Human Immunology 80 (2019) 413–414

Contents lists available at ScienceDirect

Human Immunology

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

Short population report

5-Locus high-resolution HLA allele and haplotype frequencies in Costa Ricans from the Central Valley

Esteban Arrieta-Bolaños[a,b,c,*], Juan José Madrigal-Sánchez[d], Jeremy E. Stein[b], Priscilla Órlich-Pérez[c,e], Gilbert Arrieta-Molina[f], Lizbeth Salazar-Sánchez[d], J. Alejandro Madrigal[b,g], Steven G.E. Marsh[b,g], Bronwen E. Shaw[b,h]

[a] Institute for Experimental Cellular Therapy, University Hospital, Essen, Germany
[b] Anthony Nolan Research Institute, Royal Free Hospital, London, UK
[c] Centro de Investigaciones en Hematología y Trastornos Afines (CIHATA), Universidad de Costa Rica, San José, Costa Rica
[d] Escuela de Medicina, Universidad de Costa Rica, San José, Costa Rica
[e] División de Banco de Células Madre, Laboratorio Clínico, Hospital San Juan de Dios, San José, Costa Rica
[f] Laboratorio Clínico, Hospital Nacional de Niños, San José, Costa Rica
[g] UCL Cancer Institute, Royal Free Campus, London, UK
[h] Center for International Blood and Marrow Transplant Research, Department of Medicine, Medical College of Wisconsin, Milwaukee, USA

ARTICLE INFO

Keywords:
Admixture
Ancestry
Costa Rica
Ethnicity
Frequencies
Hispanic
Human leukocyte antigen
Population

ABSTRACT

A total of 221 Costa Rican Mestizos from the Central Valley were genotyped at high-resolution for the human leukocyte antigen loci HLA-A, -B, -C, -DRB1, and -DQB1 using sequence-based typing methods. The respective allele and extended haplotype frequencies, as well as Hardy-Weinberg proportions were calculated. The most frequent extended haplotype identified was A*24:02:01-B*40:02:01-C*03:05-DRB1*08:02:01-DQB1*04:02:01, with an estimated frequency of 2.04%. No deviation from Hardy-Weinberg Equilibrium was detected at any of the loci studied. The HLA genotypic data of the population sample reported here are available publicly in the Allele Frequencies Net Database under the population name “Costa Rica Central Valley Mestizo” and the identifier AFN3606.

Costa Rica is a Central American country with an area of 51,000 km^2 and a population of approximately 5 million inhabitants. Its population is mainly Spanish-speaking and admixed (Mestizo), with previously reported regional genetic variations [1]. Its largest population, that of its Central Valley Region (CRCV), concentrates the country’s capital city and major provincial cities, which agglomerate 60% of the country’s population. This region’s population has been the subject of genetic scientific interest due to its demographic history including claims for relative isolation [2,3]. The CRCV formed approximately 400 years ago through admixture of mainly male Spanish colonizers and female Amerindians and African slaves, although low aboriginal population density and low number of indentured slaves preserved a larger European genetic component in comparison to neighboring countries and other regions within the country [4].

A total of 221 peripheral blood or saliva samples from unrelated healthy volunteer donors were included in this study. All donors traced their ancestry to the CRCV. All samples were collected in major cities of this region as part of the DNA biobank at the University of Costa Rica’s Centre for Research in Hematology and Related Disorders (Centro de Investigaciones en Hematología y Trastornos Afines, CIHATA), which continuously collects samples for the characterization of genetic variation across the country and its association with medically-relevant traits. DNA was extracted from blood or saliva by routine methods. All participants gave informed written consent as per CIHATA’s DNA biobank standard procedures. Sample collection under CIHATA’s biobank and this study were approved by the local ethics committee at the University of Costa Rica.

High-resolution HLA typing was performed at Anthony Nolan by in-house sequence-based typing methods, with generic amplification of exons 2, 3 and 4 for HLA-A, HLA-B, HLA-C. For HLA class II, exon 2 of the HLA-DRB1 and HLA-DQB1 genes was amplified using in-house allele group-specific primer pairs. Amplicons were purified and each exon

Abbreviations: AFND, Allele Frequencies Net database; CRCV, Costa Rican Mestizo from the Central Valley; EM, expectation-maximization; HWE, Hardy-Weinberg equilibrium

* Corresponding author at: Institute for Experimental Cellular Therapy, Universitätsklinikum Essen, Virchowstraße 171, Essen 45147, Germany.
E-mail address: esteban.arrieta-bolanos@uk-essen.de (E. Arrieta-Bolaños).

https://doi.org/10.1016/j.humimm.2019.05.006
Received 23 April 2019; Received in revised form 6 May 2019; Accepted 10 May 2019
Available online 23 May 2019

was sequenced on an ABI 3730xL DNA Analyzer with specific forward and reverse primers using the Big Dye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). Sequence analysis was done with Assign-SBT software (version 3.6+, Conexio Genomics, Freemantle, Australia) using IPD-IMGT/HLA database release 3.9.0 [5]. Ambiguities were further solved using allele-group- (HLA class I) or codon-86-specific (HLA class II) primer combinations and sequencing of the allelic products. For HLA class I, alleles with identical sequences at codons 2, 3, and 4 could not be distinguished, and were assigned the third-field name of the allele with the lowest numerically ordered name (which is usually the more common one). Hence, HLA class I allele frequencies given may in many cases represent the frequency of a group of alleles sharing sequences at these exons. Allele groups (G) were assigned to HLA class II alleles according to IPD-IMGT/HLA database specifications. All homozygous samples were confirmed by at least two determinations and using different techniques. Allele frequencies and compliance with Hardy-Weinberg equilibrium (HWE) were computed using the online tools available from the HLA-net platform (http://hla-net.eu/tools/) [6]. Haplotype frequencies were estimated by an in house implementation of the expectation-maximization (EM) algorithm [7].

All samples were typed for HLA-A, HLA-B, HLA-C, and HLA-DRB1, whereas a randomly selected subgroup (n = 101) was also typed for HLA-DQB1. A total of 37, 64, 25, 38, and 13 alleles were identified for HLA-A, -B, -C, -DRB1, and –DQB1, respectively. Compliance with HWE was confirmed for all loci ($p > 0.5$). The five most common alleles per locus were A*02:01:01 (17.6%), A*24:02:01 (13.6%), A*03:01:01 (11.8%), A*01:01:01 (7.9%), A*68:01:02 (5.7%); B*07:02:01 (11.8%), B*40:02:01 (9.3%), B*35:01:01 (8.8%), B*44:02:01 (5.4%), B*14:02:01 (4.7%); C*04:01:01 (17.4%), C*07:02:01 (16.7%), C*03:05 (9.5%), C*06:02:01 (9.0%), C*07:01:01 (7.5%); DRB1*13:01:01G (9.9%), DRB1*04:07:01G (9.3%), DRB1*15:01:01G (7.9%), DRB1*03:01:01G (7.5%), DRB1*01:01:01G (6.1%); DQB1*03:01:01G (22.3%), DQB1*03:02:01G (16.8%), DQB1*05:01:01G (15.8%), DQB1*06:03:01G (11.4%), DQB1*02:01:01G (10.9%). The complete lists of alleles for each locus are given in Supplementary Table 1. The HLA gene profile of the CRCV shows clear evidence of the presence of tri-ethnic admixture of its parental populations. Alleles from essentially putative European (e.g. A*25:01:01, B*37:01:01, B*35:08:01), Amerindian (e.g. A*02:22, B*35:43:01, C*03:05), and Sub-Saharan African (e.g. A*02:02, B*15:03:01, DRB1*08:04:01) origin can be found in the CRCV sample.

Based on the results for HLA typing 5-locus haplotype frequency estimations based on the expectation-maximization algorithm were generated. The number of the extended haplotypes with an estimated frequency > 1/2N was 192. Fifteen extended haplotypes have estimated frequencies of > 1%, the most common being A*24:02:01-B*40:02:01-C*03:05-DRB1*08:02:01-DQB1*04:02:01 (2.04%), A*02:01:01-B*07:02:01-C*07:02:01-DRB1*15:01:G-DQB1*06:02:01G (1.81%), A*03:01:01-B*35:01:01-C*04:01:01-DRB1*01:01:01G-DQB1*05:01:01G (1.63%), A*03:01:01-B*07:02:01-C*07:02:01-DRB1*11:01:01G-DQB1*03:01:01G (1.58%), and A*30:04:01-B*45:01-C*06:02:01-DRB1*11:02:01-DQB1*03:01:01G (1.58%). A complete list of estimated extended haplotypes is given in Supplementary Table 1. The putative continental origin of the extended haplotypes shows mixed origins in the CRCV, with a high proportion of extended haplotypes of likely European origin (66%) [8]. Overall, the HLA profile of the CRCV follows admixture proportion estimations obtained with other genetic markers for the Central Valley of Costa Rica [1,4,9,10]. All genotype, as well as haplotype and allele frequency data are available in the Allele Frequencies Net database (AFND) [11] under the population name “Costa Rica Central Valley Mestizo” and the identifier (AFN3606). Haplotype and allele data are available in “G” notation in the supplementary information accompanying this publication. However, the notation “G” is omitted in AFND due to format restrictions, and second-field level resolution is shown for the relevant alleles instead.

Declaration of Competing Interest

The authors have no conflict of interest to declare.

Acknowledgements

This work was supported by grants from University College London, the University of Costa Rica, and the Costa Rican National Council for Scientific and Technologic Research (CONICIT) to EAB.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.humimm.2019.05.006.

References

[1] B. Morera, R. Barrantes, R. Marin-Rojas, Gene admixture in the Costa Rican population, Ann. Hum. Genet. 67 (2003) 71.
[2] L.G. Carvajal-Carmona, R. Ophoff, S. Service, J. Hartiala, J. Molina, P. Leon, et al., Genetic demography of Antioquia (Colombia) and the Central Valley of Costa Rica, Hum. Genet. 112 (2003) 534.
[3] B. Morera, R. Barrantes, Is the Central Valley of Costa Rica a genetic isolate? Rev. Biol. Trop. 52 (2004) 629.
[4] S. Wang, N. Ray, W. Rojas, M.V. Parra, G. Bedoya, C. Gallo, et al., Geographic patterns of genome admixture in Latin American Mestizos, PLoS Genet. 4 (2008) e1000037.
[5] J. Robinson, J.A. Halliwell, J.D. Hayhurst, P. Flicek, P. Parham, S.G. Marsh, The IPD and IMGT/HLA database: allele variant databases, Nucl. Acids Res. 43 (2015) D423.
[6] J.M. Nunes, S. Buhler, D. Roessli, A. Sanchez-Mazas, collaboration HL-n: the HLA-net GENE[RATE] pipeline for effective HLA data analysis and its application to 145 population samples from Europe and neighbouring areas, Tissue Antigens 83 (2014) 307.
[7] H. Maldonado-Torres, J. Robinson, J.A. Madrigal, S.G.E. Marsh, Cactus, a population genetics analysis environment. ‘Genetics and The Immune Response’ Abstracts of the 35th Annual Scientific Meeting of the Australasian Society for Immunology and 14th International HLA & Immunogenetics Workshop, Tissue Antigens 66 (2005) 486.
[8] E. Arrieta-Bolanos, J.J. Madrigal-Sanchez, J.E. Stein, P. Orlich-Perez, M.J. Moreira-Espinoza, E. Paredes-Carias, et al., High-resolution HLA allele and haplotype frequencies in majority and minority populations of Costa Rica and Nicaragua: differential admixture proportions in neighboring countries, HLA 91 (2018) 514.
[9] R. Campos-Sanchez, H. Raventos, R. Barrantes, Ancestry informative markers clarify the regional admixture variation in the Costa Rican population, Hum. Biol. 85 (2013) 721.
[10] M. Segura-Wang, H. Raventos, M. Escamilla, R. Barrantes, Assessment of genetic ancestry and population substructure in Costa Rica by analysis of individuals with a familial history of mental disorder, Ann. Hum. Genet. 74 (2010) 516.
[11] E. Santos, A. McCabe, F.F. Gonzalez-Galarza, A.R. Jones, D. Middleton, Allele Frequencies Net Database: improvements for storage of individual genotypes and analysis of existing data, Hum. Immunol. 77 (2016) 238.