Human Immunology 80 (2019) 62–66

Contents lists available at ScienceDirect

Human Immunology

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

Does NGS typing highlight our understanding of HLA population diversity? Some good reasons to say *yes* and a few to say *be careful*

Alicia Sanchez-Mazas*, José Manuel Nunes*

Department of Genetics and Evolution – Anthropology Unit, University of Geneva and Institute of Genetics and Genomics in Geneva (IGE3), Switzerland

ARTICLE INFO

Keywords:
Next generation sequencing (NGS)
HLA
Population genetics
Disease associations
Molecular evolution
Nucleotide diversity
Evolutionary models

ABSTRACT

This paper discusses the advantages provided by next generation sequencing (NGS) compared to traditional typings or limited sequencing strategies for the characterization of HLA population diversity based on four documented examples. We also comment the limitations of this approach by highlighting pitfalls in interpreting NGS data.

1. Introduction

The fundamental role played by the HLA molecules both in our natural immunity and in clinical transplantation drives the researchers in immunogenetics and histocompatibility to constantly improve their sequencing facilities for characterizing the HLA molecular diversity of their samples. Several next-generation sequencing (NGS) technologies with competing achievements and costs are currently proposed [1–3], but a very precise HLA genotyping still remains particularly challenging due to the never-ending identification of new HLA alleles (their last census summing up to 18,995 according to IMGT/HLA release 3.33 in July 2018), the complexity of their molecular variation which targets different coding and non-coding regions (exons, introns and UTRs) and their huge diversity observed both within and among different human populations [4–6]. In this context, ambitious DNA sequencing projects such as 1000 Genomes [7] have not yet constituted high-coverage sequencing HLA databases that could be used as references in population and evolutionary genetics research, although important efforts are being done at the international level [8–10]. Still, precise HLA sequencing is expected to bring relevant information on human genetic variation compared to traditional typings [11] although we should be careful not to over-emphasize its potential achievements. In this manuscript, we would like to clarify to what extent NGS typing highlights our understanding of HLA population diversity first by presenting four examples where our research group and co-workers successfully used NGS data (as presented at the 17th International HLA and Immunogenetics Workshop held in Asilomar (USA), 6–10 September 2017, Symposium #S1c: JMN & ASM *Population genetics of HLA: in the era of NGS*) and then by discussing some pitfalls of interpreting impetuously NGS results.

2. Four advantages of using NGS to assess HLA population diversity

1. Increased potential for the identification of HLA alleles associated to diseases

It is widely admitted that genetic variation occurring within the HLA region may be associated to susceptibility or resistance to autoimmune and infectious diseases [12,13]. Recently, we investigated the assumed protective effect [14] of allele HLA-B*53 to *Plasmodium falciparum* malaria at the population level by testing this association in populations located in both malaria endemic and non-endemic areas in Africa [15]. The first step of this study was to use a very large set of 40 African populations genotyped for the HLA-A and -B loci at the 1st field level of resolution in a linear modelling approach taking into account several genetic, geographic and environmental parameters, among which *P. falciparum* prevalence data provided by the Malaria Atlas project [16]. The results confirmed a highly significant association of allele B*53 with malaria prevalence, but also of alleles B*78 and A*74. In a second step, we used a NGS methodology to retype all B*53, B*78 and A*74 individuals for which DNA was available to us. In all B*53-positive individuals (N = 55), this allele was identified almost exclusively as B*53:01:01 (homozygous in 3 samples), the only exception being one rare B*53:19 allele found in a small semi-nomadic population

* Corresponding authors.
E-mail addresses: Alicia.Sanchez-Mazas@unige.ch (A. Sanchez-Mazas), Jose.deAbreuNunes@unige.ch (J.M. Nunes).

https://doi.org/10.1016/j.humimm.2018.10.004
Received 4 September 2018; Received in revised form 8 October 2018; Accepted 9 October 2018
Available online 02 November 2018
0198-8859/

(the Rashaayda) living in a non-endemic area of the Sahel Belt in Sudan. By NGS typing locus HLA-C in the 55 positive individuals, we also found that 34 of them (62%) carried the HLA-C*04:01:01:01 allele, strongly suggesting that the haplotype B*53:01:01 ~ C*04:01:01:01 (homozygous in one sample) was associated to *P. falciparum* resistance in Africa. Moreover, all B*78-positive individuals were B*78:01:01, indicating that this allele was also a candidate to malaria protection. Interestingly, peculiar allotypes showing functional similarities to HLA-B*53 and -B*78 have recently been observed in pygmy chimpanzees (bonobos) which are almost not affected by malaria parasites despite living in a highly endemic area [17]. In contrast to B*53 and B*78, the NGS-retyped A*74 alleles were found to be an amalgamation of two different subtypes (A*74:01 and A*74:03) that share a functional affinity with another HLA-A allele, A*33:01, which was proposed as a susceptibility, rather than a protective factor to a severe form of malaria. In addition, the frequency of A*74 was also correlated (therefore in a confounding way with malaria prevalence) with a geographic parameter (latitude). These latter results together contradicted the idea that A*74 was also protective to this disease. In conclusion, this study is a nice example where NGS typing added a real *plus-value* to a population-based disease-association analysis by leading us not only to identify with precision two HLA-B alleles and one HLA-B ~ C haplotype putatively protective to malaria in Africa, but also by invalidating the same hypothesis regarding one HLA-A allele that revealed preliminary associations to malaria prevalence when typed with traditional techniques. Other recent studies have also shown the great interest of NGS to assess HLA associations to diseases through case-control analyses (e.g. [18–22]). Clearly, the progress done in obtaining phased sequences as a result of clonal sequencing in NGS allows to substantially reduce typing ambiguities compared to former sequencing techniques such as Sanger and in turn, to improve statistical analyses by making less assumptions and obtaining clear results for the putative associations of diseases' prevalence with allele frequencies.

2. New ways of detecting demographic signatures on population variation

Traditional HLA typings performed in hundreds of populations worldwide and their resulting allele frequency distributions have been most useful to describe HLA genetic variation within and across the different continents and to confirm the major role of human migrations in shaping the observed HLA genetic patterns [15,23–26]. In these studies, however, only a limited part of the molecular information relative to each population was taken into account, because HLA alleles were defined nominally at either the 1st or the 2nd field level of resolution without considering their molecular differences in terms of nucleotide substitutions. Now, HLA sequencing allows to use the average number of nucleotide differences among pairs of sequences in a given population, named the nucleotide diversity (π) in molecular population genetics [27], as a very useful measure for detecting signatures of demographic events and/or selective pressures, as those that affect HLA genes on populations [see for example 28]. In a previous article where we reviewed the relevance of HLA sequencing in population genetics studies [11], we provided an example where the distributions of the number of nucleotide differences between pairs of sequences at the three HLA loci A, B and DRB1 (calculated by using the molecular information of exons 2 and 3 stored in the IMGT/HLA database), named *mismatch distributions*, were plotted for two populations that underwent distinct demographic histories in the past, namely the Ami from Taiwan (data from [29]) and the Uyghur from Central Asia (data from [30]). The first population, which shows a very peculiar genetic profile [31,32], is one of the 14 still living Taiwanese (Austronesian-speaking) aboriginal tribes that diversified in the island after its colonization from China by ancestral farmers about 5500 years before present [33] and that mainly remained confined to isolated areas when Taiwan was settled by European and Chinese people since the 17th Century. The second population is a Turkic-speaking ethnic group that likely underwent extensive gene flow from both European and Asian populations [34,35]. Then, whereas the Uyghur exhibit a likely unimodal shape of their mismatch distributions at the three loci, suggesting demographic expansion, the Ami distributions are very irregular and reveal high proportions of identical sequences likely due to the presence of many homozygote individuals, reflecting population contraction. Although natural selection may create confounding effects when interpreting mismatch distributions based on coding regions, as is the case in this example, the results described above were obtained on three HLA genes that likely underwent similar selective pressures in two populations that followed very distinct demographic histories, which clearly confirms that demographic signals are easily detectable when using precise molecular information (i.e. DNA sequences) rather than nominal definitions for HLA alleles. However, as molecular variation is also found outside exons 2 and 3 (which is illustrated in the next example given below), a more detailed determination of nucleotide variation across the entire HLA genes by using NGS technologies would allow refining the results provided by this kind of approach. Indeed, the effects of either demography or natural selection on HLA diversity patterns would be easier to disentangle by comparing the nucleotide diversity of different coding regions to that of non-coding regions that would not be direct targets of selection or that would be submitted to distinct kinds of selective pressures. The generation of new NGS data for the huge set of population samples that were already HLA-typed with traditional techniques or limited sequencing strategies so far (e.g. data submitted to the IHIWs) would thus clearly be an excellent strategy to improve our knowledge on populations' history, and we strongly encourage laboratories to follow this direction.

3. Powerful tool to decipher the evolutionary processes that shaped the HLA diversity patterns

The greatest proportion of HLA molecular variation has been reported by Hughes and co-workers to lie within the exons encoding the peptide-binding region (or antigen recognition domain, hereafter ARD) of the HLA molecules, i.e. exons 2 and 3, for class I, and exon 2, for class II genes [36–38]. This is the reason why HLA typings have been predominantly applied to these exons during several decades. However, the works published by these authors more than 20 years ago used only a few samples that are unable to cover the huge diversity of the HLA region described today within and across different populations worldwide; as we have shown for supertypes [39], using only a few samples rather than extensive population data may yield substantially different results. The "high level of diversity of the ARD", which is widely recognized, therefore needs more evidence. To our knowledge, the work that we recently performed to decipher the fine nucleotide diversity across seven HLA genes in a well-documented population from sub-Saharan Africa, the Mandenka from Eastern Senegal [40], is among the first studies that compared the molecular diversity between different HLA regions at the population level by using NGS sequencing. This study first allowed us to observe a greater nucleotide diversity at the antigen-recognition-site (ARS) codons of exons 2 and 3 (for class I genes) and of exon 2 (for class II genes) compared to both the non-ARS codons and the other exons, this variation being even more pronounced when considering only non-synonymous (i.e. changing the amino-acid) substitutions. Although limited to a single population, our work thus confirms the results of Hughes and co-workers and supports the advantage of diversity for peptide presentation. Interestingly, we also found pronounced differences (although less noticeable at locus B) between HLA class I exons 2 and 3, the greatest diversity lying at exon 3, but a lesser variability of both ARS and non-ARS codons at HLA-C compared to HLA-A and –B. HLA-A exon 2 also appeared as an outlier regarding several molecular diversity indexes, most particularly Tajima's D that compares two estimates of the population parameter Θ to discriminate between the effects of, on one side, positive selection or demographic expansion and, on the other side, balancing selection or demographic contraction. In addition, the nucleotide diversity of exon 2

differed markedly between the different HLA class II genes, being the highest at DRB1 and DQB1, the lowest at DPB1 and DPA1 and intermediate at DQA1 (DRA was not considered). Such contrasts between distinct HLA class I and/or class II exons were tentatively attributed to different roles played, in the immune process, by their corresponding domains on the HLA molecules, i.e. a greater involvement of the α2 domain (coded by exon 3) of HLA class I molecules in peptide presentation, as compared to the α1 domain (coded by exon 2); a minor role played by the α1 domain of HLA-A (as compared to HLA-B) in peptide presentation and hence a more neutral evolution for this region; a main role played by the α1 domain of HLA-C in KIR interactions; and very distinct contributions of the β1 domain of diverse HLA class II molecules (possibly targeted by dissimilar selective forces) to peptide presentation. Finally, by comparing the DRB1 sequences found in the Senegalese Mandenka to those of the IPD-IMGT/HLA database, we could propose a plausible mechanism of gene conversion explaining the emergence of DRB1*13:04 (which exhibits a high frequency of 28% in the Mandenka) through an unidirectional transfer of genetic material from a donor allele (DRB1*04:05:01 or *08:06 or *13:03:01) to the recipient allele DRB1*11:02:01. This event would have been followed by a strong selective sweep of haplotype DRB1*13:04 ~ DQA1*05:05:01 ~ DQB1*03:19 (in highly significant linkage disequilibrium in the Mandenka) possibly due to a protective effect to *Onchocerca volvulus*, the parasitic worm responsible for *Onchocerciasis*, a widely spread and very deleterious infectious disease in West Africa [41]. The analysis of HLA introns' and UTRs' variability in the Mandenka as well as similar analyses of full-length HLA sequences in populations from other geographic areas are currently in progress. However, the results obtained so far demonstrate that NGS analyses applied to HLA data are most informative to understand the evolutionary processes that shaped the HLA diversity patterns observed in humans.

4. Promising data to understand the mechanisms involved in the immune process

Our last example comes from a study where we used both HLA class I sequences to estimate the nucleotide diversity of 46 populations from different continents and computer-based peptide-binding predictions to estimate the possible immune potential of each of these populations [42]. Our aim was to investigate more in depth the hypothesis of *divergent allele advantage* that is often proposed to explain the high levels of heterozygosity found at HLA genes [43–46]. This model suggests that heterozygote individuals carrying molecularly divergent alleles (most often implying divergent peptide-binding properties) would be more protected to a large variety of pathogens than heterozygotes carrying more similar alleles. Actually, we were puzzled by the fact that some small isolated populations, like the Taiwanese Ami mentioned above and many other native people from North and South America, Island Southeast Asia, Australia and the Pacific often exhibited a reduced diversity at individual HLA genes [4]; was it not contradictory with the idea that we need HLA diversity to be protected? Based on the two categories of data mentioned above, we estimated two new parameters, the relative increase of molecular distance (RIMD) and the relative gain in peptide-binding coverage (RGPBC) aiming at summarizing the immune potential of an individual. The idea is that a heterozygote would exhibit an increasingly greater potential to bind peptides compared to a homozygote if his/her HLA molecules showed either an increasingly greater molecular distance or an increasingly greater peptide-binding distance (both up to a maximum of 1) between its two alleles compared to the homozygote where both distances would be zero. To consider these measures at the population level, we averaged them on all individuals of each population sample. In addition, we estimated them both at each individual locus A, B and C and on several loci at the same time (AB, AC, BC and ABC). The results revealed contrasted values of both indexes at single loci (with low values observed in small-sized and isolated populations) but similar values of both of them in all populations when either the two loci A and B or the three loci A, B and C were considered together (the lowest variance being observed in the latter case). We thus proposed a model of *joint divergent asymmetric selection* suggesting that all populations share a similar immune potential thanks to the simultaneous involvement of the three loci A, B and C, in support to the idea that multi-locus effects play an important role in the HLA region [47].

This example shows that HLA sequence data may be highly informative to understand the mechanisms involved in the immune process. Now, to fully sketch and understand such a model of cooperation between alleles of distinct loci, we need to extend the analyses beyond the antigen recognition domain sequences, i.e. to other exons where allelic variability is also frequently reported, but also to non-coding regions that might be involved in regulatory mechanisms (e.g. gene expression levels), and this is where NGS may provide essential information. Our ongoing research using NGS data is thus focusing on nucleotide substitution patterns lying at both ARD and non-ARD regions to examine whether the *joint divergent asymmetric selection* model is only a feature of the ARD or is a general mechanism of selection acting on HLA; NGS approaches thus give us fascinating perspectives again.

3. Some pitfalls of interpreting blindly NGS data

One of the most extraordinary ability of humans is to build knowledge in a cumulative way by assimilating former information with new developments to go beyond present. The HLA world should not deviate from this strategy, i.e. it should not make a clean sweep of several decades of traditional technologies that generated thousands of genotypes used in many research fields related to immunogenetics. The examples described above show that several kinds of information – i.e. "old" and "new" - can be combined to reach new objectives in population and evolutionary genetics. Also, most of our knowledge regarding HLA diversity within the peptide-binding region has been obtained before the development of high-throughput sequencing strategies, which indicates that traditional or limited sequencing techniques hold a great potential for histocompatibility as much as for research focusing on the immune function. In addition, single nucleotide variants (SNVs) representing crucial regulatory sites governing HLA gene expression have been found [48] and are still to be discovered outside the regions that are usually typed by traditional techniques, and NGS has thus the immense power to widen our field-of-view in these domains. We should thus consider that the different technologies still play a complementary role in HLA research.

We also ought to keep in mind that even the most spectacular scientific results are worthless if not reproducible. Therefore, besides the complementarity of different typing techniques, the statistical power of any approach has to be taken seriously. NGS has the capacity to reveal very rapidly huge amounts of new alleles (e.g. [49–53] among others in 2018) because the level of resolution is considerably higher than the 2nd field. However, to accurately describe the genetic profile of a population at such high resolution levels actually needs a simultaneous increase of the sample size; otherwise several samples of equal size drawn from the same population are expected to reveal distinct HLA profiles that do not represent that of the studied population. We also find essential to recall that rare alleles discovered by sampling just a handful of people cannot be considered as unique or specific of the studied population as long as typing large population samples at the same resolution level has not been performed at a wide geographic scale (note that this comment does not apply specifically to NGS studies). Finally, many if not most HLA alleles likely display a rapid turnover which makes them appear and disappear by chance (genetic drift) or through selective processes [54]. Therefore, albeit exciting and informative to explore the mechanisms governing the evolution of the HLA region, rare alleles may not be relevant to characterize and compare populations.

NGS clearly brings new, highly relevant and original data but this should not translate in ignoring other approaches that may still be used

by small laboratories and that may provide satisfactory results for some studies. An example that comes to mind is the catalogs of common and well-documented alleles, mostly defined at the 2nd field level of resolution, that have been published by different teams [55–60]. Such catalogues might become meaningless if extended at higher (e.g. NGS) resolution levels because the probability of finding several copies of an allele defined at the 3rd or 4th field level of resolution becomes drastically low if the sample sizes are not appropriately increased: the best (resolution) may be the enemy of the good. Of course, we can always adjust the 3rd or 4th field levels to something less, and this should actually be considered as a wise decision in order to prevent a misinterpretation of the results (in particular NGS allele frequencies) when adequate statistical conditions (e.g. large sample sizes) are not fulfilled. As mentioned by a reviewer of the present perspective, to whom we are grateful, "*if we have the information and we need it, we use it. If we do not need it, we do not use it. But, if we do not have it then we cannot use it, whether we need it or not. This is, basically, the idea of NGS; have more information. If the effort, cost and time to have more information is not prohibitive, then have it*". But just keep in mind that the interpretation of NGS data has also its limitations in HLA population studies.

Acknowledgments

We thank the Swiss National Science Foundation (SNSF) grant no. 31003A_144180 and all co-authors of the published papers summarized in this manuscript for their precious work and collaboration. We are also deeply indebted to three anonymous reviewers for their useful comments.

References

[1] S. Boegel, HLA Typing: Methods and Protocols, Humana Press, Springer Nature, New York, 2018.
[2] N. Cereb, H.R. Kim, J. Ryu, S.Y. Yang, Advances in DNA sequencing technologies for high resolution HLA typing, Hum. Immunol. 76 (2015) 923.
[3] C. Gabriel, D. Furst, I. Fae, S. Wenda, C. Zollikofer, J. Mytilineos, et al., HLA typing by next-generation sequencing – getting closer to reality, Tissue Antigens 83 (2014) 65.
[4] S. Buhler, A. Sanchez-Mazas, HLA DNA sequence variation among human populations: molecular signatures of demographic and selective events, PLoS ONE 6 (2011) e14643.
[5] P.A. Gourraud, P. Khankhanian, N. Cereb, S.Y. Yang, M. Feolo, M. Maiers, et al., HLA diversity in the 1000 genomes dataset, PLoS ONE 9 (2014) e97282.
[6] O.D. Solberg, S.J. Mack, A.K. Lancaster, R.M. Single, Y. Tsai, A. Sanchez-Mazas, et al., Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies, Hum. Immunol. 69 (2008) 443.
[7] Genomes Project C, A. Auton, L.D. Brooks, R.M. Durbin, E.P. Garrison, H.M. Kang, et al., A global reference for human genetic variation, Nature 526 (2015) 68.
[8] J. Robinson, A.R. Soormally, J.D. Hayhurst, S.G.E. Marsh, The IPD-IMGT/HLA database – new developments in reporting HLA variation, Hum. Immunol. 77 (2016) 233.
[9] D. Monos, M.J. Maiers, Progressing towards the complete and thorough characterization of the HLA genes by NGS (or single-molecule DNA sequencing): consequences, opportunities and challenges, Hum. Immunol. 76 (2015) 883.
[10] C.J. Chang, K. Osoegawa, R.P. Milius, M. Maiers, W. Xiao, M. Fernandez-Vina, et al., Collection and storage of HLA NGS genotyping data for the 17th International HLA and Immunogenetics Workshop, Hum. Immunol. 79 (2018) 77.
[11] A. Sanchez-Mazas, D. Meyer, The relevance of HLA sequencing in population genetics studies, J. Immunol. Res. 2014 (2014) 971818.
[12] C.A. Dendrou, J. Petersen, J. Rossjohn, L. Fugger, HLA variation and disease, Nat. Rev. Immunol. 18 (2018) 325.
[13] J. Trowsdale, J.C. Knight, Major histocompatibility complex genomics and human disease, Annu. Rev. Genomics Hum. Genet. 14 (2013) 301.
[14] A.V. Hill, C.E. Allsopp, D. Kwiatkowski, N.M. Anstey, P. Twumasi, P.A. Rowe, et al., Common west African HLA antigens are associated with protection from severe malaria, Nature 352 (1991) 595.
[15] A. Sanchez-Mazas, V. Cerny, D. Di, S. Buhler, E. Podgorna, E. Chevallier, et al., The HLA-B landscape of Africa: signatures of pathogen-driven selection and molecular identification of candidate alleles to malaria protection, Mol. Ecol. 26 (2017) 6238.
[16] S.I. Hay, R.W. Snow, The malaria Atlas Project: developing global maps of malaria risk, PLoS Med. 3 (2006) e473.
[17] N.G. de Groot, J.M.G. Stevens, R.E. Bontrop, Does the MHC Confer Protection against Malaria in Bonobos? Trends Immunol. 39 (2018) 768.
[18] L.E. Creary, K.C. Mallempati, S. Gangavarapu, S.J. Caillier, J.R. Oksenberg, M.A. Fernandez-Vina, Deconstruction of HLA-DRB1*04:01:01 and HLA-DRB1*15:01:01 class II haplotypes using next-generation sequencing in European-Americans with multiple sclerosis, Mult. Scler. (2018) 1352458518770019.
[19] N. Samaranayake, S.D. Fernando, N.F. Neththikumara, C. Rodrigo, N.D. Karunaweera, V.H. Dissanayake, Association of HLA class I and II genes with cutaneous leishmaniasis: a case control study from Sri Lanka and a systematic review, BMC Infect. Dis. 16 (2016) 292.
[20] S. Burillo-Sanz, M.A. Montes-Cano, J.R. Garcia-Lozano, L. Ortiz-Fernandez, N. Ortego-Centeno, F.J. Garcia-Hernandez, et al., Mutational profile of rare variants in inflammasome-related genes in Behcet disease: a next generation sequencing approach, Sci. Rep. 7 (2017) 8453.
[21] S.J. Mack, J. Udell, F. Cohen, K. Osoegawa, S.K. Hawbecker, D.A. Noonan, et al., High resolution HLA analysis reveals independent class I haplotypes and amino-acid motifs protective for multiple sclerosis, Genes Immun. (2018).
[22] L.P. Zhao, S. Alshiekh, M. Zhao, A. Carlsson, H.E. Larsson, G. Forsander, et al., Next-generation sequencing reveals that HLA-DRB3, -DRB4, and -DRB5 may be associated with islet autoantibodies and risk for childhood type 1 diabetes, Diabetes 65 (2016) 710.
[23] D. Di, A. Sanchez-Mazas, Challenging views on the peopling history of East Asia: the story according to HLA markers, Am. J. Phys. Anthropol. 145 (2011) 81.
[24] D. Di, A. Sanchez-Mazas, M. Currat, Computer simulation of human leukocyte antigen genes supports two main routes of colonization by human populations in East Asia, BMC Evol. Biol. 15 (2015) 240.
[25] D. Meyer, R.M. Single, S.J. Mack, H.A. Erlich, G. Thomson, Signatures of demographic history and natural selection in the human major histocompatibility complex Loci, Genetics 173 (2006) 2121.
[26] A. Sanchez-Mazas, S. Buhler, J.M. Nunes, A new HLA map of Europe: regional genetic variation and its implication for peopling history, disease-association studies and tissue transplantation, Hum. Hered. 76 (2013) 162.
[27] M.W. Hahn, Molecular Population Genetics, Sinauer, Oxford Unversity Press, London, 2018.
[28] L. Excoffier, S. Schneider, Why hunter-gatherer populations do not show signs of pleistocene demographic expansions, Proc. Natl. Acad. Sci. U.S.A. 96 (1999) 10597.
[29] C.C. Chu, M. Lin, F. Nakajima, H.L. Lee, S.L. Chang, T. Juji, et al., Diversity of HLA among Taiwan's indigenous tribes and the Ivatans in the Philippines, Tissue Antigens 58 (2001) 9.
[30] C.M. Shen, B.F. Zhu, Y.J. Deng, S.H. Ye, J.W. Yan, G. Yang, et al., Allele polymorphism and haplotype diversity of HLA-A, -B and -DRB1 loci in sequence-based typing for Chinese Uyghur ethnic group, PLoS ONE 5 (2010) e13458.
[31] A. Sanchez-Mazas, E.S. Poloni, G. Jacques, L. Sagart, HLA genetic diversity and linguistic variation in East Asia, in: L. Sagart, R. Blench, A. Sanchez-Mazas (Eds.), The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics, vol 1, London and New York, Routledge Curzon, 2005, p. 273.
[32] B. Sewerin, F.J. Cuza, M.N. Szmulewicz, D.J. Rowold, R.L. Bertrand-Garcia, R.J. Herrera, On the genetic uniqueness of the Ami aborigines of Formosa, Am. J. Phys. Anthropol. 119 (2002) 240.
[33] L. Sagart, The expansion of *Setaria* farmers in East Asia: A linguistic and archaeological model, in: A. Sanchez-Mazas, R. Blench, M.D. Ross, I. Peiros, M. Lin (Eds.), Past Human Migrations in East Asia, vol 5, Routledge, London and New York, 2008, p. 133.
[34] Q. Feng, Y. Lu, X. Ni, K. Yuan, Y. Yang, X. Yang, et al., Genetic history of Xinjiang's Uyghurs suggests bronze age multiple-way contacts in eurasia, Mol. Biol. Evol. 34 (2017) 2572.
[35] G. He, Z. Wang, M. Wang, T. Luo, J. Liu, Y. Zhou, et al., Forensic ancestry analysis in two Chinese minority populations using massively parallel sequencing of 165 ancestry-informative SNPs, Electrophoresis (2018).
[36] A.L. Hughes, M. Yeager, Natural selection at major histocompatibility complex loci of vertebrates, Annu. Rev. Genet. 32 (1998) 415.
[37] A.L. Hughes, M. Nei, Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection, Proc. Natl. Acad. Sci. U.S.A. 86 (1989) 958.
[38] A.L. Hughes, M. Nei, Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection, Nature 335 (1988) 167.
[39] Francisco R Dos Santos, S. Buhler, J.M. Nunes, B.D. Bitarello, G.S. Franca, D. Meyer, 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 67 (2015) 651.
[40] T. Goeury, L.E. Creary, L. Brunet, M. Galan, M. Pasquier, B. Kervaire, 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 91 (2018) 36.
[41] S.J. O'Hanlon, H.C. Slater, R.A. Cheke, B.A. Boatin, L.E. Coffeng, S.D. Pion, et al., Model-based geostatistical mapping of the prevalence of onchocerca volvulus in West Africa, PLoS Negl Trop Dis 10 (2016) e0004328.
[42] S. Buhler, J.M. Nunes, A. Sanchez-Mazas, HLA class I molecular variation and peptide-binding properties suggest a model of joint divergent asymmetric selection, Immunogenetics 68 (2016) 401.
[43] P.G. Bronson, S.J. Mack, H.A. Erlich, M. Slatkin, A sequence-based approach demonstrates that balancing selection in classical human leukocyte antigen (HLA) loci is asymmetric, Hum. Mol. Genet. 22 (2013) 252.
[44] Q. Lau, Y. Yasukochi, Y. Satta, A limit to the divergent allele advantage model supported by variable pathogen recognition across HLA-DRB1 allele lineages, Tissue Antigens 86 (2015) 343.
[45] F. Pierini, T.L. Lenz, Divergent allele advantage at human MHC genes: signatures of past and ongoing selection, Mol. Biol. Evol. (2018).
[46] E.K. Wakeland, S. Boehme, J.X. She, C.C. Lu, R.A. McIndoe, I. Cheng, et al., Ancestral polymorphisms of MHC class II genes: divergent allele advantage, Immunol. Res. 9 (1990) 115.

[47] D. Meyer, VR CA, Bitarello BD, DY CB, Nunes K: a genomic perspective on HLA evolution, Immunogenetics 70 (2018) 5.
[48] K. Hosomichi, T. Shiina, A. Tajima, I. Inoue, The impact of next-generation sequencing technologies on HLA research, J. Hum. Genet. 60 (2015) 665.
[49] V. Albrecht, C. Zweiniger, V. Surendranath, K. Lang, G. Schofl, A. Dahl, et al., Dual redundant sequencing strategy: full-length gene characterisation of 1056 novel and confirmatory HLA alleles, HLA 90 (2017) 79.
[50] V. Balz, S. Krause, J. Fischer, J. Enczmann, More than 150 novel variants of HLA class I genes detected in German Stem Cell Donor Registry and UCLA International Cell Exchange samples, HLA 91 (2018) 187.
[51] S. Ferrari-Lacraz, D. Masson, J. Villard, S. Buhler, Identification of seven novel HLA class I and II alleles, HLA 92 (2018) 164.
[52] Lazaro A, Hou L, Tu B, Masaberg C, Enriquez E, Gerfen J et al. Full Gene HLA Class I Sequences of 79 Novel and 519 Mostly Uncommon Alleles from a Large United States Registry Population. HLA 2018.
[53] J. Qi, T.J. Wang, L.P. Chen, X.F. Wang, M.N. Wang, J.H. Wu, Utility of next-generation sequencing methods to identify the novel HLA alleles in potential stem cell donors from Chinese Marrow Donor Program, Int. J. Immunogenet. 45 (2018) 225.
[54] P. Cano, Evolutionary fate of new alleles, HLA 91 (2018) 329.
[55] P. Cano, W. Klitz, S.J. Mack, M. Maiers, S.G. Marsh, H. Noreen, et al., Common and well-documented HLA alleles: report of the Ad-Hoc committee of the American society for histocompatibility and immunogenetics, Hum. Immunol. 68 (2007) 392.
[56] P. Cano, S.J. Mack, M. Fernandez-Vina, The CWD allele catalog in China, Hum. Immunol. 73 (2012) 60.
[57] H.P. Eberhard, A.H. Schmidt, J. Mytilineos, K. Fleischhauer, C. Muller, Common and well-documented HLA alleles of German stem cell donors by haplotype frequency estimation, HLA 92 (2018) 206.
[58] Y. He, J. Li, W. Mao, D. Zhang, M. Liu, X. Shan, et al., HLA common and well-documented (CWD) alleles in China, HLA 92 (2018) 199.
[59] S.J. Mack, P. Cano, J.A. Hollenbach, J. He, C.K. Hurley, D. Middleton, et al., Common and well-documented HLA alleles: 2012 update to the CWD catalogue, Tissue Antigens 81 (2013) 194.
[60] A. Sanchez-Mazas, J.M. Nunes, D. Middleton, J. Sauter, S. Buhler, A. McCabe, 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 89 (2017) 104.