



OPEN ACCESS

EDITED BY

Sandra Tafalo,
Instituto Português de Sangue e da
Transplantação, Portugal

REVIEWED BY

Martin Maiers,
National Marrow Donor Program,
United States
José Manuel Nunes,
University of Geneva, Switzerland

*CORRESPONDENCE

Esteban Arrieta-Bolaños,
✉ esteban.arrieta-bolanos@uk-essen.de

SPECIALTY SECTION

This article was submitted
to Immunogenetics,
a section of the journal
Frontiers in Genetics

RECEIVED 31 January 2022

ACCEPTED 13 March 2023

PUBLISHED 23 March 2023

CITATION

Arrieta-Bolaños E,
Hernández-Zaragoza DI and Barquera R
(2023), An HLA map of the world: A
comparison of HLA frequencies in
200 worldwide populations reveals
diverse patterns for class I and class II.
Front. Genet. 14:866407.
doi: 10.3389/fgene.2023.866407

COPYRIGHT

© 2023 Arrieta-Bolaños, Hernández-Zaragoza and Barquera. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

An HLA map of the world: A comparison of HLA frequencies in 200 worldwide populations reveals diverse patterns for class I and class II

Esteban Arrieta-Bolaños^{1,2*}, Diana Iraíz Hernández-Zaragoza³ and Rodrigo Barquera⁴

¹Institute for Experimental Cellular Therapy, University Hospital Essen, Essen, Germany, ²German Cancer Consortium (DKTK), partner site Essen/Düsseldorf, Heidelberg, Germany, ³Molecular Genetics Laboratory, Escuela Nacional de Antropología e Historia (ENAH), Mexico, Mexico, ⁴Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology (MPI-EVA), Leipzig, Germany

HLA frequencies show widespread variation across human populations. Demographic factors as well as selection are thought to have shaped HLA variation across continents. In this study, a worldwide comparison of HLA class I and class II diversity was carried out. Multidimensional scaling techniques were applied to 50 *HLA-A* and *HLA-B* (class I) as well as 13 *HLA-DRB1* (class II) first-field frequencies in 200 populations from all continents. Our results confirm a strong effect of geography on the distribution of HLA class I allele groups, with principal coordinates analysis closely resembling geographical location of populations, especially those of Africa-Eurasia. Conversely, class II frequencies stratify populations along a continuum of differentiation less clearly correlated to actual geographic location. Double clustering analysis revealed finer intra-continental sub-clusters (e.g., Northern and Western Europe vs. South East Europe, North Africa and Southwest Asia; South and East Africa vs. West Africa), and HLA allele group patterns characteristic of these clusters. Ancient (Austronesian expansion) and more recent (Romani people in Europe) migrations, as well as extreme differentiation (Taiwan indigenous peoples, Native Americans), and interregional gene flow (Sámi, Egyptians) are also reflected by the results. *Barrier* analysis comparing D_{ST} and geographic location identified genetic discontinuities caused by natural barriers or human behavior explaining inter and intra-continental HLA borders for class I and class II. Overall, a progressive reduction in HLA diversity from African to Oceanian and Native American populations is noted. This analysis of HLA frequencies in a unique set of worldwide populations confirms previous findings on the remarkable similarity of class I frequencies to geography, but also shows a more complex development for class II, with implications for both human evolutionary studies and biomedical research.

KEYWORDS

human leukocyte antigen, population differentiation, allele frequencies, MHC, anthropology, population genetics, genetic drift

Introduction

Human populations share a common origin in Sub-Saharan Africa (Reyes-Centeno, 2016). However, a long history of migrations and admixture events (Hellenthal et al., 2014), as well as gene flow from other hominins (Abi-Rached et al., 2011; Wall and Yoshihara Caldeira Brandt, 2016), combined with geographic (Novembre et al., 2008) and linguistic (Baker et al., 2017) barriers have produced a complex pattern of population differentiation across the globe. This complex history is reflected and can be traced by our genetic makeup (Genomes Project et al., 2015). Indeed, analysis of polymorphism in the human genome has revealed the relationships between different human groups (Li et al., 2008; Auton et al., 2009; Genomes Project et al., 2015), and is a powerful tool to understand our populations' genetic history. The Human Leukocyte Antigens (HLA) bear the most extensive polymorphism in the human genome. Currently, according to the IPD-IMGT/HLA database (release 3.50 October 2022) (Robinson et al., 2020), more than 35,000 alleles have been discovered for this genetic system, and it is estimated that the total variation in the whole human population could reach several million alleles (Robinson et al., 2017). Importantly, allelic and haplotypic frequencies of HLA genes vary dramatically across human populations (Fernandez Vina et al., 2012; Askar et al., 2013). Knowledge of HLA allele and haplotype frequencies is of special interest for medical fields such as cellular therapy and transplantation of organs and stem cells (Tiercy and Claas, 2013; Pidala et al., 2014), disease association studies and genetic epidemiology (Ghodke et al., 2005), forensics, and pharmacogenetics (Pirmohamed et al., 2015), but can also be used to assess genetic diversity in human populations (Sanchez-Mazas et al., 2011). Of note, due to its central role in immune responses, the HLA system may be influenced both by demographic as well as by selective mechanisms, in particular balancing selection (Meyer et al., 2006). It is hence of interest to explore the interplay between selective and demographic events on population differentiation and how this has shaped HLA genetic structure across the world. In particular, a comparative analysis of class I and class II variation could also provide insights into the differential population genetics dynamics of these two functionally different components of the HLA system. Various landmark studies have previously examined HLA variation at the continental (Di and Sanchez-Mazas, 2011; 2014; Sanchez-Mazas et al., 2013; Nunes et al., 2014; Sanchez-Mazas et al., 2017a; Sanchez-Mazas et al., 2017b) and global (Middleton et al., 2000; Williams et al., 2001; Mack et al., 2007; Solberg et al., 2008; Nunes et al., 2010; Buhler and Sanchez-Mazas, 2011; Sanchez-Mazas et al., 2012a; Riccio et al., 2013; Dos Santos Francisco et al., 2015; Buhler et al., 2016; Nunes et al., 2022) level. Inspired by these previous efforts, here we aimed at contributing to the study of global patterns of HLA variation by comparing two unique datasets of human population samples from all regions of the world using multidimensional scaling (MDS) and clustering approaches, as well as genetic barrier analysis based on first-field frequencies of 50 class I and 13 class II HLA allele groups.

Materials and methods

Population data

HLA frequency data were extracted from peer-reviewed journal articles, proceedings from the International Histocompatibility

Workshops (IHWS), and/or the Allele Frequency Net Database (Gonzalez-Galarza et al., 2020; Gonzalez-Galarza et al., 2021) or the HLA-NET platform (Sanchez-Mazas et al., 2012b; Nunes et al., 2014). Population samples selected consisted of healthy individuals included in anthropology or disease association studies or bone marrow donor registries that had available first or second-field DNA-based HLA typing for either class I and/or class II. Inclusion of populations from all regions and major ethnicities of the world was sought. In addition to populations native to each of ten worldwide geographical regions (Solberg et al., 2008), migrant non-native populations were also included to investigate their relation to the extant populations from their ancestral homeland and those from the region where they migrated to. Mixed-ancestry (i.e., "Mestizo") populations were not included on account of their complex multi-continental admixtures, which have been analyzed elsewhere (Arrieta-Bolanos et al., 2012; Barquera et al., 2020a; Arrieta-Bolanos et al., 2020). Two datasets were constructed: one consisting of class I frequencies (*HLA-A* and *HLA-B*) and another one based on *HLA-DRB1* frequencies. These loci were selected since they show the highest polymorphism of the HLA genes (Robinson et al., 2020), as well the strongest geographic (Di et al., 2017) and ethnic-specific variation (Solberg et al., 2008), while maximizing also the number of populations available for analysis (Gonzalez-Galarza et al., 2020). When available, high-resolution *HLA-A*, *-B*, and *-DRB1* data were reduced to first-field (i.e., allele groups; 50 for class I and 13 for *HLA-DRB1*) in order to homogenize the resolution for all datasets for MDS and clustering analyses. Class I and class II data were compiled and analyzed independently due to their different roles in the immune system and the potential for differential effects of immune mediated selective pressures across classes. Dataset quality control included assessment of cumulative frequencies for each locus for each population to ensure it corresponded to what was reported on allelefrequencies.net or the respective publication for the allele groups investigated, and individual cross-checking of entries for selected populations. Evidence of Hardy-Weinberg equilibrium (HWE) in the population sample was not considered as a pre-requisite for inclusion in the study, but information on this was retrieved from the original publications when available. In a subset of samples for which raw data was available, HWE was tested in this work as explained below. In total, both the class I and class II datasets included 200 and 197 population samples and consisted of 10,200 and 2,561 frequency entries, respectively. In total, the population array included 712,462 chromosomes for class I and 370,794 chromosomes for class II. Details on the populations included in this study, including their nomenclature, abbreviation codes, literature source, sample size, HWE status, geographical region, and corresponding references can be found in [Supplementary Tables S1, 2](#).

Multidimensional scaling and clustering analyses

The HLA class I and class II frequency data of the populations in each dataset were analyzed by clustering analysis and Principal Coordinates Analysis (PCoA), both based on Euclidean distances, using the Multi-Variate Statistical Package (MVSP version 3.22, Kovach Computing Services, Anglesey, Wales) as explained below.

HLA-A and *HLA-B* frequencies were analyzed both separately and together, while *HLA-DRB1* frequencies were analyzed separately. A double clustering analysis was performed and dendrograms were generated for both the populations and the HLA allele groups. The clustering method was based on minimum variance of squared Euclidean distances with a randomized input order. The similarity matrices and their associated sorted data were used to construct heat maps showing the variation of HLA frequencies in the clustered data. PCoA analysis was carried out for each dataset and locus, and axes were extracted according to Kaiser's rule (Kaiser, 1960). The distance matrix generated for each dataset was further analyzed to compare intra and inter-continental subgroup distance distributions.

For subsequent comparative analyses of the MDS output results, population samples were grouped according to their geographic location into continental and subcontinental regions. Migrant populations (i.e. those descending from populations located in a different geographical region) including African-descendant populations in the Americas and European-descendant populations in the United States, Australia, and South Africa were regrouped to their ancestral region for further statistical analysis of the PCoA output and allele group diversity comparisons.

Detection of genetic discontinuity in worldwide HLA diversity

A set of populations (Supplementary Tables S3–6) for which both HLA class I (*HLA-A*, *-B*) and class II (*HLA-DRB1*) data was available was used for genetic discontinuity testing using the software *Barrier* v.2.2 (Manni et al., 2004). To construct the genetic distances matrix, we used the HLA frequency data available for the selected populations and calculated the corrected D_{st} genetic distance using the software POPTREEW (Takezaki et al., 2014) with a bootstrap of 1,000 replications. We used the reported coordinates for each population [available in the Allele Frequency Net Database (Gonzalez-Galarza et al., 2020) unless otherwise stated]. We added virtual points in the Pacific Ocean region and the Mediterranean Sea to appropriately model neighboring populations (i.e., preventing the software from assigning North African and Mediterranean populations or Beringian and Oceanian populations as neighboring clusters). To better depict the demographic and natural history of Native American, East Asian, and Oceanian populations, we shifted the longitude coordinates to include America as a continuum to Asian and Oceanian populations by adding 210° to the populations located to the west of the Greenwich meridian (0°) and subtracted 150° to the populations located to the east (Supplementary Tables S3–6), unless otherwise indicated (i.e., populations to the west of the Greenwich meridian part of Europe or Africa that otherwise would be translocated to the East). Then, we computed $k = 30$ barriers separately for each locus, for both class I loci (*HLA-A*, *HLA-B*), and for class I and *HLA-DRB1* in each of four geographical regions. Finally, to assess the statistical significance of the computed barriers, we analyzed the resampled bootstrapped ($k = 100$) matrices to test for HLA system-wide discontinuities.

Statistical analyses

HWE testing was performed for a subset of samples with available raw data (Supplementary Tables S1, 2) using Arlequin ver. 3.5 (Excoffier and Lischer, 2010). p -values < 0.016 (after Bonferroni correction) were considered statistically significant. Euclidean distances between populations calculated from the PCoA were compared using analysis of variance (ANOVA) and Tukey's method with adjustment for multiple comparisons. Similarly, the number of HLA allele groups observed in each population was compared across continental groups with ANOVA and Tukey's multiple comparisons test. For these analyses, a p -value of < 0.05 was considered statistically significant. ANOVA and descriptive statistical analyses were performed with Prism (version 9.1.2, GraphPad Software Inc., La Jolla, California).

Results

Population sample metrics

The population samples for HLA class I analysis included 200 datasets with sample sizes ranging from 36 to 50,614 individuals. For *HLA-DRB1*, the dataset included 197 population samples with sample sizes ranging from 16 to 39,689 individuals. Class I datasets were published between 1995 and 2021 with 70 (35%) published in or after 2010. For *HLA-DRB1*, datasets were published between 1992 and 2020, with 76 (39%) published from 2010 onwards. These populations samples included datasets from 10 geographical regions (Mack et al., 2007), namely Sub-Saharan Africa (SSA), Europe (EUR), North Africa (NAF), Southwest Asia (i.e. Middle East and South Asia, SWA), northern and southern East Asia (NEA, SEA), northern (NAM, including Inuit from Greenland and populations from Central America) and southern America (SAM), Oceania (OCE), and Australia (AUS). A summary of the most relevant aspects of these datasets can be found in Table 1. Overall, median sample size was comparable across continental groups and class I and class II datasets.

Clustering analysis reveals HLA worldwide population relationships

The results of the clustering analysis of the populations for class I (*HLA-A* and *HLA-B* together) and class II are shown in Figure 1 and Figure 2, respectively. For class I (Figure 1A), the biggest split in the tree separates the Native American, East Asian, AUS, and OCE populations from those of SSA, EUR, NAF, and SWA. Subsequent splits separate the Native Americans from the oriental populations, and the SSA from the cluster formed by EUR, NAF, and SWA.

Closer examination of the class I phylogenetic tree (Figures 1B–F), reveals that the EUR + NAF + SWA cluster separates into two groups defined mainly by the Alps and Carpathians (Figures 1B,C), with populations south and east of these mountain ranges (Figure 1B) clustering into four subgroups, namely 1) NAF and Arabians, 2) populations from the Near East, the Levant 3) Southeast

TABLE 1 Summary of the population datasets included in the analyses and their main characteristics.

Region	HLA-A, HLA-B				HLA-DRB1			
	Number of datasets	Median sample size	Range	Total number of individuals	Number of datasets	Median sample size	Range	Total number of individuals
SSA	29	138	36–5,928	9,900	27	102	32–5,928	9,512
NAF	7	100	73–118	674	7	98	82–118	695
SWA	27	175	42–23,000	78,467	12	155.5	72–15,542	21,663
EUR	65	154	50–39,979	120,899	46	186.5	43–39,689	104,914
NEA	11	158	50–18,604	20,209	16	185.5	50–4,128	11,442
SEA	11	236	50–50,614	96,277	30	104	25–16,807	29,155
NAM	23	100	42–8,525	14,963	23	60	16–3,000	5,316
SAM	14	81.5	40–168	1,165	20	56.5	16–144	1,273
AUS	4	147	75–891	1,260	3	103	41–191	335
OCE	9	111	49–11,499	12,417	13	65	21–292	1,095
Total	200	131	36–50,614	356,231	197	103	16–39,689	185,738

Bold values indicate the totals for each column.

Europeans and populations from the Mediterranean islands, and 4) South Asians. Interestingly, Romani populations from Hungary (Inotai et al., 2015) and Spain (Gonzalez-Galarza et al., 2020) cluster with the populations from South Asia, and the sample of Bukharan Jews (Manor et al., 2016) clusters with the populations of the Levant.

Sub-clusters to the north and west of the European mountain ranges (Figure 1C) include 1) Iberians and populations from the Balearic, Azores (Gonzalez-Galarza et al., 2020), and Canary (Romon et al., 2016) islands, and 2) the populations north of the Pyrenees, split into the British Isles and the continental Germanic (western) and Slavic-Uralic (eastern) linguistic areas. Similar results for Europe have been previously observed using higher resolution data (Nunes et al., 2010; Riccio et al., 2013). European-descendant populations from the United States (Leffell et al., 2007), South Africa (Paximadis et al., 2012), and Australia (Gonzalez-Galarza et al., 2020) also cluster together with the North-western Europeans.

The SSA cluster (Figure 1D), is split into West Africans and populations from the southern and eastern parts of the continent. Previous research has demonstrated differences between these subregions in Africa (Nunes et al., 2010). Interestingly, despite well-documented gene flow from non-African populations, African-descendant populations in the United States (Leffell et al., 2007; Maiers et al., 2007), Costa Rica (Arrieta-Bolaños et al., 2019), Martinique (Bera et al., 2001), and Trinidad and Tobago (Gonzalez-Galarza et al., 2020) cluster together among West Africans. Of note, the sample of Egyptians (Nunes et al., 2014) is found among the SSA populations, probably reflecting post-Roman SSA gene flow (Schuenemann et al., 2017).

The picture becomes more complicated for populations from East Asia, OCE and AUS (Figure 1E), and the Americas (Figure 1F).

However, some patterns emerge as New Zealand Maori (Edinur et al., 2012), Polynesians (Edinur et al., 2012), and American Samoans (Mack et al., 2007) cluster with Taiwan indigenous peoples (Gonzalez-Galarza et al., 2020), whereas the populations from the extreme north-eastern part of Asia (i.e. the Russian Nivkhis (Lou et al., 1998)) cluster with the samples from Alaskan natives (Leffell et al., 2002) and the Inuit from Greenland (Grunnet et al., 1996) (Figure 1F). Moreover, Papuans (Main et al., 2001) and New Caledonians (Main et al., 2001) cluster with the Indigenous Australians, whereas the rest of the Asian populations split into three groups formed by 1) continental and 2) insular SEA, and 3) NEA. Of note, the Russian Kalmyks (Loginova et al., 2021), although settled in Europe, cluster with other populations of Mongolian ethnic descent, and Arctic populations west of the Urals (Evseeva et al., 2002; Johansson et al., 2008; Harbo et al., 2010) are also found in the Asian cluster.

In the Americas (Figure 1F), indigenous populations form a distinct group with local (i.e. Andean, Mesoamerican, Lowlander) patterns of clustering, albeit with complex relations as previously reported (Nunes et al., 2010). Of note, the Ainu from Japan cluster with the Native Americans, something already previously observed (Bannai et al., 1996; Bannai et al., 2000).

When the clustering was performed separately with *HLA-A* or *HLA-B* frequencies, some differences were observed. For *HLA-A* (Supplementary Figure S1), the first split separated AUS and OCE populations from the rest, while the second separated NAM and SAM from SSA, NAF, SWA, EUR, NEA, and SEA. The SSA separate next, leaving the split between the East Asian populations and the EUR + NAF + SWA as the fourth one, with some South Asian populations remaining in a cluster next to the SEA. On the other hand, when only *HLA-B* frequencies were used (Supplementary Figure S2), the main splits resembled more closely those observed in the combined analysis, albeit with some reshuffling of the regional subclusters.

For *HLA-DRB1*, the clustering results roughly resemble the major patterns observed for class I, albeit with some

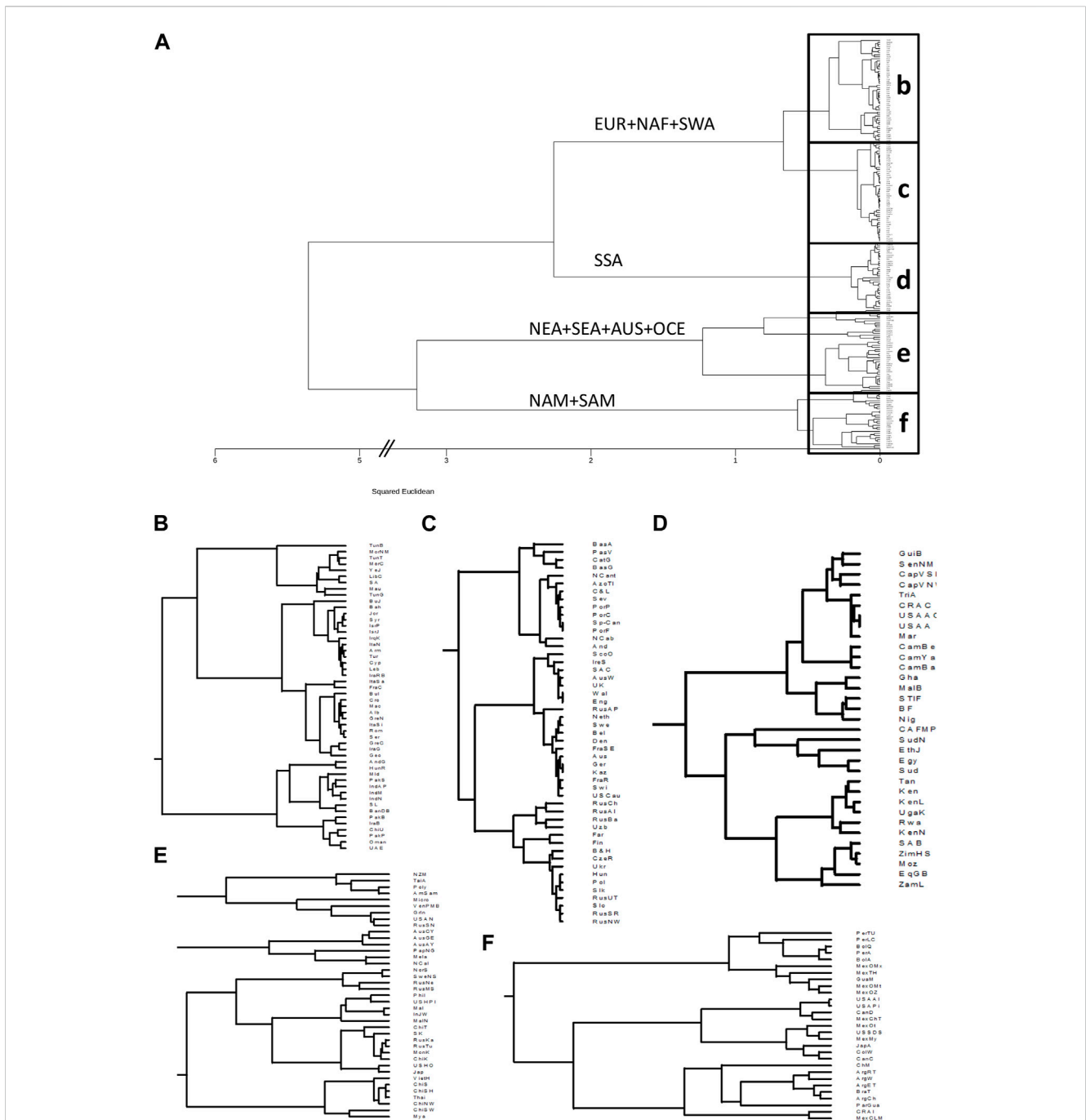


FIGURE 1 Clustering based on HLA class I frequencies identifies continental and intracontinental population similarities. **(A)**: Dendrogram generated with minimum variance clustering analysis of squared Euclidean distances calculated using 50 HLA-A and HLA-B allele group frequencies in 200 worldwide populations. Main branches are labelled according to the predominant continental origin of populations forming its clusters. b-f. Panels show the detail of the boxes in a. **(B)**: NAF, North Africa; SWA, Southwest Asia. **(C)**: EUR, Europe. **(D)**: SSA, Sub-Saharan Africa. **(E)**: NEA, Northeast Asia; SEA, Southeast Asia; AUS, Australia; OCE, Oceania. **(F)**: NAM, North America; SAM, South America.

exceptions. In this case, the samples from Native Americans appear as a first breakout group (Figures 2A,B), and AUS populations cluster among them. Of note, the remote Rapa Nui from Easter Island (Thorsby, 2012) cluster with Polynesian (Edinur et al., 2012) and Samoan (Mack et al., 2007) populations (Figure 2C).

Double clustering identifies HLA allele group variation patterns across populations

In order to illustrate worldwide HLA allele group variation patterns, we performed a double clustering procedure and integrated the results of the population and the HLA allele group

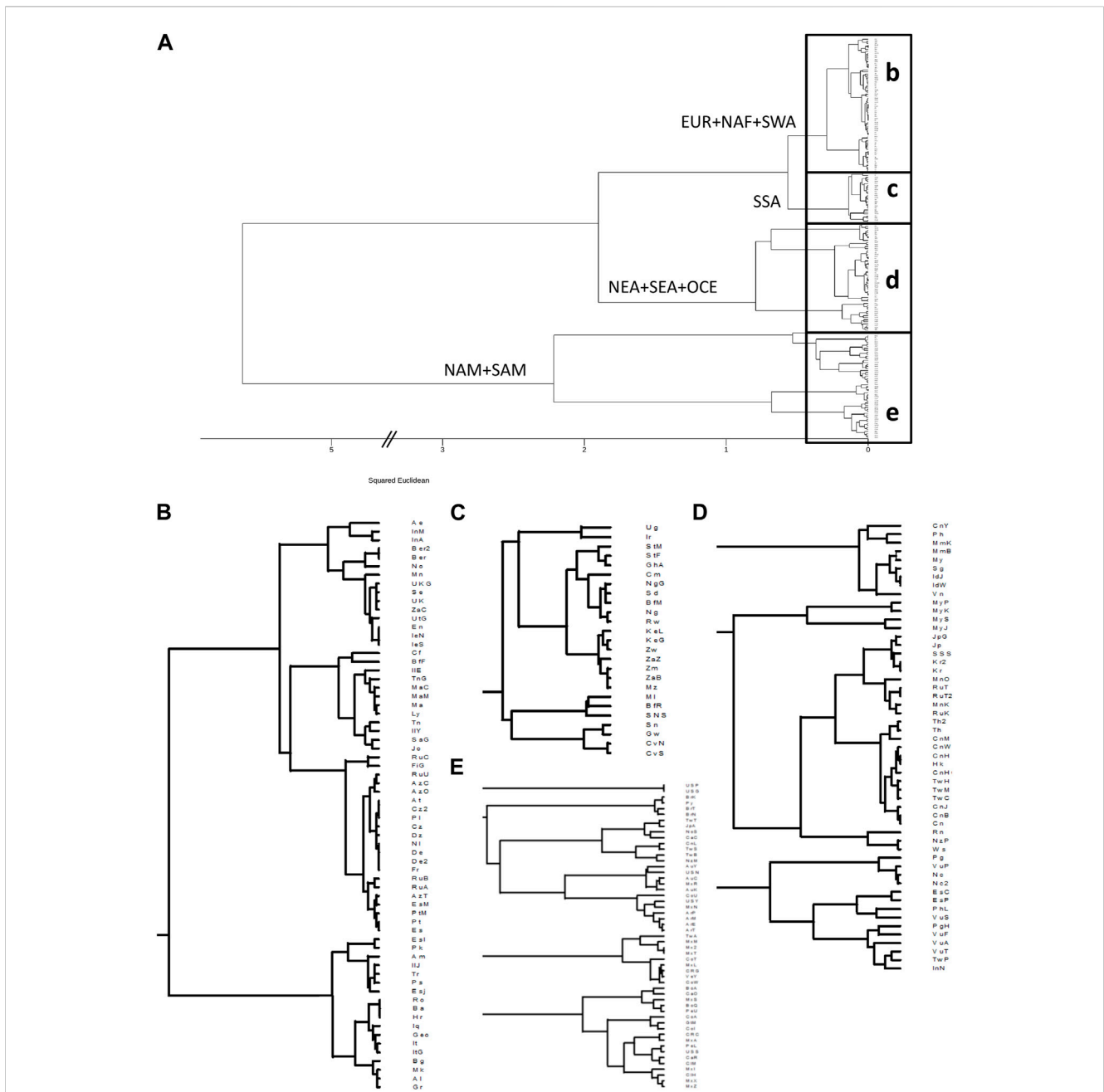


FIGURE 2 Clustering based on HLA-DRB1 frequencies identifies continental and intracontinental population similarities for HLA class II. (A): Dendrogram generated with minimum variance clustering analysis of squared Euclidean distances calculated using 13 HLA-DRB1 allele group frequencies in 197 worldwide populations. Main branches are labelled according to the predominant continental origin of populations forming its clusters. b-e. Panels show the detail of the boxes in a. (B): NAF, North Africa; SWA, Southwest Asia; EUR, Europe. (C): SSA, Sub-Saharan Africa. (D): NEA, Northeast Asia; SEA, Southeast Asia; OCE, Oceania. (E): NAM, North America; SAM, South America.

clustering in an unsupervised hierarchical HLA frequency heat maps as shown in Figure 3 and Figure 4 for class I and class II, respectively. For HLA class I (Figure 3), apart from HLA-A*02 having a high frequency in almost all human populations and together with A*24 concentrating most of the HLA-A diversity in NAM, SAM, OCE, NEA, and SEA populations, additional patterns reveal HLA signatures in specific continental groups, with high frequencies for B*15, A*30, and A*68 characterizing SSA populations and A*23,

A*33, B*53, B*58 being particularly frequent in West Africans. Northern and western EUR populations share high frequencies for A*01, A*03, B*07, and B*44, which set them apart from south-eastern EUR, NAF, and SWA populations, in turn characterized by higher B*51 frequencies. NAM and SAM populations show very high frequencies of A*31, A*68, B*35, B*39, and B*48, with B*40 and B*15 driving the split between populations in these groups. Indigenous Australians, Melanesians

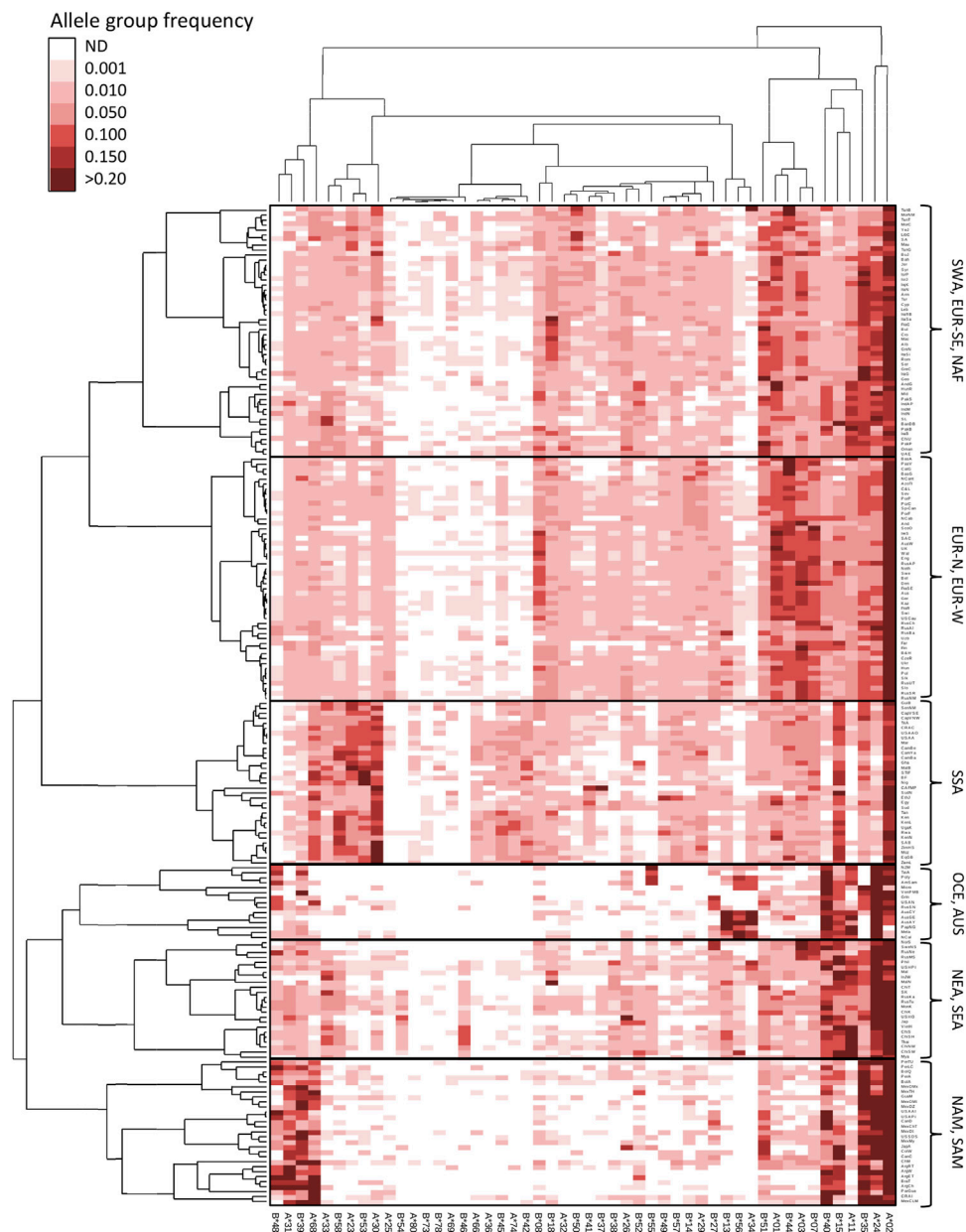


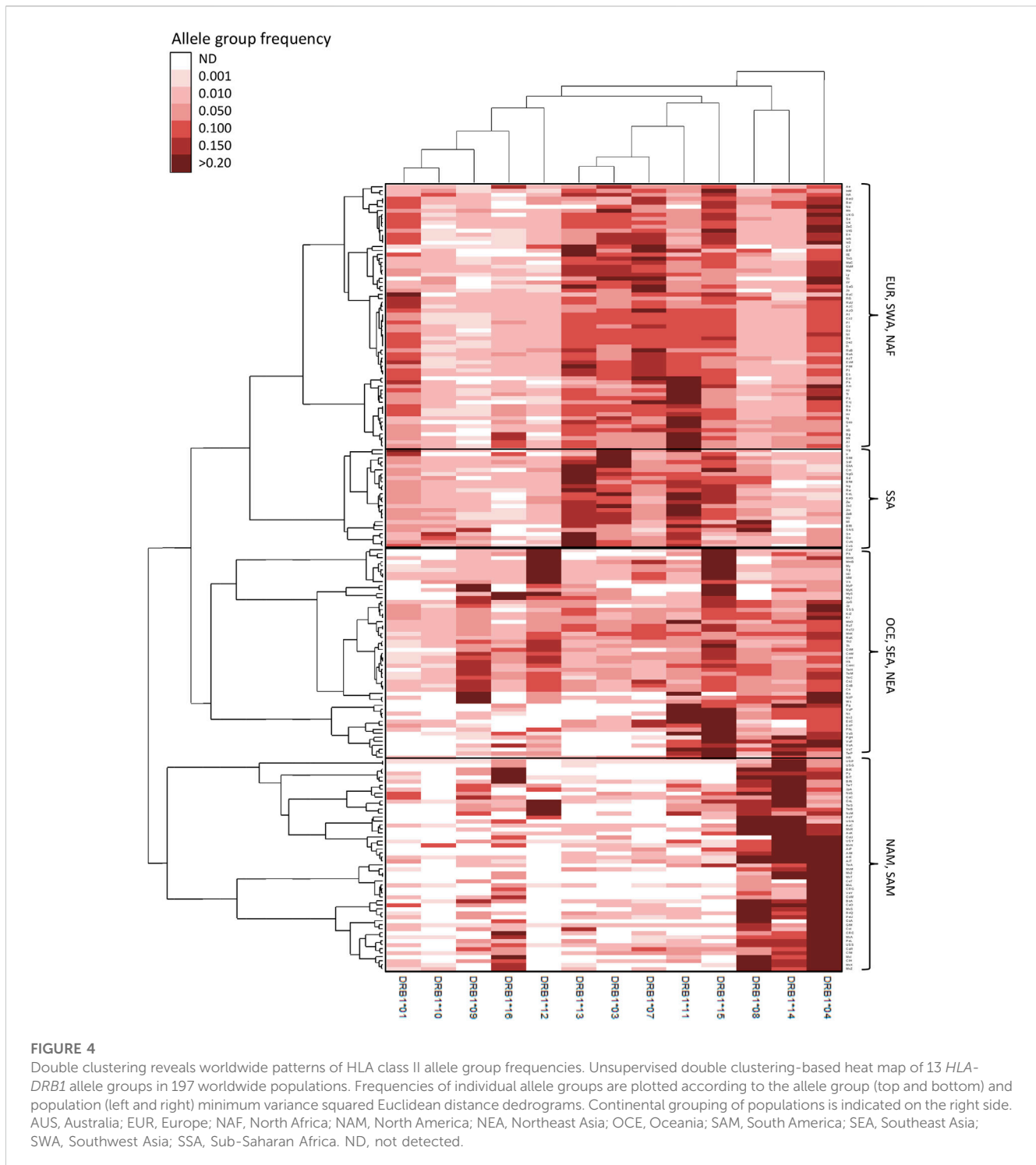
FIGURE 3

Double clustering reveals worldwide patterns of HLA class I allele group frequencies. Unsupervised double clustering-based heat map of 50 *HLA-A* and *HLA-B* allele groups in 200 worldwide populations. Frequencies of individual allele groups are plotted according to the allele group (top and bottom) and population (left and right) minimum variance squared Euclidean distance dendrograms. Continental grouping of populations is indicated on the right side. AUS, Australia; EUR-N, Northern Europe; EUR-W, Western Europe; EUR-SE, South-eastern Europe; NAF, North Africa; NAM, North America; NEA, Northeast Asia; OCE, Oceania; SAM, South America; SEA, Southeast Asia; SWA, Southwest Asia; SSA, Sub-Saharan Africa. ND, not detected.

and Papuans (Main et al., 2001) share a distinctive signature formed by A*34, B*13, and B*56 also present at lower frequencies in Polynesians (Edinur et al., 2012), Micronesians (Gao et al., 1997), and the Maori (Edinur et al., 2012). The latter three groups differentiate on account of their high B*55 frequencies, while populations from North Eastern Siberia (Lou et al., 1998), Alaska (Leffell et al., 2002), and Greenland (Grunnet et al., 1996) share very high B*27 frequencies, something only found elsewhere in Arctic Europe. Finally, NEA and SEA populations share very high frequencies of A*11, B*15, and B*40, as well as the highest

B*46 and B*56 frequencies worldwide. Of note, NEA and SEA differ starkly in their A*34 frequencies, which in the latter could represent a signature of admixture with the pre-East Asian migration substrate of the region (McCull et al., 2018).

For *HLA-DRB1* (Figure 4), distinctive patterns include very high frequencies for DRB1*04, *08, *14, and, in some cases, *16 in NAM and SAM; high DRB1*15 frequencies in OCE and insular SEA, the latter also characterized by high DRB1*12 frequencies; and DRB1*09 being highly prevalent in both East Asian and OCE populations. On the other hand, allele groups DRB1*01,



*03, *04, *07, and *15 have all high frequencies in EUR, SWA and NAF populations, with very high frequencies for DRB1*11 characterizing most SE Europeans as previously observed (Nunes et al., 2014). Finally, very high DRB1*11, *13, and *15 in addition to low DRB1*04 frequencies characterize SSA populations.

Analysis of HLA allele group frequencies patterns revealed a progressive reduction of the diversity proportional to the distance

from the African continent. This is especially evident for HLA class I, where a more balanced allele group frequency pattern in SSA, NAF, SWA, and EUR (median of 16–17 and 21–25 allele groups detected for *HLA-A* and *HLA-B*, respectively) contrasts with much more polarized frequencies in Native Americans, OCE, and AUS (medians of 8 and 10–12.5) (Figures 5A,B,D). East Asian populations have only slightly reduced numbers of detected allele groups (medians 14–15 and 20–23). A similar

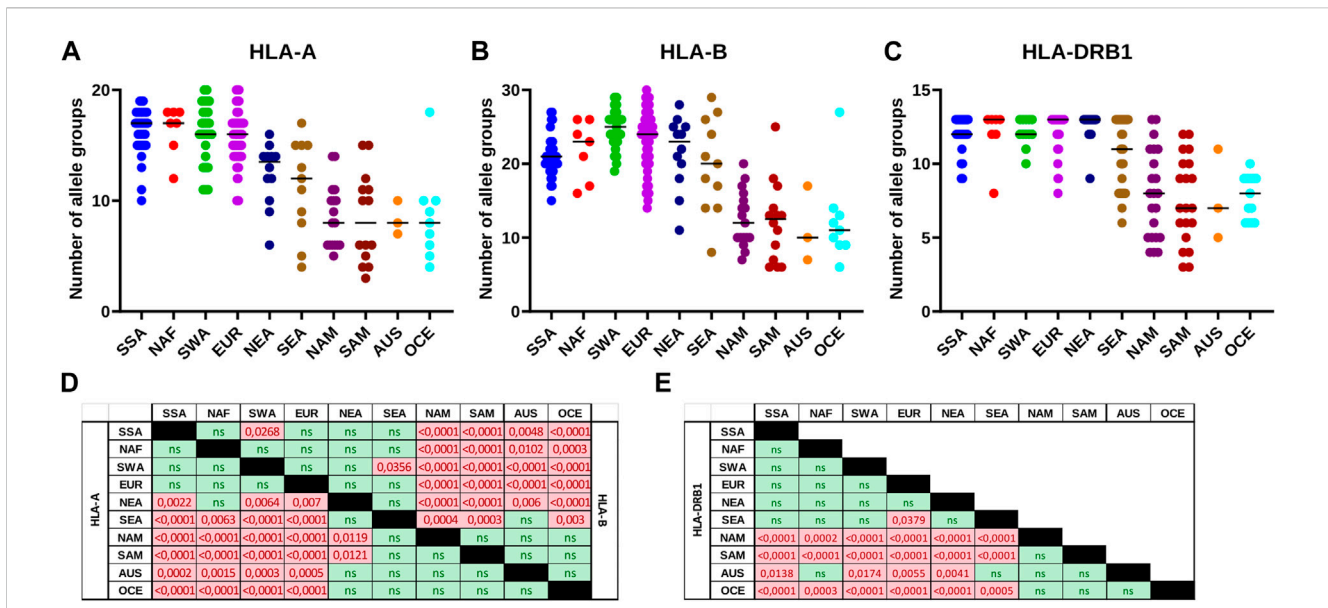


FIGURE 5
 The number of detectable HLA class I and class II allele groups differs across continental groups. (A–C): The number of HLA-A, HLA-B or HLA-DRB1 allele groups detected in each of the populations included in this study is plotted per continental group. Horizontal bar represents the median number of allele groups detected. (D,E): Matrices of statistical testing results for the differences in number of HLA allele groups detected per continental population group. (D): Class I matrix. Lower left half shows results for HLA-A, and upper right half those for HLA-B. (E): HLA-DRB1 matrix. p-values for statistically significant results in pairwise comparisons are coloured in red. AUS, Australia; EUR, Europe; NAF, North Africa; NAM, North America; NEA, Northeast Asia; OCE, Oceania; SAM, South America; SEA, Southeast Asia; SWA, Southwest Asia; SSA, Sub-Saharan Africa. ns, not significant.

significant diversity reduction is observed for HLA-DRB1 with a median of 12–13 allele groups observed in SSA, NAF, SWA, EUR, and NEA vs. only seven to eight in NAM, SAM, OCE, and AUS (Figures 5C,E). The SEA populations, previously observed to have lower HLA diversity than their NEA counterparts (Di and Sanchez-Mazas, 2011; 2014), do manifest a trend to lower numbers of detected DRB1 allele groups, albeit without reaching statistical significance.

PCoA confirms similarity between geography and HLA

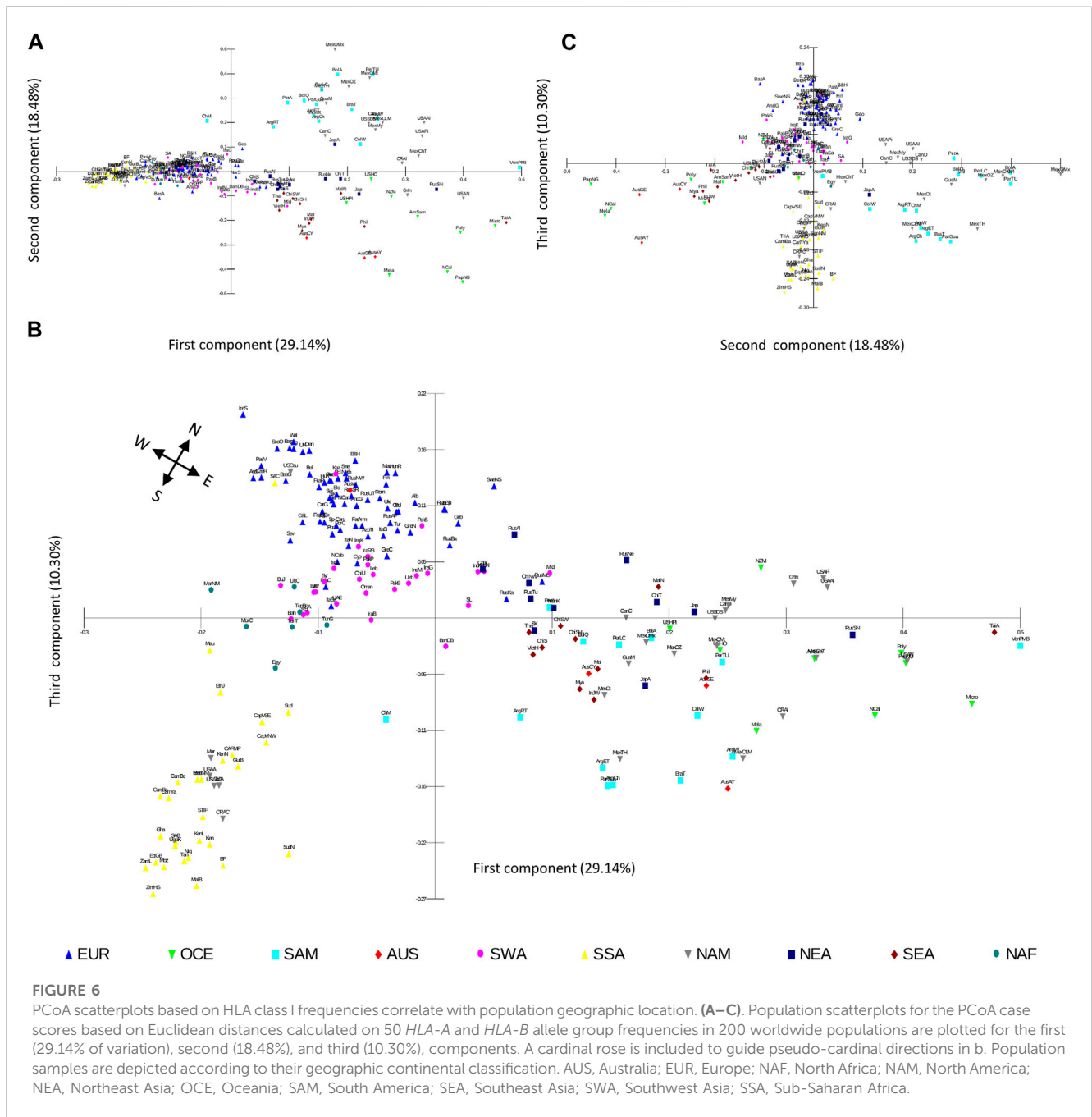
To further explore the distribution of worldwide HLA diversity we performed PCoA analysis on the class I and HLA-DRB1 allele group frequencies of the population datasets. For class I, the first component (29.14% of the variation) separates Native American, AUS, OCE, NEA, and SEA populations from SSA, NAF, SWA, and EUR populations, and the second component (18.48%) further distinguishes NAM and SAM from East Asian and OCE and AUS populations (Figure 6A). A further third component (10.30%) shows the separation of EUR, SWA, NAF, and SSA populations, while also allowing the observation of subgroups within Native American populations and a progressive scattering of NEA, SEA, OCE, and AUS populations (Figures 1B,C). Interestingly, when the first and third components of this analysis are plotted, a remarkable similarity between the geographic location of the different population samples and the PCoA scatter plot is observed (Figure 6B). This coincidence with latitudinal and longitudinal location is particularly clear for the

African continent (NAF, SSA), and Western (EUR, SWA) and Eastern (NEA, SEA) Eurasia as previously reported (Di and Sanchez-Mazas, 2011; 2014; Sanchez-Mazas et al., 2013; Sanchez-Mazas et al., 2017a).

When the PCoA is performed separately for HLA-A and HLA-B, the plots using the first three components for each gene resemble the general patterns described above, albeit with less clarity in the separation of the continental groups (Supplementary Figures S3, 4).

In turn, the PCoA based on HLA-DRB1 allele group frequencies produces somewhat different results. In this case, the first component (37.58%) also separates NAM, SAM, OCE, and AUS populations from SSA, NAF, SWA, EUR, NEA, and SEA, and although some populations in the latter region also differentiate (Supplementary Figure S5A). The second (18.74%) and third (11.99%) components allow for the separation of SEA populations and a discrete differentiation between SSA and EUR, NAF, and SWA (Supplementary Figures S5B, C), with a more continuous scattering and less clear correlation with actual geographical location.

The differences in the effects of geographic distance on HLA distribution between class I and class II are also reflected by the continental analysis of the Euclidean distances generated by the PCoA. While for class I, and in particular for HLA-A, the distances between non-SSA and SSA populations tend to increase at a constant rate proportional to geographic distance from SSA and reach a maximum for AUS and OCE (Figures 7A–C; Tables 2–4), for HLA-DRB1 they remain at lower levels for NAF, SWA, EUR, and NEA, and thereafter increase reaching a maximum for NAM and SAM (Figure 7D; Table 5). Overall, these results suggest a different dynamics of HLA diversity differentiation for class I vs. class II at



this level of resolution, with class I reflecting better geographical location and distance between populations.

Barrier analyses highlights continental HLA class I and class II genetic discontinuities

Using *Barrier* analysis, we examined the correlation between HLA genetic discontinuity and geographic location of human populations in four continental regions, namely Africa, Eurasia, East Asia-Pacific, and the Americas. Although many overlapping genetic discontinuities were observed when each of the loci were analyzed independently (Supplementary Figures S6–8), the joint

analysis of *HLA-A* and *HLA-B* (Supplementary Figure S9) and of *HLA-A*, *HLA-B*, and *HLA-DRB1* (Figure 8) was required to produce statistically significant barriers. For HLA class I, significant barriers were detected in Africa along the North Africa/Sub-Saharan/Mediterranean axis (Supplementary Figure S9A). Interestingly, this barrier runs south of some East African populations, likely reflecting gene flow from West Asia (Aamer et al., 2021). In Eurasia (Supplementary Figure S9B), in addition to some smaller barriers separating SWA from EUR and Central Asia, a major barrier was found between SEA and populations to the north and west, which was also detected in the East Asia-Pacific plot (Supplementary Figure S9C), in addition to another one separating AUS and OCE from SEA. Interestingly, the latter

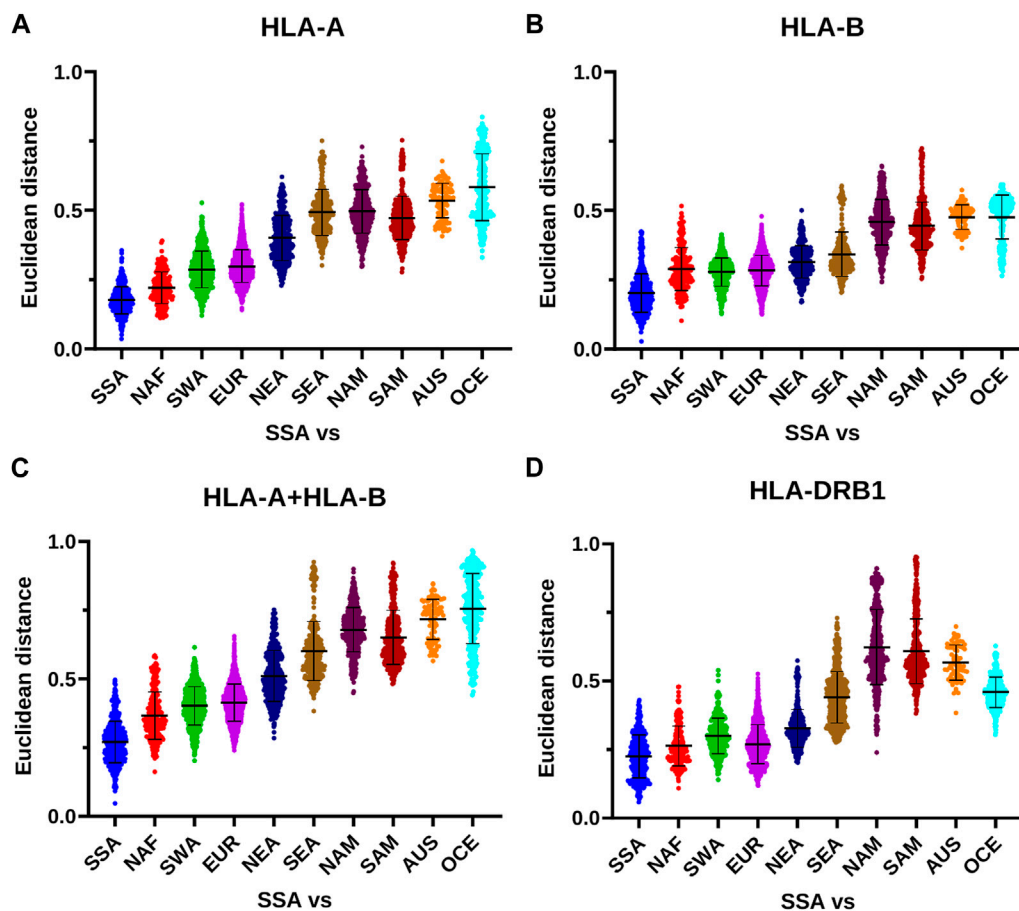


FIGURE 7

HLA class I and class II distances between SSA and other continental subgroups. (A–D) The PCoA Euclidean distances calculated on the frequencies of *HLA-A*, *HLA-B* (both independently and together), and *HLA-DRB1* allele groups among SSA populations (including SSA-descendant populations in the Americas) and between these and other non-SSA populations are plotted according to continental subgroups. Horizontal bar represents the mean distance with its standard deviation. AUS, Australia; EUR, Europe; NAF, North Africa; NAM, North America; NEA, Northeast Asia; OCE, Oceania; SAM, South America; SEA, Southeast Asia; SWA, Southwest Asia; SSA, Sub-Saharan Africa.

TABLE 2 Euclidean distances between continental groups calculated from *HLA-A* frequencies.

SSA vs.	SSA	NAF	SWA	EUR	NEA	SEA	NAM	SAM	AUS	OCE
Mean	0.176	0.221	0.287	0.299	0.400	0.492	0.495	0.473	0.535	0.583
SD	0.049	0.057	0.066	0.059	0.081	0.083	0.079	0.078	0.063	0.121
Minimum	0.036	0.112	0.121	0.141	0.229	0.301	0.297	0.277	0.407	0.330
Maximum	0.356	0.390	0.527	0.521	0.621	0.751	0.729	0.753	0.678	0.837

All intercontinental comparisons except SSAvsSEA, vs. SSAvsNAM, were statistically significant ($p < 0.01$) in the ANOVA. AUS, australia; EUR, europe; NAF, north africa; NAM, north america; NEA, northeast asia; OCE, oceania; SAM, south america; SEA, southeast asia; SWA, southwest asia; SSA, Sub-Saharan Africa; SD, standard deviation.

corresponds with the Wallace line that biogeographically divides Asia and Oceania (White et al., 2021), which could reflect ecological differences in both regions posing also different immune challenges to the inhabitants on both sides of the line. In turn, in the Americas (Supplementary Figure S9D) strong barriers are restricted to the extreme north, between Alaska Natives and Native American populations to the south, and between some SAM regions.

When both class I and *HLA-DRB1* frequencies are studied together (Figures 9A–D), statistically significant barriers mostly resemble the ones observed using class I only. However, some additional barriers become apparent including one arising between populations in Arctic Europe and those to the south (Figure 9B), and one between AUS and Near Polynesia and the rest of OCE (Figure 9C). In the Americas, the strongest barriers found are still in the extreme north, and between SAM populations

TABLE 3 Euclidean distances between continental groups calculated from *HLA-B* frequencies.

SSA vs.	SSA	NAF	SWA	EUR	NEA	SEA	NAM	SAM	AUS	OCE
Mean	0.203	0.289	0.278	0.283	0.315	0.342	0.458	0.444	0.476	0.477
SD	0.070	0.077	0.051	0.056	0.058	0.080	0.082	0.086	0.044	0.080
Minimum	0.028	0.102	0.128	0.125	0.170	0.204	0.242	0.254	0.365	0.265
Maximum	0.424	0.516	0.413	0.479	0.500	0.589	0.661	0.724	0.574	0.593

All intercontinental comparisons except SSAvsNAF, vs. SSAvsSWA, SSAvsNAF, vs. SSAvsEUR, SSAvsSWA, vs. SSAvsEUR, SSAvsAUS, vs. SSAvsOCE, and SSAvsAUS, vs. SSAvsNAM, were statistically significant ($p < 0.05$) in the ANOVA. AUS, australia; EUR, europe; NAF, north africa; NAM, north america; NEA, northeast asia; OCE, oceania; SAM, south america; SEA, southeast asia; SWA, southwest asia; SSA, Sub-Saharan Africa; SD, standard deviation.

TABLE 4 Euclidean distances between continental groups calculated from *HLA-A* and *HLA-B* frequencies.

SSA vs.	SSA	NAF	SWA	EUR	NEA	SEA	NAM	SAM	AUS	OCE
Mean	0.271	0.366	0.402	0.414	0.510	0.601	0.680	0.652	0.717	0.756
SD	0.076	0.086	0.070	0.067	0.093	0.108	0.081	0.098	0.073	0.128
Minimum	0.047	0.162	0.203	0.240	0.285	0.383	0.449	0.482	0.565	0.442
Maximum	0.496	0.585	0.615	0.655	0.752	0.925	0.900	0.922	0.847	0.967

All intercontinental comparisons were statistically significant ($p < 0.01$) in the ANOVA. AUS, australia; EUR, europe; NAF, north africa; NAM, north america; NEA, northeast asia; OCE, oceania; SAM, south america; SEA, southeast asia; SWA, southwest asia; SSA, Sub-Saharan Africa; SD, standard deviation.

TABLE 5 Euclidean distances between continental groups calculated from *HLA-DRB1* frequencies.

SSA vs.	SSA	NAF	SWA	EUR	NEA	SEA	NAM	SAM	AUS	OCE
Mean	0.226	0.263	0.300	0.270	0.328	0.441	0.624	0.608	0.567	0.459
SD	0.079	0.073	0.065	0.071	0.069	0.093	0.137	0.119	0.064	0.055
Minimum	0.058	0.109	0.140	0.119	0.203	0.276	0.239	0.382	0.384	0.304
Maximum	0.431	0.479	0.539	0.526	0.574	0.730	0.910	0.953	0.699	0.628

All intercontinental comparisons except SSAvsNAF, vs. SSAvsEUR, SSAvsSEA, vs. SSAvsOCE, and SSAvsNAM, vs. SSAvsSAM, were statistically significant ($p < 0.01$) in the ANOVA. AUS, australia; EUR, europe; NAF, north africa; NAM, north america; NEA, northeast asia; OCE, oceania; SAM, south america; SEA, southeast asia; SWA, southwest asia; SSA, Sub-Saharan Africa; SD, standard deviation.

separating those further south from Andean ones to the north, and those from populations in the South American lowlands to the east (Figure 9D).

Discussion

In this study, we performed an analysis of global patterns of HLA diversity by analyzing two unique datasets of HLA class I and class II allele group frequencies in more than 200 population samples from around the world. Our clustering and multidimensional scaling analyses show how geographic closeness correlates with similar patterns of HLA variation, especially for class I. In addition, we show that PCoA results for class I resemble geographic locations in Afro-Eurasia, while class II patterns follow a more continuous pattern, less reminiscent of actual geographic location. Furthermore, we confirm a progressive restriction of HLA diversity according to distance from the African continent, with regions with extreme population differentiation such as AUS, OCE, NAM, and SAM showing the most restricted allele group

repertoires. With *Barrier* analysis, we could also detect genetic discontinuities between well-defined ecological regions, a reminiscence of both distant and recent past evolutionary histories. Finally, we confirm HLA signatures indicative of migration patterns between non-native populations and their ancestral homelands, both ancient and more recent. A world map summarizing the main global HLA population relations derived from these analyses is shown in Figure 9.

The results of our analyses at the allele group level are congruent with a major role of demographic events in the shaping of the HLA profiles of human populations as previously reported using higher-resolution allelic frequencies or molecular variation (Prugnolle et al., 2005; Nunes et al., 2010; Buhler and Sanchez-Mazas, 2011; Sanchez-Mazas et al., 2012a; Riccio et al., 2013), especially for HLA class I. These patterns correlate well with the distance and main dispersal routes from the African continent as has been previously observed (Buhler and Sanchez-Mazas, 2011; Di and Sanchez-Mazas, 2011; 2014). Moreover, main geographic barriers between continental and intracontinental regions, such as the

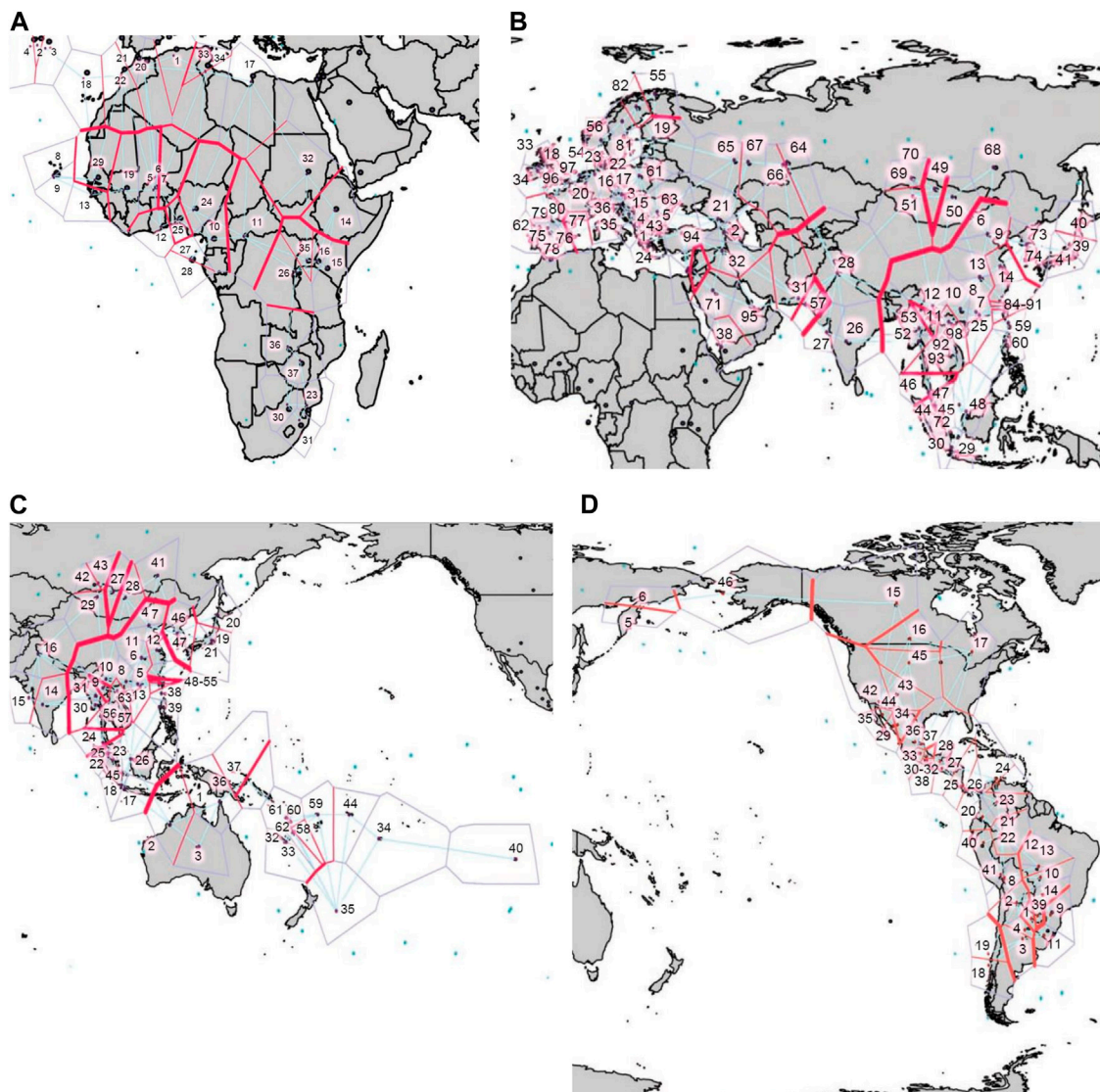


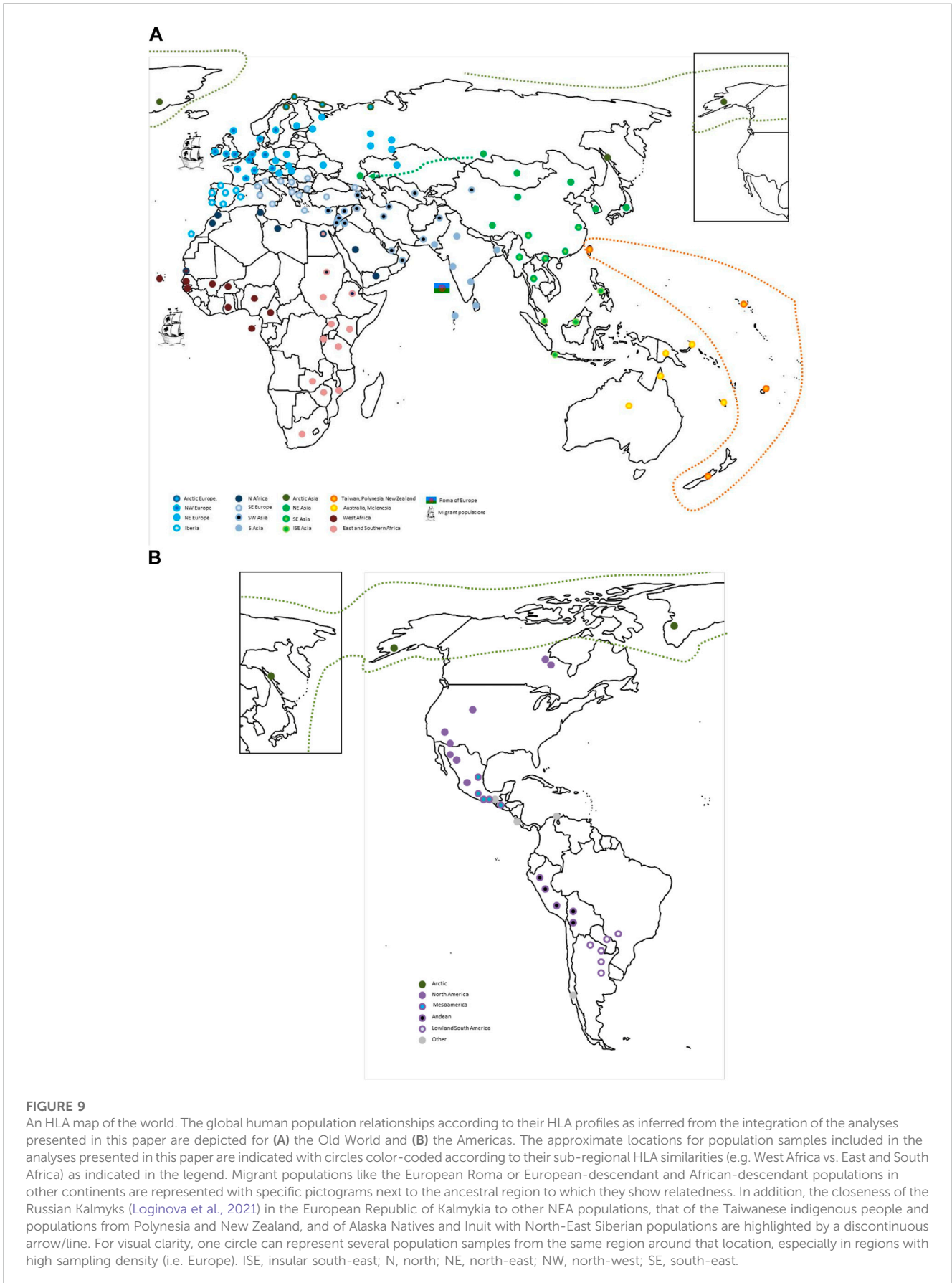
FIGURE 8

Barrier analysis for *HLA-A*, *-B*, and *-DRB1*. Plots show the statistically significant genetic barriers (thick red) determined with the joint analysis of *HLA-A*, *-B*, and *-DRB1* first-field frequencies for populations located in (A). Africa, (B). Eurasia, (C). East Asia and the Pacific, and (D). the Americas. Output of Barrier software was aligned to maps of each region to facilitate correlation with geographical context. Numbers correspond to the populations included in the Barrier analyses (Supplementary Tables S3-6).

Alps in Europe (Novembre et al., 2008; Nunes et al., 2010; Sanchez-Mazas et al., 2013) or the Saharan desert in Africa, have a clear imprint on HLA profiles analyzed here as has been previously described for both HLA and other genetic systems. However, an effect of region-specific pathogen-driven selection on the shaping of HLA population profiles has also been demonstrated (Sanchez-Mazas et al., 2017a) even at the supertype level for the *HLA-B* locus (Dos Santos Francisco et al., 2015), suggesting that HLA locus-specific demographic and selective events can interact with each other in a given geographic context (Prugnolle et al., 2005).

Our results hint to a different strength of correspondence between HLA class I and class II frequency profiles and geographic location at the first-field level, suggesting that the interplay between demographic and selective pressures can have different effects for these genes. Of

note, previous studies have found significant correlations between geographic distances and both class I and class II allelic and molecular frequencies, and pointed to a stronger similarity between the *HLA-B* and *HLA-DRB1* polymorphism and diversity (Solberg et al., 2008; Buhler and Sanchez-Mazas, 2011; Di and Sanchez-Mazas, 2014). However, previous work has also observed differential correlations between pathogen richness and HLA class I and class II diversities (Prugnolle et al., 2005; Sanchez-Mazas et al., 2012a), with *HLA-A* and *HLA-B* showing a positive correlation for heterozygosity and viral pathogens, in turn absent for *HLA-DRB1* (Sanchez-Mazas et al., 2012a). Except for the different splitting order in the clustering and the less pronounced increases in the PCoA distances across continental groups for *HLA-B*, our results examining *HLA-A* and *HLA-B* independently were mostly congruent and similar to the joint



analysis. This could be due to differences in the level of resolution and the methodologies applied in this work. It is however conceivable that the more stringent peptide-binding properties of class I molecules have led to the larger number of alleles (and broader repertoire of allele groups) resulting in a better and more similar representation of the geographic location of populations for class II than *HLA-DRB1* at this level of resolution. Indeed, a proposed model of joint divergent asymmetric selection operating between *HLA-A* and *HLA-B* could account for their increased power to discriminate human populations according to their pathogen histories (Buhler et al., 2016; Di et al., 2021), in particular when analyzed together.

Interestingly, the analysis of HLA barriers among populations provides additional insights in this respect. The analysis of individual loci detected in many cases conserved genetic discontinuity patterns, coupled with locus-specific discontinuity regions. For HLA class I (and thus, adaptation to and coevolution with intracellular pathogens) statistically significant barriers at the continental level were indeed found when both loci were analyzed together. Although we cannot rule out an effect of a reduced repertoire of first-field allele groups for *HLA-DRB1* in the lack of statistical confirmation of class II discontinuities, it is conceivable that once more the biological differences between both classes and their specific interaction with both demographic and infectious factors explain the differences in the strengths of genetic barrier development. Interestingly, several discontinuities become apparent when a combined class I + class II analysis is performed, suggesting that both genetic regions could be simultaneously affected by specific events in a given geographical context leading to system-wide barriers. Overall, the patterns observed here likely correspond to a mixture of immune adaptation intertwined with demographic and historical events. Since we are dealing with present-day data, our observations constitute just a snapshot of an intricate interplay of pathogen and human coevolving taking place for the past hundreds of thousands of years. Finally, it should be noted that the population sampling density in specific regions could affect the detection of statistically significant barriers, and that improved sampling could help confirm the genetic discontinuities observed.

The HLA frequency data analyzed here also reflect relatively recent as well as more ancient migration patterns. While the clustering and PCoA proximity of African- and European-descendant populations in the Americas, South Africa, and Australia with West African and North-western European populations, respectively, is not surprising, the fact that our analyses were also able to detect population relatedness between Jewish diaspora and European Romani populations and those of present day Eastern Mediterranean and South Asian regions, respectively, attest to the persistence of their original HLA signatures despite genetic drift and longer contact and gene flow with neighboring populations, as has been suggested by previous HLA (Klitz et al., 2010; Inotai et al., 2015) and non-HLA (Hammer et al., 2000; Atzmon et al., 2010; Bianco et al., 2020)-based studies. Similarly, the resemblance between HLA profiles of Polynesian, Rapa Nui, and Maori population samples to those of Taiwanese populations are further examples of the capacity of HLA data to detect these sequential, long-range migration patterns as previously noted (Friedlaender et al., 2008; Riccio et al., 2013; Ioannidis et al., 2021).

Apart from the recent and more ancient migration patterns, admixture events and gene flow are evident in the form of genetic clines in populations in intra or intercontinental contact zones (Di and Sanchez-Mazas, 2011; 2014; Hellenthal et al., 2014). Examples of these events that are evident from their HLA fingerprints in our frequency heat maps and their proximity in PCoA and clustering plots and that have been previously observed by others include the relatedness of the European Sámi with other Arctic and Siberian populations (Riccio et al., 2013; Lamnidis et al., 2018), the closeness between Egyptian and East African populations (Schuenemann et al., 2017), and the Australian-Papuan-Polynesian allele group signatures prevalent also in other OCE and SEA populations (Friedlaender et al., 2008).

The reduction in the numbers of HLA allele groups detected for AUS, OCE, NAM, and SAM, resulting in high frequencies in those groups reflect the intense population differentiation in these regions (Mack et al., 2007; Solberg et al., 2008). Previous research had observed this reduction in diversity at the allelic level (Solberg et al., 2008) and both in terms of expected heterozygosity and nucleotide diversity (Buhler and Sanchez-Mazas, 2011). This restriction is especially strong for Native Americans, and is reflected in the complex relation between geography and HLA profiles in the American continent, which seem less driven by geographical proximity, as previously observed (Nunes et al., 2010; Buhler and Sanchez-Mazas, 2011; Barquera et al., 2020b). One likely explanation for this situation in the Americas is the presence of different subsequent waves of colonization, all characterized by various degrees of strong founder effects and genetic drift resulting in high levels of population differentiation (Wang et al., 2007; Reich et al., 2012). Moreover, post-colonial Native American populations are also characterized by low effective population size, relative isolation (Barrantes, 1993), and differing degrees of European and African gene flow (Hollenbach et al., 2001; Arrieta-Bolanos et al., 2018; Barquera et al., 2020b), which further contribute to their complexity. Of note, these populations also show the highest number of private *HLA-DRB1* alleles (Gonzalez-Galarza et al., 2021), which account for a high proportion of their allele pools (Meyer et al., 2006) and are thought to be the result of selective forces acting in a context of severely reduced HLA diversity (Titus-Trachtenberg et al., 1994; Erlich et al., 1997; Hollenbach et al., 2001). In fact, a mixture of long-term balancing selection and a combination of episodes of more recent positive and balancing selection has been associated with simultaneously increasing intra-population diversity and inter-population differentiation in Native Americans (Single et al., 2020; Nunes et al., 2021).

Our study includes a unique sample of human populations and HLA frequency data providing a comprehensive view on worldwide HLA variation patterns; however some aspects remain as limitations for the approach followed. Although sufficient for identifying the main patterns of population differentiation at the global level (Di et al., 2017), the use of first-field molecular frequency data could have masked finer variation patterns characterizing specific ethnic groups or populations at higher levels of resolution (Creary et al., 2021; Nunes et al., 2022). In addition, a lack of genotypic data for many of the datasets precludes assessment of other relevant population genetics parameters (Sanchez-Mazas et al., 2012b). In particular, evidence of HWE was not used as criterion for population

sample inclusion. Despite this, the majority of the populations included (61% and 63% for class I and *HLA-DRB1*, respectively) were in HWE, and only a small minority showed deviations (5% and 10%, respectively). These deviations were in most cases for one of the loci, at higher resolution, and had low statistical significance that may have disappeared if correction for multiple testing had been applied by in the original studies. Exclusion of those samples from the analyses did not impact the main observations made at the global and continental levels (not shown). Despite this, the results for samples with deviations and of those in which testing was not done/not reported by the original authors (34% and 25%, for class I and *HLA-DRB1*, respectively) should be taken with caution. An examination of whether the deviations found arise from an effect of sampling or as a consequence of a biological or demographic phenomenon are of interest but out of the scope of the present work.

Due to their significant variation across human populations (Maiers et al., 2007; Gragert et al., 2013) and their conserved nature, it is expected that HLA extended haplotype frequencies be more powerful than allele or allele group frequencies in segregating human populations (Askar et al., 2013; Hollenbach et al., 2015; Arrieta-Bolanos et al., 2018). However, a significant proportion of published HLA frequency data is still of low or intermediate resolution (Gonzalez-Galarza et al., 2021), especially for class I, and haplotype frequency data is much more scarce and subject to important technical limitations (Single et al., 2002; Eberhard et al., 2013; Nunes, 2021). Hence, the homogenization of HLA frequency data to the first field allows for the inclusion of many more datasets, while also reducing the potential confounding effects of sampling, variations in sample size, and different epochs of HLA typing. Newer third and fourth-field high-resolution HLA data obtained with next-generation sequencing methods (Creary et al., 2021; Nunes et al., 2022) will likely further refine the population comparisons based on this genetic system. Moreover, although the inclusion of as many populations from as many geographical regions was striven for, some regions are still much better represented than others due to better sampling. This is particularly important for genetically complex regions like Central Asia (Yunusbayev et al., 2015; Jeong et al., 2019), where paucity of HLA frequency data, of which we have included some recent studies (Shen et al., 2010; Hajjej et al., 2020; Wang et al., 2021), remains a major gap (Gonzalez-Galarza et al., 2021). Hence, future inclusion of additional populations from that and other underrepresented regions will likely improve the comprehensiveness of the HLA global variation picture. Lastly, because our results are based on modern data the effects of recent migration, relatively recent selection events, and the demographic history of the last five centuries cannot be completely dissected from more ancient events. With the aid of ancient DNA and paleogenomics (Quintana-Murci, 2019; Barquera and Krause, 2020), we may be able to better study true selection events and the effects of past pandemics and epidemics on the HLA profiles of human populations across time and space.

In conclusion, this large comparison of worldwide HLA first-field frequencies contributes to our understanding of the global

population relations for this genetic system, confirming previous results based on high-resolution data and further demonstrating their remarkable similarity to geography, albeit with different intensities between class I and class II loci. In addition, the unique HLA dataset analyzed here also reflects specific patterns of extreme differentiation, genetic barriers and admixture, and ancient migrations and diasporas, all of which have implications for both human evolutionary studies and biomedical research.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <http://www.allelefrequencies.net/hla.asp>.

Author contributions

EA-B, DH-Z, and RB collected the datasets, performed analyses, interpreted the results, and wrote the manuscript.

Acknowledgments

The authors are grateful to Cosimo Posth and Rita Radzeviciute for their comments, which are part of the final version of this manuscript. The authors want to acknowledge the support received by the Open Access Publication Fund of the University of Duisburg-Essen.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2023.866407/full#supplementary-material>

References

- Aamer, W., Hassan, H. Y., Nakaoka, H., Hosomichi, K., Jaeger, M., Tahir, H., et al. (2021). Analysis of HLA gene polymorphisms in East Africans reveals evidence of gene flow in two Semitic populations from Sudan. *Eur. J. Hum. Genet.* 29 (8), 1259–1271. doi:10.1038/s41431-021-00845-6
- Abi-Rached, L., Jobin, M. J., Kulkarni, S., McWhinnie, A., Dalva, K., Gragert, L., et al. (2011). The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science* 334 (6052), 89–94. doi:10.1126/science.1209202
- Arrieta-Bolanos, E., Madrigal, J. A., and Shaw, B. E. (2012). Human leukocyte antigen profiles of Latin American populations: Differential admixture and its potential impact on hematopoietic stem cell transplantation. *Bone Marrow Res.* 2012, 136087. doi:10.1155/2012/136087
- Arrieta-Bolanos, E., Madrigal-Sanchez, J. J., Stein, J. E., Arrieta-Molina, G., Grant, S., Salazar-Sanchez, L., et al. (2019). 4-Locus high-resolution HLA allele and haplotype frequencies in Costa Ricans from African-Caribbean descent. *Hum. Immunol.* 80 (7), 411–412. doi:10.1016/j.humimm.2019.05.007
- Arrieta-Bolanos, E., Madrigal-Sanchez, J. J., Stein, J. E., Orlich-Perez, P., Moreira-Espinoza, M. J., Paredes-Carias, E., et al. (2018). 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 (6), 514–529. doi:10.1111/tan.13280
- Arrieta-Bolanos, E., Oliveira, D. C., and Barquera, R. (2020). Differential admixture, human leukocyte antigen diversity, and hematopoietic cell transplantation in Latin America: Challenges and opportunities. *Bone Marrow Transpl.* 55 (3), 496–504. doi:10.1038/s41409-019-0737-4
- Askar, M., Daghestani, J., Thomas, D., Leahy, N., Dunn, P., Claas, F., et al. (2013). 16(th) IHW: Global distribution of extended HLA haplotypes. *Int. J. Immunogenet.* 40 (1), 31–38. doi:10.1111/iji.12029
- Atzmon, G., Hao, L., Pe'er, I., Velez, C., Pearlman, A., Palamara, P. F., et al. (2010). Abraham's children in the genome era: Major Jewish diaspora populations comprise distinct genetic clusters with shared middle eastern ancestry. *Am. J. Hum. Genet.* 86 (6), 850–859. doi:10.1016/j.ajhg.2010.04.015
- Auton, A., Bryc, K., Boyko, A. R., Lohmueller, K. E., Novembre, J., Reynolds, A., et al. (2009). Global distribution of genomic diversity underscores rich complex history of continental human populations. *Genome Res.* 19 (5), 795–803. doi:10.1101/gr.088898.108
- Baker, J. L., Rotimi, C. N., and Shriner, D. (2017). Human ancestry correlates with language and reveals that race is not an objective genomic classifier. *Sci. Rep.* 7 (1), 1572. doi:10.1038/s41598-017-01837-7
- Bannai, M., Ohashi, J., Hariharu, S., Takahashi, Y., Juji, T., Omoto, K., et al. (2000). Analysis of HLA genes and haplotypes in Ainu (from Hokkaido, northern Japan) supports the premise that they descent from Upper Paleolithic populations of East Asia. *Tissue Antigens* 55 (2), 128–139. doi:10.1034/j.1399-0039.2000.550204.x
- Bannai, M., Tokunaga, K., Imanishi, T., Hariharu, S., Fujisawa, K., Juji, T., et al. (1996). HLA class II alleles in Ainu living in Hidaka District, Hokkaido, northern Japan. *Am. J. Phys. Anthropol.* 101 (1), 1–9. doi:10.1002/(SICI)1096-8644(199609)101:1<1::AID-AJPA1>3.0.CO;2-Z
- Barquera, R., Hernandez-Zaragoza, D. I., Bravo-Acevedo, A., Arrieta-Bolanos, E., Clayton, S., Acuna-Alonzo, V., et al. (2020a). The immunogenetic diversity of the HLA system in Mexico correlates with underlying population genetic structure. *Hum. Immunol.* 81, 461–474. doi:10.1016/j.humimm.2020.06.008
- Barquera, R., and Krause, J. (2020). An ancient view on host pathogen interaction across time and space. *Curr. Opin. Immunol.* 65, 65–69. doi:10.1016/j.coi.2020.05.004
- Barquera, R., Zuniga, J., Flores-Rivera, J., Corona, T., Penman, B. S., Hernandez-Zaragoza, D. I., et al. (2020b). Diversity of HLA class I and class II blocks and conserved extended haplotypes in lacandon mayans. *Sci. Rep.* 10 (1), 3248. doi:10.1038/s41598-020-58897-5
- Barrantes, R. (1993). Genetic diversity and racial mixture in Amerindians from Costa Rica and Panama. *Rev. Biol. Trop.* 41 (3A), 379–384.
- Bera, O., Cesaire, R., Quelvenec, E., Quillivic, F., de Chavigny, V., Ribal, C., et al. (2001). HLA class I and class II allele and haplotype diversity in Martinicans. *Tissue Antigens* 57 (3), 200–207. doi:10.1034/j.1399-0039.2001.057003200.x
- Bianco, E., Laval, G., Font-Porterias, N., Garcia-Fernandez, C., Dobon, B., Sabido-Vera, R., et al. (2020). Recent common origin, reduced population size, and marked admixture have shaped European roma genomes. *Mol. Biol. Evol.* 37 (11), 3175–3187. doi:10.1093/molbev/msaa156
- Buhler, S., Nunes, J. M., and Sanchez-Mazas, A. (2016). HLA class I molecular variation and peptide-binding properties suggest a model of joint divergent asymmetric selection. *Immunogenetics* 68 (6-7), 401–416. doi:10.1007/s00251-016-0918-x
- Buhler, S., and Sanchez-Mazas, A. (2011). HLA DNA sequence variation among human populations: Molecular signatures of demographic and selective events. *PLoS One* 6 (2), e14643. doi:10.1371/journal.pone.0014643
- Creary, L. E., Sacchi, N., Mazzocco, M., Morris, G. P., Montero-Martin, G., Chong, W., et al. (2021). High-resolution HLA allele and haplotype frequencies in several unrelated populations determined by next generation sequencing: 17th international HLA and Immunogenetics workshop joint report. *Hum. Immunol.* 82 (7), 505–522. doi:10.1016/j.humimm.2021.04.007
- Di, D., Nunes, J. M., Jiang, W., and Sanchez-Mazas, A. (2021). Like wings of a bird: Functional divergence and complementarity between HLA-A and HLA-B molecules. *Mol. Biol. Evol.* 38 (4), 1580–1594. doi:10.1093/molbev/msaa325
- Di, D., Nunes, J. M., and Sanchez-Mazas, A. (2017). The influence of HLA resolution level on population comparisons strongly depends on loci and geographic ranges. *HLA* 89 (6), 371.
- Di, D., and Sanchez-Mazas, A. (2011). Challenging views on the peopling history of East Asia: The story according to HLA markers. *Am. J. Phys. Anthropol.* 145 (1), 81–96. doi:10.1002/ajpa.21470
- Di, D., and Sanchez-Mazas, A. (2014). HLA variation reveals genetic continuity rather than population group structure in East Asia. *Immunogenetics* 66 (3), 153–160. doi:10.1007/s00251-014-0757-6
- Dos Santos Francisco, R., Buhler, S., Nunes, J. M., Bitarello, B. D., Franca, G. S., Meyer, D., et al. (2015). HLA supertype variation across populations: New insights into the role of natural selection in the evolution of HLA-A and HLA-B polymorphisms. *Immunogenetics* 67 (11-12), 651–663. doi:10.1007/s00251-015-0875-9
- Eberhard, H. P., Madbouly, A. S., Gourraud, P. A., Balere, M. L., Feldmann, U., Gragert, L., et al. (2013). Comparative validation of computer programs for haplotype frequency estimation from donor registry data. *Tissue Antigens* 82 (2), 93–105. doi:10.1111/tan.12160
- Edinur, H. A., Dunn, P. P., Hammond, L., Selwyn, C., Velickovic, Z. M., Lea, R. A., et al. (2012). Using HLA loci to inform ancestry and health in Polynesian and Maori populations. *Tissue Antigens* 80 (6), 509–522. doi:10.1111/tan.12026
- Erlich, H. A., Mack, S. J., Bergstrom, T., and Gyllensten, U. B. (1997). HLA class II alleles in amerindian populations: Implications for the evolution of HLA polymorphism and the colonization of the Americas. *Hereditas* 127 (1-2), 19–24. doi:10.1111/j.1601-5223.1997.00019.x
- Evsееva, I., Spurkland, A., Thorsby, E., Smerdel, A., Tranebjaerg, L., Boldyreva, M., et al. (2002). HLA profile of three ethnic groups living in the North-Western region of Russia. *Tissue Antigens* 59 (1), 38–43. doi:10.1034/j.1399-0039.2002.590107.x
- Excoffier, L., and Lischer, H. E. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under linux and windows. *Mol. Ecol. Resour.* 10 (3), 564–567. doi:10.1111/j.1755-0998.2010.02847.x
- Fernandez Vina, M. A., Hollenbach, J. A., Lyke, K. E., Szein, M. B., Maiers, M., Klitz, W., et al. (2012). Tracking human migrations by the analysis of the distribution of HLA alleles, lineages and haplotypes in closed and open populations. *Philos. Trans. R. Soc. Lond B Biol. Sci.* 367 (1590), 820–829. doi:10.1098/rstb.2011.0320
- Friedlaender, J. S., Friedlaender, F. R., Reed, F. A., Kidd, K. K., Kidd, J. R., Chambers, G. K., et al. (2008). The genetic structure of Pacific Islanders. *PLoS Genet.* 4 (1), e19. doi:10.1371/journal.pgen.0040019
- Gao, X., Lester, S., Boetcher, B., and McCluskey, J. (1997). "Diversity of HLA genes in populations of Australia and the Pacific." in Proceedings of the Twelfth International Histocompatibility Workshop and Conference (EDK).
- Genomes Project, C., Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., et al. (2015). A global reference for human genetic variation. *Nature* 526 (7571), 68–74. doi:10.1038/nature15393
- Ghodke, Y., Joshi, K., Chopra, A., and Patwardhan, B. (2005). HLA and disease. *Eur. J. Epidemiol.* 20 (6), 475–488. doi:10.1007/s10654-005-5081-x
- Gonzalez-Galarza, F. F., McCabe, A., Melo Dos Santos, E. J., Jones, A. R., and Middleton, D. (2021). A snapshot of human leukocyte antigen (HLA) diversity using data from the Allele Frequency Net Database. *Hum. Immunol.* 82 (7), 496–504. doi:10.1016/j.humimm.2020.10.004
- Gonzalez-Galarza, F. F., McCabe, A., Santos, E., Jones, J., Takeshita, L., Ortega-Rivera, N. D., et al. (2020). Allele frequency net database (AFND) 2020 update: Gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Res.* 48 (D1), D783–D788. doi:10.1093/nar/gkz1029
- Gragert, L., Madbouly, A., Freeman, J., and Maiers, M. (2013). Six-locus high resolution HLA haplotype frequencies derived from mixed-resolution DNA typing for the entire US donor registry. *Hum. Immunol.* 74 (10), 1313–1320. doi:10.1016/j.humimm.2013.06.025
- Grunnet, N., Steffensen, R., Varming, K., and Jersild, C. (1996). "HLA class I allele typed by PCR-ARMS in Greenlander of Eskimo origin," in Twelfth International Histocompatibility Workshop and Conference, Paris.
- Hajjaj, A., Abdrakhmanova, S., Turganbekova, A., and Almawi, W. Y. (2020). Distribution of HLA Class I and Class II alleles and haplotypes in German and Uzbek minorities in Kazakhstan, and relationship to other populations. *HLA* 96 (5), 615–620. doi:10.1111/tan.14057
- Hammer, M. F., Redd, A. J., Wood, E. T., Bonner, M. R., Jarjanazi, H., Karafet, T., et al. (2000). Jewish and Middle Eastern non-Jewish populations share a common pool of

- Y-chromosome biallelic haplotypes. *Proc. Natl. Acad. Sci. U. S. A.* 97 (12), 6769–6774. doi:10.1073/pnas.100115997
- Harbo, H. F., Riccio, M. E., Lorentzen, A. R., Utsi, E., Myhr, K. M., Mellgren, S. I., et al. (2010). Norwegian Sami differs significantly from other Norwegians according to their HLA profile. *Tissue Antigens* 75 (3), 207–217. doi:10.1111/j.1399-0039.2009.01425.x
- Hellenthal, G., Busby, G. B., Band, G., Wilson, J. F., Capelli, C., Falush, D., et al. (2014). A genetic atlas of human admixture history. *Science* 343 (6172), 747–751. doi:10.1126/science.1243518
- Hollenbach, J. A., Saperstein, A., Albrecht, M., Vierra-Green, C., Parham, P., Norman, P. J., et al. (2015). Race, ethnicity and ancestry in unrelated transplant matching for the national marrow donor Program: A comparison of multiple forms of self-identification with genetics. *PLoS One* 10 (8), e0135960. doi:10.1371/journal.pone.0135960
- Hollenbach, J. A., Thomson, G., Cao, K., Fernandez-Vina, M., Erlich, H. A., Bugawan, T. L., et al. (2021). HLA diversity, differentiation, and haplotype evolution in Mesoamerican Natives. *Hum. Immunol.* 62 (4), 378–390. doi:10.1016/s0198-8859(01)00212-9
- Inotai, D., Szilvasi, A., Benko, S., Boros-Major, A., Illes, Z., Bors, A., et al. (2015). HLA genetic diversity in Hungarians and Hungarian gypsies: Complementary differentiation patterns and demographic signals revealed by HLA-A, -B and -DRB1 in central Europe. *Tissue Antigens* 86 (2), 115–121. doi:10.1111/tan.12600
- Ioannidis, A. G., Blanco-Portillo, J., Sandoval, K., Hagelberg, E., Barberena-Jonas, C., Hill, A. V. S., et al. (2021). Paths and timings of the peopling of Polynesia inferred from genomic networks. *Nature* 597 (7877), 522–526. doi:10.1038/s41586-021-03902-8
- Jeong, C., Balanovsky, O., Lukianova, E., Kahbatkyy, N., Flegontov, P., Zaporozhchenko, V., et al. (2019). The genetic history of admixture across inner Eurasia. *Nat. Ecol. Evol.* 3 (6), 966–976. doi:10.1038/s41559-019-0878-2
- Johansson, A., Ingman, M., Mack, S. J., Erlich, H., and Gyllenstein, U. (2008). Genetic origin of the Swedish Sami inferred from HLA class I and class II allele frequencies. *Eur. J. Hum. Genet.* 16 (11), 1341–1349. doi:10.1038/ejhg.2008.88
- Kaiser, H. (1960). The application of electronic computers to factor analysis. *Psychol. Meas.* 20, 141–151. doi:10.1177/001316446002000116
- Klitz, W., Gragert, L., Maiers, M., Fernandez-Vina, M., Ben-Naeh, Y., Benedek, G., et al. (2010). Genetic differentiation of Jewish populations. *Tissue Antigens* 76 (6), 442–458. doi:10.1111/j.1399-0039.2010.01549.x
- Lamnidis, T. C., Majander, K., Jeong, C., Salmela, E., Wessman, A., Moiseyev, V., et al. (2018). Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. *Nat. Commun.* 9 (1), 5018. doi:10.1038/s41467-018-07483-5
- Leffell, M. S., Cherkh, W. S., Land, G., and Zachary, A. A. (2007). Improved definition of human leukocyte antigen frequencies among minorities and applicability to estimates of transplant compatibility. *Transplantation* 83 (7), 964–972. doi:10.1097/01.tp.0000258588.09356.d7
- Leffell, M. S., Fallin, M. D., Erlich, H. A., Fernandez-Vijna, M., Hildebrand, W. H., Mack, S. J., et al. (2002). HLA antigens, alleles and haplotypes among the yup'ik Alaska natives: Report of the ashii minority Workshops, Part II. *Hum. Immunol.* 63 (7), 614–625. doi:10.1016/s0198-8859(02)00415-9
- Li, J. Z., Absher, D. M., Tang, H., Southwick, A. M., Casto, A. M., Ramachandran, S., et al. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319 (5866), 1100–1104. doi:10.1126/science.1153717
- Loginova, M., Smirnova, D., Kutuyavina, S., Makhova, O., Kashin, K., and Paramonov, I. (2021). Kalmyks from republic of Kalmykia, Russia. *HLA* 97 (2), 177–179. doi:10.1111/tan.14177
- Lou, H., Li, H. C., Kuwayama, M., Yashiki, S., Fujiyoshi, T., Suehara, M., et al. (1998). HLA class I and class II of the Nivkhi, an indigenous population carrying HTLV-I in Sakhalin, Far Eastern Russia. *Tissue Antigens* 52 (5), 444–451. doi:10.1111/j.1399-0039.1998.tb03071.x
- Mack, S., Tsai, Y., Sanchez-Mazas, A., and Erlich, H. A. (2007). “Chapter 3: Anthropology/human genetic diversity population reports. 13th international histocompatibility workshop anthropology/human genetic diversity joint report,” in *Immunobiology of the human MHC: Proceedings of the 13th international histocompatibility workshop and conference*. Editor J. A. Hansen (Seattle, WA: IHWG Press), 580–652.
- Maiers, M., Gragert, L., and Klitz, W. (2007). High-resolution HLA alleles and haplotypes in the United States population. *Hum. Immunol.* 68 (9), 779–788. doi:10.1016/j.humimm.2007.04.005
- Main, P., Attenborough, R., Chelvanayagam, G., Bhatia, K., and Gao, X. (2001). The peopling of new Guinea: Evidence from class I human leukocyte antigen. *Hum. Biol.* 73 (3), 365–383. doi:10.1353/hub.2001.0036
- Manni, F., Guerard, E., and Heyer, E. (2004). Geographic patterns of (genetic, morphologic, linguistic) variation: How barriers can be detected by using monmonier's algorithm. *Hum. Biol.* 76 (2), 173–190. doi:10.1353/hub.2004.0034
- Manor, S., Halagan, M., Shriki, N., Yaniv, I., Zisser, B., Maiers, M., et al. (2016). High-resolution HLA A~B~DRB1 haplotype frequencies from the ezer mizion bone marrow donor registry in Israel. *Hum. Immunol.* 77 (12), 1114–1119. doi:10.1016/j.humimm.2016.09.004
- McColl, H., Racimo, F., Vinner, L., Demeter, F., Gakuhari, T., Moreno-Mayar, J. V., et al. (2018). The prehistoric peopling of Southeast Asia. *Science* 361 (6397), 88–92. doi:10.1126/science.aat3628
- Meyer, D., Single, R. M., Mack, S. J., Erlich, H. A., and Thomson, G. (2006). Signatures of demographic history and natural selection in the human major histocompatibility complex Loci. *Genetics* 173 (4), 2121–2142. doi:10.1534/genetics.105.052837
- Middleton, D., Williams, F., Meenagh, A., Daar, A. S., Gorodezky, C., Hammond, M., et al. (2000). Analysis of the distribution of HLA-A alleles in populations from five continents. *Hum. Immunol.* 61 (10), 1048–1052. doi:10.1016/s0198-8859(00)00178-6
- Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A. R., Auton, A., et al. (2008). Genes mirror geography within Europe. *Nature* 456 (7218), 98–101. doi:10.1038/nature07331
- Nunes, J. M. (2021). A comment on estimating HLA haplotype frequencies from homozygous individuals. *Int. J. Immunogenet* 48, 496–497. doi:10.1111/iji.12555
- Nunes, J. M., Buhler, S., Roessli, D., Sanchez-Mazas, A., and collaboration, H. L.-n. (2014). 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 (5), 307–323. doi:10.1111/tan.12356
- Nunes, J. M., Di, D., Gerbault, P., Faye, N. K., Almawi, W., Andreani, M., et al. (2022). A refined topography of the HLA molecular landscape across populations. *HLA* 99 (5), 441.
- Nunes, J. M., Riccio, M. E., Buhler, S., Di, D., Currat, M., Ries, F., et al. (2010). Analysis of the HLA population data (AHPD) submitted to the 15th International Histocompatibility/Immunogenetics Workshop by using the Gene[r]ate computer tools accommodating ambiguous data (AHPD project report). *Tissue Antigens* 76 (1), 18–30. doi:10.1111/j.1399-0039.2010.01469.x
- Nunes, K., Maia, M. H. T., Dos Santos, E. J. M., Dos Santos, S. E. B., Guerreiro, J. F., Petzl-Erler, M. L., et al. (2021). How natural selection shapes genetic differentiation in the MHC region: A case study with native Americans. *Hum. Immunol.* 82 (7), 523–531. doi:10.1016/j.humimm.2021.03.005
- Paximadis, M., Mathebula, T. Y., Gentle, N. L., Vardas, E., Colvin, M., Gray, C. M., et al. (2012). Human leukocyte antigen class I (A, B, C) and II (DRB1) diversity in the black and Caucasian South African population. *Hum. Immunol.* 73 (1), 80–92. doi:10.1016/j.humimm.2011.10.013
- Pidalá, J., Lee, S. J., Ahn, K. W., Spellman, S., Wang, H. L., Aljurf, M., et al. (2014). Nonpermissive HLA-DPB1 mismatch increases mortality after myeloablative unrelated allogeneic hematopoietic cell transplantation. *Blood* 124 (16), 2596–2606. doi:10.1182/blood-2014-05-576041
- Pirmohamed, M., Ostrov, D. A., and Park, B. K. (2015). New genetic findings lead the way to a better understanding of fundamental mechanisms of drug hypersensitivity. *J. Allergy Clin. Immunol.* 136 (2), 236–244. doi:10.1016/j.jaci.2015.06.022
- Prugnolle, F., Manica, A., Charpentier, M., Guegan, J. F., Guernier, V., and Balloux, F. (2005). Pathogen-driven selection and worldwide HLA class I diversity. *Curr. Biol.* 15 (11), 1022–1027. doi:10.1016/j.cub.2005.04.050
- Quintana-Murci, L. (2019). Human immunology through the lens of evolutionary genetics. *Cell* 177 (1), 184–199. doi:10.1016/j.cell.2019.02.033
- Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., et al. (2012). Reconstructing Native American population history. *Nature* 488 (7411), 370–374. doi:10.1038/nature11258
- Reyes-Centeno, H. (2016). Out of Africa and into Asia: Fossil and genetic evidence on modern human origins and dispersals. *Quat. Int.* 416, 249–262. doi:10.1016/j.quaint.2015.11.063
- Riccio, M. E., Buhler, S., Nunes, J. M., Vangenot, C., Cuenod, M., Currat, M., et al. (2013). 16th IHIW: Analysis of HLA population data, with updated results for 1996 to 2012 workshop data (AHPD project report). *Int. J. Immunogenet* 40 (1), 21–30. doi:10.1111/iji.12033
- Robinson, J., Barker, D. J., Georgiou, X., Cooper, M. A., Flicek, P., and Marsh, S. G. E. (2020). IPD-IMGT/HLA database. *Nucleic Acids Res.* 48 (D1), D948–D955. doi:10.1093/nar/gkz950
- Robinson, J., Guethlein, L. A., Cereb, N., Yang, S. Y., Norman, P. J., Marsh, S. G. E., et al. (2017). Distinguishing functional polymorphism from random variation in the sequences of >10,000 HLA-A, -B and -C alleles. *PLoS Genet.* 13 (6), e1006862. doi:10.1371/journal.pgen.1006862
- Romon, I., Montes, C., Ligeiro, D., Trindade, H., Sanchez-Mazas, A., Nunes, J. M., et al. (2016). Mapping the HLA diversity of the iberian peninsula. *Hum. Immunol.* 77 (10), 832–840. doi:10.1016/j.humimm.2016.06.023
- Sanchez-Mazas, A., Buhler, S., and Nunes, J. M. (2013). A new HLA map of Europe: Regional genetic variation and its implication for peopling history, disease-association studies and tissue transplantation. *Hum. Hered.* 76 (3-4), 162–177. doi:10.1159/000360855
- Sanchez-Mazas, A., Cerny, V., Di, D., Buhler, S., Podgorna, E., Chevallier, E., et al. (2017a). The HLA-B landscape of Africa: Signatures of pathogen-driven selection and molecular identification of candidate alleles to malaria protection. *Mol. Ecol.* 26 (22), 6238–6252. doi:10.1111/mec.14366

- Sanchez-Mazas, A., Fernandez-Vina, M., Middleton, D., Hollenbach, J. A., Buhler, S., Di, D., et al. (2011). Immunogenetics as a tool in anthropological studies. *Immunology Trans. R. Soc. Lond B Biol. Sci.* 367 (1590), 830–839. doi:10.1098/rstb.2011.0312
- Sanchez-Mazas, A., Lemaitre, J. F., and Currat, M. (2012a). Distinct evolutionary strategies of human leucocyte antigen loci in pathogen-rich environments. *Philos. Trans. R. Soc. Lond B Biol. Sci.* 367 (1590), 830–839. doi:10.1098/rstb.2011.0312
- Sanchez-Mazas, A., Nunes, J. M., Middleton, D., Sauter, J., Buhler, S., McCabe, A., et al. (2017b). Common and well-documented hla alleles over all of Europe and within European sub-regions: A catalogue from the European federation for Immunogenetics. *HLA* 89 (2), 104–113. doi:10.1111/tan.12956
- Sanchez-Mazas, A., Vidan-Jeras, B., Nunes, J. M., Fischer, G., Little, A. M., Bekmane, U., et al. (2012b). Strategies to work with HLA data in human populations for histocompatibility, clinical transplantation, epidemiology and population genetics: HLA-NET methodological recommendations. *Int. J. Immunogenet* 39 (6), 459–472; quiz 473–476. doi:10.1111/j.1744-313X.2012.01113.x
- Schuenemann, V. J., Peltzer, A., Welte, B., van Pelt, W. P., Molak, M., Wang, C. C., et al. (2017). Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods. *Nat. Commun.* 8, 15694. doi:10.1038/ncomms15694
- Shen, C. M., Zhu, B. F., Deng, Y. J., Ye, S. H., Yan, J. W., Yang, G., et al. (2010). Allele polymorphism and haplotype diversity of HLA-A, -B and -DRB1 loci in sequence-based typing for Chinese Uyghur ethnic group. *PLoS One* 5 (11), e13458. doi:10.1371/journal.pone.0013458
- Single, R. M., Meyer, D., Hollenbach, J. A., Nelson, M. P., Noble, J. A., Erlich, H. A., et al. (2002). Haplotype frequency estimation in patient populations: The effect of departures from hardy-weinberg proportions and collapsing over a locus in the HLA region. *Genet. Epidemiol.* 22 (2), 186–195. doi:10.1002/gepi.0163
- Single, R. M., Meyer, D., Nunes, K., Francisco, R. S., Hunemeier, T., Maiers, M., et al. (2020). Demographic history and selection at HLA loci in Native Americans. *PLoS One* 15 (11), e0241282. doi:10.1371/journal.pone.0241282
- Solberg, O. D., Mack, S. J., Lancaster, A. K., Single, R. M., Tsai, Y., Sanchez-Mazas, A., et al. (2008). Balancing selection and heterogeneity across the classical human leukocyte antigen loci: A meta-analytic review of 497 population studies. *Hum. Immunol.* 69 (7), 443–464. doi:10.1016/j.humimm.2008.05.001
- Takezaki, N., Nei, M., and Tamura, K. (2014). Poptree: Web version of POPTREE for constructing population trees from allele frequency data and computing some other quantities. *Mol. Biol. Evol.* 31 (6), 1622–1624. doi:10.1093/molbev/msu093
- Thorsby, E. (2012). The polynesian gene pool: An early contribution by amerindians to easter island. *Philos. Trans. R. Soc. Lond B Biol. Sci.* 367 (1590), 812–819. doi:10.1098/rstb.2011.0319
- Tiercy, J. M., and Claas, F. (2013). Impact of HLA diversity on donor selection in organ and stem cell transplantation. *Hum. Hered.* 76 (3-4), 178–186. doi:10.1159/000358798
- Titus-Trachtenberg, E. A., Rickards, O., De Stefano, G. F., and Erlich, H. A. (1994). Analysis of HLA class II haplotypes in the cayapa Indians of Ecuador: A novel DRB1 allele reveals evidence for convergent evolution and balancing selection at position 86. *Am. J. Hum. Genet.* 55 (1), 160–167.
- Wall, J. D., and Yoshihara Caldeira Brandt, D. (2016). Archaic admixture in human history. *Curr. Opin. Genet. Dev.* 41, 93–97. doi:10.1016/j.gde.2016.07.002
- Wang, H. D., Jin, X. Y., Yin, S. S., Zhang, Q., Su, J. X., Shen, C. M., et al. (2021). Diversities of HLA-A, -B, -C, -DRB1 and -DQB1 loci in Chinese kazak population and its genetic relatedness dissection with multiple populations: A comparative study. *Hum. Immunol.* 82 (4), 215–225. doi:10.1016/j.humimm.2020.06.006
- Wang, S., Lewis, C. M., Jakobsson, M., Ramachandran, S., Ray, N., Bedoya, G., et al. (2007). Genetic variation and population structure in native Americans. *PLoS Genet.* 3 (11), e185. doi:10.1371/journal.pgen.0030185
- White, A. E., Dey, K. K., Stephens, M., and Price, T. D. (2021). Dispersal syndromes drive the formation of biogeographical regions, illustrated by the case of Wallace's Line. *Global Ecol Biogeogr.* 30, 685–696. doi:10.1111/geb.13250
- Williams, F., Meenagh, A., Darke, C., Acosta, A., Daar, A. S., Gorodezky, C., et al. (2001). Analysis of the distribution of HLA-B alleles in populations from five continents. *Hum. Immunol.* 62 (6), 645–650. doi:10.1016/s0198-8859(01)00247-6
- Yunusbayev, B., Metspalu, M., Metspalu, E., Valeev, A., Litvinov, S., Valiev, R., et al. (2015). The genetic legacy of the expansion of Turkic-speaking nomads across Eurasia. *PLoS Genet.* 11 (4), e1005068. doi:10.1371/journal.pgen.1005068