Human Immunology 80 (2019) 157–162

Contents lists available at ScienceDirect

Human Immunology

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

Exploring the ancestry and admixture of Mexican Oaxaca Mestizos from Southeast Mexico using next-generation sequencing of 11 *HLA* loci

B.A. González-Quezada[a,b], L.E. Creary[c], A.J. Munguia-Saldaña[a], H. Flores-Aguilar[a,b], M.A. Fernández-Viña[c], C. Gorodezky[a,b,*]

[a] Dept. of Immunology & Immunogenetics, InDRE, Secretary of Health, Mexico City, Mexico
[b] Fundación Comparte Vida, A.C., Mexico City, Mexico
[c] Department of Pathology, Stanford University School of Medicine, Palo Alto, CA, USA

ARTICLE INFO

Keywords:
NGS
High throughput *HLA* genotyping
Mestizos from Oaxaca
Mexicans

ABSTRACT

The Mestizos of Oaxaca resulted from the admixture of Zapotecan Natives with Spaniards and Africans. We selected 112 donors from Oaxaca and applied next-generation sequencing to characterize exon and intron variants in complete or extended *HLA* genes. Some alleles found, are unique to Mexican Natives and most likely will be absent in most major ethnicities, namely: Caucasians, Africans or Asians. Among these are *HLA-A*68:03:01, HLA-A*68:05:01, HLA-C*03:04:01:02, HLA-C*15:09, HLA-C*3:05, HLA-C*03:06:01, HLA-B*39:05:01, HLA-B*35:14:01, HLA-B*35:12:01, HLA-B*35:43:01, HLA-B*40:05, HLA-B:40:08, HLA-B*51:02:01, HLA-B*35:24:01* and *HLA-B*39:08*. *HLA-DQA1*05:05:01:05* and some *HLA-DRB1* alleles were only present in Amerindians/Mestizos. Three haplotypes are unique to Mexican Natives, five to Middle-Eastern and Sephardi-Jews. We detected a novel *HLA-DQA1*04:01:01* exon 4 variant. Any novel allele may have been positively selected to enlarge the peptide-binding repertoire, and some, like *HLA-B*39:02:02* and *HLA-B*39:05:01* were found with unique haplotype associations, suggesting convergent evolution events and/or allele lineage diversification. The allele frequencies were fairly evenly distributed in most *HLA* loci with the exception of *HLA-DPB1*. The application of NGS in Oaxaca is novel and will lead to better use in the clinical setting. It offers deep knowledge on the population structure, origins, migration, and discovery of new alleles and haplotypes that other techniques did not achieve.

1. Introduction

Human Leucocyte Antigen (*HLA*) genotyping is beyond doubt, fundamental in the immunogenetics and histocompatibility field [1]. In the past, the *HLA* genotyping methods have evolved from serology to molecular techniques based on polymerase chain reaction (PCR) such as sequence specific primers (SSP), sequence specific oligonucleotide probes (SSOP) and sequence-based typing (SBT) which usually includes amplification and sequencing of exons 2–3 for classic class I and exon 2 for class II genes [2,3]. Since the *HLA* complex is the most polymorphic system of the human genome [4] currently with 20,088 variants described in the IPD-IMGT/HLA Database [5], it has become very challenging for the genotyping technology [1]. Therefore, Next-generation sequencing (NGS) methods developed for high throughput *HLA* genotyping have been emerging [2,3]. NGS allows the clonal massively parallel sequencing of a single DNA molecule using long-range PCR that produces large amounts of reads of each *HLA* gene at a lower cost and saving time, since many samples may be typed in the same run [1,4]. The application of this method worldwide through the use of several commercial kits or in-house methods has revealed the extensive allele diversity of *HLA* on a global scale and has allowed the constant discovery of novel alleles. In addition, NGS has improved the accuracy of *HLA* genotyping and characterization of haplotypes which is beneficial for donor selection worldwide, as well as, other areas of clinical

Abbreviations: AF, Allele frequency; DONORMO, The Mexican Unrelated Donors Registry; EWH, Ewens-Waterson Homozygosity; HLA, Human Leukocyte Antigens; HF, haplotype frequency; HWE, Hardy-Weinberg Equilibrium; LD, linkage disequilibrium; NGS, Next-Generation Sequencing; PCR, Polymerase Chain Reaction; SBT, sequence-based typing

* Corresponding author at: Francisco P. Miranda 177, Laboratory Building, 2nd floor, Col. Lomas de Plateros, Mexico City 01480 CDMX, Mexico.
E-mail addresses: bgonzalez@fundacioncompartevida.org.mx (B.A. González-Quezada), lcreary@stanford.edu (L.E. Creary), orionica@yahoo.com (A.J. Munguia-Saldaña), hflores@fundacioncompartevida.org.mx, hilario.flores@salud.gob.mx (H. Flores-Aguilar), marcelof@stanford.edu (M.A. Fernández-Viña), clarag@unam.mx, clara.gorodezky@salud.gob.mx, cgorodea@fundacioncompartevida.org.mx (C. Gorodezky).

https://doi.org/10.1016/j.humimm.2019.01.004
Received 20 June 2018; Received in revised form 3 December 2018; Accepted 17 January 2019
Available online 29 January 2019

Table 1
Allele frequencies of classic *HLA* class I loci: *HLA-A, HLA-B* and *HLA-C*.

HLA-A (k = 28)			*HLA-C* (k = 29)			*HLA-B* (k = 48)		
Allele	**Counts**	Freq	Allele	Counts	Freq	Allele	Counts	Freq
A*02:01:01:01	51	0.2440	C*07:02:01:01	50	0.2347	**B*39:05:01†**	21	0.0963
A*24:02:01:01	38	0.1818	C*04:01:01:01	39	0.1831	B*35:01:01:02	19	0.0872
A*02:06:01:01	24	0.1148	C*01:02:01	16	0.0751	B*39:02:02	18	0.0826
A*68:03:01†	13	0.0622	C*07:01:01:01	14	0.0657	B*39:06:02	11	0.0505
A*68:01:02:01	12	0.0574	C*03:03:01	10	0.0469	B*35:17:01	10	0.0459
A*01:01:01:01	11	0.0526	**C*03:04:01:02†**	10	0.0469	B*51:01:01:01	10	0.0459
A*31:01:02:01	11	0.0526	C*07:02:01:03	8	0.0376	B*07:02:01	9	0.0413
A*11:01:01:01	6	0.0287	C*08:01:01	8	0.0376	**B*52:01:02†**	8	0.0367
A*32:01:01	6	0.0287	C*02:02:02:01	6	0.0282	B*08:01:01	7	0.0321
A*02:05:01	5	0.0239	C*06:02:01:01	5	0.0235	**B*35:14:01†**	7	0.0321
A*03:01:01:01	4	0.0191	C*06:02:01:03	5	0.0235	B*40:02:01	6	0.0275
A*23:01:01	4	0.0191	C*12:03:01:01	5	0.0235	B*44:03:01:01	6	0.0275
A*30:01:01	4	0.0191	**C*15:09†**	5	0.0235	B*15:01:01:01	5	0.0229
A*29:02:01:02	3	0.0144	C*16:01:01:01	5	0.0235	B*45:01:01	5	0.0229
A*68:05:01†	3	0.0144	C*03:04:01:01	4	0.0188	B*48:01:01	5	0.0229
A*30:02:01:03	2	0.0096	**C*03:05†**	4	0.0188	B*27:05:02	4	0.0183
A*33:03:01	2	0.0096	C*08:02:01:01	4	0.0188	B*39:01:01:03	4	0.0183
A*02:02:01:01	1	0.0048	C*07:18	3	0.0141	B*49:01:01	4	0.0183
A*02:02:01:02	1	0.0048	C*12:02:02	2	0.0094	B*57:01:01	4	0.0183
A*02:17:02	1	0.0048	C*15:02:01:01	2	0.0094	B*14:02:01:01	4	0.0183
A*25:01:01	1	0.0048	C*03:02:02:01	1	0.0047	B*15:15	3	0.0138
A*29:02:01:01	1	0.0048	**C*03:06:01†**	1	0.0047	B*15:30	3	0.0138
A*30:02:01:02	1	0.0048	C*04:01:01:06	1	0.0047	B*18:01:01:02	3	0.0138
A*33:01:01	1	0.0048	C*05:01:01:01	1	0.0047	B*35:03:01	3	0.0138
A*36:01	1	0.0048	C*05:01:01:02	1	0.0047	**B*35:12:01†**	3	0.0138
A*68:01:01:02	1	0.0048	C*07:27:01	1	0.0047	**B*35:43:01†**	3	0.0138
A*68:05:01	1	0.0048	C*16:02:01	1	0.0047	B*38:01:01	3	0.0138
A*69:01	1	0.0048	C*17:03	1	0.0047	**B*40:05†**	3	0.0138
						B*53:01:01	3	0.0138
						B*58:01:01:01	3	0.0138
						B*13:02:01	2	0.0092
						B*40:08†	2	0.0092
						B*41:01:01	2	0.0092
						B*51:02:01†	2	0.0092
						B*52:01:01:02	2	0.0092
						B*15:48	1	0.0046
						B*18:01:01:01	1	0.0046
						B*35:02:01	1	0.0046
						B*35:24:01†	1	0.0046
						B*35:241	1	0.0046
						B*39:08†	1	0.0046
						B*41:02:01	1	0.0046
						B*44:02:01:01	1	0.0046
						B*51:168	1	0.0046
						B*55:01:01	1	0.0046
						B*57:03:01	1	0.0046

† Alleles found in Mexican Natives and other Amerindians and most likely absent in the major ethnic groups: Caucasians, Africans and Asians.

practice [1,3,6].

Numerous studies have shown the importance of accurate *HLA* genotyping for successful transplantation outcomes, for disease association studies, pharmacogenetics and population genetics [4,6]. The *HLA* alleles may be typed up to four fields: according to the official nomenclature described in the IPD-IMGT/HLA Database, the first field before the colon corresponds to the allele group; the second field depicts variation at protein sequence level; the third denotes synonymous nucleotide substitutions in the exons and the fourth field shows differences in the non-coding intron and flanking regions of the gene [7]. In the clinical practices such as transplantation and disease association studies, *HLA* genotyping characterized at the two-field is sufficient. In comparison, molecular anthropology, population diversity, evolution studies, the fourth fields HLA data are relevant [4].

The history of the people from Oaxaca dates back to around 10,000 years BC with the arrival of nomads to the central valleys of the state. The most ancient data of corn seeding dates to 4000 BCE, and later to bean seeding, giving rise to the settlement of the native groups. At that time, linguistic diversification from the Otomangüe started and around 1500 BCE nine linguistic families were present; four of them were Zapotecana, Mixtecana, Popolocana, Chias and the Hanantecana [8,9]. Since then, differentiation occurred and 16 linguistic groups were formed as time went by.

During Spanish colonization, governmental structures and cultural patterns were formed and prevail up to the present day in the Native communities. The Mestizo communities that populated the City of Oaxaca were the result of the admixture of the local Indian groups and the Spanish conquerors that arrived to the city in the 15th Century. These cultures showed an important versatility to adapt themselves to the new changes of the last 50 years. They incorporated in their day to day life, new agricultural, dressing, education, feeding and health activities whilst, keeping their ancestral traditions: features that make the Mestizos from Oaxaca, culturally different from the rest of the Mexicans. The European component came almost entirely from Spain, mainly from Castilla, Extremadura and Andulacía. The Amerindian component that finally settled in Oaxaca corresponds to a very early occupation of Asian descent groups in the Paleolithic period. These Asian groups arrived in Mexico through different waves of immigration

Table 2
HLA-DRB1/3/4/5 allele frequencies.

HLA-DRB3/DRB4/DRB5 (k = 16)			*HLA-DRB1* (k = 36)		
Allele	Counts	Freq	Allele	Counts	Freq
DRB4*01:03:01:01	54	32.7	**DRB1*04:07:01**†	29	0.1408
DRB3*01:01:02:01	34	20.6	**DRB1*08:02:01**†	26	0.1262
DRB5*02:02	20	12.1	**DRB1*16:02:01:02**†	19	0.0922
DRB3*02:02:01:02	15	9.1	DRB1*04:04:01	13	0.0631
DRB3*02:02:01:01	10	6.1	**DRB1*14:06:01**†	13	0.0631
DRB5*01:01:01	10	6.1	DRB1*03:01:01:01	10	0.0485
DRB4*01:03:02	8	4.8	DRB1*04:03:01	8	0.0388
DRB3*03:01:01	7	4.2	DRB1*11:04:01	8	0.0388
DRB4*01:01:01:01	2	1.2	DRB1*13:02:01	8	0.0388
DRB5*01:02	2	1.2	**DRB1*14:02:01**†	8	0.0388
DRB3*02:24	1	0.6	DRB1*13:01:01	7	0.0340
DRB4*01:03:01:02	1	0.6	DRB1*04:05:01	6	0.0291
DRB4*01:03:01:03	1	0.6	DRB1*07:01:01:01	5	0.0243
			DRB1*11:01:01	5	0.0243
			DRB1*04:11:01†	4	0.0194
			DRB1*15:01:01:01	4	0.0194
			DRB1*01:01:01	3	0.0146
			DRB1*13:03:01	3	0.0146
			DRB1*15:01:01:03	3	0.0146
			DRB1*01:02:01	2	0.0097
			DRB1*07:01:01:02	2	0.0097
			DRB1*10:01:01	2	0.0097
			DRB1*13:04	2	0.0097
			DRB1*14:54:01	2	0.0097
			DRB1*15:01:01:02	2	0.0097
			DRB1*15:02:01	2	0.0097
			DRB1*01:03	1	0.0049
			DRB1*04:02:01	1	0.0049
			DRB1*04:08:01	1	0.0049
			DRB1*08:06	1	0.0049
			DRB1*08:10	1	0.0049
			DRB1*09:01:02	1	0.0049
			DRB1*12:01:01	1	0.0049
			DRB1*13:05:01	1	0.0049
			DRB1*14:01:01	1	0.0049
			DRB1*15:03:01:02	1	0.0049

† Alleles found in Mexican Natives and other Amerindians and most likely absent in the major ethnic groups: Caucasians, Africans and Asians.

in different parts of Mexico, giving rise to the 16 different linguistic groups settled in Oaxaca. These groups were mated with the Spanish conquerors resulting in the admixed mestizo population in Oaxaca [9–11].

Due to the fact that there are 69 different ethnicities in Mexico, and each of them is genetically different, and due to admixture from Europe and the Middle east, since over 500 hundred years up to date, it is very important to map thoroughly the different geographical areas of Mexico and to unravel the distinct kind of admixtures [11–13]. We typed 11 *HLA* genes from the well-known anthropologically characterized Oaxaca population using NGS to decipher the allele diversity at the maximum resolution possible, to better understand their *HLA* nucleotide polymorphism, evolution and demographic history. In addition, the NGS *HLA* data generated would be useful for bone marrow transplantation to search for the most appropriate unrelated donors, for patients with this type of ancestry, anywhere in the world.

2. Material and methods

2.1. Samples

Peripheral blood samples collected in EDTA tubes were obtained from 112 Mexican Mestizo healthy Mexican Mestizo subjects, born and living in the city of Oaxaca, that were recruited for the **DONORMO-The Mexican Unrelated Donors Registry**. Of the total, 59% were females and 41% males with a median age of 28.4 and 32.5 years respectively. Each subject, registered as an unrelated donor and signed an informed consent according to the Declaration of Helsinki [14]. None of the donors reported personal history of cancer, or any HLA associated disease, including chronic infections or autoimmune diseases.

2.2. DNA extraction

DNA was extracted of all the blood samples with the Maxwell16 instrument (Promega Corporation, Madison, WI, USA) according to the technical manual.

Table 3
HLA-DQA1, *HLA-DQB1*, *HLA-DPA1* and *HLA-DPB1* allele frequencies.

HLA-DQA1 (k = 23)			*HLA-DQB1* (k = 22)			*HLA-DPA1* (k = 13)			*HLA-DPB1* (k = 26)		
Allele	Counts	Freq	Allele	Counts	Freq	Allele	Counts	Freq	Allele	Counts	Freq
DQA1*03:01:01	51	0.2500	DQB1*03:02:01	52	0.2667	DPA1*01:03:01:05	81	0.4551	DPB1*04:02:01:02	71	38.6
DQA1*04:01:01	25	0.1225	DQB1*03:01:01:01	38	0.1949	DPA1*01:03:01:02	40	0.2247	DPB1*04:01:01:01	30	16.3
DQA1*05:03	21	0.1029	DQB1*04:02:01	26	0.1333	DPA1*02:01:01	19	0.1067	DPB1*02:01:02	19	10.3
DQA1*05:05:01:01	21	0.1029	DQB1*03:01:01:03	15	0.0769	DPA1*01:03:01:03	11	0.0618	DPB1*03:01:01	21	11.4
DQA1*05:05:01:05†	13	0.0637	DQB1*02:01:01	9	0.0462	DPA1*01:03:01:04	9	0.0506	DPB1*04:02:01:01	13	7.1
DQA1*01:03:01:02	8	0.0392	DQB1*06:02:01	8	0.0410	DPA1*02:02:02	5	0.0281	DPB1*01:01:01	6	3.3
DQA1*03:03:01	8	0.0392	DQB1*06:03:01	8	0.0410	DPA1*01:03:01:01	4	0.0225	DPB1*17:01:01	5	2.7
DQA1*05:05:01:02	8	0.0392	DQB1*02:02:01:01	9	0.0462	DPA1*02:01:08	3	0.0169	DPB1*05:01:01	4	2.2
DQA1*01:02:01:04	7	0.0343	DQB1*06:09:01	4	0.0205	DPA1*02:01:02	2	0.0112	DPB1*14:01:01	4	2.2
DQA1*02:01	7	0.0343	DQB1*05:01:01:03	3	0.0154	DPA1*01:04	1	0.0056	DPB1*01:01:02	3	1.6
DQA1*05:01:01:02	7	0.0343	DQB1*05:02:01	3	0.0154	DPA1*02:01:04	1	0.0056	DPB1*104:01	3	1.6
DQA1*01:02:01:03	6	0.0294	DQB1*06:04:01	3	0.0154	DPA1*03:01	1	0.0056	DPB1*10:01:01	3	1.6
DQA1*01:02:01:01	4	0.0196	DQB1*03:01:01:02	2	0.0103	DPA1*04:01	1	0.0056	DPB1*13:01:01	2	1.1
DQA1*01:01:01	3	0.0147	DQB1*03:03:02:01	2	0.0103				DPB1*04:01:01:02	1	0.5
DQA1*01:04:01:02	3	0.0147	DQB1*03:04:01	2	0.0103				DPB1*105:01	1	0.5
DQA1*05:01:01:01	3	0.0147	DQB1*03:19	2	0.0103				DPB1*131:01	1	0.5
DQA1*01:01:02	2	0.0098	DQB1*05:01:01:01	2	0.0103				DPB1*138:01	1	0.5
DQA1*01:03:01:01	2	0.0098	DQB1*05:01:01:02	2	0.0103				DPB1*15:01:01	1	0.5
DQA1*01:05:01	2	0.0098	DQB1*05:03:01:01	3	0.0154				DPB1*40:01	1	0.5
DQA1*01:02:01:02	1	0.0049	DQB1*06:01:01	2	0.0103						
DQA1*04:01:01eExon4Var	1	0.0049									
DQA1*06:01:01	1	0.0049									

† Alleles found in Mexican Natives and other Amerindians and most likely absent in the major ethnic groups: Caucasians, Africans and Asians.

Table 4
The top 20 *HLA* -A-C-B-DRB3/DRB4/DRB5-DRB1-DQA1-DQB1-DPA1-DPB1 haplotypes.

No.	A1-C1-B1-DRB3/4/5-DRB1-DQA1-DQB1-DPA1-DPB1	HF
1	**A*68:03:01**-C*07:02:01:01-**B*39:05:01**-DRB4*01:03:01:01-**DRB1*04:07:01**-DQA1*03:01:01-DQB1*03:02:01-DPA1*01:03:01:05-DPB1*04:02:01:02	0.0205
2	A*02:01:01:01-C*07:02:01:01-B*39:02:02-DRB5*02:02-**DRB1*16:02:01:02-DQA1*05:05:01:05**-DQB1*03:01:01:01-DPA1*01:03:01:05-DPB1*04:02:01:02	0.0164
3	A*24:02:01:01-C*07:02:01:01-B*39:06:02-DRB4*01:03:01:01-**DRB1*14:06:01**-DQA1*03:01:01-DQB1*03:01:01:01-DPA1*01:03:01:05-DPB1*04:02:01:02	0.0164
4	**A*68:03:01**-C*07:02:01:01-**B*39:05:01**-DRB4*01:03:02-**DRB1*04:07:01**-DQA1*03:01:01-DQB1*03:02:01-DPA1*01:03:01:05-DPB1*04:02:01:02	0.0164
5	A*02:01:01:01-C*07:02:01:03-B*07:02:01-DRB5*01:01:01-DRB1*15:01:01:01-DQA1*01:02:01:01-DQB1*05:02:01-DPA1*01:03:01:05-DPB1*04:02:01:02	0.0123
6	A*02:06:01:01-C*07:02:01:01-**B*39:05:01**-DRB4*01:03:01:01-**DRB1*04:07:01**-DQA1*03:01:01-DQB1*03:02:01-DPA1*01:03:01:05-DPB1*04:02:01:02	0.0123
7	A*11:01:01:01-C*02:02:02:01-B*27:05:02-DRB3*02:02:01:02-DRB1*13:01:01:01-DQA1*01:03:01:02-DQB1*06:03:01-DPA1*01:03:01:02-DPB1*104:01	0.0123
8	A*24:02:01:01-C*04:01:01:01-**B*35:14:01**-DRB5*02:02-**DRB1*16:02:01:02-DQA1*05:05:01:05**-DQB1*03:01:01:01-DPA1*01:03:01:05-DPB1*04:02:01:02	0.0123
9	**A*68:03:01-C*03:05**-B*35:01:01:02-DRB4*01:03:01:01-**DRB1*04:07:01**-DQA1*03:01:01-DQB1*03:02:01-DPA1*01:03:01:05-DPB1*04:02:01:02	0.0123
10	A*01:01:01:01-C*07:01:01:01-B*08:01:01:01-DRB*Abs-**DRB1*08:02:01**-DQA1*04:01:01-DQB1*04:02:01-DPA1*01:03:01:05-DPB1*04:02:01:02	0.0082
11	A*01:01:01:01-C*07:01:01:01-B*49:01:01-DRB3*03:01:01-DRB1*13:02:01-DQA1*01:02:01:04-DQB1*06:04:01-DPA1*01:03:01:02-DPB1*04:01:01:01	0.0082
12	A*02:01:01:01-C*04:01:01:01-B*35:17:01-DRB*Abs-**DRB1*08:02:01**-DQA1*04:01:01-DQB1*04:02:01-DPA1*01:03:01:05-DPB1*04:02:01:02	0.0082
13	A*02:01:01:01-C*07:02:01:01-B*39:02:02-DRB5*02:02-**DRB1*16:02:01:02-DQA1*05:05:01:05**-DQB1*03:01:01:01-DPA1*01:03:01:02-DPB1*04:01:01:01	0.0082
14	A*02:01:01:01-C*07:02:01:01-B*39:02:02-DRB4*01:03:01:01-DRB1*04:04:01-DQA1*03:01:01-DQB1*03:02:01-DPA1*01:03:01:05-DPB1*04:02:01:02	0.0082
15	A*02:01:01:01-C*07:02:01:01-**B*39:05:01**-DRB4*01:03:01:01-**DRB1*04:07:01**-DQA1*03:01:01-DQB1*03:02:01-DPA1*01:03:01:05-DPB1*04:02:01:02	0.0082
16	A*02:06:01:01-C*07:02:01:01-B*39:02:02-DRB*Abs-**DRB1*08:02:01**-DQA1*04:01:01-DQB1*04:02:01-DPA1*01:03:01:05-DPB1*04:02:01:02	0.0082
17	A*23:01:01-C*07:01:01:01-B*49:01:01-DRB4*01:03:01:01-DRB1*04:05:01-DQA1*03:03:01:01-DQB1*03:02:01-DPA1*02:01:01:01-DPB1*17:01	0.0082
18	A*24:02:01:01-**C*03:04:01:02**-B*39:01:01:03-DRB4*01:03:01:01-DRB1*04:04:01-DQA1*03:01:01-DQB1*03:02:01-DPA1*01:03:01:05-DPB1*04:02:01:02	0.0082
19	A*24:02:01:01-C*07:02:01:01-**B*39:05:01**-DRB5*02:02-**DRB1*16:02:01:02-DQA1*05:05:01:05**-DQB1*03:01:01:01-DPA1*01:03:01:05-DPB1*04:02:01:02	0.0082
20	A*24:02:01:01-C*07:02:01:01-B*39:06:02-DRB3*01:01:02:01-**DRB1*14:06:01**-DQA1*05:03-DQB1*03:01:01:01-DPA1*01:03:01:05-DPB1*04:02:01:02	0.0082

Note: The bold alleles are Mexican Amerindian alleles.

2.3. HLA NGS genotyping

Genotyping at high resolution for 11 loci: HLA class I (*HLA-A*, *HLA-C*, *HLA-B*) and class II (*HLA-DRB1*, *HLA-DRB3*, *HLA-DRB4*, *HLA-DRB5*, *HLA-DQA1*, *HLA-DQB1, HLA-DPA1, HLA-DPB1*) loci was performed using the MIA FORA NGS low-throughput semi-automated typing protocol (Immucor, Inc., Norcross, GA, USA) according to the manufacturer's instructions [15]. The PerkinElmer NGS Express (PerkinElmer, Waltham, MA, USA) and Biomek 4000 liquid handling workstations (Beckman Coulter, Inc., Indianapolis, IN, USA) were used to perform all the PCR and most of the post-PCR library operations respectively. Briefly, long range PCR was performed with full gene coverage for *HLA-A, HLA-B and HLA-C* (greater than200 bp 5′UTR to 3′UTR ~200–400 bp), *HLA-DQA1* (~200 bp of the 5′UTR to ~200 bp of the 3′UTR) and *HLA-DQB1* (~70 bp of the 5′UTR to ~100 bp of the 3′UTR) genes. For *HLA-DPA1* coverage ranged from exon 1 through exon 4 and for *HLA-DPB1* from exon 2 to exon 4. The coverage for *HLA-DRB1/3/4* loci included the 5′UTR to the first −270 bp of intron-1 and end of intron-1 (−250 bp) to exon-6. Amplification was achieved for the *HLA-DRB5* gene exons 2 to exon 6. Each PCR containing 100 ng of genomic DNA for each *HLA* locus was amplified using Veriti Thermal Cyclers (Applied Biosystems/Thermo Fisher Scientific, Waltham, MA, USA). PCR products were quantified using a Victor X plate reader (Perkin Elmer, Waltham, MA, USA) with a PicoGreen assay (Invitrogen/Thermo Fisher Scientific, Waltham, MA, USA). Amplicons for each sample were pooled and purified using Agencourt AMPure XP beads (Beckman Coulter, Fullerton, CA, USA). The barcoded library construction was performed by enzymatic cleavage (generating 300–500 bp fragments), enzymatic end repair, purification (using Agencourt AMPure beads) and ligation of unique adaptor indices to each pooled sample. All adaptor ligated samples were pooled into a single tube, purified (using Agencourt AMPure XP beads) and 400–500 bp DNA fragments were size selected using the Blue Pippin system (Sage Science, Inc., Beverly, MA, USA). The eluted pooled samples were enriched using Illumina primers by a short PCR cycle and purified (using Agencourt AMPure XP beads). The quality of the pooled samples were checked using the Agilent 2200 Tape Station instrument (Agilent Technologies, Inc., Santa Clara, CA, USA). The library was quantified using the Qubit™ dsDNA BR Assay kit and the Qubit Fluorometer (ThermoFisher Scientific, Waltham, MA, USA). The samples were denatured with sodium hydroxide, and sequenced at a final concentration of 12 pM spiked with 0.2% PhiX Control v3 (Illumina, Inc., San Diego, CA, USA) on the Illumina MiSeq instrument using 300 cycle paired-end kits (Illumina, Inc., San Diego, CA, USA).

2.4. HLA sequence data analysis

FASTQ files generated by the Illumina instrument were analyzed with the MIA FORA FLEX v3.0 alignment software (Immucor, Norcross, GA, USA) assign *HLA* genotypes. The MIA FORA software demultiplexes FASTQ files according to each unique index and uses two complementary informatics algorithms; mapping of sequences to references and de novo assembly to construct one or two phased consensus sequences.

2.5. NGS of HLA ambiguities analysis

A great advantage of using NGS for *HLA* genotyping instead of SBT is the capability of assign unambiguous results at highest resolution. However, some ambiguities were also found due to; (i) the presence of short tandem repeats (STR) in introns of some class II genes that cannot be assessed accurately by the sequencing methodology, (ii) the lack of genomic coverage, (iii) phasing ambiguities due to low intronic variation across intron 1 of the *HLA-DPB1* gene. Due to these limitations, indistinguishable allele groups at the 4-field resolution were merged to the lowest numbered allele according to IPD-IMGT/HLA Database v3.25.0.

2.6. Population statistics

Allele frequency (AF) at each locus, Hardy-Weinberg Equilibrium (HWE), Ewens-Waterson Homozygosity (EWH) test of neutrality [16], Haplotype frequency estimation and linkage disequilibrium (LD) locus and allele level were performed using the PyPop version 0.7.0 software [17]. Standardized LD values range between 0 (which means equilibrium), and 1 (linkage). Extended *HLA* haplotype frequencies (including/or not including the *HLA-DPA1* and *HLA-DPB1* loci) was estimated using an Expectation algorithm implemented in the Hapl-o-Mat, v. 1.1 software [18]. The authors have followed the STREIS statement and its checklist [19].

3. Results

Hardy Weinberg probability values and Ewens-Waterson

Homozygosity EWH test of neutrality at *HLA* loci, showed no deviation from HWE for any locus tested (results not shown). Table 1 depicts the AF of class I alleles *HLA-A, HLA-B and HLA-C*, showing that the most frequent alleles frequencies greater than 5% are alleles of putative Amerindian origin. The lowest frequent alleles are of putative Mediterranean and African origin. In Table 2, the allele frequencies of *HLA-DRB3/DRB4/DRB5* and *HLA DRB1* are listed. Table 3 includes the allele distribution frequencies of *HLA-DQA1, HLA-DQB1, HLA-DPA1* and *HLA-DPB1*. Finally, in Table 4, the 20 most frequent haplotypes estimated are shown.

It is important to mention that a novel allele at *HLA-DQA1* locus was detected: an exon 4 variant in allele *HLA-DQA1*04:01:01* (codon 200C > T, Leu > Leu).

4. Discussion

We selected in this study a very interesting Mexican population located in Southeast Mexico with admixture of diverse Amerindian groups admixed with Mediterranean and African components [10,11]. According to the Institute of Anthropology, the status of Mexican Mestizos is defined to be tested exploring with each individual, his ethnic background back to the third generation (parents, grandparents and grand grandparents) should have Hispanics last name and must have been born in Mexico with Mediterranean ancestry (mainly Spain), the African component which came from the African slaves brought by the conquerors from West Africa mainly Bantu and the San Thomé Island [20–26]. As shown from the results of alleles and haplotypes frequencies, (Tables 1 through to 4), the *HLA* diversity of the Mestizos of Oaxaca is very high. One third of the population inhabiting the whole state speaks any of the 16 different native languages; they maintain their unique social structure, their cultural features, religious habits and their identity that is probably stronger than in other Mexican native groups. The name of the city is Oaxaca de Juárez, originally named by the Spaniards, Antequera, was founded in 1529, and was occupied by a group of Zapotecan Indians. The city is an example of the colonialism of the XVI Century and Spaniards brought with them the complex Mediterranean component as well as the Middle-Eastern one, represented by Arabs and, Sephardi Jews these different gene ancestries are evident in the alleles shown in Tables 1, 2 and 3 and in the haplotypes (Table 4) numbered as 5,7, 10, 11, 14, 17, and 18 [11,27,28]. By analyzing deeply the nucleotide diversity of 11 *HLA* loci in a very well-known admixed population of the Southeast, using the high throughput technology of NGS, it was fascinating to understand the different ancestral ethnic contributions to the Oaxaca population. It can be seen that 27, 46 and 29 different alleles were detected for *HLA-A*, *HLA-C* and *HLA-B* loci respectively, and the Mexican Amerindian background was clear in the most frequent alleles; although Semitic and Mediterranean components are also present. The most frequent five *HLA-A*, *HLA-C* and *HLA-B* alleles are of Mexican Amerindian origin such as *HLA-B*35:17:01* being a typical example [29]. An interesting example is the *HLA-B*39* allele and haplotype diversification, shown in these Mestizos. Of the top 20 haplotypes, 12 have different combinations with *HLA-B*39*, suggesting that convergent evolution events and allele lineage diversification may have occurred. Also possibly recombination, selection and bottleneck effects, as well as different admixture events occurred that gave rise to the current haplotype structure. We observe in Table 4, that the first 4 haplotypes that include *HLA-B*39:05:01*, *HLA-B*39:02:02, HLA-B*39:06:02*, *HLA-B*39:05:01* and the 6th most frequent haplotype that bears *HLA-B*39:05:01*, have different *HLA-A* locus alleles except two; almost all have different *HLA-DRB1* alleles. These haplotypes are present in Native Americans, Hispanics, Asians and some of them mainly in Mexican Mestizos. Most of them have *HLA-C*07:02:01:01* but interestingly, the 18th haplotype is a combination of a class I haplotype present in Mexicans and class II haplotype found in Amerindians but also in Slavic populations such as Russians and Polish people [30]. These findings which suggest that diversification may have occurred due to the admixture with different groups that migrated from East Asia in diverse waves at different times in the past. The *HLA-DRB1* ($k = 34$), *HLA-DQA1* ($k = 20$) and *HLA-DQB1* ($k = 21$) alleles with frequencies over 5% are found in some Mexican Amerindian groups as well as in Hispanics from many countries in Latin-America and from USA, but, other ethnic influences are also present, such as Asian, Jewish, African and Mediterranean [30]. The most frequent *HLA-DPA1* ($k = 14$) and *HLA-DPB1* ($k = 16$) with frequencies over 5% are the ancestral genes found in most populations [30].

As shown by the recent study by Goeury et al [6], it is clear that NGS is a powerful technique that has led to the fast discovery of new alleles between April 2016 and April 2017. The knowledge of population diversity is beyond doubt increasing our knowledge on origins, migration routes, admixture, genetic drift or bottleneck effects, We detected class II alleles that were never shown before in our Mestizos such as *HLA-DRB1*08:10*, *HLA-DRB1*04:08:01*, *HLA-DPB1*131:01* and *HLA-DPB1*105:01* although their cumulated frequency is less than 2%.

The analysis of the haplotypes, the combinations of class I and class II haplotypes and alleles present, gives new insights into the possible mechanisms that drove the evolution and actual polymorphisms of this ethnic group. The extended 11 loci haplotypes (table 4, it becomes evident that the haplotypes 2nd; 8th; 14th; 15th; 16th; 19th are found in Mexican Mestizos, in Hispanics populations worldwide and in Oaxaca, but the 1st and 4th are only present in the Oaxaca group.

Moreover, there are other haplotypes shown in populations different than these Mestizos, such as Asians and Mediterraneans and of course, other Hispanics [31–33]. Undoubtedly additional level of diversity was identified at non-coding segments and a high level of heterozygosity at all loci, with the exception of *HLA-DPA1* and *HLA-DPB1* loci. We did show unique alleles and haplotypic associations and predominant native alleles, not found in other ethnic groups.

Finally, in our small sample we observed different haplotypes that differentiated at multiple loci due to recombination and convergent evolution such seen in as the first four frequent haplotypes in Table 4. The first haplotype was class I of Native American putative origin, combined with a class II Hispanic and Native American origin. The second one was of Asian origin, combined with a class II Native American because of the ancestry of Asian groups present in Natives.

In conclusion, NGS permits the characterization of rare genetic variants that we could not detect using traditional *HLA* molecular typing methods. This technology characterizes more than one common allele associated with the same protein. A deep knowledge of population structure will enhance our understanding of the remarkable *HLA* genetic system. The new information revealed by NGS will open new roads to explore the biological meaning of the *HLA* polymorphism and its role in our innate and adaptive system. Furthermore, we will be able to know how protection against infections has developed, and which HLA molecular or immunological mechanisms trigger autoimmune diseases and cancer.

Declaration of interest

none

Acknowledgements

We acknowledge sincerely the contribution of the donors that participated in this study.

Funding

This work was supported by The Fundación Comparte Vida, A.C., Mexico and by the National Institutes of Health; grant U19AI119350, USA.

References

[1] R.P. Milius, M. Heuer, D. Valiga, K.J. Doroschak, C.J. Kennedy, Y.-T. Bolon, J. Schneider, J. Pollack, H.R. Kim, N. Cereb, J.A. Hollenbach, S.J. Mack, M. Maiers, Histoimmunogenetics Markup Language 1.0: Reporting next generation sequencing-based HLA and KIR genotyping, Hum. Immunol. 76 (2015) 963–974, https://doi.org/10.1016/j.humimm.2015.08.001.

[2] M. Alizadeh, A. Walencik, C. Frassati, P. Moskovtchenko, X. Lafarge, F. Verite, G. Semana, Evidence for a higher resolution of HLA genotyping by a new NGS-based approach, Transfus. Clin. Biol. 24 (2017) 120–123, https://doi.org/10.1016/j.tracli.2017.05.011.

[3] A. Larjo, R. Eveleigh, E. Kilpeläinen, T. Kwan, T. Pastinen, S. Koskela, J. Partanen, Accuracy of Programs for the Determination of Human Leukocyte Antigen Alleles from Next-Generation Sequencing Data, Front. Immunol. 8 (2017) 1815, https://doi.org/10.3389/fimmu.2017.01815.

[4] R. Carapito, M. Radosavljevic, S. Bahram, Next-Generation Sequencing of the HLA locus: Methods and impacts on HLA typing, population genetics and disease association studies, Hum. Immunol. 77 (2016) 1016–1023, https://doi.org/10.1016/j.humimm.2016.04.002.

[5] Statistics < IMGT/HLA < IPD < EMBL-EBI, (n.d.). https://www.ebi.ac.uk/ipd/imgt/hla/stats.html (accessed November 21, 2018).

[6] T. Goeury, L.E. Creary, L. Brunet, M. Galan, M. Pasquier, B. Kervaire, A. Langaney, J.-M. Tiercy, M.A. Fernández-Viña, J.M. Nunes, A. Sanchez-Mazas, Deciphering the fine nucleotide diversity of full HLA class I and class II genes in a well-documented population from sub-Saharan Africa, HLA 91 (2018) 36–51, https://doi.org/10.1111/tan.13180.

[7] HLA Nomenclature @ hla.alleles.org, (n.d.). http://hla.alleles.org/nomenclature/naming.html (accessed June 4, 2018).

[8] J.W. Hopkins, Irrigation & the Cuicatec Ecosystem: A Study of Agriculture & Civilization in North Central Oaxaca, University of Michigan, Museum of Anthropology, 1984.

[9] Nahmad Salomón, Las Raíces de las Culturas de Oaxaca, in: CIESAS, Montealbán, Oaxaca, México, 2011.

[10] Bonfil Batalla, Guillermo, Panorama Étnico y Cultural de México, in: Política Cultural Para Un País Multiétnico in Stavenhagen, Rodolfo y Margarita Nolasco (Coords), Secretaría de Educación Pública, México, D.F:, 1988: pp. 61–68.

[11] Carlos Serrano Sánchez, Mestizaje e historia de la población en México (con un esbozo antropológico de los lacandones de Chiapas), in: Ángel Martín Municio, Pedro García Barreno (Eds.), Polimorfismo génico (HLA) en poblaciones Hispanoaméricanas, Real Academia de Ciencias Exactas, Físicas y Naturales, Madrid, 1996: pp. 173–193.

[12] González García de Alba, L., y Monterrubio, M., Tendencias en la dinámica y la distribución de la población, 1970-1992, in: El Poblamiento de México, Una Visión Histórico-Demográfica, 4, México En El Siglo XX., Consejo Nacional de Población, México, 1993: pp. 154–187.

[13] J Faulhaber, Anthropology of living Indians, in: Handbook of Middle American Indians, University of Texas Press, Austin, 1970, pp. 82–104.

[14] General Assembly of the World Medical Association, World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects, J. Am. Coll. Dent. 81 (2014) 14–18.

[15] C. Wang, S. Krishnakumar, J. Wilhelmy, F. Babrzadeh, L. Stepanyan, L.F. Su, D. Levinson, M.A. Fernandez-Viña, R.W. Davis, M.M. Davis, M. Mindrinos, High-throughput, high-fidelity HLA genotyping with deep sequencing, Proc. Natl. Acad. Sci. USA 109 (2012) 8676–8681, https://doi.org/10.1073/pnas.1206614109.

[16] G.A. Watterson, The homozygosity test of neutrality, Genetics 88 (1978) 405–417.

[17] PyPop update – a software pipeline for large-scale multilocus population genomics - Lancaster - 2007 - Tissue Antigens - Wiley Online Library, (n.d.). https://onlinelibrary.wiley.com/doi/full/10.1111/j.1399-0039.2006.00769.x (accessed June 5, 2018).

[18] Hapl-o-Mat: open-source software for HLA haplotype frequency estimation from ambiguous and heterogeneous data, (n.d.). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5450239/ (accessed June 5, 2018).

[19] J.A. Hollenbach, S.J. Mack, P.-A. Gourraud, R.M. Single, M. Maiers, D. Middleton, G. Thomson, S.G.E. Marsh, M.D. Varney, Immunogenomics Data Analysis Working Group, A community standard for immunogenomic data reporting and analysis: proposal for a STrengthening the REporting of Immunogenomic Studies statement, Tissue Antigens 78 (2011) 333–344, https://doi.org/10.1111/j.1399-0039.2011.01777.x.

[20] Aguirre Beltrán, Gonzalo, La población negra en México: Estudio etnohistórico, FCE, México, 1946.

[21] Aguirre Beltrán, Gonzalo, Cuijla, FCE, Esbozo etnográfico de un pueblo negro, México, 1958.

[22] C. Gorodezky, A. Escobar-Gutiérrez, M. Salazar-Mallén, Distribution of Some of the HL-A System Lymphocyte Antigens in Mexicans, Vox Sang. 23 (1972) 439–443, https://doi.org/10.1111/j.1423-0410.1972.tb03832.x.

[23] C. Gorodezky, L. Terán, A. Escobar-Gutiérrez, HLA frequencies in a Mexican Mestizo population, Tissue Antigens 14 (1979) 347–352.

[24] C. Serrano, Rubén Lisker, Estructura genética de la población mexicana. Aspectos médicos y antropológicos. México: Salvat Mexicana de Ediciones, 158 p, Anales de Antropología - UNAM 19 (1982) (1981) 197–198.

[25] C. Gorodezky, Genetic difference between Europeans and Indians: tissue and blood types, Allergy Proc. 13 (1992) 243–250.

[26] Ángel Martín Municio, Pedro García Barreno, Polimorfismo génico (HLA) en poblaciones hispanoamericanas, Real Academia de Ciencias Exactas, Físicas y Naturales, 1996.

[27] Spanish Mexico: Windows to Oaxaca - Migratory Impacts and Leadership on Sustainable Developments Goals | Spanish and Portuguese Studies | University of Washington, (n.d.). https://spanport.washington.edu/study-abroad/spanish-mexico-windows-oaxaca-migratory-impacts-and-leadership-sustainable-developments (accessed June 15, 2018).

[28] U.W.H. Centre, Historic Centre of Oaxaca and Archaeological Site of Monte Albán, UNESCO World Heritage Centre. (n.d.). https://whc.unesco.org/en/list/415/ (accessed June 15, 2018).

[29] C. Gorodezky, C. Alaez, M.N. Vázquez-García, G. de la Rosa, E. Infante, S. Balladares, R. Toribio, E. Pérez-Luque, L. Muñoz, The genetic structure of Mexican Mestizos of different locations: tracking back their origins through MHC genes, blood group systems, and microsatellites, Hum. Immunol. 62 (2001) 979–991.

[30] The Allele Frequency Net Database [Search HLA Allele Frequencies], (n.d.). http://www.allelefrequencies.net/hla6006a.asp (accessed June 15, 2018).

[31] K. Cao, A.M. Moormann, K.E. Lyke, C. Masaberg, O.P. Sumba, O.K. Doumbo, D. Koech, A. Lancaster, M. Nelson, D. Meyer, R. Single, R.J. Hartzman, C.V. Plowe, J. Kazura, D.L. Mann, M.B. Sztein, G. Thomson, M.A. Fernández-Viña, Differentiation between African populations is evidenced by the diversity of alleles and haplotypes of HLA class I loci, Tissue Antigens 63 (2004) 293–325.

[32] M. Tshabalala, J. Mellet, M.S. Pepper, Human Leukocyte Antigen Diversity: A Southern African Perspective, J Immunol Res. 2015 (2015) 746151, , https://doi.org/10.1155/2015/746151.

[33] F.C. Arnett, P. Gourh, S. Shete, C.W. Ahn, R.E. Honey, S.K. Agarwal, F.K. Tan, T. McNearney, M. Fischbach, M.J. Fritzler, M.D. Mayes, J.D. Reveille, Major histocompatibility complex (MHC) class II alleles, haplotypes and epitopes which confer susceptibility or protection in systemic sclerosis: analyses in 1300 Caucasian, African-American and Hispanic cases and 1000 controls, Ann. Rheum. Dis. 69 (2010) 822–827, https://doi.org/10.1136/ard.2009.111906.