1.2-mm punch was taken from each card and amplified by direct PCR using the

✉ Carlos David Neyra-Rivera
carlosdavidmp@outlook.es

1 Universidad Científica del Sur, Carr. Panamericana Sur 19, Villa EL Salvador, 15067 Lima, Peru

2 Posgrado de Ciencias Biológicas, Universidad Nacional Mayor de San Marcos, Ciudad Universitaria Cercado de Lima,, 15081 Lima, Peru

3 Universidad Nacional Mayor de San Marcos, Ciudad Universitaria Cercado de Lima,, 15081 Lima, Peru

4 Center for Human Identification, University of North Texas Health Science Center, Ft Worth, TX 76107, USA

Verifiler Express kit (Life Technologies) following the manufacturer's instructions. Amplified products were detected in the Applied Biosystem™ 3500XL Genetic Analyzer (Life Technologies) following the manufacturer's recommended protocol. After capillary electrophoresis, the data were imported into the GeneMapper® ID-X v1.5 software to generate the genetic profiles.

The power of discrimination (PD), the content of polymorphic information (PIC), the probability of exclusion (PE), the observed heterozygosity (Ho), and the match probability (PM) were calculated with Power Stats V1.2 software [13]. The expected heterozygosity (He) and detection of departures from Hardy-Weinberg equilibrium, the combined PD, the combined PE, and the population distance (using Fst and Reynold's genetic distance) were performed with Arlequin software [14]. The population studied ($n = 190$) was compared with other populations [4, 15–17] using 15 STR markers that were common among the different studies (D3S1358, vWA, D16S539, CSF1PO, TPOX, D8S1179, D21S11, D18S51, D19S433, TH01, FGA, D5S818, D13S317, D7S820, and D2S1338).

The allele frequencies and other relevant population parameters for the 23 autosomal STR markers for the Aymara sample population are shown in Supplemental Table 1. All markers were highly polymorphic [18]. There were no detectable departures from Hardy-Weinberg equilibrium expectations except for the FGA locus ($p = 0.0248$). After applying the Bonferroni correction, the FGA marker no longer departed significantly from expectations.

The average PD per locus was 0.8668 and ranged from 0.6768 to 0.9738. The marker that had the highest value was PENTA E (0.9738, 15 alleles) and the lowest value was observed at the D2S441 locus (0.6768, 7 alleles). The loci with the highest PD and PIC values were PENTA E (0.8793) and D2S441 (0.4312). The PE ranged from 0.7847 (PENTA E) to 0.1532 (D2S441). The PM values ranged from 0.3232 (D2S441) and 0.0262 (PENTA E). He and Ho were similar within a locus, and the averages were 0.7217 and 0.7204, respectively. The He ranged between 0.4718 (D2S441) and 0.8917 (PENTA E), while the Ho range was 0.4579 (D2S441) and 0.8947 (PENTA E). Under the assumption of independence, the combined PD was greater than 0.99999999, and the combined PE was 0.99999994.

Population distances are shown applying Fst and Reynold's genetic distance [14] in Supplementary Table 2. In this study, a Fst value of less than 0.01 was considered to have little population subdivision, and a value greater than 0.01 was considered indicative of notable population substructure [19, 20]. The Aymara population, the Bolivian population [15], and the Peruvian population [4] all cluster and are close to each other with the greatest distance being with the Ashaninca [16] and US Hispanics [17] (Supplemental Fig. 1). When separating out the Aymara population sample by region, the three regions—the Aymara Island, Bolivian border population, and the population not near the border of Bolivia—are similar (Supplemental Table 3 and Supplemental Fig. 2). The data suggest that Aymara is similar to Peruvian and other Bolivian population data [4, 15] and that the Ashaninca population [16] is notably different from all other populations analyzed herein.

This study provides STR population data from the Aymara, one of the ethnic groups of Peru. The results support the use of the population data for statistical calculations in human identity testing cases in Peru.

Compliance with ethical standards

References

1. Moretti T, Moreno L, Smerick J, Pignone M, Hizon R, Buckleton J, Bright J, Onorato A (2016) Population data on the expanded CODIS core STR loci for eleven populations of significance for forensic DNA analyses in the United States. Forensic Sci Int Genet 25:175–181. https://doi.org/10.1016/j.fsigen.2016.07.022

2. Delgado E, Neyra CD (2018) Frecuencias alélicas de 21 marcadores STR autosómicos en una población mestiza peruana aplicado a la práctica forense. Rev Esp Med Legal 45:92–97. https://doi.org/10.1016/j.reml.2018.09.002

3. Neyra CD, Suarez M, Cueva E, Bailón H, Gutierrez E (2019) Identificación genética de recién nacidos en Perú: un estudio piloto. Rev Chil Pediatr 90(1):26–35. https://doi.org/10.32641/rchped.v90i1.730

4. Neyra Rivera CD, Delgado Ramos E, Robles Mamani CS, Velasquez Reinoso MRE, Caceres Rey OA, Budowle B (2020) Population genetic study of a Peruvian population using human identification STRs. Int J Legal Med. https://doi.org/10.1007/s00414-020-02418-6

5. Instituto Nacional de Desarrollo de Pueblos Andinos, Amazónicos y Afroperuanos (INDEPA) (2010) Mapa etnolingüístico del Perú. Rev Peru Med Exp Salud Publica 27(2):288–291. https://doi.org/10.17843/rpmesp.2010.272.1477

6. Instituto Nacional de Estadística e Informática (INEI). Perú: Perfil Sociodemográfico Informe Nacional Censos Nacionales 2017: XII de Población, VII de Vivienda y III de Comunidades Indigenas [Internet]. Lima: INEI; 2018 [revised 2017-2018; cited 2020 Sep 04]. Available from: https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1539/

7. Instituto Nacional de Estadística e Informática (INEI). PERÚ-BOLIVIA Perfil Sociodemográfico y Económico de los Distritos y Municipios Fronterizos, 2015. Lima: INEI, 2015 [revised 2015; cited 2020 Sep 04]. Available from: https://www.inei.gob.pe/

media/MenuRecursivo/publicaciones_digitales/Est/Lib1383/index.html

8. Córdova J, Fujita R, Sandoval J, Descailleaux J, Velásquez M, Távara C, Barletta C (2011) Divergencia genética en poblaciones peruanas detectada a partir de las frecuencias haplotípicas del mtDNA y del gen nuclear MBL. An Fac Med 72(1):51–59. https://doi.org/10.15381/anales.v72i1.1103

9. Acosta O, Solano L, Huerta D, Oré D, Sandoval J, Figueroa J (2012) Variabilidad genética de la respuesta inflamatoria. I. Polimorfismo -511 C/T en el gen *IL1β* en diferentes subpoblaciones peruanas. An Fac Med 3:221–225. https://doi.org/10.15381/anales.v73i3.868

10. Bisso-Machado R, Ramallo V, Paixao-Cortes V, Acuna-Alonzo V, Demarchi D, Sandoval J, Granara A, Salzano F, Hunemeier T, Bortolini M (2016) NAT2 gene diversity and its evolutionary trajectory in the Americas. Pharm J 16(6):559–565. https://doi.org/10.1038/tpj.2015.72

11. Córdova JH, Fujita R, Sandoval J, Velásquez M, Távara C, Cotos D, Vásquez J, Barletta C, Descailleaux J (2008) Poblamiento del continente americano y del Perú sugerido de un análisis filogeográfico de haplogrupos del mtDNA en etnias nativas. I: inferencias primarias. Arch Biol Andina 14(1):23–39

12. Sandoval J, Salazar-Granara A, Acosta O, Castillo-Herrera W, Fujita R, Pena S, Santos F (2013) Tracing the genomic ancestry of Peruvians reveals a major legacy of pre-Columbian ancestors. J Hum Genet 58(9):627–634. https://doi.org/10.1038/jhg.2013.73

13. Tereba A (1999) Tools for analysis of population statistics. Profiles DNA 2:3–5

14. Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinformatics Online 1:47–50

15. Rocabado O, García O, Yurrebaso I, Inda F, Taboada P (2009) Population genetic data for 15 STR loci (Identifiler™ kit) in Bolivia. Legal Med 11(6):302–304

16. Tineo DH (2017) Caracterización genética de una muestra poblacional ashaninka en el distrito de Puerto Bermúdez, Pasco-Perú, empleando marcadores STR autosómico y del cromosoma Y [Tesis de Maestría]. Universidad Nacional Mayor de San Marcos, Lima

17. Hill C, Duewer D, Kline M, Coble M, Butler J (2013) US population data for 29 autosomal STR loci. Forensic Sci Int Genet 7(3): e82–e83

18. Foissac A, Crouau R, Fauré S, Thomsen M, Cambon-Thomsen A (1997) Microsatellites in the HLA región: an overview. Tissue Antigens 49:197–214. https://doi.org/10.1111/j.1399-0039.1997.tb02740.x

19. Budowle B, Shea B, Niezgoda S, Chakraborty R (2001) CODIS STR loci data from 41 sample populations. J Forensic Sci 46(3): 453–489

20. De la Cruz AF (2006) Análisis de la variabilidad genética en una muestra poblacional de la provincia de Andahuaylas–Apurímac, Perú–con los marcadores microsatélites TPOX