Best Practice & Research Clinical Haematology 37 (2024) 101559

Contents lists available at ScienceDirect

Best Practice & Research Clinical Haematology

journal homepage: www.elsevier.com/locate/issn/15216926

The most frequent HLA alleles around the world: A fundamental synopsis

Alicia Sanchez-Mazas [a,*,1], José Manuel Nunes [a,**,1], PGAE HLA Consortium of the 18th International HLA and Immunogenetics Workshop [2]

[a] *Laboratory of Anthropology, Genetics and Peopling History (AGP), Department of Genetics and Evolution & Institute of Genetics and Genomics in Geneva (IGE3), University of Geneva, 30 Quai Ernest-Ansermet, 1205, Geneva, Switzerland*

ARTICLE INFO

Keywords:
HLA allele frequencies
HLA population genetics
HLA workshops
CWD catalogues
HLA databases
HLA biostatistical tools

ABSTRACT

A comprehensive knowledge of human leukocyte antigen (HLA) molecular variation worldwide is essential in human population genetics research and disease association studies and is also indispensable for clinical applications such as allogeneic hematopoietic cell transplantation, where ensuring HLA compatibility between donors and recipients is paramount. Enormous progress has been made in this field thanks to several decades of HLA population studies allowing the development of helpful databases and bioinformatics tools. However, it is still difficult to appraise the global HLA population diversity in a synthetic way. We thus introduce here a novel approach, based on approximately 2000 data sets, to assess this complexity by providing a fundamental synopsis of the most frequent HLA alleles observed in different regions of the world. This new knowledge will be useful not only as a fundamental reference for basic research, but also as an efficient guide for clinicians working in the field of transplantation.

1. Introduction - 30 years of HLA population studies: key achievements

1.1. The huge molecular diversity of the HLA genomic region

The Human Leukocyte Antigen (HLA) genomic region, comprising more than 140 coding genes on the short arm of chromosome 6, is one of the most polymorphic regions of the human genome, with a particularly high density of single nucleotide polymorphisms (SNPs) within the HLA genes [1]. The number of named HLA alleles according to the IPD-IMGT/HLA nomenclature is constantly increasing and has reached up to 26,610 Class I and 11,398 Class II alleles in 2024 [IPD-IMGT/HLA Release 3.55–2024-01, 2]. Each allele itself is made up of multiple SNPs, both synonymous and non-synonymous, and often multi-allelic, which can make them difficult to analyse [3]. Differences between humans at the level of HLA proteins (2nd field of resolution) are key to understanding specific immune responses to pathogens and adaptations to environmental stresses, explaining in great part the highly heterogeneous allele frequencies observed in populations [4], as described later in this article. However, these differences are also central in clinical

* Corresponding author.
** Corresponding author.
E-mail addresses: alicia.sanchez-mazas@unige.ch (A. Sanchez-Mazas), jose.deabreununes@unige.ch (J.M. Nunes).
[1] These two authors contributed equally to this work.
[2] List of collaborators (see Online version for the complete list of authors' names and affiliations).

https://doi.org/10.1016/j.beha.2024.101559
Received 23 April 2024; Received in revised form 10 June 2024; Accepted 27 June 2024
Available online 8 July 2024

settings, such as allogeneic hematopoietic cell transplantation (HCT), where they can induce antibody-mediated responses and increase the risk of graft-versus-host-disease (GVHD) in recipients of allogeneic transplants [5,6]. In addition, the 3rd and 4th field allelic variation (corresponding to silent mutations within and outside the HLA coding regions, respectively), while providing essential information for evolutionary studies [7], are also thought to play a role in HCT outcomes [8,9] by potentially regulating HLA expression levels [10,11]. Therefore, a thorough characterisation of HLA variation within and between populations is fundamental across several fields related to immunogenetics.

1.2. Collective projects and useful resources in HLA population genetics

A comprehensive understanding of the distribution of HLA variation within and across global populations has been achieved thanks to more than 30 years of HLA anthropological studies, much of which has been fostered by the International HLA and Immunogenetics Workshops (IHIWs). The first Anthropology Component of the HLA Workshops was launched at the 11th IHIW held in Yokohama, Japan, in 1991, where Imanishi's papers [12,13] became the first key references for populations' HLA frequencies and genetic relationships of populations worldwide. In 1996, the Anthropology Component of the 12th IHIW held in St-Malo (France) devoted a special attention to HLA diversity in African populations [14] under the leadership of Julia Bodmer [15]. The joint report of the Anthropology Component of the 13th IHIW held in Seattle (USA) in 2002 was presented in several complementary chapters providing

Table 1
Main resources related to HLA variation in populations.

Name & Reference(s)	Specificity	Strengths	Weaknesses
DATABASES			
Allelefrequencies.net (Gonzalez-Galarza et al., 2021 [4])	AFND repository of allele frequencies, mostly for HLA data.	- Big quantity of data sets; - Gold Standard datasets fulfil 3 minimal criteria of quality: 1) sums of allele frequencies = 1 ± 0.015 or ±0.05; 2) sample sizes >50 individuals; 3) HLA alleles defined at least at the 2nd-field level of resolution.	- Silver & Bronze Standard datasets (~60 % of the data) do not fulfil the 3 minimal criteria of quality; - HWE not required for data inclusion; - Heterogeneous data quality.
DATABASES AND STATISTICAL TOOLS			
hla-net.eu **Ihla-net.eu** (Nunes et al., 2014 [21]; Nunes 2016 [31]; https://hla-net.eu/ihlanet)	- Gene[VA] repository of HLA genotypic, allelic and haplotypic data from consecutive IHIWs; - Statistical tools for HLA data analysis: allele & haplotype frequency estimation, tests for HWE, LD & selective neutrality; - Tools for population comparisons: genetic distances & MDS.	- Validated allele frequencies recomputed from genotype data using uniform statistical tools; - HWE and convergence of EM frequency estimation taken as quality criteria for data inclusion; - User-friendly statistical tools for HLA data analysis, providing publishable tables, graphs & maps.	- Moderate quantity of available datasets due to data quality filtering.
PyPop 1.0.0 (Solberg et al., 2008 [17]; Lancaster et al., 2024 [29])	- Repository of HLA allele frequencies - Statistical tools for HLA data analysis: allele & haplotype frequency estimation, tests for HWE, LD & selective neutrality.	- HLA allele frequency maps - Useful statistical tools for HLA data analysis.	- Only 2nd-field level alleles reported in maps
PGG.MHC (Zhao et al., 2023 [32])	- Repository of HLA allele frequencies in worldwide populations - Tools for HLA imputation and case/control analyses	- Useful data visualization tools (allele population prevalence at 3 geographic levels, population haplotype structures)	- Vast majority of data for Chinese and other East Asian populations - Very low sample sizes for other populations
CATALOGUES OF COMMON (, INTERMEDIATE) AND WELL-DOCUMENTED ALLELES (CWD, CIWD)			
ASHI CWD 2.0.0 (Mack et al., 2013 [30])	CWD alleles detected at 9 loci (A, B, C, DRB1, DRB3/4/5, DQA1, DQB1, DPA1, DPB1) in 6 geographic regions (Australia, Brazil, China, France, Netherlands, USA).	Dataset of 139,961 SBT observations	- Vast majority of data (99 %) come from the USA (84 %) and China (15 %)
EFI CWD 1.0 (Sanchez-Mazas et al., 2017 [23])	CWD alleles detected at 7 loci (A, B, C, DRB1, DQA1, DQB1, DPB1) in Europe and 5 European sub-regions.	Large dataset of 21,571–3,966,984 individuals based on 3 repositories: GENE [VA], DKMS and AFND;	- Only 2nd-field level alleles reported
DKMS CWD (Eberhard et al., 2018 [26])	CWD alleles detected at 6 loci (A, B, C, DRB1, DQB1, DPB1) in the German population.	Large dataset of 5,104,477 individuals from 26 German HSC donor centers	- Only 2nd-field level alleles reported
CMDP CWD (He et al., 2018 [27])	CWD alleles detected at 5 loci (A, B, C, DRB1, DQB1) in the Chinese population.	Dataset of 3296 individuals	- Only 2nd-field level alleles reported
CIWD 3.0.0 (Hurley et al., 2020 [28])	CIWD alleles detected at 7 loci (A, B, C, DRB1, DRB3/4/5, DQB1, DPB1) in 7 population groups (US classifications).	Large dataset of 8,077,802 individuals from 20 donor registries	- Vast majority of data from individuals of European descent (>70 %), Asian/Pacific (>9 %) and unknown/multiple (>9 %) ancestry

methods as well as a substantial set of analyses based on high-resolution DNA-based genotyping of 95 populations worldwide [16]. Data from both the 12th and 13th IHIWs were further included in a key article on HLA population genetics [17], still highly cited today. The 14th IHIW held in Melbourne (Australia) in 2005 [18], and the 15th IHIW held in Buzios (Brazil) in 2008 [19], substantially increased the datasets of HLA-typed populations and stimulated the implementation of two major initiatives, allelefrequencies.net [20] and https://hla-net.eu [21]. The first aimed to create and maintain a public database of HLA allele frequency data collected from published papers, while the second was the main outcome of a COST Action (https://www.cost.eu/actions/BM0803/), which worked for 4 years (2009–2013) to develop a bioinformatics platform providing useful tools and guidelines for handling and analysing HLA population data in different disciplines involved in HLA research [19,22]. The implementation of *hla-net.eu* also led to the creation of a *Population Genetics Working Group* within the EFI Scientific Committee in 2014, whose first achievement was the publication of a European catalogue of Common and Well-Documented (CWD) HLA alleles [23]. The 16th IHIW held in Liverpool (UK) in 2012 [24] was key to the completion and release of the *hla-net.eu* platform, which is now widely used to estimate basic statistics for HLA population genetics studies. The 17th IHIW held in Asilomar (USA) in 2017 focused on HLA Next-Generation Sequencing (NGS) [25], which encouraged many laboratories to adopt NGS devices and (re)type population samples at very high resolution. A significant amount of such data was analysed by the *Population Genetics, Anthropology and Evolution (PGAE)* Component of the 18th IHIW held in Noordwijkerhout (The Netherlands) in 2022, where the new *ihla-net* tool was achieved to handle large amounts of highly complex HLA data (https://hla-net.eu/ihlanet). All these efforts have resulted in the creation and maintenance of large HLA population data repositories, CWD/CIWD catalogues and bioinformatics and data analysis tools [4,21,23,26–33], which have helped to better understand how populations differ from each other for each HLA gene and whether such differences diverge from those observed for other parts of the genome. These resources are presented with some additional comments in Table 1.

1.3. HLA population differentiations at different loci: what we have learned

As a result of significant advances in population genetic analyses focusing both on specific genomic regions or on genome-wide data, our scientific understanding supports the assertion that the observed molecular diversity between human populations is predominantly shaped by geographic factors [34]. Allele frequencies tend to vary continuously from sub-Saharan Africa to North Africa, the Middle-East and Europe, then to Northeast Asia and the Americas across the Bering Strait on one side, and to Southeast Asia and Oceania across the Pacific on the other side, with Northeast and Southeast Asian populations also becoming genetically distinct from each other along continuous clines maintained by gene flow [35]. Overall, this global pattern suggests greater genetic relatedness between geographically closer populations, consistent with the expansion of our species *Homo sapiens* around the world (either by one or two main waves [36]) from a common origin that the fossil record places in Africa [37,38]. In addition, some extended eco-environmental barriers, such as the Sahara Desert, create zones of sharper genetic change by limiting population migration, explaining, for example, why North Africans are, on average, genetically closer to Europeans than to sub-Saharan Africans [39]. Similarly, indigenous populations from the Americas and Oceania often have distinctive genetic profiles due to strong historical bottlenecks [40] that led to geographic isolation and small effective population sizes, two factors that are known to accelerate genetic drift. In addition, cultural boundaries have a significant influence on human relationships, sometimes generating extensive genetic variation between neighbouring populations that speak different languages [41] and/or follow different lifestyles [42].

Demographic processes thus have a major impact on human molecular variation, and similar conclusions have been drawn from analyses of HLA genes at both global and continental scales [17,43–51]. However, these mechanisms are not the only source of genetic differentiation observed worldwide. Many genes – including HLA – are also susceptible to natural selection, which creates additional genetic differences between populations depending on local environmental conditions [52]. For example, strong selective effects have been shown to affect the evolution of genes involved in skin pigmentation in response to UV light [53,54], metabolic adaptation to altitude [55,56], or lactose digestion in response to milk consumption [57,58], among many others [59]. Numerous genetic polymorphisms have also been associated with resistance or susceptibility to disease, including erythrocyte components to malaria [60,61] and HLA variants to multiple autoimmune and infectious diseases [62–65].

In this context, population genetic studies have estimated how HLA genetic diversity is distributed on a global scale [49,66,67] and have also attempted to understand how demographic and selective forces influence the evolution of the different HLA genes [17,44,49]. It has been shown that, although HLA genes resemble the rest of the genome in that most of their molecular diversity is found within populations [68,69], their allelic distributions deviate significantly from neutral expectations towards an excess of heterozygosity. This has led researchers to propose heterozygous advantage as the main selective model for HLA [17,44,49,70–72], further refining it by considering both the molecular divergence of the two HLA alleles carried by a heterozygote for a given gene [73,74] and the complementary roles of various HLA loci at the functional level [75,76]. Actually, different HLA genes are subject to different modes and intensities of selection (as previously described in Ref. [77] and consistent with most of the references cited above), which result in locus-specific patterns of population differentiations that do not always align with simple geographic structures. Briefly, the most highly polymorphic HLA-B locus probably owes its enormous allelic diversity to small protective effects conferred by multiple HLA alleles, consistent with a « soft selective sweep » model [78]. HLA-A would be less influenced by natural selection, showing highly correlated genetic and geographic distances between populations except for some isolated populations displaying extreme allele frequencies (e.g. 86 % for A*24:02 in native Taiwanese Paiwan), likely due to genetic drift. This gene also shows fewer deviations from neutrality and no significant impact of functional diversity (as estimated by « supertypes ») on population variation [79]. The evolution of HLA-C appears to be weakly influenced by geographical differentiation, possibly due to strong and complex selective effects associated with the specialisation of this locus for NK cell recognition via HLA-KIR interactions [80,81]. Among the HLA Class II genes, DRB1 shows the highest heterozygosity within populations suggesting overall heterozygous advantage, but also highly significant

correlations between populations' genetic and geographic distances, making this locus particularly informative for reconstructing human demographic history. HLA-DQA1 and -DQB1, which are both highly polymorphic (this is not the case for HLA-DRA), are thought to be strongly selected, showing high frequencies of several alleles in most populations and patterns of genetic differentiation that are only weakly correlated with geography. Finally, although most HLA-DPB1 allelic profiles observed in populations are consistent with neutral expectations based on classical neutrality tests [17,44], the high DPB1 allele frequencies observed in some populations – e.g. African – indicate strong pathogen-driven selection [82] and Goeury et al. submitted]. Overall, these findings

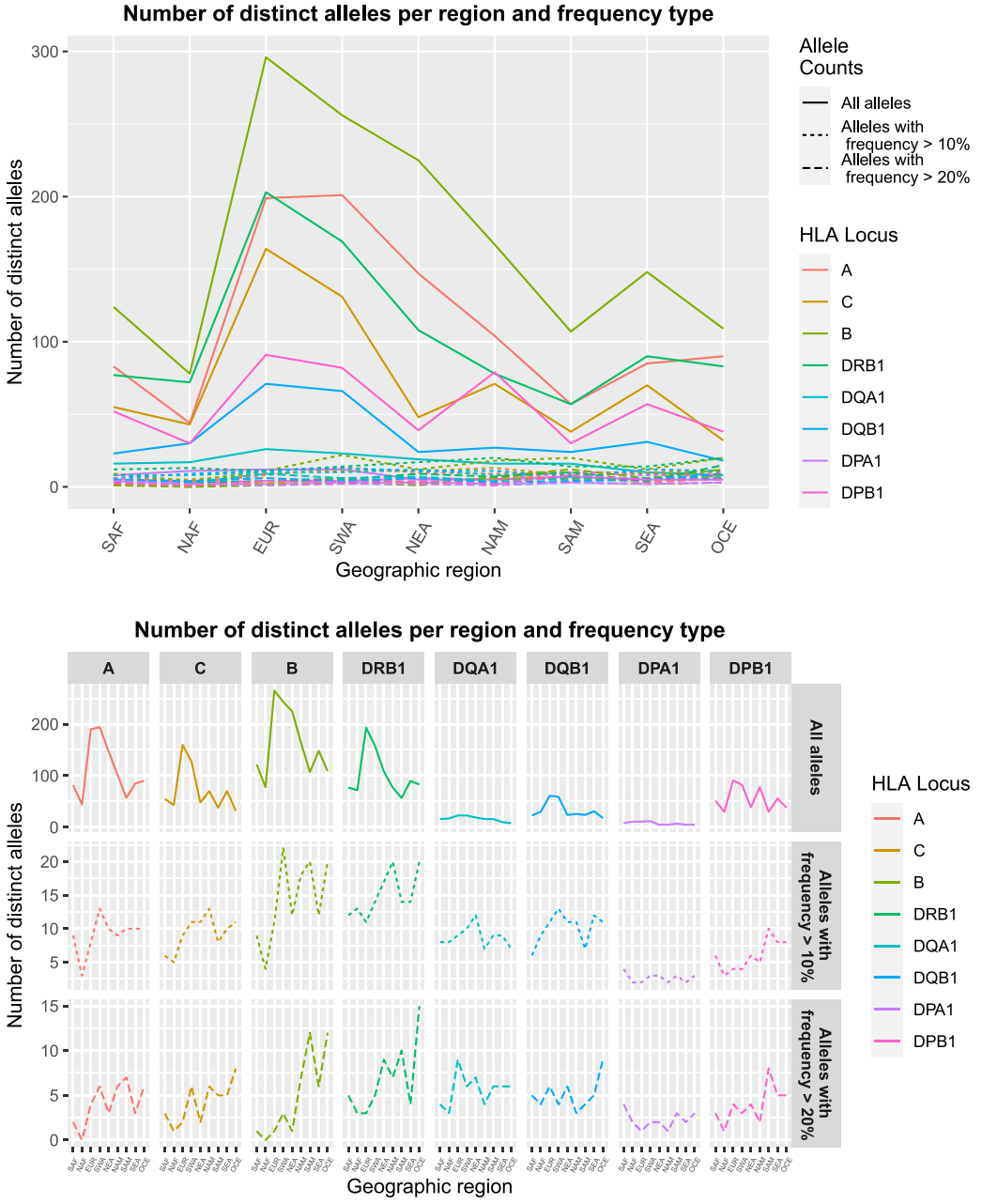

Fig. 1. Absolute numbers of distinct alleles detected at each of the 8 HLA loci A, C, B, DRB1, DQA1, DQB1, DPA1 and DPB1 in each of 9 main (sub-) continental regions of the world, and for 3 frequency types, i.e., all alleles ($AF > 0$, solid lines), frequent alleles ($AF > 10$ %, short dashed lines) and very frequent alleles ($AF > 20$ %, long dashed lines), respectively. Note that very frequent (AF > 20 %) alleles are also considered as frequent (AF > 10 %) alleles. The 3 frequency types are shown together (Fig. 1a, top) and separately (Fig.1b, bottom). SAF: sub-Saharan Africa; NAF: North Africa; EUR: Europe; SWA: South & West & Central Asia; NEA: North-East Asia; NAM: North & Central America; SAM: South America; SEA: South-East Asia (SEA); OCE: Oceania. AF: allele frequency. See Supplementary Table S2 for the number of population samples that have been considered in each region and at each locus

underscore that geography may not always be a reliable predictor of HLA genetic relatedness, as each HLA gene responds differently to selective pressures. This observation has broad implications for HLA research, including in the context of allogeneic transplantation where a nuanced understanding of HLA diversity can guide the search for unrelated donors.

2. The most frequent HLA alleles around the world: a fundamental synopsis

The resources described in part 1.2 of this Review have provided many opportunities to assess and analyse HLA population data, as illustrated by the major achievements summarized in part 1.3. However, with the rapid accumulation of data, it is still difficult to obtain a synthetic view of how HLA allelic variation is distributed throughout the world. A major problem is the heterogeneous representativeness and very uneven sample sizes of the HLA-typed populations, resulting in incomplete, biased and/or unstable descriptions of HLA variation by population-specific frequency profiles or regional lists of CWD alleles, the latter being particularly prone to change with increasing amounts of data. We have therefore decided to introduce a novel approach for looking at the data, i.e. through the prism of the distribution of frequent alleles. Instead of comparing precise allele inventories and frequencies between populations, we considered two allele frequency thresholds – above 10 % and above 20 % - and estimated the abundance of such alleles in the different regions of the world.

On the basis of both HLA-typed population datasets of consecutive IHIWs and data retrieved from the allelefrequencies.net (AF) database (a total of 1971 well-characterised datasets represented by 260 population samples typed at locus A, 258 at locus B, 201 at locus C, 46 at locus DPA1, 230 at locus DPB1, 199 at locus DQA1, 354 at locus DQB1 and 423 at locus DRB1), we first calculated the absolute numbers of distinct alleles detected at each of the 8 HLA loci in each of the 9 main (sub-)continental regions of the world, i.e. Sub-Saharan Africa (SAF), North Africa (NAF), Europe (EUR), South & West & Central Asia (SWA), North East Asia (NEA), North & Central America (NAM), South America (SAM), South East Asia (SEA) and Oceania (OCE) and for 3 frequency types, i.e. all alleles (when allele frequency, hereafter *AF*, was greater than 0), « frequent » alleles (when *AF* was greater than 10 %) and « very frequent »

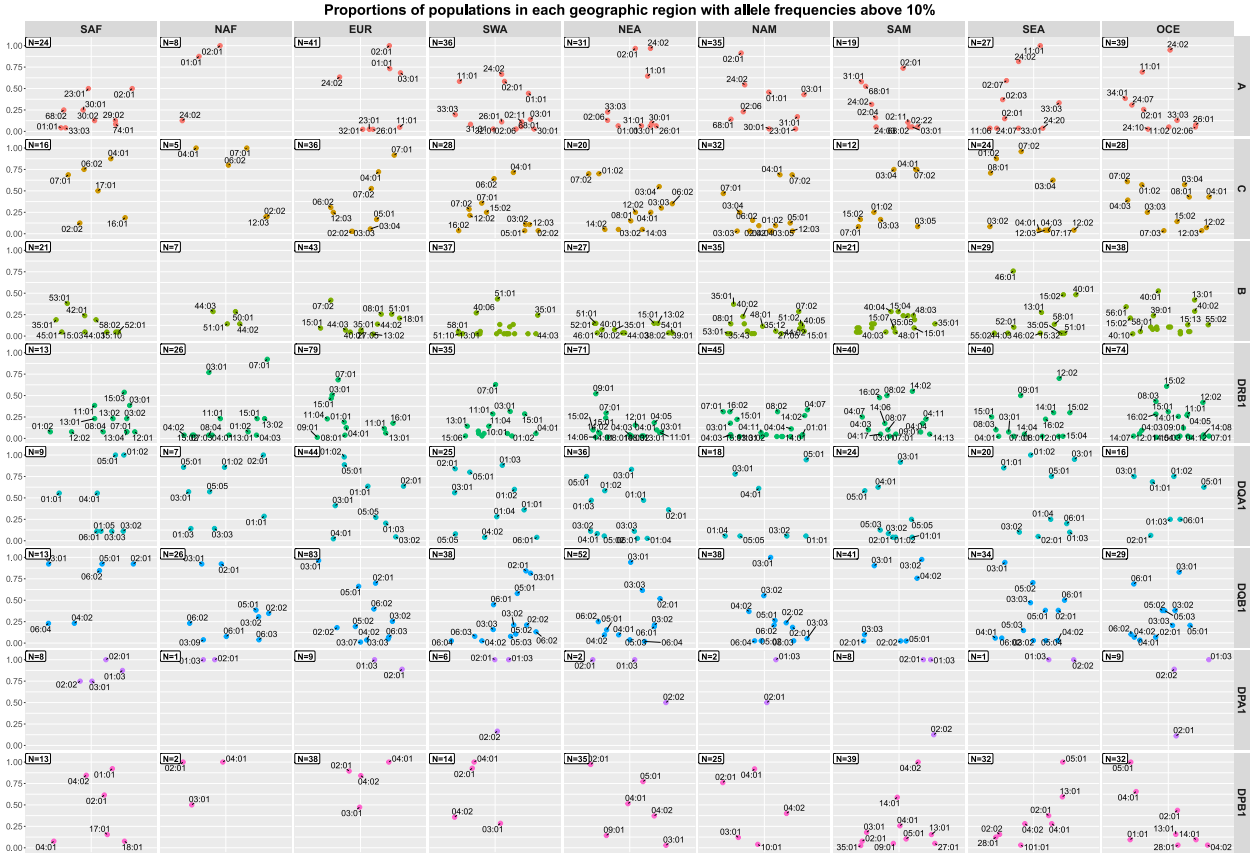

Fig. 2. Proportion of populations, in each geographic region, displaying a frequency above 10 % (frequent alleles, Fig.2a, top) and above 20 % (very frequent alleles, Fig.2b, bottom) for the alleles shown at each of the 8 HLA loci A, C, B, DRB1, DQA1, DQB1, DPA1 and DPB1 in each of 9 main (sub-) continental regions of the world. Note that very frequent (AF > 20 %) alleles are also considered as frequent (AF > 10 %) alleles. Regions are SAF: sub-Saharan Africa; NAF: North Africa; EUR: Europe; SWA: South & West & Central Asia; NEA: North-East Asia; NAM: North & Central America; SAM: South America; SEA: South-East Asia (SEA); OCE: Oceania. N (boxed in each quadrant): number of populations represented in each region and for each locus. AF: allele frequency. Larger versions of Fig. 2a and b are available in Supplementary Figs. S1a and S1b, respectively.

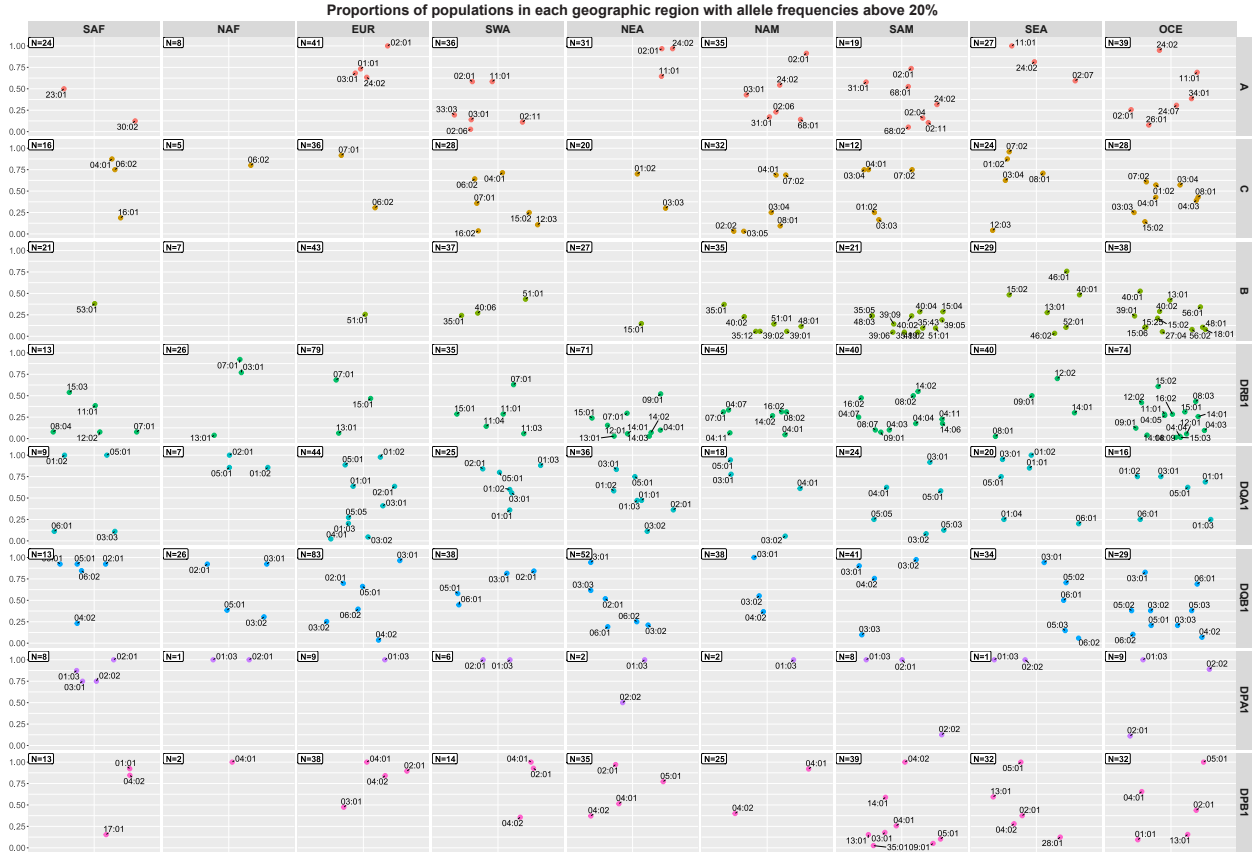

Fig. 2. (*continued*).

alleles (when *AF* was greater than 20 %), respectively. The results are presented graphically in Fig. 1a, for the 3 frequency types together, and in Fig. 1b, for the 3 frequency types considered separately.

We have then calculated the proportions of populations with each « frequent » or « very frequent » allele in each of the 9 regions. These are shown graphically in Fig. 2a, for « frequent» and in Fig. 2b, for « very frequent » alleles, respectively (larger figures are available in Supplementary Figs. S1a and S1b, respectively, and numerical data are reported in Supplementary Table S1).

These four graphs summarise the geographical distribution of the frequent and very frequent alleles in the world, allowing one to see at a glance how they are distributed both across different geographical regions, for each HLA locus, and across different HLA loci, for each geographical region (note, however, that the number of populations represented is low for most loci in the NAF region and for most regions at the DPA1 locus, which limits some interpretations).

A few observations can be made from these plots:

- Fig. 1a shows that the vast majority of distinct alleles observed are « infrequent » (term that we will use thereafter to name alleles with frequencies below 10 %). The large variance between regions mostly reflects the unequal number of populations represented (e.g. many more alleles are observed in Europe where most populations have been typed, see Supplementary Table S2). However, we note that SAM and OCE have a low number of distinct alleles despite being well represented in terms of population numbers (e. g. in most cases better than for SAF). Allelic diversity therefore appears to be lower in these regions.
- In contrast, Fig. 1b shows that OCE and SAM and, to a lesser extent, NAM and SWA, have higher numbers of frequent ($AF > 10$ %) and/or very frequent ($AF > 20$ %) alleles at most loci (with extreme values at HLA-B and DRB1) compared to the other regions. Together with the low allelic diversity mentioned above, this suggests that the HLA genetic profiles are both more homogeneous *within* the populations of these regions (i.e. they contain a limited number of (very) frequent alleles) and heterogeneous *between* different populations of the same region (i.e. they present different set of (very) frequent alleles).
- Fig. 2a and b shows that frequent and very frequent alleles are observed in all regions, respectively. The only exception is North Africa at loci A and B, which is easily explained by the small number of populations represented. In addition, there are some notable differences between regions and/or between loci (Supplementary Tables S2 and S3). Very frequent HLA Class I (A, B and C) alleles are abundant in OCE (26 in total), followed by SAM (24) and NAM (19) (Fig. 2b & Supplementary Table S2). This is also the case for DRB1 very frequent alleles (15, 10 and 7 in the 3 regions, respectively), the latter being also abundant in NEA (9). On the contrary, DQA1 very frequent alleles are abundant in EUR (9), followed by NEA (7). These two regions also have the highest number of DQB1

very frequent alleles (6), after OCE (9). Finally, whereas DPB1 very frequent alleles are mostly found in SAM (8), only 1 to 4 very frequent DPA1 alleles are observed in distinct regions, but this result is based on very few data. Overall, very frequent alleles are, on average, most abundant in OCE and SAM (Supplementary Table S3). The differences between regions are less pronounced for frequent alleles (Fig. 2a & Supplementary Table S2). If we compare the different loci, very frequent alleles are, on average, most abundant at the DRB1 locus (6.78), followed by DQA1 (5.67) and DQB1 (5.11), while frequent alleles are, on average, most abundant at loci DRB1 (15) and B (14.22), with the highest standard deviation for locus B in both cases (Supplementary Table S3). We also see that some alleles are very frequent in a high proportion of the populations of some regions at all loci except B and DRB1, consistent with the high degree of polymorphism of these two genes. However, some of these abundant alleles are shared by several regions while others are less widespread.

Based on Fig. 2a and b and Supplementary Table S1, we can then classify the alleles into the following categories (the detailed data are reported in Supplementary Table S4).

1) **Universally frequent alleles**, i.e., alleles with frequencies above 10 % (but below 20 %) in at least one population of the majority of (i.e. more than 4) regions. A total of 39 alleles meet this criterion. They are listed in Table 2.
2) **Universally very frequent alleles**, i.e., alleles with frequencies above 20 % in at least one population of the majority of (i.e. more than 4) regions. A total of 20 alleles (none at the HLA-B locus) meet this criterion. They are listed in Table 3 with additional relevant information. The precise distribution of each universally very frequent allele is also illustrated in the maps presented in Supplementary Fig. S2.

We checked that all these universally frequent and very frequent alleles were classified as Common in all population groups considered in the CWD/CIWD catalogues (Supplementary Table 3a of Hurley et al. [28]). Recently, Silva et al. [83] also found that, at locus B, B*35:01, B*51:01, B*07:02 and B*44:03 were the most widespread, with a frequency greater than 5 % in all geographical regions they studied. This is in agreement with our results, since, with the exception of B*07:02, these alleles belong to the list of universally frequent (but not very frequent) alleles (20 % > *AF* > 10 %) that we reported in Table 2.

3) **Locally frequent or very frequent alleles:** by contrast, some alleles are either frequent (*AF* > 10 %) or very frequent (*AF* > 20 %) in a single region (see detailed list in Supplementary Table S4). We have reported the number of alleles falling into this category in Table 4, considering three cases:
   - Alleles that are frequent (*AF* > 10 %) in a single region and infrequent (*AF* < 10 %) elsewhere;
   - Alleles that are very frequent (*AF* > 20 %) in a single region and frequent (20 % > *AF* > 10 %) elsewhere;
   - Alleles that are very frequent (*AF* > 20 %) in a single region and infrequent (*AF* < 10 %) elsewhere.

The precise distribution of locally very frequent alleles is also illustrated in the map presented in Supplementary Fig. S3. The results reported above, which considerably improve our previous list of the four most frequent HLA alleles in 10 world regions [84], are very

Table 2
List of universally frequent (for **very frequent, see Table 3) HLA alleles**.These alleles are frequent but not very frequent (20 % > AF > 10 %) in at least one population of most (i.e. more than 4) regions.

Allele (Class I)	Regions where allele is found frequent (AF > 10 %) in at least one population	Number of regions	Allele (Class II)	Regions where allele is found frequent (AF > 10 %) in at least one population	Number of regions
A*01:01	SAF NAF EUR SWA NEA NAM	6	DRB1*03:01	SAF NAF EUR SWA NEA NAM SAM SEA	8
A*03:01	EUR SWA NEA NAM SAM	5	DRB1*04:01	NAF EUR SWA NEA NAM SEA	6
A*11:01	EUR SWA NEA SEA OCE	5	DRB1*04:03	NAF SWA NEA NAM SAM OCE	6
A*33:03	SAF SWA NEA SEA OCE	5	DRB1*09:01	EUR NEA NAM SAM SEA OCE	6
C*01:02	NEA NAM SAM SEA OCE	5	DRB1*11:01	SAF NAF EUR SWA NEA NAM OCE	7
C*02:02	SAF NAF EUR SWA NAM	5	DRB1*13:01	SAF NAF EUR SWA NEA	5
C*03:03	EUR NEA NAM SAM OCE	5	DRB1*15:01	NAF EUR SWA NEA NAM SEA OCE	7
C*03:04	EUR NEA NAM SAM SEA OCE	6	DRB1*15:02	NAF SWA NEA SEA OCE	5
C*06:02	SAF NAF EUR SWA NEA NAM	6	DQA1*01:03	NAF EUR SWA NEA SEA OCE	6
C*07:01	SAF NAF EUR SWA NAM SAM	6	DQA1*02:01	NAF EUR SWA NEA SAM SEA OCE	7
C*07:02	EUR SWA NEA NAM SAM SEA OCE	7	DQA1*03:02	SAF EUR NEA NAM SAM SEA	6
C*12:03	NAF EUR SWA NAM SEA OCE	6	DQA1*04:01	SAF EUR NEA NAM SAM	5
B*15:01	EUR SWA NEA NAM SAM	5	DQA1*05:05	NAF EUR SWA NAM SAM	5
B*35:01	SAF EUR SWA NEA NAM SAM	6	DQA1*06:01	SAF SWA NEA SEA OCE	5
B*40:01	EUR SWA NEA SEA OCE	5	DQB1*02:02	NAF EUR SWA NAM SAM	5
B*40:02	SWA NEA NAM SAM OCE	5	DQB1*03:03	EUR SWA NEA NAM SAM SEA OCE	7
B*44:03	SAF NAF EUR SWA NEA SEA	6	DQB1*05:02	EUR SWA NAM SEA OCE	5
B*51:01	NAF EUR SWA NEA NAM SAM SEA	7	DQB1*05:03	SWA NEA NAM SEA OCE	5
B*52:01	SAF SWA NEA SAM SEA	5	DQB1*06:01	NAF SWA NEA SEA OCE	5
			DPB1*03:01	NAF EUR SWA NEA NAM SAM	6

Regions: SAF: Sub-Saharan Africa; NAF: North Africa; EUR: Europe; SWA: South-West, Central and South Asia; NEA: North-East Asia; NAM; North America; SAM: South America; SEA: South-East Asia; OCE: Oceania (Pacific, New-Guinea & Australia). AF: allele frequency.

Table 3
List of universally very frequent HLA alleles. These alleles are very frequent (AF > 20 %) in at least one population of most (i.e. more than 4) regions.

Allele	Regions where allele is found very frequent (AF > 20 %) in at least one population	Number of regions	Additional comment(s)
A*02:01	EUR SWA NEA NAM SAM OCE	6	This allele is very frequent (AF > 20 %) in all EUR populations and about 75 % of NAM and SAM populations. It is also frequent (AF > 10 %) in all NAF, NEA and NAM, 75 % of SAM and about 50 % of SAF and SWA.
A*24:02	EUR NEA NAM SAM SEA OCE	6	This allele is also frequent (>10 %) in the majority of populations of all regions but SAF.
C*04:01	SAF SWA NAM SAM OCE	5	This allele is also frequent (AF > 10 %) in about or more than 75 % of populations of most regions.
DRB1*07:01	SAF NAF EUR SWA NEA NAM	6	This allele is also frequent (AF > 10 %) in all regions, with large proportions of NAF, EUR and SWA populations.
DQA1*01:01	EUR SWA NEA SEA OCE	5	This allele is also frequent (AF > 10 %) in large proportions of SAF, EUR, NEA, SEA and OCE populations.
DQA1*01:02	SAF NAF EUR SWA NEA SEA OCE	7	This allele is very frequent (AF > 20 %) in almost all SAF populations. It is also frequent (AF > 10 %) in very large proportions of populations of all regions except NAM and SAM.
DQA1*03:01	EUR SWA NEA NAM SAM SEA OCE	7	This allele is very frequent (AF > 20 %) in large proportions of NEA, NAM and SAM populations. It is also frequent (AF > 10 %) in large or very large proportions of populations of all regions except SAF.
DQA1*05:01	SAF NAF EUR SWA NEA NAM SAM SEA OCE	9	This allele is also frequent (AF > 10 %) in most populations of all regions.
DQB1*02:01	SAF NAF EUR SWA NEA	5	This allele is also frequent (AF > 10 %) in almost all regions and in large proportions of SAF, NAF, EUR, SWA and NEA populations.
DQB1*03:01	SAF NAF EUR SWA NEA NAM SAM SEA OCE	9	This allele is very frequent (AF > 20 %) in all regions with more than 50 % of populations of most regions.
DQB1*03:02	NAF EUR NEA NAM SAM OCE	6	This allele is also frequent (AF > 10 %) in large proportions of NAM and SAM populations.
DQB1*04:02	SAF EUR NAM SAM OCE	5	This allele is also frequent (AF > 10 %) in a large proportion of SAM populations.
DQB1*05:01	SAF NAF EUR SWA OCE	5	This allele is also frequent (AF > 10 %) in large proportions of SAF, EUR and SWA populations.
DQB1*06:02	SAF EUR NEA SEA OCE	5	This allele is very frequent (AF > 20 %) in a large proportion of SAF populations.
DPA1*01:03	SAF NAF EUR SWA NEA NAM SAM SEA OCE	9	This allele is very frequent (AF > 20 %) in almost all populations tested of all regions. DPA1*02:01 and DPA1*02:02 are also frequent (AF > 10 %) in most regions.
DPA1*02:01	SAF NAF SWA SAM OCE	5	This allele is also frequent (AF > 10 %) in most regions.
DPA1*02:02	SAF NEA SAM SEA OCE	5	This allele is also frequent (AF > 10 %) in most regions.
DPB1*02:01	EUR SWA NEA SEA OCE	5	This allele is also frequent (AF > 10 %) in a large majority of SAF, NAF, EUR, SWA, NEA and NAM populations.
DPB1*04:01	NAF EUR SWA NEA NAM SAM OCE	7	This allele is very frequent (AF > 20 %) in almost all NAF, EUR, SWA and NAM populations. It is also frequent (AF > 10 %) in all regions.
DPB1*04:02	SAF EUR SWA NEA NAM SAM SEA	7	This allele is very frequent (AF > 20 %) in almost all SAM populations and a large proportion of SAF. It is also frequent (AF > 10 %) in almost all EUR.

Regions: SAF: Sub-Saharan Africa; NAF: North Africa; EUR: Europe; SWA: South-West, Central and South Asia; NEA: North-East Asia; NAM; North America; SAM: South America; SEA: South-East Asia; OCE: Oceania (Pacific, New-Guinea & Australia). AF: allele frequency.

informative from an evolutionary point of view. First, they show that very frequent HLA alleles exist everywhere, even at the most polymorphic HLA-B and DRB1 loci. Many of them are either universal or present in multiple geographic regions, suggesting that the HLA allelic pool is widely shared among populations on a global scale, as are the majority of genome-wide SNPs [85]. These universally (very) frequent alleles were probably present in the ancestral population of all modern humans. Exploring the relationship of this allelic pool with the MHC allelic repertoire of our closest relatives, the chimpanzees (i.e. *Patr* polymorphism), would be most informative to better understand the mechanisms of MHC molecular evolution, as the latter appears to be highly conserved between the two species [86].

Locally frequent or very frequent HLA alleles (i.e. those that are (very) frequent in one region and infrequent in others) also exist and could be related either to local demographic history or to some disease associations in the regions concerned. For example, DPB1*17:01 has been proposed to protect against malaria in SAF [82] and Goeury et al. submitted] and DRB1*14:03 against systemic lupus erythematosus in NEA [87], where these alleles are very frequent in the corresponding regions. Interestingly, DRB1*12:02, which may have undergone a dramatic increase in frequency in Mongolia due to strong selective pressure [88], is very frequent in several regions (SEA, OCE and SAF), consistent with this allele being considered as a generalist HLA Class II variant protecting against a wide range of pathogens [89]. Most locally very frequent alleles, however, are observed in SAM (including native populations from South America) and OCE (including isolated populations from the Pacific, Taiwan, New Guinea and Australia) populations. Together with the highest average numbers of very frequent alleles observed in these regions (Supplementary Table S3), this is likely to be the result of rapid genetic drift randomly increasing the frequencies of different sets of alleles in small isolated populations and reducing their heterozygosity (although selection in native Americans would not be excluded [90]). On the other hand, almost all locally (very) frequent alleles are observed at low frequencies in other - often multiple - regions (Table 4) and are classified as either Common,

Table 4
Number and list of locally very frequent (AF > 20 %, see also Supplementary Fig. S3) and frequent (20 % > AF > 10 %) HLA alleles.

Single region	Very frequent (AF > 20 %) only in the single region and infrequent (AF < 10 %) elsewhere		Very frequent (AF > 20 %) only in the single region and frequent but not very frequent (20 > AF > 10 %) elsewhere		Frequent but not very frequent (20 % > AF > 10 %) in only the single region and infrequent (AF < 10 %) elsewhere		Detected in other regions
SAF	4	A*30:02 C*16:01 DPA1*03:01 DPB1*17:01	4	A*23:01 B*53:01 DRB1*08:04 DQA1*03:03	12	A*29:02 A*74:01 C*17:01 B*15:03 B*15:10 B*42:01 B*45:01 B*58:02 DRB1*03:02 DRB1*13:04 DQA1*01:05 DPB1*18:01	All
NAF	0		1	DRB1*03:01	3	DRB1*04:02 DRB1*07:03 DQB1*03:09	All
EUR	0		1	A*01:01	1	DQB1*03:07	All
SWA	3	C*16:02 B*40:06 DRB1*11:03	2	A*33:03 DRB1*11:04	9	B*18:07 B*35:03 B*38:01 B*57:01 B*44:06 B*51:10 DRB1*10:01 DRB1*15:06 DQA1*04:02	All
NEA	1	DRB1*14:03	1	B*15:01	4	C*14:02 C*14:03 B*54:01 DQA1*05:02	All
NAM	2	B*39:02 B*35:12	2	C*02:02 C*03:05	6	C*04:04 B*27:03 B*51:02 B*40:05 DRB1*08:11 DPB1*10:01	All
SAM	11	A*02:04 B*39:06 B*15:04 B*39:09 B*40:04 B*48:02 B*48:03 B*35:19 DRB1*08:07 DQA1*05:03 DPB1*35:01	7	A*68:02 B*35:43 B*39:05 B*35:05 DRB1*14:06 DPB1*09:01 DPB1*14:01	8	A*24:03 A*02:22 B*15:07 B*35:06 B*40:03 DRB1*04:17 DRB1*14:13 DPB1*27:01	All
SEA	2	A*02:07 B*46:02	5	B*52:01 B*46:01 DRB1*08:01 DQA1*01:04 DPB1*28:01	10	A*33:01 A*02:03 A*11:06 A*24:20 C*07:17 B*15:32 DRB1*15:04 DQB1*05:04 DPB1*02:02 DPB1*101:01	All except DPB1*101:01
OCE	8	A*34:01 B*56:01 B*15:25 B*27:04 B*56:02 B*15:06 DRB1*14:08 DRB1*14:09	7	A*26:01 A*24:07 C*04:03 B*18:01 DRB1*04:05 DRB1*08:03 DRB1*15:02	9	A*11:02 A*24:10 C*07:03 B*15:13 B*15:21 B*40:10 DRB1*14:05 DRB1*14:07 DRB1*04:12	All except DRB1*04:12
Total	31		30		62		All except 2

Regions: SAF: Sub-Saharan Africa; NAF: North Africa; EUR: Europe; SWA: South-West, Central and South Asia; NEA: North-East Asia; NAM; North America; SAM: South America; SEA: South-East Asia; OCE: Oceania (Pacific, New-Guinea & Australia). AF: allele frequency.

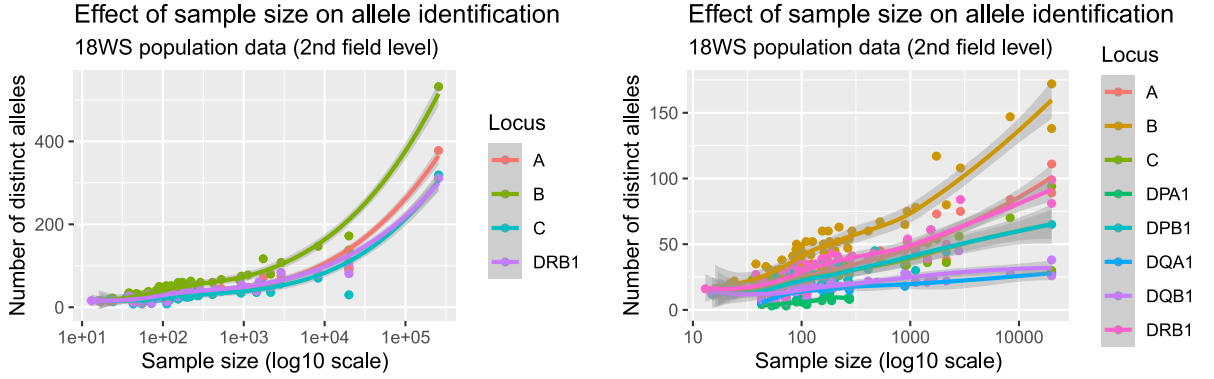

Fig. 3. Number of distinct HLA 2nd field alleles detected as a function of sample size (log10 scale) for the data submitted to the 18th IHIW PGAE Component. Fig.3a, left: whole data set. Fig. 3b, right: after exclusion of the largest sample.

Intermediate or Well-Documented in the existing CWD/CIWD catalogues [28]. Therefore, the early concept of « private » [91], now also replaced by « endemic » [51], used to characterise alleles that would be unique to some populations or groups of populations such as Native Americans [51,92] is clearly limited to infrequent or rare alleles. The latter, which are likely to have accumulated worldwide as a result of the explosive population growth that has occurred over the last 10′000 years [93], may only be informative to describe recently shared ancestry between populations, and, even in this case, only when the datasets used are composed of very high sample sizes [94].

Another remarkable result of our study is the predominance of some alleles not only in most regions, but also in most populations of these regions. For example, A*02:01, DQB1*03:01, DPA1*01:03 and DPB1*04:01 are among the top universally very frequent alleles according to these criteria (Fig. 2b) and a number of frequent alleles are also abundant in most populations (Fig. 2a). Although some of these alleles have been identified as disease risk variants in some studies [64], we might assume, from an evolutionary point of view, that they are or have been globally protective, e.g. as generalist (or promiscuous) alleles [95], against pathogen exposure throughout

human history. This stimulates further analyses to estimate the peptide-binding promiscuity of universally or locally (very) frequent HLA alleles, possibly also related to allele-specific expression levels [95] and pathogen diversity [89]. Such intriguing perspectives highlight the importance of focusing on alleles that are representative of the global genetic variation, rather than on those that are predominantly found in populations of European descent, as is common practice in biomedical research [96,97].

3. Concluding remarks and take-home messages

On the way to full-length HLA sequences As mentioned in this Review, there has been a huge effort over the last few decades to identify, report, name and classify the human HLA alleles in various ways. The popular IPD-IMGT/HLA graphs of the number of alleles named per year [2] as well as the large number of "New Allele Alerts" published regularly in the *HLA* Journal (e.g. a total of 192 such papers appeared in the first 4 issues of Volume 103 between January and April 2024), indicate that the reporting of new HLA alleles is never-ending. In fact, most of the new alleles reported in recent years have been defined at the 3rd and 4th field levels of resolution, as more and more full-length HLA sequences are submitted to the IPD-IMGT/HLA database (S.GE Marsh, personal communication). The analysis of such sequences will add a new dimension to HLA evolutionary studies by allowing more thorough comparisons of the HLA region with the rest of the genome [98], and will also help to refine HLA matching in clinical transplantation in the future, as HCT is still currently based on alleles defined at the protein level [99]. In this context, both evolutionary and clinical studies will soon benefit from the remarkable efforts made by HLA-typing laboratories at the international level, which is already illustrated, for example, by almost 50 % of the new population samples submitted to the 18th IHIW PGAE Component being unambiguously reported at the 3rd or 4th field levels for at least one locus, respectively (Nunes et al. in preparation).

The eternal problem of the sample size In the meantime, studies using HLA data can benefit from the excellent knowledge that has been acquired over the last decades on high-resolution (2nd field) allelic profiles in populations and from the sophisticated biostatistical methods that have been developed to analyse them (summarized in Table 1). However, even in this case, the data should be interpreted with caution. Indeed, as shown in this Review, the vast majority of currently known 2nd field alleles are infrequent in populations (Fig. 1a) and their detection is highly dependent on sample size, as shown in Fig. 3a and b for the data submitted to the 18th IHIW PGAE Component. This limits the effectiveness of CWD/CIWD catalogues in detecting the presence of infrequent or rare alleles, as is often needed in HCT. As the status of each allele as Common, Intermediate, or Well-Documented is also likely to change very rapidly with increasing high-throughput HLA sequencing, international efforts should now focus on maintaining a unique, constantly updated online repository of CWD/CIWD alleles. As long as population sampling remains a major statistical problem, the presence/absence of infrequent HLA alleles will also remain difficult to interpret in population genetics and evolutionary studies. With the increasing number of known alleles, the minimum sample sizes to allow such discussions – for example those focusing on « private or endemic alleles » - should ideally be increased by a factor of 10, 100 or even more.

Advantages of a focus on (very) frequent alleles Focusing on (very) frequently observed HLA alleles, as has been done in the present Review, may seem irrelevant in the age of powerful high-throughput sequencing technologies where the rush for new alleles has almost become the rule. However, there are two main reasons for this choice. First, there has been so far no synthetic or easily useable description of which HLA alleles are frequent or very frequent in the world and how they are distributed geographically, despite constant need for such information in almost every HLA-related research study. Second, after some 30 years of HLA population studies using high-resolution typing at most HLA loci, encouraged by successive International HLA and Histocompatibility Workshops, we were confident that a global distribution of (very) frequent HLA alleles based on such data would provide a fairly robust picture of how HLA diversity is distributed across different geographical regions for the major classical HLA loci (although some regions, such as North Africa, and some loci, such as those encoding the α chains of HLA Class II molecules, still need to be better represented). Our Review is indeed robust to imbalances in sample size between studies and even to possible redundancies in the data (although we have removed clear cases of duplicate samples between the different datasets used, we cannot exclude that some data are not independent), because our lists of universally or locally (very) frequent alleles are based on alleles that are (very) frequent in *at least one* population per region. Therefore, only a small error rate in the proportions of populations showing each frequent or very frequent allele in the different regions could be considered, but the status of each allele is unaffected. We therefore believe that this fundamental synopsis, by providing an immediate and consistent check of whether and where any given HLA allele is (very) frequently observed in the world, rigorously complements the information we need for a comprehensive panorama of HLA diversity worldwide.

Benefit for clinical practice As shown above, the present synopsis of HLA variation around the world has allowed identifying HLA alleles that are universally widespread with (very) high frequencies, but also alleles that are much more likely to be observed in specific geographic regions, being infrequent or rare in other parts of the world. In some regions (e.g. the Americas), these « locally (very) frequent » alleles are expected to show contrasted frequencies between native populations (e.g. native Americans) and more recent immigrants (e.g. Americans from European descent) that are numerically predominant today. In other areas, some of these alleles are expected to exhibit significant frequency differences between communities (often designated as *ethnic minorities*) that migrated to other regions relatively recently (e.g. South Asians in United Kingdom) and the general population of those regions. These situations suggest that both HLA typing and donor recruitment strategies for allogeneic HCT ought to be adapted, e.g. by actively including individuals from ethnic minorities living in each specific country, to enhance the chances of finding HLA-matched donors for patients from any geographic region or ethnic background. Our knowledge of HLA variant distribution around the world is most useful in this context.

Overall, by providing a different and complementary approach to the existing CWD/CIWD catalogues, the present synopsis of universal and local HLA variation around the world increases the diversity of potential resources for clinical practice, which is particularly welcome where saving lives is a priority.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Practice points

- HLA alleles are highly diverse and exhibit varying frequencies in populations around the world. While some alleles are universally frequent, others are only frequent within specific geographic regions or populations, particularly outside Europe, North-East Asia, and North America. These variations affect the chances of finding HLA-matched donors in allogeneic hematopoietic cell transplantation (HCT).
- Based on our knowledge of HLA allelic variation worldwide, both HLA typing and donor recruitment strategies for HCT ought to be adapted to actively include volunteers from diverse geographic origins and/or ethnic backgrounds. This approach is expected to increase the likelihood of finding compatible donors, especially for patients from ethnic minorities.
- Population samples typed for HLA are still unequally represented across different geographic regions, and the available datasets are also heterogeneous across different HLA loci. Efforts should be made in HLA typing at the international level to enhance both population and locus representativeness.

Research agenda

The depth and accuracy of HLA population data should be improved by addressing the following needs.

- Future HLA typing should concentrate on under-represented HLA-typed populations, such as sub-Saharan and North African populations, and on under-typed HLA genes, such as genes encoding the alpha chain of HLA class II molecules.
- HLA full-gene sequencing should be prioritised to improve the quality of the population data sets and to add a new dimension to HLA evolutionary studies.

CRediT authorship contribution statement

Alicia Sanchez-Mazas: Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **José Manuel Nunes:** Writing – review & editing, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Acknowledgements

This work was supported by the Swiss National Science Foundation (grant #310030_188820 to ASM). We are most indebt to Katharina Fleischhauer for her support to the Population Genetics Working Group of the EFI Scientific Committee.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.beha.2024.101559.

References

[1] Kulski JK, Suzuki S, Shiina T. Human leukocyte antigen super-locus: nexus of genomic supergenes, SNPs, indels, transcripts, and haplotypes. Hum Genome Var 2022;9:49. https://doi.org/10.1038/s41439-022-00226-5.
[2] Barker DJ, Maccari G, Georgiou X, Cooper MA, Flicek P, Robinson J, et al. The IPD-IMGT/HLA database. Nucleic Acids Res 2023;51:D1053–60. https://doi.org/10.1093/nar/gkac1011.
[3] Campbell IM, Gambin T, Jhangiani S, Grove ML, Veeraraghavan N, Muzny DM, et al. Multiallelic positions in the human genome: challenges for genetic analyses. Hum Mutat 2016;37:231–4. https://doi.org/10.1002/humu.22944.
[4] Gonzalez-Galarza FF, McCabe A, Melo Dos Santos EJ, Jones AR, Middleton D. A snapshot of human leukocyte antigen (HLA) diversity using data from the Allele Frequency Net Database. Hum Immunol 2021;82:496–504. https://doi.org/10.1016/j.humimm.2020.10.004.
[5] Fleischhauer K, Hien Tran T, Meisel R, Mytilineos J, Dreger P, Kröger N. Donor selection for allogeneic hematopoietic cell transplantation. Dtsch Arztebl Int 2023;120:261–8. https://doi.org/10.3238/arztebl.m2023.0031.
[6] Petersdorf EW. Genetics of graft-versus-host disease: the major histocompatibility complex. Blood Rev 2013;27:1–12. https://doi.org/10.1016/j.blre.2012.10.001.
[7] Sanchez-Mazas A, Meyer D. The relevance of HLA sequencing in population genetics studies. Journal of Immunology Research 2014;2014:1–12. https://doi.org/10.1155/2014/971818.
[8] Petersdorf EW, Gooley TA, Malkki M, Bacigalupo AP, Cesbron A, Du Toit E, et al. HLA-C expression levels define permissible mismatches in hematopoietic cell transplantation. Blood 2014;124:3996–4003. https://doi.org/10.1182/blood-2014-09-599969.
[9] Petersdorf EW, Bengtsson M, De Santis D, Dubois V, Fleischhauer K, Gooley T, et al. Role of HLA-DP expression in graft-versus-host disease after unrelated donor transplantation. J Clin Orthod 2020;38:2712–8. https://doi.org/10.1200/JCO.20.00265.
[10] Carey BS, Poulton KV, Poles A. Factors affecting HLA expression: a review. Int J Immunogenet 2019;46:307–20. https://doi.org/10.1111/iji.12443.

[11] Kaur G, Gras S, Mobbs JI, Vivian JP, Cortes A, Barber T, et al. Structural and regulatory diversity shape HLA-C protein expression levels. Nat Commun 2017;8:15924. https://doi.org/10.1038/ncomms15924.
[12] Imanishi T, Akaza T, Kimura A, Tokunaga K, Gojobori T. Allele and haplotype frequencies for HLA and complement loci in various ethnic groups. In: Tsuji K, Aizawa M, Sasazuki T, editors. HLA 1991, proceedings of the eleventh international histocompatibility workshop and conference, vol. 1. Oxford: Oxford Science Publications; 1992. p. 1065–220.
[13] Imanishi T, Akaza T, Kimura A, Tokunaga K, Gojobori T. Genetic relationships among various human populations indicated by MHC polymorphisms. HLA 1991. In: Tsuji K, Aizawa M, Sasazuki T, editors. Proceedings of the eleventh international histocompatibility workshop and conference, vol. 1. Oxford: Oxford Science Publications; 1992. p. 627–32.
[14] Hammond MG, Du Toit ED, Sanchez-Mazas A, Andrien M, Coluzzi M, De Pablo MR, et al. HLA in sub-saharan Africa: 12th international histocompatibility workshop: SSAF report. In: Genetic diversity of HLA: functional and medical implications. vol. 2; 1997. p. 345–53. Sèvres, France: EDK.
[15] Bodmer JG, Cambon-Thomsen A, Hors J, Piazza A, Sanchez-Mazas A. Anthropology report: introduction. Genetic diversity of HLA: functional and medical implications 1997;2:269–84. Sèvres, France: EDK.
[16] Mack SJ, Tsai Y, Sanchez-Mazas A, Erlich HA. 13th international histocompatibility workshop anthropology/human genetic diversity joint report: chapter 3 - Anthropology/human genetic diversity population reports. Int. Histocompatibility Working Group Press; 2006. p. 580–652.
[17] Solberg OD, Mack SJ, Lancaster AK, Single RM, Tsai Y, Sanchez-Mazas A, et al. Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies. Hum Immunol 2008;69:443–64. https://doi.org/10.1016/j.humimm.2008.05.001.
[18] Single RM, Meyer D, Mack SJ, Lancaster A, Erlich HA, Thomson G. 14th International HLA and Immunogenetics Workshop: report of progress in methodology, data collection, and analyses. Tissue Antigens 2007;69:185–7. https://doi.org/10.1111/j.1399-0039.2006.00767.x.
[19] Nunes JM, Riccio ME, Buhler S, Di D, Currat M, Ries F, et al. Analysis of the HLA population data (AHPD) submitted to the 15th International Histocompatibility/Immunogenetics Workshop by using the Gene[rate] computer tools accommodating ambiguous data (AHPD project report). Tissue Antigens 2010;76:18–30. https://doi.org/10.1111/j.1399-0039.2010.01469.x.
[20] Gonzalez-Galarza FF, Christmas S, Middleton D, Jones AR. Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. Nucleic Acids Res 2011;39:D913–9. https://doi.org/10.1093/nar/gkq1128.
[21] Nunes JM, Buhler S, Roessli D, Sanchez-Mazas A. The HLA-net 2013 collaboration. 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 2014;83:307–23. https://doi.org/10.1111/tan.12356.
[22] Sanchez-Mazas A, Vidan-Jeras B, Nunes JM, Fischer G, Little A-M, Bekmane U, et al. Strategies to work with HLA data in human populations for histocompatibility, clinical transplantation, epidemiology and population genetics: HLA-NET methodological recommendations. Int J Immunogenet 2012;39:459–72. https://doi.org/10.1111/j.1744-313X.2012.01113.x. quiz 473–472.
[23] Sanchez-Mazas A, Nunes JM, Middleton D, Sauter J, Buhler S, McCabe A, et al. Common and well-documented HLA alleles over all of Europe and within European sub-regions: a catalogue from the European Federation for Immunogenetics. HLA 2017;89:104–13. https://doi.org/10.1111/tan.12956.
[24] Riccio ME, Buhler S, Nunes JM, Vangenot C, Cuénod M, Currat M, et al. 16th IHIW: analysis of HLA Population Data, with updated results for 1996 to 2012 workshop data (AHPD project report). Int J Immunogenet 2013;40:21–30. https://doi.org/10.1111/iji.12033.
[25] Osoegawa K, Vayntrub TA, Wenda S, De Santis D, Barsakis K, Ivanova M, et al. Quality control project of NGS HLA genotyping for the 17th international HLA and immunogenetics workshop. Hum Immunol 2019;80:228–36. https://doi.org/10.1016/j.humimm.2019.01.009.
[26] Eberhard H-P, Schmidt AH, Mytilineos J, Fleischhauer K, Müller CR. Common and well-documented HLA alleles of German stem cell donors by haplotype frequency estimation. HLA 2018;92:206–14. https://doi.org/10.1111/tan.13378.
[27] He Y, Li J, Mao W, Zhang D, Liu M, Shan X, et al. HLA common and well-documented alleles in China. HLA 2018;92:199–205. https://doi.org/10.1111/tan.13358.
[28] Hurley CK, Kempenich J, Wadsworth K, Sauter J, Hofmann JA, Schefzyk D, et al. Common, intermediate and well-documented HLA alleles in world populations: CIWD version 3.0.0. HLA 2020;95:516–31. https://doi.org/10.1111/tan.13811.
[29] Lancaster AK, Single RM, Mack SJ, Sochat V, Mariani MP, Webster GD. PyPop: a mature open-source software pipeline for population genomics. https://doi.org/10.5281/zenodo.10602940; 2024.
[30] Mack SJ, Cano P, Hollenbach JA, He J, Hurley CK, Middleton D, et al. Common and well-documented HLA alleles: 2012 update to the CWD catalogue. Tissue Antigens 2013;81:194–203. https://doi.org/10.1111/tan.12093.
[31] Nunes JM. Using uniformat and gene[rate] to analyze data with ambiguities in population genetics. Evol Bioinform Online 2016;11:19–26. https://doi.org/10.4137/EBO.S32415.
[32] Zhao X, Ma S, Wang B, Jiang X, Initiative The Han100K, Xu SPGG. MHC: toward understanding the diversity of major histocompatibility complexes in human populations. Nucleic Acids Res 2023;51:D1102–8. https://doi.org/10.1093/nar/gkac997.
[33] Cano P, Klitz W, Mack SJ, Maiers M, Marsh SGE, Noreen H, et al. Common and well-documented HLA alleles: report of the ad-hoc committee of the American society for histocompatiblity and immunogenetics. Hum Immunol 2007;68:392–417. https://doi.org/10.1016/j.humimm.2007.01.014.
[34] Peter BM, Petkova D, Novembre J. Genetic landscapes reveal how human genetic diversity aligns with geography. Mol Biol Evol 2020;37:943–51. https://doi.org/10.1093/molbev/msz280.
[35] Serre D, Pääbo S. Evidence for gradients of human genetic diversity within and among continents. Genome Res 2004;14:1679–85. https://doi.org/10.1101/gr.2529604.
[36] Vizzari MT, Benazzo A, Barbujani G, Ghirotto S. A revised model of anatomically modern human expansions out of Africa through a machine learning approximate bayesian computation approach. Genes 2020;11:1510. https://doi.org/10.3390/genes11121510.
[37] Hublin J-J, Ben-Ncer A, Bailey SE, Freidline SE, Neubauer S, Skinner MM, et al. New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens. Nature 2017;546:289–92. https://doi.org/10.1038/nature22336.
[38] Vidal CM, Lane CS, Asrat A, Barfod DN, Mark DF, Tomlinson EL, et al. Age of the oldest known Homo sapiens from eastern Africa. Nature 2022;601:579–83. https://doi.org/10.1038/s41586-021-04275-8.
[39] Lorente-Galdos B, Lao O, Serra-Vidal G, Santpere G, Kuderna LFK, Arauna LR, et al. Whole-genome sequence analysis of a Pan African set of samples reveals archaic gene flow from an extinct basal population of modern humans into sub-Saharan populations. Genome Biol 2019;20:77. https://doi.org/10.1186/s13059-019-1684-5.
[40] Tournebize R, Chu G, Moorjani P. Reconstructing the history of founder events using genome-wide patterns of allele sharing across individuals. PLoS Genet 2022;18:e1010243. https://doi.org/10.1371/journal.pgen.1010243.
[41] Barbieri C, Blasi DE, Arango-Isaza E, Sotiropoulos AG, Hammarström H, Wichmann S, et al. A global analysis of matches and mismatches between human genetic and linguistic histories. In: Proceedings of the national academy of sciences, vol. 119; 2022, e2122084119. https://doi.org/10.1073/pnas.2122084119.
[42] Černý V, Kulichová I, Poloni ES, Nunes JM, Pereira L, Mayor A, et al. Genetic history of the african sahelian populations. HLA 2018;91:153–66. https://doi.org/10.1111/tan.13189.
[43] Arrieta-Bolaños E, Hernández-Zaragoza DI, Barquera R. 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 2023;14.
[44] Buhler S, Sanchez-Mazas A. HLA DNA sequence variation among human populations: molecular signatures of demographic and selective events. PLoS One 2011;6:e14643. https://doi.org/10.1371/journal.pone.0014643.
[45] Cao K, Moormann Am, Lyke Ke, Masaberg C, Sumba Op, Doumbo Ok, et al. Differentiation between African populations is evidenced by the diversity of alleles and haplotypes of HLA class I loci. Tissue Antigens 2004;63:293–325. https://doi.org/10.1111/j.0001-2815.2004.00192.x.
[46] Di D, Sanchez-Mazas A, Currat M. Computer simulation of human leukocyte antigen genes supports two main routes of colonization by human populations in East Asia. BMC Evol Biol 2015;15:240. https://doi.org/10.1186/s12862-015-0512-0.
[47] Di D, Sanchez-Mazas A. Challenging views on the peopling history of East Asia: the story according to HLA markers. Am J Phys Anthropol 2011;145:81–96. https://doi.org/10.1002/ajpa.21470.

[48] Mack Sj, Bugawan Tl, Moonsamy Pv, Erlich Ja, Trachtenberg Ea, Paik Yk, et al. Evolution of Pacific/Asian populations inferred from HLA class II allele frequency distributions. Tissue Antigens 2000;55:383–400. https://doi.org/10.1034/j.1399-0039.2000.550501.x.
[49] Meyer D, Single RM, Mack SJ, Erlich HA, Thomson G. Signatures of demographic history and natural selection in the human major histocompatibility complex loci. Genetics 2006;173:2121–42. https://doi.org/10.1534/genetics.105.052837.
[50] Sanchez-Mazas A, Buhler S, Nunes JM. A new HLA map of Europe: regional genetic variation and its implication for peopling history, disease-association studies and tissue transplantation. Hum Hered 2014;76:162–77. https://doi.org/10.1159/000360855.
[51] Single RM, Meyer D, Nunes K, Francisco RS, Hünemeier T, Maiers M, et al. Demographic history and selection at HLA loci in Native Americans. PLoS One 2020; 15:e0241282. https://doi.org/10.1371/journal.pone.0241282.
[52] Rees JS, Castellano S, Andrés AM. The genomics of human local adaptation. Trends Genet 2020;36:415–28. https://doi.org/10.1016/j.tig.2020.03.006.
[53] Feng Y, McQuillan MA, Tishkoff SA. Evolutionary genetics of skin pigmentation in African populations. Hum Mol Genet 2021;30:R88–97. https://doi.org/10.1093/hmg/ddab007.
[54] Rocha J. The evolutionary history of human skin pigmentation. J Mol Evol 2020;88:77–87. https://doi.org/10.1007/s00239-019-09902-7.
[55] Basak N, Thangaraj K. High-altitude adaptation: role of genetic and epigenetic factors. J Biosci 2021;46:107. https://doi.org/10.1007/s12038-021-00228-5.
[56] Bigham AW. Genetics of human origin and evolution: high-altitude adaptations. Curr Opin Genet Dev 2016;41:8–13. https://doi.org/10.1016/j.gde.2016.06.018.
[57] Anguita-Ruiz A, Aguilera CM, Gil Á. Genetics of lactose intolerance: an updated review and online interactive world maps of phenotype and genotype frequencies. Nutrients 2020;12:2689. https://doi.org/10.3390/nu12092689.
[58] Gerbault P, Moret C, Currat M, Sanchez-Mazas A. Impact of selection and demography on the diffusion of lactase persistence. PLoS One 2009;4:e6369. https://doi.org/10.1371/journal.pone.0006369.
[59] Fan S, Hansen MEB, Lo Y, Tishkoff SA. Going global by adapting local: a review of recent human adaptation. Science 2016;354:54–9. https://doi.org/10.1126/science.aaf5098.
[60] Band G, Le QS, Clarke GM, Kivinen K, Hubbart C, Jeffreys AE, et al. Insights into malaria susceptibility using genome-wide data on 17,000 individuals from Africa, Asia and Oceania. Nat Commun 2019;10:5732. https://doi.org/10.1038/s41467-019-13480-z.
[61] Kariuki SN, Williams TN. Human genetics and malaria resistance. Hum Genet 2020;139:801–11. https://doi.org/10.1007/s00439-020-02142-6.
[62] Dendrou CA, Petersen J, Rossjohn J, Fugger L. HLA variation and disease. Nat Rev Immunol 2018;18:325–39. https://doi.org/10.1038/nri.2017.143.
[63] Matzaraki V, Kumar V, Wijmenga C, Zhernakova A. The MHC locus and genetic susceptibility to autoimmune and infectious diseases. Genome Biol 2017;18:76. https://doi.org/10.1186/s13059-017-1207-1.
[64] Sanchez-Mazas A. A review of HLA allele and SNP associations with highly prevalent infectious diseases in human populations. Swiss Med Wkly 2020;150: w20214. https://doi.org/10.4414/smw.2020.20214. w20214.
[65] Trowsdale J, Knight JC. Major histocompatibility complex genomics and human disease. Annu Rev Genom Hum Genet 2013;14:301–23. https://doi.org/10.1146/annurev-genom-091212-153455.
[66] Maróstica AS, Nunes K, Castelli EC, Silva NSB, Weir BS, Goudet J, et al. How HLA diversity is apportioned: influence of selection and relevance to transplantation. Phil Trans Biol Sci 2022;377:20200420. https://doi.org/10.1098/rstb.2020.0420.
[67] Sanchez-Mazas A. An apportionment of human HLA diversity. Tissue Antigens 2007;69:198–202. https://doi.org/10.1111/j.1399-0039.2006.00802.x.
[68] Barbujani G, Magagni A, Minch E, Cavalli-Sforza LL. An apportionment of human DNA diversity. Proc Natl Acad Sci USA 1997;94:4516–9. https://doi.org/10.1073/pnas.94.9.4516.
[69] Lewontin RC. The apportionment of human diversity. In: Dobzhansky T, Hecht MK, Steere WC, editors. Evolutionary biology, vol. 6. New York, NY: Springer US; 1972. p. 381–98. https://doi.org/10.1007/978-1-4684-9063-3_14.
[70] Begovich Ab, Moonsamy Pv, Mack Sj, Barcellos Lf, Steiner Ll, Grams S, et al. Genetic variability and linkage disequilibrium within the HLA-DP region: analysis of 15 different populations. Tissue Antigens 2001;57:424–39. https://doi.org/10.1034/j.1399-0039.2001.057005424.x.
[71] Meyer D, Thomson G. How selection shapes variation of the human major histocompatibility complex: a review. Ann Hum Genet 2001;65:1–26. https://doi.org/10.1046/j.1469-1809.2001.6510001.x.
[72] Valdes AM, McWEENEY SK, Meyer D, Nelson MP, Thomson G. Locus and population specific evolution in HLA class II genes. Ann Hum Genet 1999;63:27–43. https://doi.org/10.1046/j.1469-1809.1999.6310027.x.
[73] Pierini F, Lenz TL. Divergent allele advantage at human MHC genes: signatures of past and ongoing selection. Mol Biol Evol 2018;35:2145–58. https://doi.org/10.1093/molbev/msy116.
[74] Wakeland EK, Boehme S, She JX, Lu C-C, McIndoe RA, Cheng I, et al. Ancestral polymorphisms of MHC class II genes: divergent allele advantage. Immunol Res 1990;9:115–22. https://doi.org/10.1007/BF02918202.
[75] Buhler S, Nunes JM, Sanchez-Mazas A. HLA class I molecular variation and peptide-binding properties suggest a model of joint divergent asymmetric selection. Immunogenetics 2016;68:401–16. https://doi.org/10.1007/s00251-016-0918-x.
[76] Di D, Nunes JM, Jiang W, Sanchez-Mazas A. Like wings of a bird: functional divergence and complementarity between HLA-A and HLA-B molecules. Mol Biol Evol 2020;38:1580–94. https://doi.org/10.1093/molbev/msaa325.
[77] Yasukochi Y, Satta Y. Current perspectives on the intensity of natural selection of MHC loci. Immunogenetics 2013;65:479–83. https://doi.org/10.1007/s00251-013-0693-x.
[78] Sanchez-Mazas A, Černý V, Di D, Buhler S, Podgorná E, Chevallier E, et al. The HLA-B landscape of Africa: signatures of pathogen-driven selection and molecular identification of candidate alleles to malaria protection. Mol Ecol 2017;26:6238–52. https://doi.org/10.1111/mec.14366.
[79] dos Santos Francisco R, Buhler S, Nunes JM, Bitarello BD, França GS, Meyer D, et al. HLA supertype variation across populations: new insights into the role of natural selection in the evolution of HLA-A and HLA-B polymorphisms. Immunogenetics 2015;67:651–63. https://doi.org/10.1007/s00251-015-0875-9.
[80] Augusto DG, Petzl-Erler ML. KIR and HLA under pressure: evidences of coevolution across worldwide populations. Hum Genet 2015;134:929–40. https://doi.org/10.1007/s00439-015-1579-9.
[81] Velastegui E, Vera E, Vanden Berghe W, Muñoz MS, Orellana-Manzano A. HLA-C: evolution, epigenetics, and pathological implications in the major histocompatibility complex. Front Genet 2023;14. https://doi.org/10.3389/fgene.2023.1206034.
[82] Goeury T, Creary LE, Brunet L, Galan M, Pasquier M, Kervaire B, et al. Deciphering the fine nucleotide diversity of full HLA class I and class II genes in a well-documented population from sub-Saharan Africa. HLA 2018;91:36–51. https://doi.org/10.1111/tan.13180.
[83] Silva N do, Souza A da S, Andrade H de S, Pereira RN, Castro CFB, Vince N, et al. Immunogenetics of HLA-B: SNP, allele, and haplotype diversity in populations from different continents and ancestry backgrounds. HLA 2023;101:634–46. https://doi.org/10.1111/tan.15043.
[84] Sanchez-Mazas A, Fernandez-Viña M, Middleton D, Hollenbach JA, Buhler S, Di D, et al. Immunogenetics as a tool in anthropological studies. Immunology 2011; 133:143–64. https://doi.org/10.1111/j.1365-2567.2011.03438.x.
[85] Consortium T1000GP. A global reference for human genetic variation. Nature 2015;526:68. https://doi.org/10.1038/nature15393.
[86] Vangenot C, Nunes JM, Doxiadis GM, Poloni ES, Bontrop RE, de Groot NG, et al. Similar patterns of genetic diversity and linkage disequilibrium in Western chimpanzees (Pan troglodytes verus) and humans indicate highly conserved mechanisms of MHC molecular evolution. BMC Evol Biol 2020;20:119. https://doi.org/10.1186/s12862-020-01669-6.
[87] Furukawa H, Kawasaki A, Oka S, Ito I, Shimada K, Sugii S, et al. Human leukocyte antigens and systemic lupus erythematosus: a protective role for the HLA-DR6 alleles DRB1*13:02 and *14:03. PLoS One 2014;9:e87792. https://doi.org/10.1371/journal.pone.0087792.
[88] Sun H, Yang Z, Lin K, Liu S, Huang K, Wang X, et al. The adaptive change of HLA-DRB1 allele frequencies caused by natural selection in a Mongolian population that migrated to the South of China. PLoS One 2015;10:e0134334. https://doi.org/10.1371/journal.pone.0134334.
[89] Manczinger M, Boross G, Kemény L, Müller V, Lenz TL, Papp B, et al. Pathogen diversity drives the evolution of generalist MHC-II alleles in human populations. PLoS Biol 2019;17:e3000131. https://doi.org/10.1371/journal.pbio.3000131.

[90] Nunes K, Maia MHT, dos Santos EJM, dos Santos SEB, Guerreiro JF, Petzl-Erler ML, et al. How natural selection shapes genetic differentiation in the MHC region: a case study with Native Americans. Hum Immunol 2021;82:523–31. https://doi.org/10.1016/j.humimm.2021.03.005.
[91] Slatkin M, Takahata N. The average frequency of private alleles in a partially isolated population. Theor Popul Biol 1985;28:314–31. https://doi.org/10.1016/0040-5809(85)90032-2.
[92] Wang S, Lewis CM, Jakobsson M, Ramachandran S, Ray N, Bedoya G, et al. Genetic variation and population structure in native Americans. PLoS Genet 2007;3:e185. https://doi.org/10.1371/journal.pgen.0030185.
[93] Keinan A, Clark AG. Recent explosive human population growth has resulted in an excess of rare genetic variants. Science 2012;336:740–3. https://doi.org/10.1126/science.1217283.
[94] Bergström A, McCarthy SA, Hui R, Almarri MA, Ayub Q, Danecek P, et al. Insights into human genetic variation and population history from 929 diverse genomes. Science 2020;367:eaay5012. https://doi.org/10.1126/science.aay5012.
[95] Chappell PE, Meziane EK, Harrison M, Magiera Ł, Hermann C, Mears L, et al. Expression levels of MHC class I molecules are inversely correlated with promiscuity of peptide binding. Elife 2015;4:e05345. https://doi.org/10.7554/eLife.05345.
[96] Randolph HE, Aracena KA, Lin Y-L, Mu Z, Barreiro LB. Shaping immunity: the influence of natural selection on population immune diversity. Immunol Rev 2024. https://doi.org/10.1111/imr.13329.
[97] Slieker RC, Warmerdam DO, Vermeer MH, van Doorn R, Heemskerk MHM, Scheeren FA. Reassessing human MHC-I genetic diversity in T cell studies. Sci Rep 2024;14:7966. https://doi.org/10.1038/s41598-024-58777-2.
[98] Meyer D, Aguiar Vr C, Bitarello Bd C, Brandt DY, Nunes K. A genomic perspective on HLA evolution. Immunogenetics 2018;70:5–27. https://doi.org/10.1007/s00251-017-1017-3.
[99] Dehn J, Spellman S, Hurley CK, Shaw BE, Barker JN, Burns LJ, et al. Selection of unrelated donors and cord blood units for hematopoietic cell transplantation: guidelines from the NMDP/CIBMTR. Blood 2019;134:924–34. https://doi.org/10.1182/blood.2019001212.