Title

Sequence-based HLA-A, B, C, DP, DQ, and DR typing of 496 adults from San Diego, California, USA

Journal

Human Immunology, 79(12)

ISSN

0198-8859

Authors

Moore, Eugene
Grifoni, Alba
Weiskopf, Daniela
et al.

Publication Date

2018-12-01

DOI

10.1016/j.humimm.2018.09.008

Peer reviewed

Published in final edited form as:

Hum Immunol. 2018 December ; 79(12): 821–822. doi:10.1016/j.humimm.2018.09.008.

Sequence-based HLA-A, B, C, DP, DQ, and DR typing of 496 adults from San Diego, California, USA

Eugene Moore[a,*], **Alba Grifoni**[a], **Daniela Weiskopf**[a], **Veronique Schulten**[a], **Cecilia S. Lindestam Arlehamn**[a], **Michael Angelo**[a], **John Pham**[a], **Shay Leary**[b], **John Sidney**[a], **David Broide**[e], **April Frazier**[a], **Elizabeth Phillips**[b,c], **Simon Mallal**[b,c], **Steven J. Mack**[d], and **Alessandro Sette**[a,e]

[a]Division of Vaccine Discovery, La Jolla Institute for Allergy & Immunology, La Jolla, CA 92037, USA

[b]Institute for Immunology and Infectious Diseases, Murdoch University, Perth, Western Australia 6150, Australia

[c]Vanderbilt University School of Medicine, Nashville, TN 37232, USA

[d]Children's Hospital Oakland Research Institute, Oakland, CA 94609, USA

[e]Department of Medicine, University of California San Diego, La Jolla, CA 92093, USA

Abstract

DNA sequence-based typing at the HLA-A, -B, -C, -DPB1, -DQA1, -DQB1, and -DRB1 loci was performed on 496 healthy adult donors from San Diego, California, to characterize allele frequencies in support of studies of T cell responses to common allergens. Deviations from Hardy Weinberg proportions were detected at each locus except A and C. Several alleles were found in more than 15% of individuals, including the class II alleles DPB1*02:01, DPB1*04:01, DQA1*01:02, DQA1 *05:01, DQB1*03:01, and the class I allele A*02:01. Genotype data will be available in the Allele Frequencies Net Database (AFND 3562).

Keywords

HLA alleles; HLA typing; San Diego; California; Allergy

San Diego, located on the western coast of the United States, is California's second largest city and third largest metropolitan area, boasting a population of over 3 million (2010 census). The principal economies are based in military and defense, tourism, manufacturing, pharmaceutical research and development, and health and bio-technology research centers.

The Greater San Diego metropolitan area has a multi-ethnic population, of which the largest fraction (about 45%; US Census Bureau 2010; www.census.gov, San Diego County, CA) is represented by White, non-Hispanics. Closely following is Hispanic or Latino, comprising

*Corresponding author. emoore@lji.org (E. Moore).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.humimm.2018.09.008.

about 34% of the population. Other significant demographic representations are Asian (12.5%), and African-American (5.5%). Though the indigenous Kumeyaay people are cited to have occupied the area for about 12,000 years, indigenous people have only a minor presence in present day San Diego with an estimated Native American population of about 1.3% [1]. The principal language spoken publicly throughout San Diego County is English (American) while Spanish, Asian and Pacific Islander languages comprise approximately 80% of languages spoken in households that speak languages other than English at home (www.cpehn.org).

Blood donations from 496 adults were obtained from the University of California, San Diego (UCSD), General Clinical Research Center (GCRC), the Clinical Studies Core at the La Jolla Institute for Allergy and Immunology (LJI) and the San Diego Blood Bank as previously described [2,3]. Donors were recruited for studies examining immune responses to various common allergens, and included non-allergic controls. Donors were of both sexes, from the general population, and between 18 and 60 years old. Because many of the donors were sourced from a university and may not be representative of the more static local population, we analyzed the ethnicities of the donor population and compared them to both those of UCSD's undergraduate enrollment and the local population. The San Diego cohort consisted of a population of 50% White non-Hispanic, 12% Hispanic/Latino, 22% Asian, 6.3% African American and 1.5% Native American donors with 8% not reporting. The cohort ethnicities do not comport with the reported ethnicities of the UCSD student body population, suggesting our cohort consists of a mix of both local and UCSD faculty/student populations.

In accordance with local policies, samples collected from recruited donors were subject to human subject review and, accordingly, the institutional review board (IRB) of LJI approved all protocols. Peripheral blood mononuclear cells (PBMCs) and serum were purified by density gradient centrifugation (Ficoll-Paque Premium, GE Healthcare Biosciences, Kowloon, Hong Kong). Cells were then re-suspended in fetal bovine serum (FBS, Gemini Bio-Products, Sacramento, California, USA), containing 10% dimethyl sulfoxide (DMSO, Sigma Life Science) and cryopreserved in liquid nitrogen [2].

HLA-A, -B, -C, -DPB1, -DQA1, -DQB1, and -DRB1 genotyping using locus-specific PCR amplification on genomic DNA was performed for 292 donors by an American Society for Histocompatibility and Immunogenetics (ASHI) accredited laboratory at The Institute for Immunology and Infectious Diseases (IIID) at Murdoch University, Western Australia. The assay was adapted from a previously published protocol for Barcoded-PCR method [4] with modifications to the primer sequences (Supplemental Table I), and performed with genomic DNA isolated from donor PBMCs [5]. Reads were quality-filtered, separated by MID tags and alleles called using an in-house accredited HLA allele caller software pipeline that minimizes the influence of sequencing errors. Alleles were called using the IMGT HLA allele database v.3.21.0 (www.ebi.ac.uk/ipd/imgt/hla) as the reference library [6]. Ambiguities were resolved during the original typing using a proprietary allele-calling algorithm and analysis pipeline and the latest IMGT HLA allele database.

Four-digit HLA typing was performed for the remaining 204 donors at LJI as detailed previously [7], using either high-resolution Luminex-based typing or next-generation sequencing (NGS). Briefly, High resolution Luminex-based typing was performed according to the manufacturer's instructions (Sequence-Specific Oligonucleotides (SSO) typing; One Lambda, Canoga Park, CA). Where indicated, PCR based methods were used to provide high resolution sub-typing. (Sequence-Specific Primer (SSP) typing; One Lambda, Canoga Park, CA). Genomic DNA was isolated from PBMCs using standard techniques and amplified using whole-genome amplification (REPLI-g; Qiagen). Amplicons for HLA class I and class II genes were generated using PCR and locus-specific primers. Amplicons of the correct size were purified using Zymo DNA Clean-up Kit, according to the manufacturer's instructions (Zymo Research, Irvine, CA). Sequencing libraries were prepared using Nextera XT reagents (Illumina), according to manufacturer's instructions. The libraries were purified using AMPure XP (Beckman Coulter) with a ratio of 0.5:1 beads to DNA (v/v). The libraries were pooled in equimolar amounts and loaded at 5.4 pM on one MiSeq flowcell containing 1% phiX spike (MiSeq Reagent Kit v3). Paired-end sequencing was performed with 300 cycles in each direction. HLA typing calls were made using HLATyphon (https://github.com/LJI-Bioinformatics/HLATyphon).

For historical reasons, not all donors were typed at all loci. Of the 496 donors reported in this study, 192 donors were typed at all 7 loci. In total, the number of donors typed at the A, B, C, DPB1, DQA1, DQB1, and DRB1 loci were 310, 312, 312, 495, 364, 492 and 496, respectively. In order to provide consistent and compatible genotyping data, all donor genotypes are being reported to four-digit resolution despite the increased resolution afforded by NGS-based typing. In order to validate new methods as they were adopted, some donor samples were retyped and confirmed.

Allele frequencies for each locus were determined by direct counting (Supplemental Table II). The most frequent specificities (gene frequency > 0.15) were the class II alleles DPB1*02:01, DPB1*04:01, DQA1*01:02, DQA1*05:01 and DQB1*03:01; and the class I allele A*02:01. The most frequently occurring alleles at other loci were Class II DRB1*07:01 (0.122) and DRB1*15:01 (0.113); and Class I B*08:01 (0.072) and C*04:01 (0.146). No significant frequency differences were noted in any of the cohort subsets collected for any of several different allergy research projects, including control donors [8].

Haplotype frequencies (i.e. A~B~C~DPB1~DQA1~DQB1~DRB1, Supplemental Table III) for the subset of 192 donors typed at all loci were estimated using an iterative Expectation-Maximization (EM) algorithm implemented in BIGDAWG (version 1.16) [9]. From a total of 322 unique haplotypes identified, *A*01:01-B*08:01-C*07:01-DPB1*04:01-DQA1*05:01-DQB1*02:01-DRB1*03:01* (frequency 2.4%) was the most common. Three other haplotypes were also present with frequencies greater than or equal to 1.3%.

Deviation from Hardy-Weinberg equilibrium proportions ($p < 0.05$) was detected at each locus, except A and C. The number of unique HLA-DPB1, -DQA1, -DQB1 and -DRB1 alleles was 38, 23, 24, and 56 respectively, and the number of unique HLA-A, -B and -C alleles was 39, 78, 29, respectively (see Supplemental Table II). In total, 286 unique alleles were identified. Because the donors used for the study described herein were drawn from

multiple populations over a period of 10 years, the Hardy-Weinberg deviations are likely the result of the highly admixed nature of the combined population. As a result, both genotype frequencies and haplotype frequencies should be used with caution in any comparative population studies.

The frequency and genotype data are available in the Allele Frequency Net Database (http://allelefrequencies.net), population USA San Diego (AFND population identifier 3562) [10].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

This work was supported by the U.S. National Institutes of Health (NIH) under the following contracts and grants: HHSN272200700048C, HHSN27220140045C, U19AI135731, UM1AI114271, R01AI077653 and U19AI100275.

References

[1]. Erlandson JM, Rick Torben C., Jones Terry L., Porcasi Judith F., One if by land, two if by sea: who were the first Californians? in: Jones TL, Klar Kathryn A. (Eds.), California Prehistory: Colonization, Culture and Complexity, AltaMira Press, Lanham, MD, 2010, p. 53.

[2]. Weiskopf D, Angelo MA, de Azeredo EL, Sidney J, Greenbaum JA, Fernando AN, et al., Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells, Proc. Natl. Acad. Sci. U.S.A. 110 (2013) E2046. [PubMed: 23580623]

[3]. Oseroff C, Sidney J, Kotturi MF, Kolla R, Alam R, Broide DH, et al., Molecular determinants of T cell epitope recognition to the common Timothy grass allergen, J. Immunol. 185 (2010) 943. [PubMed: 20554959]

[4]. Erlich RL, Jia X, Anderson S, Banks E, Gao X, Carrington M, et al., Next-generation sequencing for HLA typing of class I loci, BMC Genomics 12 (2011) 42. [PubMed: 21244689]

[5]. Grifoni A, Sidney J, Carpenter C, Phillips E, Mallal S, Scriba TJ, et al., Sequence-based HLA-A, B, C, DP, DQ, and DR typing of 159 individuals from the Worcester region of the Western Cape province of South Africa, Hum. Immunol. 79 (2018) 143. [PubMed: 29398200]

[6]. Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SG, The IPD and IMGT/HLA database: allele variant databases, Nucleic Acids Res. 43 (2015) D423. [PubMed: 25414341]

[7]. McKinney DM, Fu Z, Le L, Greenbaum JA, Peters B, Sette A, Development and validation of a sample sparing strategy for HLA typing utilizing next generation sequencing, Hum. Immunol. 76 (2015) 917. [PubMed: 26027778]

[8]. Paul S, Arlehamn CSL, Schulten V, Westernberg L, Sidney J, Peters B, et al., Experimental validation of the RATE tool for inferring HLA restrictions of T cell epitopes, BMC Immunol. 18 (2017) 20. [PubMed: 28681704]

[9]. Pappas DJ, Marin W, Hollenbach JA, Mack SJ, Bridging immunogenomic data analysis workflow gaps (BIGDAWG): an integrated case-control analysis pipeline, Hum. Immunol. 77 (2016) 283. [PubMed: 26708359]

[10]. Gonzalez-Galarza FF, Takeshita LY, Santos EJ, Kempson F, Maia MH, da Silva AL, et al., Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations, Nucleic Acids Res. 43 (2015) D784. [PubMed: 25414323]